International Journal of Reconfigurable and Embedded Systems (IJRES)
Vol. 15, No. 1, March 2026, pp. 42~53
ISSN: 2089-4864, DOI: 10.11591/ijres.v15.i1.pp42-53

Deployment and evaluation of facial expression recognition on Android and Temi V3 in controlled settings

Mohamad Hariz Nazamid[1], Rozita Jailani[1], Nur Khalidah Zakaria[1], Anwar P. P. Abdul Majeed[2]
[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]Faculty of Engineering and Technology, Sunway University, Bandar Sunway, Malaysia

Article Info

Article history:

Received Jan 21, 2025
Revised Dec 11, 2025
Accepted Jan 11, 2026

Keywords:

Facial expression recognition
Image processing
Lightweight model
Robot integration
Temi V3

ABSTRACT

Facial expression recognition (FER) is vital for improving human-robot interaction (HRI). This study presents the deployment and evaluation of an optimized FER model on android devices, specifically tested on the Temi V3 robot in controlled environments. Trained using FER+ and CK+ datasets and optimized with TensorFlow Lite (TFLite) and MobileNetV2, the model achieved a validation accuracy of 92.32%. Its performance was assessed on the Temi V3 robot and a Samsung A52 smartphone, focusing on CPU usage, memory, and power consumption. Cross-device compatibility and real-time performance challenges were addressed through model quantization and thread optimization. Real-time testing on the Temi V3 showed an overall accuracy of 82.28%, with emotion-specific accuracies ranging from 46.19% to 92.28%. This study offers practical insights for optimizing FER systems across android platforms, with potential applications in education, healthcare, and customer service. The results support the feasibility of implementing FER models as backends in android applications, enabling more intuitive and responsive HRI. Future work will focus on improving model efficiency for lower-end devices and exploring on-device learning techniques to boost accuracy in diverse real-world environments.

Corresponding Author:

Rozita Jailani
Faculty of Electrical Engineering, Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia
Email: rozitaj@uitm.edu.my

1. INTRODUCTION

Human-robot interaction (HRI) plays a significant role in various sectors, including healthcare and HRI. One notable application is assistive robots that utilize artificial intelligence (AI) to provide emotional support in elderly care settings [1], [2]. HRI has advanced considerably with the integration of AI technologies. For instance, natural language processing (NLP) technologies have enabled seamless communication with users across diverse accents, languages, and speech patterns [3]. Additionally, computer vision technologies facilitate the recognition of human gestures and expressions, allowing for more natural interactions [4], [5]. Future developments in HRI are expected to incorporate multi-modal interaction capabilities, combining speech, gestures, facial expressions, and touch, to enable even more natural and intuitive interactions.

Facial expression recognition (FER) lightweight models can be categorized into two classes: regression models and classification models. This study focuses on classification models, which are considered to be more reliable. A comparison of different models conducted by other researchers is shown in Table 1. Based on this table, the success rate of FER detection among various lightweight models ranges from approximately 55% to 57% for TensorFlow Lite (TFLite) models. The highest success rate observed was 56.89%, achieved using a TFLite model with the MnasNet network [6].

Table 1. The comparison of different lightweight models [6]

Model	Batch	Dropout	Optimizer	LR	Success rate (%)	
					TF	TFLite
MnasNet A=1.5 D=3	8	None	Adam	0.0001	56.864	56.889
EfficientNetB0	8	None	SGD	0.01	56.164	56.164
MnasNet A=1.0 D=7	8	None	Adam	0.0001	55.989	56.014
EfficientNetB0	8	None	Adam	0.001	55.914	55.939
EfficientNetB0	8	None	Adam	0.001	55.689	55.689
DenseNet121	8	None	Adam	0.0001	55.639	55.639
EfficientNetB1	8	None	SGD	0.01	55.489	55.489
EfficientNetB0	4	None	Adam	0.001	55.264	55.264
MnasNet A=1.0 D=3	8	None	Adam	0.0001	55.114	55.114
EfficientNetB0	4	None	SGD	0.01	55.089	55.089
MobileNetV2	8	0.2	Adam	0.001	54.839	54.839
GhostNet	8	None	Adam	0.0001	54.714	54.714
MnasNet A=1.0 D=1	8	None	Adam	0.0001	54.589	54.564
MobileNetV3Large A=1.25	16	0.5	Adam	0.001	54.489	54.389
EfficientNetV2B1	16	None	SGD	0.01	54.464	54.439
MobileNetV3Large A=1.25	16	0.2	Adam	0.001	54.364	54.339
EfficientNetV2B0	8	None	SGD	0.01	54.264	54.239
SqueezeNet C=1.0	8	0.2	Adam	0.0001	54.139	54.114
EfficientNetV2B1	8	None	SGD	0.01	54.064	54.039
MobileNetV3Small A=1.25	16	0.5	Adam	0.001	53.963	53.938

Meanwhile, robot applications have become indispensable in various sectors, particularly within mental health, education, and services [7]-[11], due to their dedicated functionality and resource efficiency. The demand for lighter, more efficient models suitable for robotic applications has led to innovative solutions such as the local sliding window attention network (SWA-Net), which achieves an accuracy of 90.03% on RAF-DB datasets [12], and recent transformer-based FER approaches, such as the attentive pooling vision transformer by Xue *et al*. [13], report an accuracy of 89.94% on the RAF-DB dataset, showing strong generalization in real-world scenarios. These advancements deliver impressive results despite reduced model sizes and resource demands.

Another important component of the FER model is the datasets used for training. Based on the study, there are multiple types of datasets available for training FER models. One of the most commonly used benchmarks is FER 2013 and its extended version, FER+ [14]. FER+ datasets are considered cleaner and can provide better accuracy. Additionally, the CK+ dataset [15], originally comprising video sequences, has been adapted by some researchers into 48×48 grayscale images for use in image classification training. FER+ has demonstrated good accuracy in pooling vision transformer research, achieving a validation accuracy of 89.94% [13]. Meanwhile, the CK+ dataset is one of the best, offering an accuracy of 100% [16]. The high accuracy of the CK+ dataset can be attributed to the consistency of the data in image classification. FER+ also benefits from being a clean version of the FER 2013 dataset [17], with some noise removed. The noise in the data can be seen in Figure 1.

Figure 1. The noise in the FER 2013 datasets

The implementation of lightweight FER models on different mobile devices has been widely studied, with multiple models being deployed across various mobile platforms, as illustrated in Table 2 in the practical review [6] of the FER system in the Figure 1. While in the microcontroller, it has high capability in real time application embedded in therapeutic systems [18]. However, the integration of these models into real robots still requires testing to evaluate their performance in real-time detection and the accuracy of FER in a real-time environment. Currently, the application of FER on the Temi V3 robot is still under exploration, as there is limited research on its use in real-life implementations. In FER systems for mobile devices or smartphones, evaluation metrics include latency in milliseconds. Additionally, assessing the accuracy of the models during implementation involves comparing the model's output with actual human emotions. These assessment metrics determine the efficiency of the model's processing during interactions with humans and evaluate the model's reliability in real-life implementations.

Table 2. Testing of different model on different android device [6]

Model	Size (MB)	Device latency (ms)			
		ZenFone 5	Pixel 6	Galaxy A40	Raspberry Pi 4
DenseNet121	15.39	663	171	751	3298
EfficientNetB0	2.4	87	29	79	333
MnasNet A=1.0 D=1	5.77	104	39	86	461
MnasNet A=1.5 D=3	12.24	249	66	227	562
MobileNetV2	6.76	94	37	106	472
MobileNetV3Large A=1.0 MINI	4.6	74	28	78	402
MobileNetV3Large A=1.0	7.6	83	32	91	746
ShuffleNet CH=64	0.78	76	36	84	74
ShuffleNet CH=128	1.82	177	70	170	186
ShuffleNet CH=200	3.41	351	152	341	338
ShuffleNetV2 SF=0.5	4.47	57	25	63	1257
ShuffleNetV2 SF=1.0	9.66	164	60	142	1270
ShuffleNetV2 SF=1.5	17.66	268	74	253	1321
SqueezeNet C=1.0	2.41	92	38	94	397

The Temi V3, illustrated in Figure 2, is an advanced service robot featuring smart path planning, face tracking, NLP, text-to-speech (TTS), speech-to-text (STT), and camera capabilities. While the robot supports various customization approaches, this study focuses on android application development by integrating the existing software development kit (SDK) provided by Temi. The android application was installed on the Temi V3 using the android debug bridge (ADB) [19], enabling seamless deployment and optimization of the robot’s advanced functionalities to align with the objectives of the research.

Figure 2. Robot Temi V3

In this paper, the author will contribute to a FER model using a combination of the FER+ dataset and CK+48 dataset, employing the TensorFlow framework and the pre-trained MobileNetV2 [20] network. The paper's contributions include developing and optimizing an android application for implementation on android devices and the Temi V3 robot. The android application will be developed using the Robot Temi SDK [19] and OpenCV API [21] for real-time detection. Optimization efforts will focus on establishing guidelines for developing an optimized android application. It will begin with background information and research on the implementation of lightweight models in android systems, along with the development of models using the TFLite framework [22]. Following this, the paper will detail the methodology and approach used to implement the FER model into a custom android application, as well as the approach employed to collect and analyze real-time detection data using the FER system on the Temi V3. The subsequent section will present the results of the performance analysis and offer optimization suggestions for app development to enhance the backend. Following this, the paper will discuss the testing of the FER system on the Temi V3, focusing on controlled subjects and environments. Finally, the author will conclude the findings and suggest improvements and directions for future development.

2. RESEARCH METHOD

This section explains the methodology adopted in this study in detail. It begins with an introduction to the custom FER TFLite model, the android application integration with the custom TFLite model, and the approach used to collect data for controlled subjects for the real-time detection test.

2.1. Facial expression recognition TensorFlow Lite model

The datasets used to train the FER model were a combination of the FER+ [14] and modified CK+ [15] with the specification of 48×48 in grayscales datasets. Together, they contain 8.159 black-and-white (grayscale) photos with a resolution of 48 by 48 pixels. For training and testing, the dataset is divided into two sets: 7.343 photos for training and 816 photos for testing. The images represent seven fundamental facial emotions: anger, fear, sadness, neutrality, happiness, surprise, and disgust. The number of images varies across these seven distinct emotion categories. Figure 3 illustrates the proposed model architecture for the FER system. The input layer receives pre-processed grayscale facial images of size 48×48 pixels, which are then resized to 96×96 and converted to RGB format to match the input requirements of the pre-trained MobileNetV2 [20] model. This model, with include top=False, serves as a powerful feature extractor. An attention layer is applied to enhance the model's focus on salient facial regions. The extracted features are then passed through a global average pooling (GAP) layer to reduce dimensionality. To mitigate overfitting, two Dropout layers with a dropout rate of 0.5 are used. A fully connected dense layer with 128 units and ReLU activation introduces non-linearity, followed by a final dense output layer with a sigmoid activation for binary classification. This architecture enables the model to predict one of two facial expression classes. The trained model is later converted to TFLite [22] format to enable efficient deployment on edge devices.

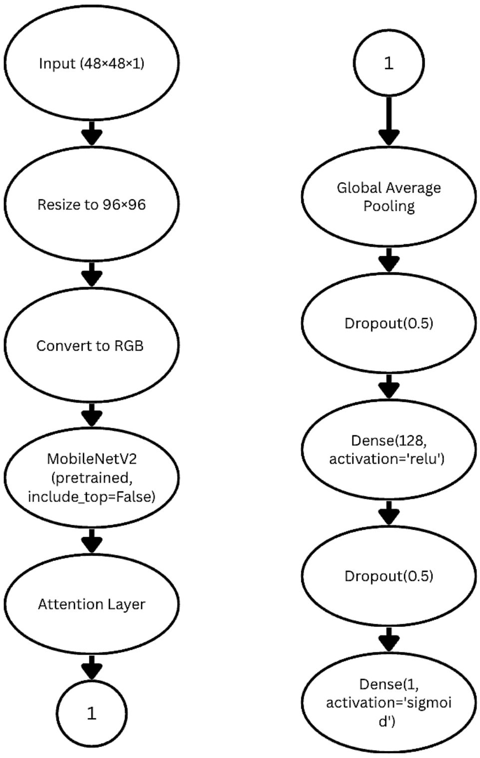
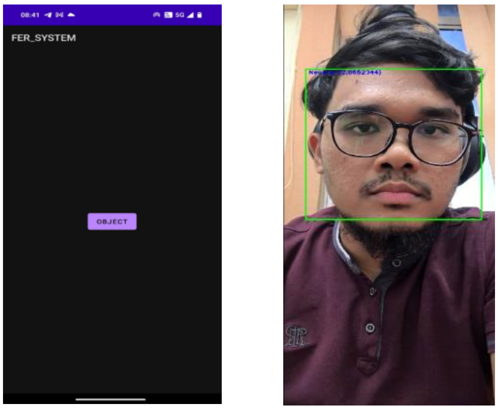


Figure 3. The FER model networks

2.2. Android application for facial expression recognition

The FER system design in android Studio, as illustrated in Figure 4, begins with initialization using the Haar Cascade model and OpenCV [20] to detect the presence of a face on the screen before determining the expression. The developed model is a single-output system, producing outputs in the range of 0 to 6.5, with each range corresponding to a specific emotion classification. This system was developed in android studio to facilitate implementation on multiple devices. The application consists of three main activities: MainActivity, CameraActivity, and FacialExpressionActivity. The MainActivity serves as the landing page, featuring a button that navigates to the CameraActivity and FacialExpressionActivity for real-time FER detection. The interface for the MainActivity is shown in Figure 4(a) on the left. Meanwhile, the CameraActivity and FacialExpressionActivity handle real-time face detection and FER processing, as depicted in the flowchart in Figure 5, with their interface displayed in Figure 4(b) on the right. The development of the application was guided by the RoboTemi documentation [19], OpenCV documentation [21], and android developer documentation [23], [24].



(a) (b)

Figure 4. The application view for; (a) MainActivity and (b) CameraActivity and FacialExpressionActivity

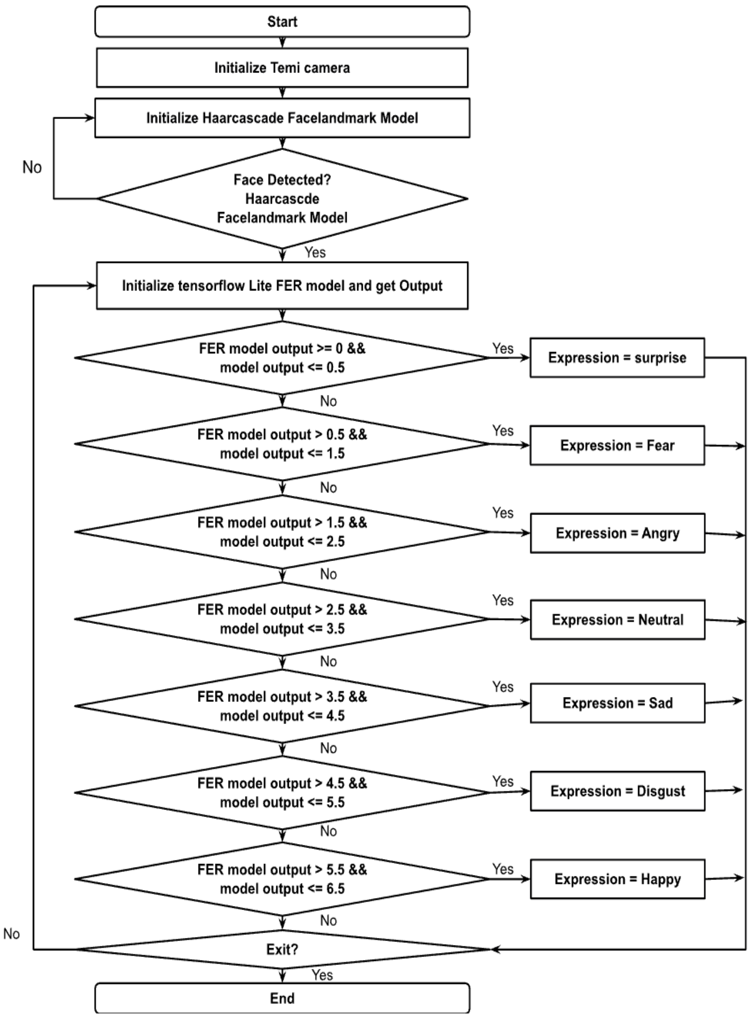


Figure 5. The flowchart for android application developed on android studio

2.3. Data collection on performance analysis

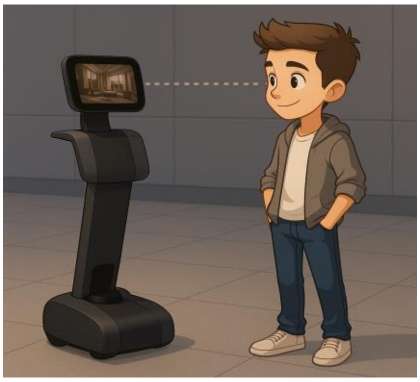
The algorithm underwent systematic evaluation across diverse devices, including the Temi V3 and the Samsung Galaxy A52 2021. The Temi V3 is equipped with an ARM Hexa-Core processor and 4 GB of RAM [25], [26], while the Samsung Galaxy A52 2021 features an Octa-Core CPU (2x2.3 GHz Kryo 465 Gold & 6x1.8 GHz Kryo 465 Silver), 8 GB of RAM, and an Adreno 618 GPU [27], [28]. These evaluations aimed to assess performance metrics such as latency and computational efficiency, ensuring the algorithm's compatibility and effectiveness across different hardware platforms.

The testing procedure for the algorithm involved initializing the program on the specified devices, followed by a continuous run duration of approximately twenty minutes. Participants positioned themselves in front of the camera and interacted with the devices to test the FER system. They navigated to the main page and accessed the FER system by activating the designated button within the application. Additionally, a non-detection scenario was tested by having participants exit the camera frame, ensuring no subject remained within its boundaries while the FER system continued to operate.

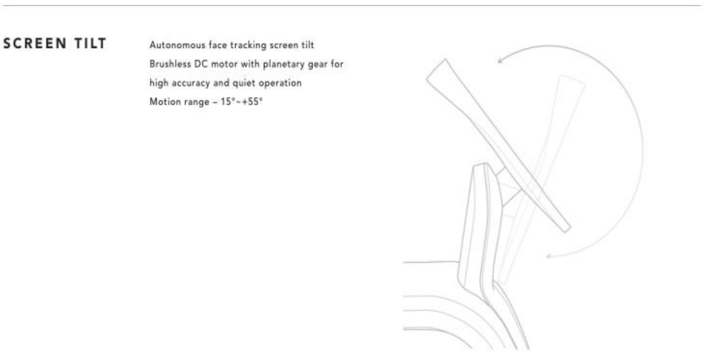
This evaluation aims to assess power, CPU, and RAM consumption to guide application optimization for backend deployment. RAM usage is measured under two conditions: non-detection (no user in frame) and detection (user present), helping identify opportunities to release unnecessary activities and reduce memory load. CPU consumption is analyzed to understand its impact on system latency, with comparisons between two devices offering insights based on hardware capabilities. Power consumption is also evaluated to minimize energy usage during continuous operation, ensuring the application remains efficient when running in the background of android systems.

2.4. Data collection on controlled settings

In this study, the data collection was done for controlled settings (well lighted room) as illustrated in Figure 6. The line of sight of the subject will face the screen of Temi V3 as in Figure 6(a), while in some cases, there are differences in the height, the Temi V3 screen will tilt to make sure the line of sight of the subject faces the screen as shown in Figure 6(b).



(a)



(b)

Figure 6. Data collection setup; (a) the controlled settings for data collection and (b) Temi V3 capabilities for face tracking based on different human height

3. RESULTS AND DISCUSSION

This section will detail the results obtained from the study, divided into system optimization and analysis, FER TFLite model result, and FER real time detection testing on controlled subjects.

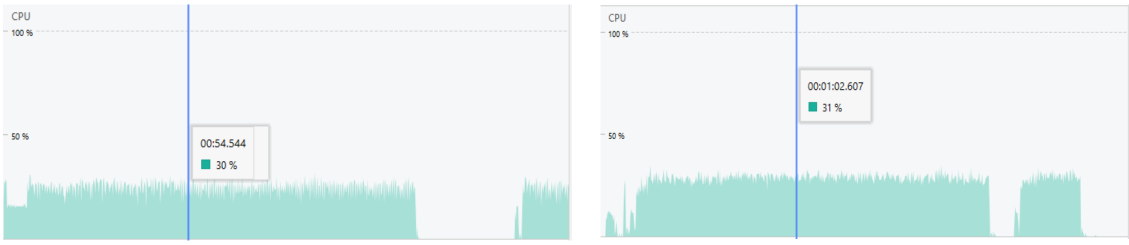
3.1. Lightweight facial expression recognition model

The lightweight model developed using the network described in the methodology achieved a validation accuracy of 92.86%. This accuracy is considered fair when compared to previous studies, such as the validation accuracy of 89.94% was reported by Xue *et al.* [13] using a transformer-based model, and the CK+ dataset, which reported a validation accuracy of 100% [16], [29], [30].

3.2. Android application performance analysis and optimization (Temi V3 vs Samsung A52 2021)

3.2.1. CPU performance

The Temi V3 CPU performance peaks at 30% usage, demonstrating efficient operation comparable to standard applications. Figure 7 illustrates intervals of idle CPU usage, coinciding with periods when FER activity ceases after participants discontinue detection. These findings suggest that the FER system can effectively operate concurrently with other applications, delivering feedback based on FER mode decisions. The CPU is inversely proportional to the latency of the application. Meanwhile, the Samsung A52 exhibits a peak CPU consumption of 31%, as depicted in Figure 7(b), which is comparable to the performance of the Temi V3. Notably, CPU usage drops almost to zero when the activity is terminated, such as when a participant returns to the main page. Notability happens in the Temi V3 as in Figure 7(a). The CPU power is inversely proportional to the latency of the application. By knowing this, power optimization can be done by destroying the activity and navigating it to Main Activity when there is no FER detection happening.

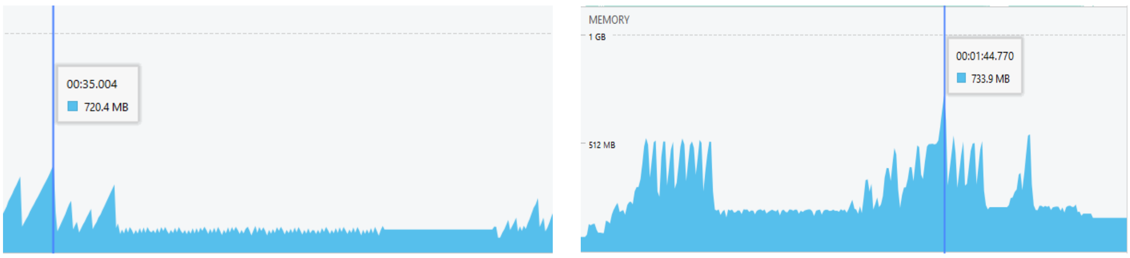


(a) (b)

Figure 7. CPU performance of; (a) Temi V3 and (b) Samsung A52

3.2.2. Memory consumption

Figure 8 shows the RAM consumption for Temi V3 and Samsung A52. The Temi V3 exhibits a peak RAM usage of 720.4 MB out of its available 4 GB, as depicted in Figure 8(a). This peak occurs notably when participants move out of frame or distance themselves from the device, underscoring the need to optimize the application to trigger FER selectively, such as by using a button or detecting a face in front of the camera. Meanwhile, Figure 8(b) displays the memory consumption of the Samsung A52 during FER system operation. Peaks in the graph occur when participants move out of frame or distance themselves from the device, requiring increased processing memory to detect facial expressions. These issues can be overcome by destroying the activity and navigating it to Main Activity when there is no FER detection happening.

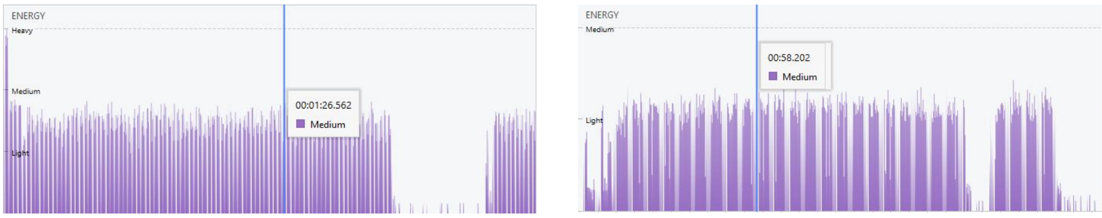


(a) (b)

Figure 8. Memory consumption; (a) Temi V3 and (b) Samsung A52

3.2.3. Energy consumption

Energy consumption data, illustrated in Figure 9, is crucial for optimizing battery usage during application development. Intervals of no battery consumption coincide with the cessation of FER activity, suggesting that effective management of FER activity can significantly extend the Temi V3's battery life as shown in Figure 9(a). In contrast, Figure 9(b) depicts the energy consumption of the FER system on the Samsung A52, showing moderate power consumption at its peak. Similar to CPU usage, energy consumption decreases significantly when the FER activity stops, as shown in Figure 9(b).



(a) (b)

Figure 9. Energy consumption; (a) Temi V3 and (b) Samsung A52

3.2.4. Optimization

Meanwhile, the latency of each device is as follows: Temi V3 with 120 ms and Samsung Galaxy A52 2021 with 92 ms. Based on the data presented, optimizing the FER system implementation can be achieved through several key characteristics. First, conditional activation ensures the FER system is only active when a face is detected, utilizing models like OpenCV or a button-triggered mechanism to initiate it. Second, Resource Management is critical, requiring proper termination in the application code to prevent unnecessary power, RAM, or CPU consumption during idle periods. Lastly, device performance highlights efficient latency and operational effectiveness on both the Temi V3 robot and Samsung Galaxy A52, effectively managing CPU, RAM, and energy usage. Further considerations for developing the FER system in an android application include addressing noise issues that arise when users are distant from the device's camera, positioned outside the frame, or when no face is detected. These challenges can be effectively managed by integrating a face detection model into the backend of the android application, ensuring that the FER system activates only upon detecting a face [31], [32].

To conclude this part of the results, performance analysis was conducted to optimize the activity lifecycle management (such as activity creation and destruction) in Java-based android application development. The analysis revealed an inverse relationship between CPU quality and application latency: better CPUs result in lower latency. Despite the varying CPU performance, overall CPU consumption remains constant, as the application is the same. The comparison between the Samsung A52 and Temi V3 devices was undertaken to observe the success of optimization efforts. Further testing of real-time FER detection on the Temi V3 is planned to assess the model's accuracy in real-time scenarios [33], [34].

3.3. Controlled subject data collection

The FER system demonstrated variable performance across 30 subjects in a controlled environment, as illustrated in Figure 10. The system achieved an average accuracy of 82.28%, with individual subject accuracies ranging from 39.13% to 100%. Notably, five subjects (18, 21, 22, 23, and 30) exhibited perfect recognition rates, while the majority (approximately 20 out of 30) showed accuracies above 80%. However, the system struggled with a few subjects, particularly subjects 4, 5, and 14, where accuracy fell below 70%. This wide performance range suggests that while the FER system is highly effective for most subjects, it may be sensitive to individual facial characteristics or expression nuances. The presence of both perfect scores and significantly lower accuracy highlights the need for further investigation into the factors influencing recognition accuracy. These findings underscore the importance of refining the algorithm to enhance consistency across diverse subjects, potentially through improved feature extraction or the incorporation of adaptive learning techniques to account for individual variations in facial expressions.

The FER system demonstrated varying levels of accuracy across different expressions when tested on 30 subjects in a controlled environment. As illustrated in Figure 11, the system excelled in recognizing happy (92.28%) and neutral (91.82%) expressions, showed moderate performance for sad (75.89%), and struggled with angry (61.42%) and surprise (46.18%) expressions. The overall average accuracy was 73.52%. These results highlight a significant disparity in the system's ability to detect different emotional states, with particular challenges in distinguishing between surprise and anger. The high accuracy for happy and neutral expressions suggests robust performance for these common states, which is promising for many real-world applications. However, the lower accuracy for angry and surprise expressions, along with reported instances of misclassification between sad/angry and neutral expressions, indicates areas requiring improvement. These findings underscore the need for further refinement of the FER algorithms, especially in differentiating between easily confused expressions, to enhance the system's reliability across a broader range of emotional states.

However, there were challenges noted in the recognition of surprise and anger causing the accuracy to be low, which sometimes confused with each other. Specifically, angry expressions were occasionally mistaken for sad or surprised, and vice versa. Additionally, misclassifications of sad and angry expressions sometimes resulted in a neutral facial expression.

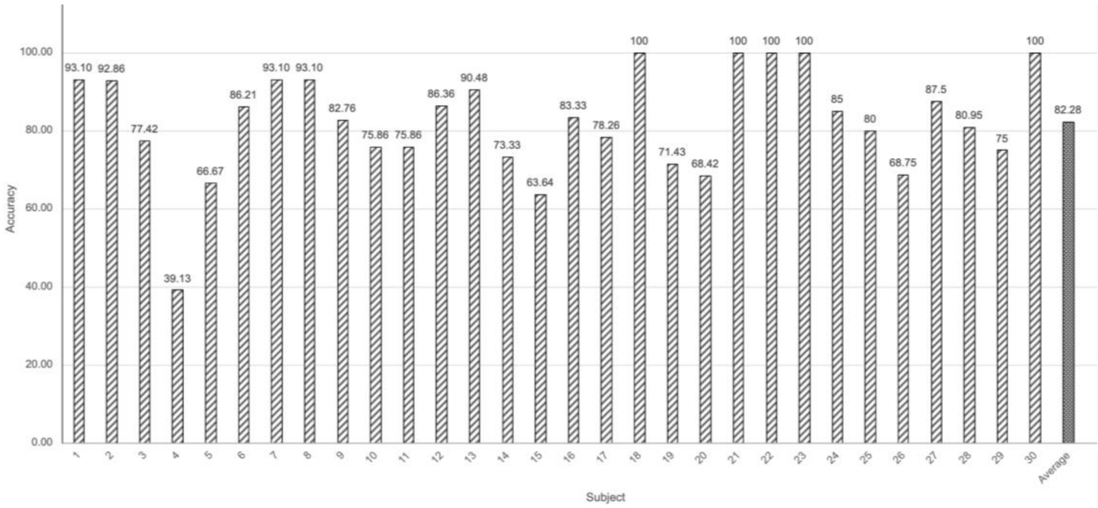


Figure 10. FER detection accuracy for 30 subjects

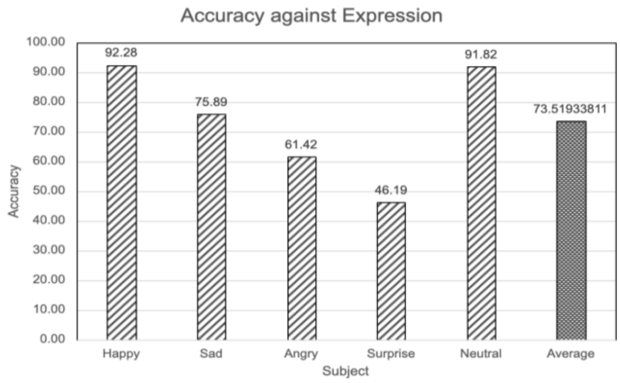


Figure 11. FER detection accuracy for 5 different expressions

4. CONCLUSION

The study aimed to evaluate the performance of a FER system on mobile devices and the Temi V3 robot, focusing on accuracy, latency, and real-time applications. The FER system achieved a validation accuracy of 92.86%, surpassing the Vision Transformer model’s 89.94% accuracy. Performance analysis revealed that CPU power is inversely proportional to application latency, highlighting the need for optimization techniques to enhance power efficiency and RAM management in android applications. The optimization suggestion had been proposed in the result to further optimize the android application developed. This paper has presented the preliminary result on the test of the android application developed into Temi V3 on the controlled subject. This part of the result will be further discussed on the continuation of this paper.

For future work, expanding the testing of the FER system across a broader range of model architectures, including MobileNetV2, MobileNetV3, and EfficientNet, is recommended to evaluate trade-offs between accuracy and resource efficiency. Incorporating larger and more diverse datasets, such as AffectNet will help improve the generalization of the model across age groups, ethnicities, and lighting conditions. Further, the adoption of TensorFlow Lite and other lightweight deployment frameworks should continue to be explored for optimizing inference on edge and embedded systems. Future efforts may also involve benchmarking the FER system on additional embedded platforms such as Raspberry Pi, Edge TPU, or NVIDIA Jetson Nano to assess cross-platform performance under real-world constraints. This would enable the development of more robust, adaptive FER systems capable of handling variable environments. These improvements will enhance the system's reliability and usability across application domains such as assistive robotics, emotion-aware learning environments, and mobile security systems. The findings of this study lay a solid foundation for such advancements, underscoring the transformative potential of FER in enabling more responsive and intuitive HRI.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Universiti Teknologi MARA (UiTM) for providing research support and grant facilities; Kizzu Kids Rehabilitation Centre for generously sponsoring the Temi V3 robot, which played a pivotal role in testing and software development; and EUREKA Robotics Centre, Cardiff Metropolitan University, UK, for their valuable research collaboration.

FUNDING INFORMATION

This work was funded by the Strategic Research Partnership (SRP) grant (100-RMC 5/3/SRP INT (047/2023) and 100-RMC 5/3/SRP INT (048/2023)) from the Universiti Teknologi MARA (UiTM), Malaysia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohamad Hariz Bin Nazamid	✓	✓	✓	✓	✓	✓		✓	✓		✓			
Rozita Jailani	✓			✓	✓		✓			✓	✓	✓	✓	✓
Nur Khalidah Zakaria					✓					✓		✓	✓	✓
Anwar P. P. Abdul Majeed	✓			✓			✓			✓			✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the Universiti Teknologi MARA (UiTM) Ethics Committee under approval number of REC/06/2024 (PG/FB/21).

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [initials, RJ] on request.

REFERENCES

[1] A. Obaigbena *et al.*, "AI and human-robot interaction: A review of recent advances and challenges," *GSC Advanced Research and Reviews*, vol. 18, no. 2, pp. 321–330, 2024, doi: 10.30574/gscarr.2024.18.2.0070.

[2] B. Sawik *et al.*, "Robots for Elderly Care: Review, Multi-Criteria Optimization Model and Qualitative Case Study," *Healthcare*, vol. 11, no. 9, 2023, doi: 10.3390/healthcare11091286.

[3] Y. Dong and X. Zhou, "Advancements in AI-driven multilingual comprehension for social robot interactions: An extensive review," *Electronic Research Archive*, vol. 31, no. 11, pp. 6600–6633, 2023, doi: 10.3934/era.2023334.

[4] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives," *Frontiers in Robotics and AI*, vol. 7, 2020, doi: 10.3389/frobt.2020.532279.

[5] H. Dong, "Research Progress and Review on Service Interaction between Intelligent Service Robots and Customers," in *Proceedings - 2023 International Conference on Service Robotics, ICoSR 2023*, 2023, pp. 78–85, doi: 10.1109/ICoSR59980.2023.00025.

[6] M. Krumnikl and V. Maiwald, "Facial Emotion Recognition for Mobile Devices: A Practical Review," *IEEE Access*, vol. 12, pp. 15735–15747, 2024, doi: 10.1109/ACCESS.2024.3358455.

[7] P. Suarez-Brito, E. O. Lopez-Caudana, and J. J. Baena-Rojas, “Co-Teaching with Humanoid Robots to Develop Complex Thinking: Appreciation of ICT Usage in Higher Education Students,” in *Proceedings - 2023 5th International Workshop on Artificial Intelligence and Education, WAIE 2023*, Tokyo, Japan, 2023, pp. 1–6, doi: 10.1109/WAIE60568.2023.00008.

[8] H. Abdollahi, M. H. Mahoor, R. Zandie, J. Siewierski, and S. H. Qualls, “Artificial Emotional Intelligence in Socially Assistive Robots for Older Adults: A Pilot Study,” *IEEE Transactions on Affective Computing*, vol. 14, no. 3, 2023.

[9] J. M. Anson, L. Leo, J. Mijohn T, R. Milton, J. Davies, and D. Devassy, “Exode: Humanoid Healthcare Robot,” in *2nd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2023 - Proceedings*, Erode, India, 2023, pp. 967–972, doi: 10.1109/ICSCDS56580.2023.10104997.

[10] Z. Zhao *et al.*, “Applications of robotics, artificial intelligence, and digital technologies during COVID-19: A review,” D*isaster Medicine and Public Health Preparedness*, vol. 16, no. 4, pp. 1634–1644, Aug. 2022, doi: 10.1017/dmp.2021.9.

[11] K. Özdemir and Ç. Taşkın, “Service robots in the restaurants: A systematic literature review using TCCM framework,” *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, vol. 28, pp. 498–516, Nov. 2025, doi: 10.29249/selcuksbmyd.1642664.

[12] S. Qiu, G. Zhao, X. Li, and X. Wang, “Facial Expression Recognition Using Local Sliding Window Attention,” *Sensors*, vol. 23, no. 7, 2023, doi: 10.3390/s23073424.

[13] F. Xue, Q. Wang, Z. Tan, Z. Ma, and G. Guo, “Vision Transformer With Attentive Pooling for Robust Facial Expression Recognition,” *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 3244–3256, 2023, doi: 10.1109/TAFFC.2022.3226473.

[14] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, “Training deep networks for facial expression recognition with crowd-sourced label distribution,” in *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction*, Tokyo, Japan, 2016, pp. 279–283, doi: 10.1145/2993148.2993165.

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, 2010, pp. 94–101, doi: 10.1109/CVPRW.2010.5543262.

[16] J. Le Ngwe, K. M. Lim, C. P. Lee, T. S. Ong, and A. Alqahtani, “PAtt-Lite: Lightweight Patch and Attention MobileNet for Challenging Facial Expression Recognition,” *IEEE Access*, vol. 12, pp. 79327–79341, 2024, doi: 10.1109/ACCESS.2024.3407108.

[17] I. J. Goodfellow *et al.*, “Challenges in representation learning: A report on three machine learning contests,” *Neural Networks*, vol. 64, pp. 59–63, 2015, doi: 10.1016/j.neunet.2014.09.005.

[18] J. Zhang *et al.*, “A Real-Time and Privacy-Preserving Facial Expression Recognition System Using an AI-Powered Microcontroller,” *Electronics (Switzerland)*, vol. 13, no. 14, 2024, doi: 10.3390/electronics13142791.

[19] “Overview — Temi V3 Guide 1.0.0 documentation,” Temi Guide, [Online]. Available: https://temi-guide-v3.readthedocs.io/en/latest/developer/01_overview.html, (Accessed: Apr. 20, 2024).

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[22] “TensorFlow Lite | ML for mobile and edge devices,” TensorFlow, [Online]. Available: https://www.tensorflow.org/lite, (Accessed: Apr. 20, 2024).

[23] “Developer guides,” Android Developers, [Online], Available: https://developer.android.com/guide/, (Accessed: Apr. 20, 2024).

[24] A. M. Pascual, E. C. Valverde, J. I. Kim, J. W. Jeong, Y. Jung, S. H. Kim, and W. Lim, "Light-FER: A Lightweight Facial Emotion Recognition System on Edge Devices," *Sensors*, vol. 22, no. 23, p. 9524, Dec. 2022, doi: 10.3390/s22239524.

[25] Temi, “Get Temi the personal robot for your business | Robotemi.com,” Temi Robot, [Online]. Available: https://www.robotemi.com/product/temi/, (Accessed: Apr. 20, 2024).

[26] C. Chang, B. Shao, Y. Li, and Y. Zhang, “Factors influencing consumers’ willingness to accept service robots: Based on online reviews of Chinese hotels,” *Frontiers in Psychology*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.1016579.

[27] Gsmarena, “Samsung Galaxy A52 - Full phone specifications,” GSMArena, Jul. 18, 2024, [Online]. Available: https://www.gsmarena.com/samsung_galaxy_a52-10641.php, (Accessed: Apr, 20. 2024).

[28] A. Ignatov, R. Timofte, W. Chou, M. W. Ke Wang, T. Hartley, and L. Van Gool, “AI Benchmark: Running Deep Neural Networks on Android Smartphones,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 288–314.

[29] H. Andrew *et al.*, “Searching for mobilenetv3,” in *Proceedings of the IEEE International Conference on Computer Vision*, Seoul, South Korea, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[30] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *36th International Conference on Machine Learning, ICML 2019*, Long Beach, CA, USA, 2019, pp. 10691–10700.

[31] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras,“Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets,” *Information*, vol. 15, no. 3, p. 135, 2024, doi: 10.3390/info15030135.

[32] M. Keinert, S. Pistrosch, A. Mallol-Ragolta, B. W. Schuller, and M. Berking, “Facial emotion recognition of 16 distinct emotions from smartphone videos: Comparative study of machine learning and human performance,” *Journal of Medical Internet Research*, vol. 27, p. e68942, Jul. 2025, doi: 10.2196/68942.

[33] K. Rao, J. Wang, S. Yalamanchili, Y. Wardi, and H. Ye, “Application-specific performance-aware energy optimization on Android mobile devices,” in *2017 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, Austin, TX, USA, 2017, pp. 169–180, doi: 10.1109/HPCA.2017.32.

[34] D. K. Alqahtani, A. Cheema, and A. N. Toosi, "Benchmarking Deep Learning Models for Object Detection on Edge Computing Devices," *arXiv preprint*, 2024, doi: 10.48550/arXiv.2409.16808.

BIOGRAPHIES OF AUTHORS

Mohamad Hariz Bin Nazamid received his B.Eng. degree in Electronics at Engineering Control System, Robotics and Automation at Universiti Teknologi Mara (UiTM). Currently pursuing M.Sc. in Electrical Engineering in the field of rehabilitation technology and robotics at UiTM. He can be contacted at email: hariznazamid@gmail.com.

Rozita Jailani received her Ph.D. in Automatic Control and System Engineering from Sheffield University, UK. She is currently a Professor at the Faculty of Electrical Engineering, and a research fellow at the Integrative Pharmacogenomics Institute (iPROMISE), Universiti Teknologi MARA, Malaysia. Her research interests include intelligent control system, rehabilitation engineering, assistive technology, instrumentation, artificial intelligence, advanced signal, and image processing techniques. She can be contacted at email: rozita@ieee.org or rozitaj@uitm.edu.my.

Nur Khalidah Zakaria received her Ph.D. in Electrical Engineering from Universiti Teknologi MARA. She is currently a senior lecturer at the Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. Her research interests include artificial intelligence, internet of things, advanced signal, and image processing techniques. She can be contacted at email: nkhalidah@uitm.edu.my.

Anwar P. P. Abdul Majeed is an Associate Professor at the Faculty of Engineering and Technology, Sunway University with an experience in transnational education. He obtained his Ph.D. in Rehabbilitation Robotics from Universiti Malaysia Pahang Al-Sultan Abdullah. He is also a Chartered Engineer with IMechE, a member of IET, and a Senior Member of IEEE. With over 200 publications and editorial roles in prestigious journals and book series, he is committed to advancing knowledge in the application of machine learning and deep learning in different domains. He can be contacted at email: anwarm@sunway.edu.my.