Asian Journal of
Science, Technology, Engineering, and Art

e-ISSN: 3025-4507
p-ISSN: 3025-5287
Index: Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/AJSTEA.v4i1.8252

Comparison of CNN and CNN-LSTM Performance in Facial Expression Classification Based on FER2013 Dataset

Putu Ananda Adi Savitri, Agus Aan Jiwa Permana, Ni Putu Novita Puspa Dewi
Universitas Pendidikan Ganesha, Indonesia
ananda.adi@undiksha.ac.id; agus.aan@undiksha.ac.id

Article Info:

Submitted:	Revised:	Accepted:	Published:
Nov 24, 2025	Dec 19, 2025	Dec 31, 2025	Jan 5, 2026

Abstract

Although facial expression recognition (FER) using deep learning has received increasing attention in prior studies, research specifically addressing the comparative effectiveness of sequential modeling on static image data remains limited. This study aims to evaluate and compare the performance of a pure Convolutional Neural Network (CNN) model and a hybrid CNN–Long Short-Term Memory (CNN-LSTM) model in classifying seven basic facial expressions using the static FER2013 dataset. A quantitative experimental approach with a comparative study design was employed, utilizing the publicly available FER2013 dataset and two custom deep learning architectures. Data were obtained from FER2013 and model performance was evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The findings indicate that the pure CNN model significantly outperformed the CNN-LSTM model, achieving a testing accuracy of 63.25% compared to 46.82% for the hybrid model; the CNN provided strong discrimination for visually distinct classes but continued to struggle with visually similar expressions. These results contribute to the theoretical development of deep learning architecture selection and expand understanding of the application of sequence models to static data. The study concludes that data characteristics (static versus temporal) play a crucial role in determining model effectiveness, and that for static datasets such as FER2013, a pure CNN constitutes the more appropriate choice. The implications of this research include theoretical contributions to the growing literature on deep learning-based FER and practical recommendations for developers to prioritize CNN architectures for non-temporal image classification tasks, while also highlighting opportunities for future research on transfer learning and attention mechanisms to better capture subtle expression nuances.

Keywords: CNN; CNN-LSTM; FER2013; Deep Learning; Image Classification.

INTRODUCTION

Facial expressions are one of the most fundamental and universal forms of nonverbal communication in humans. They intrinsically reflect a person's emotional state, intentions, and cognitive responses. (Ekman & Friesen, 1978)

The ability to automatically recognize and interpret facial expressions has paved the way for a variety of innovative applications across various sectors. In the field of human-computer interaction, systems that can understand user emotions can create more personalized and adaptive experiences, for example in virtual assistants or social robotics. (Sabrina Salsabila, 2025) In the security sector, expression detection can certainly be an early indicator of suspicious behavior. In education, monitoring student expressions can help teachers gauge levels of understanding and engagement. (Viola & Jones, 2001) (Li et al., 2022).

Despite its broad potential applications, automatic facial expression recognition still faces significant challenges. Inherent variations in lighting, viewing angle, occlusion, and individual differences in how emotions are expressed all contribute to the complexity of this research. Traditional facial expression recognition methods often rely on manual feature extraction and shallow machine learning, which tend to be less robust to this variability and require a high degree of domain expertise. (Ekawati et al., 2024)

Rapid advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in computer vision tasks, including image classification. CNNs are highly effective in extracting hierarchical features from image data (Athanasiadou et al., 2018). CNNs also are highly effective in extracting hierarchical features from image data. Their proven capability in complex spatial classification tasks includes applications like identifying diseases in potato leaves and classifying land imagery (Lesmana et al., 2022; Pangestu et al., 2020). The capability of CNNs to isolate discriminative spatial features without manually engineered features makes it an ideal method for static image classification tasks, such as facial expression recognition (FER). However, facial expressions are dynamic, involving sequential changes in facial features over time. To capture this temporal aspect, a model capable of processing sequential data is essential. (Dewi & Ismawan, 2021).

Long Short-Term Memory (LSTM), as a type of Recurrent Neural Network (RNN), is well-suited for modeling long-term dependencies in sequential data (Sedana et al., 2024). The efficacy of this hybrid architecture in handling sequential data is evident in studies (Zhang et al., 2021). The CNN-LSTM combination, potentially combining the spatial feature extraction capabilities of CNN with the sequential modeling capabilities of LSTM, is expected to improve the accuracy of facial expression classification, especially on datasets that capture spatial and temporal variations. (Altiarika & Sari, 2023).

The rapid development of technology in the field of artificial intelligence (AI), particularly deep learning, has opened new vistas in human behavior analysis. The ability of deep learning to learn complex patterns from large amounts of raw data, without the need for manually engineered features, makes it an ideal candidate for facial expression detection tasks (Montaha et al., 2022). This study aims to compare the performance of a pure CNN model with a CNN-LSTM model in the task of facial expression classification on the FER2013 dataset, thus, it is expected to provide insight into the effectiveness of each architecture in handling the complexity of facial expression classification (Alfandi & Sihite, 2022).

The ability of deep learning to learn complex patterns makes it an ideal candidate for facial expression detection tasks. Recent efforts have utilized advanced techniques like YOLOv5 for real-time emotion detection and comparative studies using architectures like RepVGG to achieve high accuracy (Mulyana et al., 2023; Saputri et al., 2022) . Furthermore, the importance of rigorously comparing classification algorithms is highlighted in the systematic review by Marti, Puspa Dewi, & Erawati Dewi (2024). Their findings, derived from the analysis of land use classification using remote sensing technology, underscore that the **choice and meticulous comparison of algorithms** are crucial factors in maximizing predictive accuracy across diverse image-based and sensor-based datasets. (Satriawan et al., 2024)

METHODS

Research Type

This research employs an experimental quantitative approach. The quantitative methodology is justified by the objective measurement of model performance using numerical metrics such as accuracy, precision, recall, and F1-score. The experimental nature is defined by the systematic construction, training, testing, and comparative evaluation of multiple custom deep learning architectures (CNN and CNN-LSTM) to test a specific hypothesis about their effectiveness in facial expression classification (Baek et al., 2020; Indraswari et al., 2022; Primawati et al., 2023).

Figure 1 LSTM Architecture

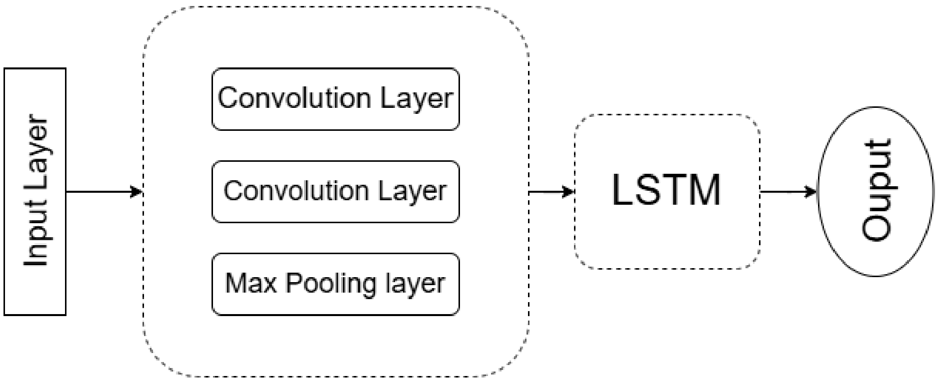

Figure 2 CNN-LSTM Architecture

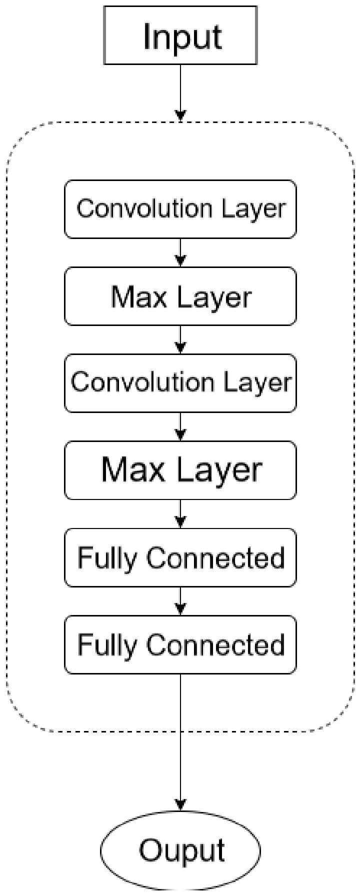

Figure 3 CNN Architecture

Research Design

The study utilizes a comparison-based research design (quasi-experiment) to assess the relative efficacy of two distinct deep learning methods, CNN and CNN-LSTM, on a standardized image dataset. This design is appropriate because the research aims directly to answer a comparative question: "How does the performance of CNN and CNN-LSTM models compare...?". The research flow is systematic, beginning with problem identification and literature review, followed by data preprocessing, model architecture design, model training, model testing, and concluding with model analysis and comparison. This design allows for empirical hypothesis testing regarding the suitability of sequence models (LSTM) for static image data.

Population and Sample (Research Data)

The population in this study is the set of all facial images labelled with basic emotions. The sample used is the FER2013 (Facial Expression Recognition 2013) Dataset. The dataset comprises 35,557 grayscale facial images (48x48 pixels).

• Sample Size: The full dataset is used, divided into a training set (28,821 images, 80%) and a test set (7,066 images, 20%).

• Emotion Classes: The data is classified into seven basic emotions: 'Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', and 'Neutral'.

• Data Imbalance: The distribution is imbalanced, with 'Happy' having the most samples (7,164 training) and 'Disgust' the fewest (436 training).

Tabel 1. Data Distribution

Expression	Number of Training Data	Number of Test Data
Angry	3,993	960
Disgusting	436	111
Afraid	4.103	1,018
Like	7,164	1,825
Sad	4,938	1,139
Surprised	3,205	797
Neutral	4,932	1,216
Total	28,821	7,066

Tabel 2. Dataset Visualization

Angry	Sad	Happy	Surprised	Disgusting	Afraid	Neutral
						
						

Data Collection Instruments and Techniques

The data collection instrument is the publicly available FER2013 dataset, accessed in CSV format. The technique involves secondary data analysis (Johnston, 2014) since the dataset was crowdsourced and manually annotated prior to this study. The raw data undergoes several pre-processing steps:

- **Data Reading and Separation:** Loading pixel data and labels from the CSV and separating them into training and test sets.
- **Image Normalization:** Pixel values are divided by 255 to normalize the range from [0, 255] to **[0, 1]** to ensure stable gradients during training.
- **Input Reshaping:** Specifically for CNN-LSTM, the image is reshaped to (48, 48), treating pixel rows as time steps for sequential input.

Data Analysis

The data analysis focuses on evaluating the performance of the trained deep learning models on the test dataset (Ihsan et al., 2021).

- **Analytical Techniques:** The primary technique is **inferential statistics** based on model performance metrics.
- **Evaluation Metrics:** The calculated metrics include:
  - Accuracy: Measures the proportion of correct predictions overall.
  - Precision, Recall, and F1-Score: Used to evaluate performance across individual, potentially imbalanced classes.
  - Confusion Matrix: A visualization to understand class-specific errors (True Positives, False Positives).
  - ROC-AUC (Receiver Operating Characteristics and Area Under Curve): Assesses the model's ability to discriminate between positive and negative classes across all thresholds.
- **Software Used:** The models were implemented using a Deep Learning framework, specifically leveraging the Python programming language and its relevant libraries (e.g., Keras/TensorFlow) for model construction and metric calculation.

RESULTS

In this section, the research findings are presented systematically, detailing the performance results of the Convolutional Neural Network (CNN) model and the hybrid CNN-Long Short-Term Memory (CNN-LSTM) model in classifying facial expressions on the FER2013 dataset. The results are presented based on the findings from the model testing phase that directly addresses the research formulations.

CNN Model Classification Results

A series of experiments were conducted using various CNN architectures to find the optimal configuration. The overall training results are summarized in Table 3.

Tabel 3. All CNN Training Architectures

No	architecture	epoch	training results	testing results	Note
1	2 conv2D (32, 64), 1 dense (256 units), dropout 0.4	100	90.70%	56.94%	early stopping occurred in epoch 29
2	4 conv2D (32, 64, 128, 256), 2 dense (256 & 512 units), dropout 0.5		80.08%	63.25%	early stopping occurred in epoch 37
3	3 conv2D (32, 64, 128), 1 dense (256), dropout 0.5		81.66%	61.59%	early stopping occurred at epoch 33
4	4 conv2D (64, 128, 256, 512), 1 dense (512), dropout 0.3	50	98.73%	51.39%	without augmentation
5			98.59%	51.55%	without augmentation
6			98.57%	51.45%	without augmentation
7			98.31%	51.09%	without augmentation
8			98.56%	58.71%	without augmentation
9		100	99.52%	52.37%	without augmentation
10			58.55%	59.40%	With minimal augmentation
11			58.40%	58.94%	
12			53.38%	54.17%	Minimal augmentation+ class weight
13	4 conv2D (64, 128, 256, 512), 1 dense (512), dropout 0.5		99.15%	57.33%	without augmentation
14	4 conv2D (64, 128, 256, 512), 1 dense (128), dropout 0.5 1e-4		99.37%	54.00%	Without Augmentation

15	4 conv2D (64, 128, 256, 512), 1 dense (128), dropout 0.5		65.62%	62.80%	Augmentation + class weight + early stopping (did not happen)
16	4 conv2D (64, 128, 256, 512), 1 dense (512), dropout 0.5, 1e-4		99.15%	57.33%	without augmentation
17	4 conv2D (64, 128, 256, 512), 1 dense (512), dropout 0.5		99.37%	54.00%	early stoping occurred at epoch 25

The significantly low accuracy achieved by the CNN-LSTM model, reaching only 46.82%, is primarily attributable to the static nature of the FER2013 dataset, which fundamentally lacks a temporal dimension. This architectural failure stands in direct contrast to the established success of CNN-LSTM models in studies involving genuine sequential or time-dependent data. The Long Short-Term Memory (LSTM) component is specifically engineered to model long-term dependencies across sequences, relying on intricate gating mechanisms (input, forget, and output gates) to selectively retain or discard information over time. Since the FER2013 dataset consists of single, non-sequential images, the key advantage of LSTM, its ability to model sequential dependencies, cannot be exploited. Consequently, the inclusion of the LSTM layer merely increases the computational overhead and the total number of trainable parameters, thereby raising model complexity without contributing any useful feature enhancement. Furthermore, the study confirms the established theoretical understanding that the CNN-LSTM architecture is sub-optimal and highly inefficient for non-temporal data, aligning with the known limitations of recurrent neural networks when applied outside their intended domain of modelling sequential input (Musa et al., 2023).

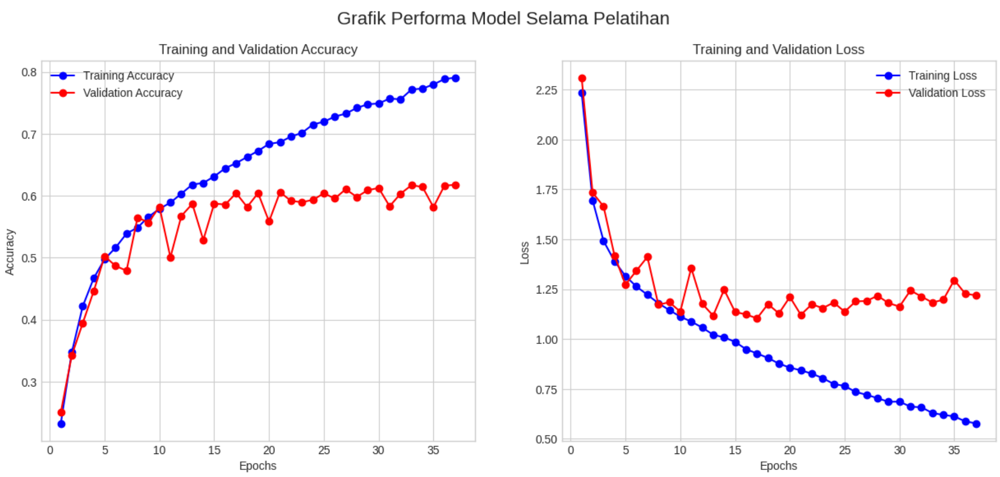

Figure 2 CNN Architecture Performance

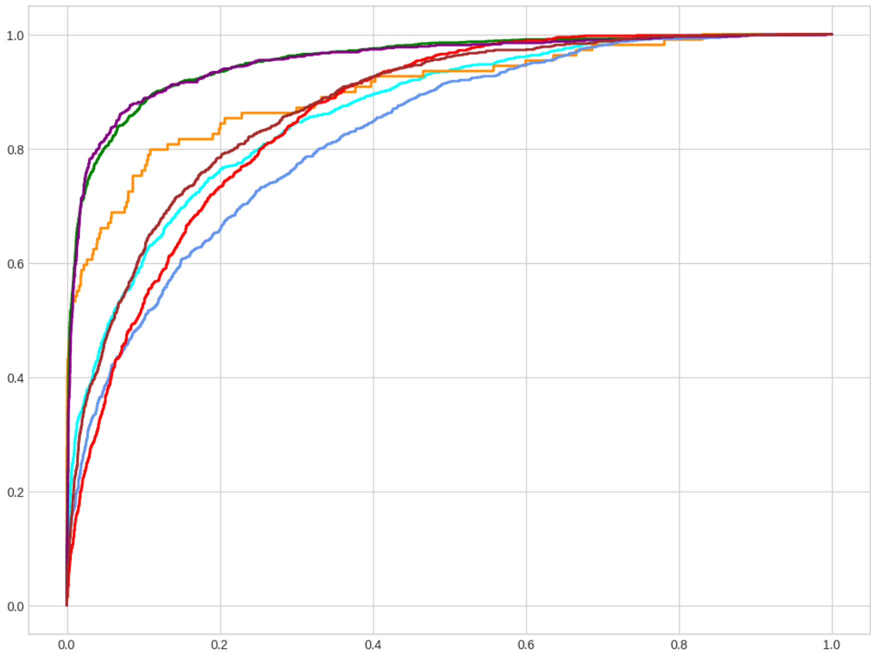

Figure 3 CNN ROC AUC Graph of Architecture

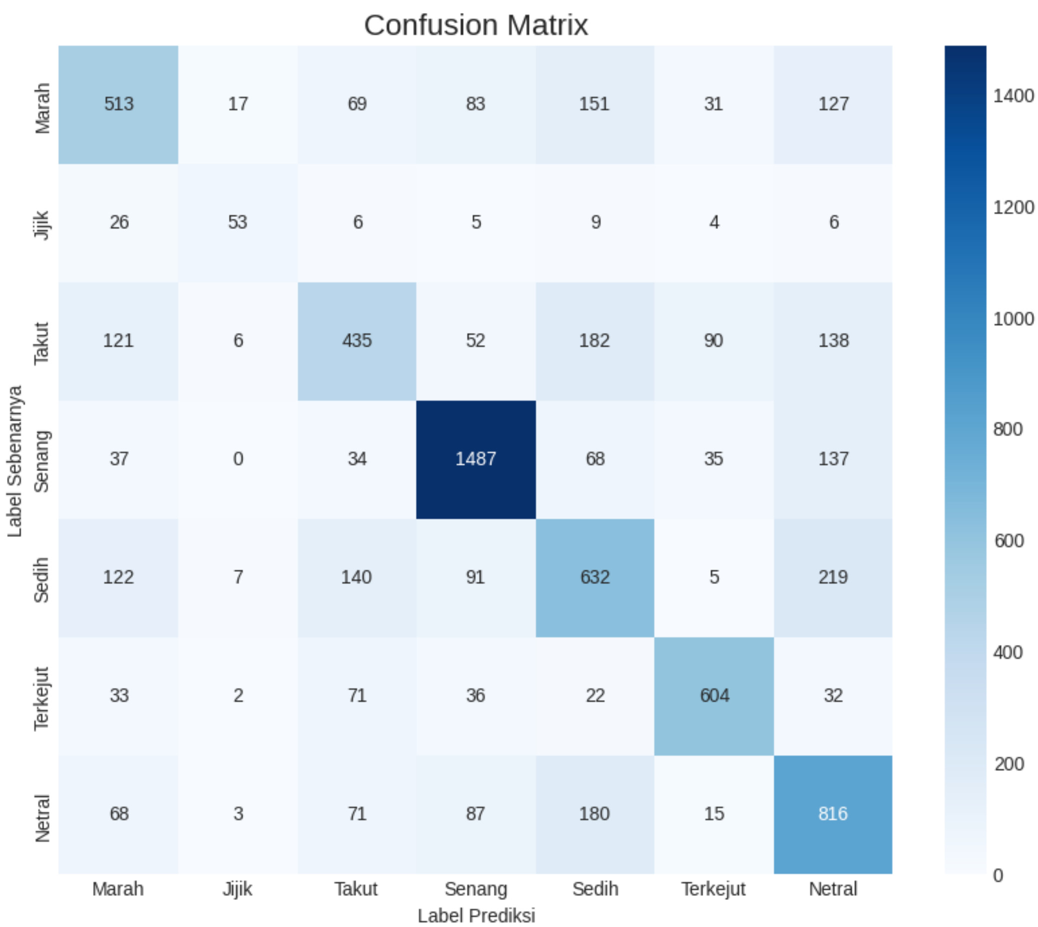

Figure 4 CNN Confusion Matrix Architecture

CNN-LSTM Model Classification Results

The CNN-LSTM model was built with an experimental architecture involving an Input Layer (48x48), followed by a TimeDistributed Dense (64 units), an LSTM (128 units), and a Dense output layer.

Tabel 4. CNN-LSTM Archictecture Training Result

No	architecture	epoch	training results	testing results	Note
1	input layer (48x48), CNN TimeDistributed Dense 64, LSTM 128, output Dense 7	100	46.67%	46.82%	without augmentation, early stopping occurs at epoch 46

The CNN-LSTM model achieved a training accuracy of 46.67% and a testing accuracy of only 46.82%. This accuracy is significantly lower compared to the best pure CNN architecture.

The significantly low accuracy achieved by the CNN-LSTM model, reaching only, is primarily attributable to the static nature of the FER2013 dataset, which fundamentally lacks a temporal dimension. This architectural failure stands in direct contrast to the established success of CNN-LSTM models in studies involving genuine sequential or time-dependent data. The Long Short-Term Memory (LSTM) component is specifically engineered to model long-term dependencies across sequences, benefiting from frame-to-frame or step-to-step relationships in inputs like videos or time series. Since the FER2013 dataset consists of single, non-sequential images, the key advantage of LSTM, its complex sequential modeling capability, cannot be exploited. Consequently, the poor performance here validates the conclusion that the LSTM layer's critical function is essentially nullified when sequential context is absent. Furthermore, the study confirms the established theoretical understanding that the CNN-LSTM architecture is sub-optimal and highly inefficient for non-temporal data, which perfectly aligns with the known limitations of recurrent neural networks when applied outside their intended domain of modeling sequential input. (Hermanto et al., 2021).

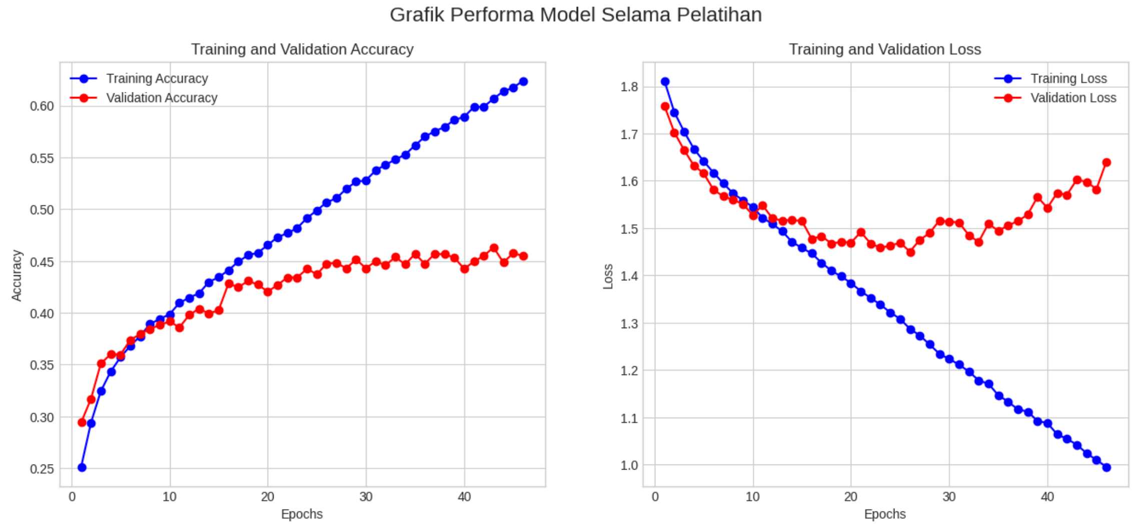

Figure 5 CNN-LSTM Architecture Performance Graph

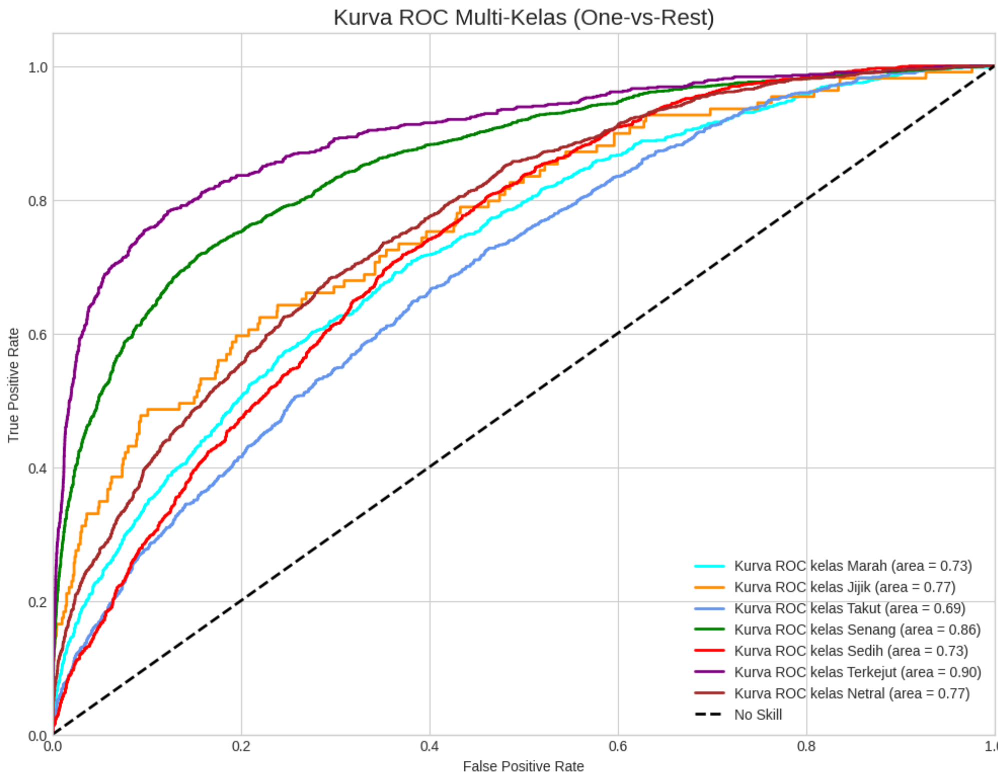

Figure 6 ROC AUC CNN-LSTM Architecture

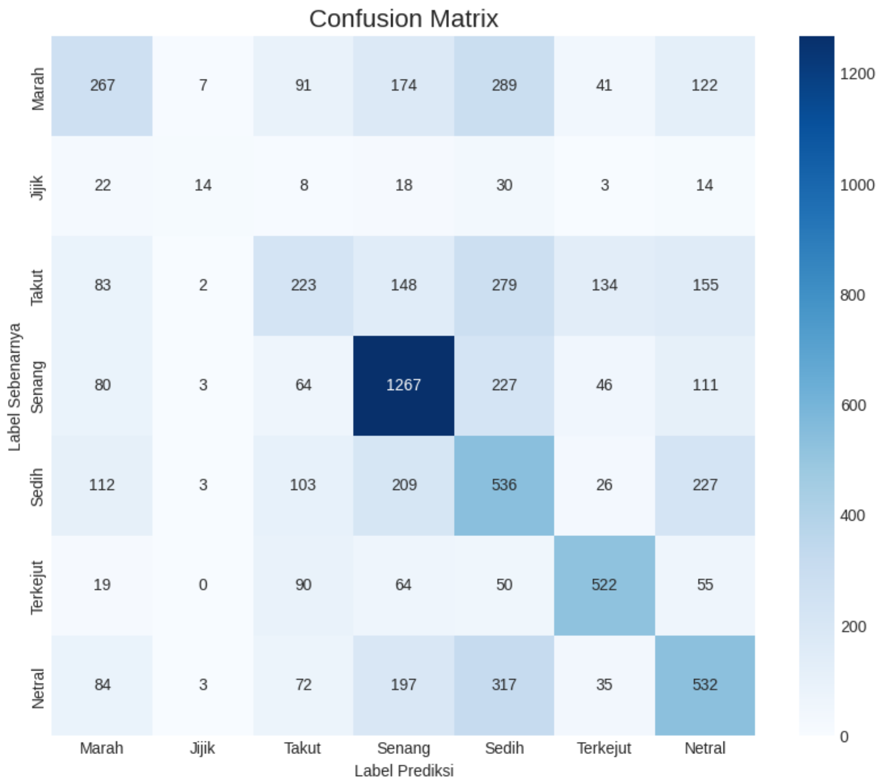

Figure 7 Confusion Matrix CNN-LSTM Architecture

DISCUSSION

Analysis and Interpretation of Findings

The experimental comparison between the pure Convolutional Neural Network (CNN) model and the hybrid CNN-Long Short-Term Memory (CNN-LSTM) model on the static FER2013 dataset yielded clear and conclusive results. The best CNN architecture achieved a significantly higher testing accuracy of 63.25%, while the CNN-LSTM model managed only 46.82%. This substantial difference directly addresses the research objective concerning the comparative performance of the two architectures.

The superior performance of the pure CNN model is directly attributable to its inherent capability to efficiently extract spatial features from facial images through its convolution and pooling layers. These core layers are highly effective at capturing key spatial indicators of expression, such as eyebrow position, lip contour, and eye shape, which are essential for classification.

Conversely, the significantly lower accuracy of the CNN-LSTM is primarily explained by the static nature of the FER2013 dataset. The Long Short-Term Memory

(LSTM) layer is intrinsically designed to process sequential or temporal data, with its efficacy dependent on benefiting from the relationship between frames in a video or time series. Since FER2013 consists of single, non-sequential images, the key advantage of LSTM. Its ability to model sequential dependencies, cannot be exploited. The experimental approach taken, treating pixel rows as an artificial time sequence, coupled with the unconventional use of a TimeDistributed Dense layer instead of the standard Conv2D for feature extraction, proved ineffective. Instead of gaining sequential context, the architecture merely increased model complexity without improving performance, resulting in the model experiencing genuine learning difficulties rather than traditional overfitting. This finding validates the conclusion that the LSTM layer's critical function is nullified when sequential context is truly absent.

Despite the CNN model's overall success, its main limitation remains the inherent inability to distinguish subtle nuances between expressions that have significant visual overlap, such as Fear and Sad. This persistent challenge, highlighted by misclassifications in the confusion matrix, requires a more sophisticated approach than the custom CNN can provide. The importance of continuous evaluation and comparison of new and specialized algorithms is essential for maximizing predictive performance on image data, a necessity further highlighted by the systematic review conducted (Satriawan et al., 2024).

Comparison with Previous Studiess

The CNN results of 63.25% are competitive with findings from other standard CNN-based studies on FER2013, which typically report accuracies in the 60%–70%$ range. Specifically, this result is comparable to Ihsan et al. (2021) and Musa et al. (2023). However, the result is lower than those achieved by models employing advanced techniques like YOLO with RepVGG (95%) or using pre-trained architectures like AlexNet (94.81%). This suggests that while the custom CNN architecture is fundamentally sound, integrating transfer learning or more complex architectures may be necessary to achieve state-of-the-art accuracy.

Furthermore, the study confirms the established understanding that the CNN-LSTM architecture is sub-optimal for non-temporal data, which aligns with the theoretical limitations of recurrent neural networks when applied outside their intended domain of sequential data.

Theoretical Contributions

This research empirically contributes to the literature on deep learning architecture selection by providing a clear benchmark and confirmation that architectural suitability to data structure is paramount. It demonstrates that for spatial classification tasks (single images), augmenting the CNN with sequential components (LSTM) without a true temporal dimension is detrimental to performance.

Practical Contributions

The findings offer direct, actionable guidance for developers and practitioners working on computer vision tasks. For applications strictly involving static image data (e.g., classifying photos from a database, identifying emotions in historical media, or simple diagnostics), a pure, well-tuned Convolutional Neural Network (CNN) is the unequivocally recommended and most resource-effective architecture.

This recommendation implies a strategic allocation of computational and human resources: Resources should be consciously invested in refining the intrinsic CNN components (optimizing filters, tuning dropout rates, implementing advanced augmentation, and adjusting regularization) rather than introducing unnecessary complexity through sequential layers (Fadilla et al., 2023). Incorporating complex sequential architectures like LSTM on static data, while theoretically appealing for dynamic problems, proved ineffective in this study, confirming that they only increase training time and parameter count without enhancing classification performance. This strategic focus is vital for developing efficient and deployable systems, such as the implementation of CNN models in mobile environments like Android applications for real-time detection.

CONCLUSION

Summary of Key Findings

This study successfully implemented and evaluated the performance of the pure CNN and hybrid CNN-LSTM models for classifying seven basic facial expressions using the static FER2013 dataset. The pure CNN model proved to be the more effective and accurate classifier for static image data. The best configuration achieved a test accuracy of 63.25%. Classification was strongest for expressions with distinct visual features like Happy and. The greatest challenge lay in distinguishing visually ambiguous classes like Fear and Sad.

While The CNN-LSTM model exhibited significantly lower performance, achieving a test accuracy of only 46.82%. This poor result confirms that the CNN-LSTM architecture is unsuitable for single-image classification tasks due to the absence of a temporal dimension for the LSTM layer to exploit.

Scientific Contribution

This study makes several scientific contributions: The Theoretical Contribution where It provides empirical evidence confirming the architectural incompatibility between sequence models (LSTM) and static datasets (FER2013), reinforcing the principle that data structure must dictate model choice in deep learning for effective feature extraction. While the Methodological Contribution details a comprehensive evaluation methodology using a full suite of metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC curves per class), offering a more robust performance analysis beyond simple accuracy metrics reported in some prior studies.

Recommendations for Future Research

The analysis derived from the recommendations indicates that while the CNN model successfully demonstrated proficiency in static spatial feature extraction (achieving 63.25% accuracy), the study encountered a critical limitation in distinguishing between visually ambiguous emotions (e.g., Fear and Sad), which could not be resolved by the custom architecture. A major finding is the architectural incompatibility of the CNN-LSTM model: since the FER2013 dataset is static, the LSTM's temporal function was neutralized, leading to significantly lower performance (46.82%). Future research must therefore follow two distinct paths: first, to truly leverage the CNN-LSTM's potential, subsequent work must use video-based datasets or image sequences to capture genuine temporal dynamics of expression change; and second, to improve static classification accuracy for ambiguous emotions, the methodology must incorporate state-of-the-art models such as Transfer Learning or Attention Mechanisms to enhance micro-feature analysis.

REFERENCES

Alfandi, M., & Sihite, A. M. H. (2022). Penerapan Metode CNN-LSTM dalam Memprediksi Hujan pada Wilayah Medan. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, *6*(1), 490–499. https://ejurnal.stmik-budidarma.ac.id/index.php/komik/article/view/5713

Altiarika, E., & Sari, W. P. (2023). Pengembangan Deteksi Realtime untuk Bahasa Isyarat Indonesia dengan Menggunakan Metode Deep Learning Long Short Term Memory dan Convolutional Neural Network. *Jurnal Teknologi Informatika dan Komputer*, *9*(1), 1–13.

Athanasiadou, E., Geradts, Z., & van Eijk, E. (2018). Camera recognition with deep learning. *Forensic Sciences Research*, *3*(3), 210–218. https://doi.org/10.1080/20961790.2018.1485198

Baek, S. S., Pyo, J., & Chun, J. A. (2020). Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water*, *12*(12), 3399. https://doi.org/10.3390/w12123399

Dewi, N., & Ismawan, F. (2021). Implementasi Deep Learning Menggunakan CNN untuk Sistem Pengenalan Wajah. *Faktor Exacta*, *14*(1), 34–43. https://doi.org/10.30998/faktorexacta.v14i1.8989

Ekawati, I., Putra, F. N. R., Sumadyo, M., & Whidhiasih, R. N. (2024). Deteksi Emosi Menggunakan Convolutional Neural Network Berdasarkan Ekspresi Wajah. *Journal of Students' Research in Computer Science*, *5*(1), 73–82. https://doi.org/10.31599/h0kayy31

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Consulting Psychologists Press.

Fadilla, M. A., Setiawan, H., & Ramadhan, M. (2023). Implementasi Metode Convolutional Neural Network (CNN) pada Sistem Deteksi Emosi dari Ekspresi Wajah Manusia dengan Aplikasi Android sebagai Antarmuka Pengguna. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKSI)*, *9*(4), 126–138.

Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online. *Creative Information Technology Journal*, *8*(1), 64. https://doi.org/10.24076/citec.2021v8i1.264

Ihsan, M., Niswatin, R. K., & Swanjaya, D. (2021). Deteksi Ekspresi Wajah Menggunakan Tensorflow. *Joutica*, *6*(1), 428. https://doi.org/10.30736/jti.v6i1.554

Indraswari, R., Herulambang, W., & Rokhana, R. (2022). Deteksi Penyakit Mata pada Citra Fundus Menggunakan Convolutional Neural Network (CNN). *Techno.Com*, *21*(2), 378–389. https://doi.org/10.33633/tc.v21i2.6162

Johnston, M. (2014). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries*, *3*, 619–626.

Lesmana, A. M., Fadhillah, R. P., & Rozikin, C. (2022). Identifikasi Penyakit pada Citra Daun Kentang Menggunakan Convolutional Neural Network (CNN). *Jurnal Sains dan Informatika*, *8*(1), 21–30. https://doi.org/10.34128/jsi.v8i1.377

Li, J., Yap, M. H., Cheng, W.-H., See, J., Hong, X., Li, X., & Wang, S.-J. (2022). FME '22: 2nd workshop on facial micro-expression: Advanced techniques for multi-modal facial expression analysis. In *Proceedings of the 30th ACM International Conference on*

Multimedia (pp. 7397–7399). Association for Computing Machinery. https://doi.org/10.1145/3503161.3554777

Montaha, S., Azam, S., Rafid, A. K. M. R. H., Hasan, M. Z., Karim, A., & Islam, A. (2022). TimeDistributed-CNN-LSTM: A hybrid approach combining CNN and LSTM to classify brain tumor on 3D MRI scans performing ablation study. *IEEE Access*, *10*, 60039–60059. https://doi.org/10.1109/ACCESS.2022.3179577

Mulyana, D. I., Sufriman, A., & Yel, M. B. (2023). Implementasi Deteksi Emosional pada Wajah Menggunakan Deep Learning–YOLOv5. *JUTECH: Journal Education and Technology*, *4*(1), 12–22. https://doi.org/10.31932/jutech.v4i1.2174

Musa, P., Anam, W. K., Musa, S. B., Aryunani, W., Senjaya, R., & Sularsih, P. (2023). Pembelajaran Mendalam Pengklasifikasi Ekspresi Wajah Manusia dengan Model Arsitektur Xception pada Metode Convolutional Neural Network. *Rekayasa*, *16*(1), 65–73. https://doi.org/10.21107/rekayasa.v16i1.16974

Pangestu, A. R., Basuki Rahmat, & Fetty Tri Anggraeny. (2020). Implementasi Algoritma CNN untuk Klasifikasi Citra Lahan dan Perhitungan Luas. *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, *1*(1), 166–174.

Primawati, A., Sitanggang, I. S., Annisa, A., & Astuti, D. A. (2023). Perbandingan Kinerja LSTM dan Prophet untuk Prediksi Deret Waktu (Studi Kasus Produksi Susu Sapi Harian). *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, *9*(3), 428. https://doi.org/10.26418/jp.v9i3.72031

Sabrina Salsabila. (2025). Teman AI: Peran Character AI dalam Kehidupan Sosial Generasi Z. *Jurnal Yudistira: Publikasi Riset Ilmu Pendidikan dan Bahasa*, *3*(2), 42–49. https://doi.org/10.61132/yudistira.v3i2.1626

Saputri, A. P., Taqwa, A., & Soim, S. (2022). Analisis Deteksi Objek Citra Digital Menggunakan Algoritma YOLO dan CNN dengan Arsitektur RepVGG pada Sistem Pendeteksian dan Pengenalan Ekspresi Wajah. *Syntax Literate: Jurnal Ilmiah Indonesia*, *7*(9), 13069–13080. https://doi.org/10.36418/syntax-literate.v7i9.9393

Satriawan, P. R., Ferdinand, G. M., I Nyoman Putra Satya Natha, I Gst Ayu Pradnya Saci Devi Sastrawan, Marti, N. W., & Ni Putu Novita Puspa Dewi. (2024). Evaluasi dan Perbandingan Algoritma Klasifikasi dalam Analisis Penggunaan Lahan dengan Teknologi Remote Sensing: Sebuah Kajian Sistematik. *INSERT: Information System and Emerging Technology Journal*, *5*(2), 97–109.

Sedana, N. M. K., Wijaya, I. N. S. W., & Arthana, I. K. R. (2024). Analisis Sentimen Berbahasa Inggris dengan Metode LSTM Studi Kasus Berita Online Pariwisata Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(6), 1325–1334.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)* (Vol. 1, pp. I-511–I-518). IEEE Computer Society. https://doi.org/10.1109/CVPR.2001.990517

Zhang, F., Bajwa, U. I., & Ahmed, M. U. (2021). Deception detection in videos using the facial action coding system [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2105.13659