Original Research Article

Statistical Validation of GRMobiNet: A Lightweight Deep Neural Network Surpassing MobileNetV2 in Classification Accuracy

Raafi Careem[1, *], Md Gapar MD Johar[2], Ali Khatibi[3]

[1] School of Graduate Studies, Postgraduate Centre ,Management and Science University, Shah Alam, Selangor, Malaysia
[2] Software Engineering and Digital Innovation Centre, Management and Science University, Shah Alam, Malaysia
[3] Postgraduate Center, Management and Science University, Shah Alam, Malaysia

Citation R. Careem, M.G.MD Johar, A. Khatibi. Statistical Validation of GRMobiNet: A Lightweight Deep Neural Network Surpassing MobileNetV2 in Classification Accuracy. *J. Eng. Ind. Res.* **2025**, 6 (4):310-323.

https://doi.org/10.48309/jeires.2025.530840.1246

Article info:
Submitted: 2025-06-21
Revised: 2025-07-21
Accepted: 2025-08-05
ID: JEIRES-2506-1246

Keywords:
Deep neural networks; GRMobiNet; Image classification; MobileNet; Resource constrained environment; t-test.

ABSTRACT

This study presents a statistical evaluation of GRMobiNet, a novel lightweight deep neural network model designed for efficient image classification in resource-constrained environments (RCEs). Built upon the MobileNet architecture, GRMobiNet introduces targeted modifications to enhance predictive accuracy without increasing model size and computational complexity. To validate the performance gains, a series of controlled experiments were conducted using a custom dataset comprising ten image categories, with each model evaluated across ten repeated runs on identical test sets. A paired samples t-test was applied to compare the classification accuracy of GRMobiNet and MobileNetV2 under identical experimental conditions. Results indicate that GRMobiNet achieved a mean accuracy of 80%, significantly outperforming MobileNetV2's 57%, with a mean improvement of 2.3 correctly predicted images per run. The observed p-value of 0.019 confirms statistical significance at the 95% confidence level. Moreover, GRMobiNet exhibited lower variance and a reduced standard error of the mean, indicating greater stability across trials. These findings confirm that GRMobiNet offers not only computational efficiency, but also statistically validated performance superiority, making it highly suitable for real-world deployment in domains such as mobile diagnostics, precision agriculture, and embedded surveillance systems. The statistical rigor of this validation underscores GRMobiNet's robustness and reliability as an advancement over existing lightweight architectures.

***Corresponding Author: Raafi Careem** (mraafi@gmail.com)

Introduction

In the rapidly evolving field of deep learning, lightweight deep neural networks (DNNs) have emerged as a critical area of research, particularly for applications involving resource-constrained environments (RCEs) [1-4]. These environments such as mobile phones, embedded devices, and Internet-of-Things (IoT) platforms pose unique challenges due to limited computational resources, memory, and energy availability [5-9]. Despite these constraints, the demand for intelligent, real-time applications has surged across various sectors, including mobile healthcare diagnostics [10,11], smart agriculture [12-15], autonomous vehicles [16-19], and surveillance systems [20-22].

These applications often require on-device inference, making it imperative to develop DNNs that are both computationally efficient and accurate. One of the most prominent models in this domain is MobileNet [23-27], which leverages architectural innovations such as depthwise separable convolutions [28-30] and inverted residuals [31] to decrease model complexity and improve execution speed. MobileNet has become a baseline for many edge artificial intelligent (AI) deployments due to its small footprint and relatively high classification performance.

However, as real-world applications demand greater accuracy, robustness, and consistency, MobileNet often falls short particularly in complex or noisy environments where prediction stability is essential. Furthermore, MobileNet exhibits variability in its performance across runs, especially when deployed on diverse hardware configurations or tested with varying datasets. To meet these limitations, this study introduces GRMobiNet, a modified lightweight DNN designed to enhance predictive accuracy and consistency, while maintaining a small model size and low computational burden. Built upon the architectural foundations of MobileNet, GRMobiNet integrates advanced optimization strategies such as compound scaling [32-34] and quantization [35-38].

These enhancements aim to improve the model's inference stability, decrease its sensitivity to data and hardware variation, and ensure better generalization across diverse use cases. Unlike several previous studies that rely solely on average performance metrics, this research emphasizes the importance of statistical rigor in validating the claimed performance improvements. The central motivation of this research lies in the growing recognition that traditional benchmarking approaches reporting mean accuracy or performance on a single test set may be insufficient for evaluating the true reliability of a model. In mission-critical domains like healthcare or autonomous systems, even small improvements in prediction consistency can translate to significantly better real-world outcomes. Therefore, beyond merely reporting superior accuracy, the present study aim to statistically validate the performance improvements of GRMobiNet over MobileNetV2, an advanced version of MobileNet through controlled experiments and formal hypothesis testing. Specifically, the evaluation strategy involves repeated trials on a standardized dataset to account for variability in performance due to stochastic elements inherent in model training and deployment. A paired samples t-test [39,40] is subsequently employed to assess whether the observed differences in classification accuracy between GRMobiNet and MobileNetV2 are statistically significant. This methodological approach ensures that the results are not due to random chance, but represent a genuine improvement in model performance. The use of statistical testing in deep learning model comparison is gaining traction, especially in domains where robustness and reliability are paramount. By applying this rigorous validation methodology, our study contributes to the broader discussion about the best practices in evaluating lightweight DNNs for deployment on RCEs. While many studies focus on architectural novelty, few offer statistically grounded evidence to support claims of superiority. This work seeks to bridge this gap by providing both architectural insights and empirical validation rooted in sound statistical reasoning. Another critical dimension of this research is the practical relevance of GRMobiNet. The model is designed not only as a theoretical construct, but also as a solution aimed at real-world deployment. Its architecture supports inference on a range of commercial smartphones and edge computing devices [41-43] without the need for specialized hardware accelerators such as GPUs or TPUs [44,45].

This makes GRMobiNet an attractive option for developers seeking to embed intelligent vision capabilities into consumer-grade devices. The model's compact footprint, fast inference times, and high accuracy make it particularly suitable for applications such as real-time crop monitoring, point-of-care diagnostics, and urban surveillance, where computational overhead and latency are critical constraints. Moreover, the research methodology used in this study aligns closely with practical development workflows in applied AI [46].

By employing a custom image classification dataset composed of ten representative categories and conducting ten independent test runs per model, we simulate the kind of variability encountered in production scenarios. The controlled experimental setup ensures that both models are tested under identical conditions, removing confounding variables and enabling a fair and meaningful comparison. The inclusion of key performance metrics such as standard error of the mean (SEM) and inference variability further strengthens the evaluation robustness. Initial results show that GRMobiNet outperforms MobileNetV2 not only in terms of mean accuracy, but also in terms of predictive consistency across runs. The observed p-value of 0.019, derived from a paired t-test, confirms that the performance gain is statistically significant at the 95% confidence level. Additionally, the model exhibits a lower standard deviation and reduced SEM, indicating improved stability. These findings affirm GRMobiNet's suitability as a drop-in replacement for MobileNetV2 in edge-based image classification tasks that demand both efficiency and reliability. In summary, this study presents a statistically grounded performance analysis of GRMobiNet, positioning it as a robust and efficient alternative to existing lightweight DNN architectures. By coupling architectural optimization with formal statistical validation, we provide compelling evidence that GRMobiNet offers meaningful improvements over MobileNetV2 in real-world classification tasks. The study highlights the significance of rigorous experimental design and statistical validation in AI research, particularly when addressing deployment challenges in resource-constrained environments. Recent advancements in lightweight DNNs have been instrumental in enabling image classification on RCEs. Among these models, MobileNet [23,25,26,47] has gained widespread adoption due to its use of depthwise separable convolutions [23,30,48] and inverted residuals [31], providing a favorable trade-off between accuracy and computational efficiency. However, it still suffers from accuracy limitations and performance variability in complex real-world scenarios. Alternative architectures such as EfficientNet [49-52], ShuffleNet [53], and SqueezeNet [53,54] have introduced innovations like compound scaling [32-34] and parameter-efficient modules to enhance deployment on edge devices. Despite these efforts, many existing solutions continue to struggle with the inherent accuracy-efficiency trade-off or require extensive post-processing optimizations such as pruning or quantization [35,36].

Furthermore, performance assessments in prior studies [4,51] often rely solely on average metrics without rigorous statistical validation, making it difficult to judge the true significance of reported improvements. In contrast, this study introduces GRMobiNet, a MobileNetV2-based model enhanced with compound scaling and quantization, and provides a statistically grounded evaluation using paired t-tests to establish its performance superiority with reliability and confidence. The remainder of this article is organized as follows: Section 2 details the experimental materials and methodology. Section 3 presents the results of the empirical evaluation and statistical analysis. Section 4 provides a discussion of findings and their implications for real-world deployment. Finally, Section 5 concludes the paper with a summary and directions for future work.

Experimental

This section outlines the design and implementation of the experimental framework used to validate the performance improvements of GRMobiNet over MobileNetV2. The methodology includes a description of the model architecture, dataset, image pre-processing, the experimental design ensuring repeatability and control, and the statistical procedures applied to determine the significance of observed differences in classification accuracy.

Model Architecture

GRMobiNet is light weight convolutional neural network (CNN) architecture [4] specifically optimized for deployment in RCEs [4,41-43] such as mobile devices, embedded systems, and edge AI platforms. It is designed as a refined extension of the well-known MobileNet architecture, leveraging its fundamental principles, while introducing key enhancements that improve classification performance and robustness without increasing model complexity or inference latency. Figure 1 displays the overall architecture of GRMobiNet, highlighting its key components and data flow. The design builds upon the MobileNet backbone, integrating several modifications aimed at improving accuracy while maintaining a light-weight footprint. At its core, GRMobiNet retains MobileNet's architectural foundation, which includes the use of depthwise separable convolutions. Depthwise separable convolutions are a crucial innovation that reduce computational cost by factorizing a standard convolution into two separate operations: a depthwise convolution, which applies a single filter per input channel, and a pointwise (1×1) convolution, which combines the outputs of the depthwise layer. This technique drastically reduces the number of parameters and floating-point operations (FLOPs), making it ideal for low-power devices.

GRMobiNet preserves its structural efficiency while extending the design through a combination of scaling strategies, pooling optimizations, and quantization techniques that improve its adaptability to more diverse datasets and hardware environments. A major enhancement in GRMobiNet is the integration of compound scaling, a technique originally introduced in the EfficientNet family of models. Compound scaling involves the coordinated scaling of model depth (number of layers), width (number of channels), and input image resolution, instead of arbitrarily increasing one dimension. This balanced scaling ensures that each layer can learn richer features while keeping the overall model size and computational footprint within acceptable limits.

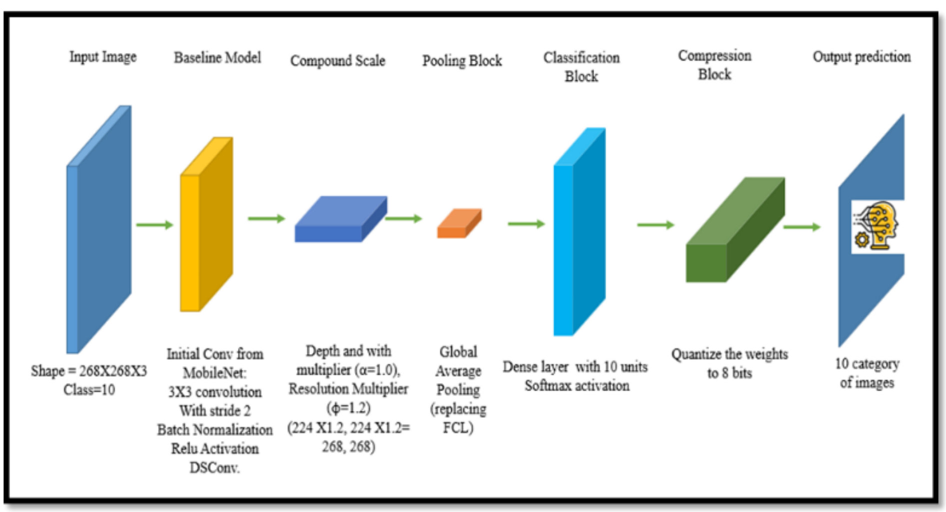


Figure 1: Architecture of GRMobiNet

In GRMobiNet, the compound scaling is applied conservatively to avoid the pitfalls of overfitting or exceeding the resource capacities typical of RCEs. The scaling coefficients were empirically tuned through ablation studies to optimize accuracy while maintaining efficiency. Besides compound scaling, GRMobiNet incorporates an improved global average pooling (GAP) [53] layer prior to the final classification block. GAP replaces traditional fully connected layers, significantly decreasing the number of parameters while maintaining feature abstraction. This pooling approach also helps minimize overfitting by enforcing spatial invariance and allows the model to generalize better across varying input contexts. The GAP layer in GRMobiNet is coupled with a dropout layer for additional regularization, enhancing the model's resilience to noisy or imbalanced data. Another defining characteristic of GRMobiNet is its native support for 8-bit quantization [35], a model compression strategy that reduces the precision of weights and activations from 32-bit floating point (FP32) to 8-bit integers (INT8). Quantization reduces memory footprint and allows the model to execute faster on processors without floating-point units, which are common in embedded and mobile systems. To further support deployment flexibility, GRMobiNet is designed with modularity and scalability in mind. It can be easily adjusted to accommodate different hardware constraints by toggling parameters related to resolution, depth multiplier, and quantization thresholds. This makes the architecture highly adaptable across use cases, ranging from real-time disease detection in agricultural settings [12,15] to mobile health diagnostics [5] and surveillance [20-22]. GRMobiNet enhances the efficiency-first philosophy of MobileNet with principled improvements in architectural scaling, quantization, and regularization. These enhancements allow the model to achieve higher classification accuracy and greater predictive stability, especially when deployed in low-resource environments. Crucially, these gains are achieved without inflating the model's size or compromising its runtime performance, making GRMobiNet a compelling solution for edge AI deployments requiring both intelligence and efficiency.

Dataset

To evaluate and compare the performance of GRMobiNet and MobileNetV2 under controlled and consistent conditions, both models were trained and tested on a carefully curated image classification dataset derived from the ImageNet1K [54] benchmark. ImageNet1K is a well-established large-scale dataset consisting of over 1.2 million training images and 50,000 validation images, organized into 1,000 distinct object categories. Due to the high computational cost associated with training and evaluating models on the full ImageNet1K set, especially in the context of lightweight models and resource-constrained experimentation, a focused subset, referred to as ImageNet10 [42], was created for the purposes of this study. ImageNet10 is a representative subset consisting of 10 distinct image categories selected to cover a range of semantic classes, visual complexity, and color variability. The categories were chosen to reflect diversity in object shapes, textures, and backgrounds to challenge the models' generalization capabilities without overwhelming them with extreme inter-class similarities. Table 1 presents the class categories of ImageNet10 along with their corresponding class IDs. The selected categories consist of common and visually distinguishable objects, including aeroplane, balloon, cock, dog, and shark. These classes were selected to balance practical relevance and experimental manageability, allowing efficient model training and evaluation while ensuring statistical validity across multiple runs. The ImageNet10 dataset comprises a total of 11,000 images, systematically divided into three subsets to support effective model development and testing. Specifically, 7,700 images (70% of the dataset) are allocated to the training set, which is employed to teach the model and enable it to learn relevant patterns and features. The validation set consists of 1,650 images (15%), serving to fine-tune model parameters and monitor performance during training, thereby helping to prevent overfitting.

Table 1: Class categories of the ImageNet10 dataset

Class ID	Class Name
0	aeroplane
1	balloon
2	cock
3	dog
4	goldfish
5	horse
6	lion
7	panda
8	parrot
9	shark

The remaining 1,650 images (15%) form the test set, which is used exclusively for test the model's final performance on unseen data. This structured distribution follows standard practices for machine learning workflows and is consistent with prior dataset configurations [55]. For statistical evaluation, a fixed test set of 10 real-world images per class was used, resulting in 100 test images per evaluation run. This test size was selected based on a trade-off between computational efficiency and statistical reliability. Running multiple evaluation iterations using the same 100-image set allowed for controlled pairwise comparison of model performance, while keeping the inference workload manageable across 10 independent runs per model.

Image pre-processing

To ensure consistency across training and testing, standard image preprocessing techniques were applied uniformly to all images. Each image was resized to 224×224 pixels, a common input dimension for lightweight DNN models, to maintain compatibility with both MobileNetV2 and GRMobiNet architectures. Data augmentation techniques were also employed during training to enhance the generalization ability of both models. These included random cropping, horizontal flipping, rotation, and brightness adjustments, which helped reduce overfitting and improved model robustness to variations in image orientation and lighting. All images were converted to the RGB color space, regardless of the original format (*e.g.,* grayscale, CMYK, or RGBA). This ensured consistent channel input for all convolutional layers, which expect three input channels.

Experimental design

To ensure the reliability and reproducibility of results, both GRMobiNet and MobileNetV2 were subjected to 10 independent evaluation runs. In each run, the models were used to classify the same 100 test real-world images, allowing for a paired-sample design that directly compares the models under identical input conditions. This setup controls for variation in dataset composition and eliminates confounding variables introduced by data shuffling (re-arrange the order of training dataset) [56]. The comparison analysis was carried out using a laptop with an Intel Celeron CPU N2920 @ 1.86 GHz (4 cores) and 4 GB of RAM in a resource-constrained setting. Both models, GRMobiNet and MobileNetV2, were set up on the laptop in this experiment to evaluate how well they performed in actual classification tasks. A total of 100 real-world images, ten in each of the ImageNet10 categories shown in Table 1, were chosen for experiment. Each model was run separately to classify pictures which were locally stored on the device. Model outputs were captured as discrete prediction results, and the number of correctly predicted samples per run was recorded. Thus, for each of the 10 evaluation runs, a score ranging from 0 to 100 was logged per model, representing the total count of correct classifications out of 100 images. This experimental framework emphasizes the significance of measuring not just average accuracy, but consistency and repeatability of model predictions key requirements in real-world applications where stable performance across multiple runs and devices is crucial.

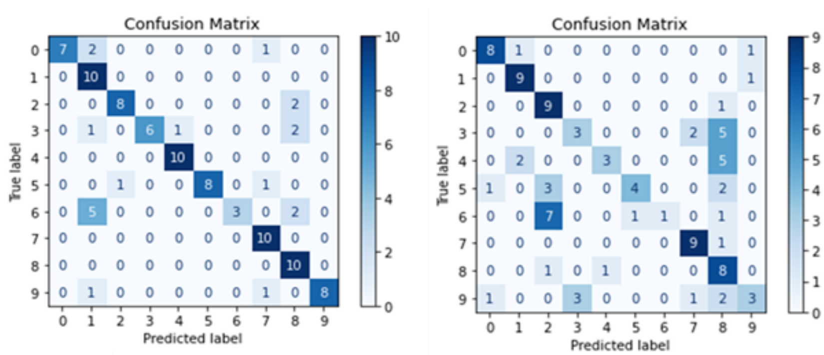
Statistical test

To evaluate whether the observed accuracy improvements of GRMobiNet over MobileNetV2 were statistically significant, a paired samples t-test [39,40] was conducted. This statistical method is appropriate for analyzing dependent samples in this case, matched pairs of model outputs from identical test sets across 10 runs. The paired design ensures that variance due to data differences is minimized, isolating model performance as the primary variable of interest. The test compared the mean number of correctly predicted images by each model over the 10 runs. GRMobiNet's performance was hypothesized to be superior to MobileNetV2, with statistical significance defined at the 95% confidence level ($p < 0.05$) [40,46]. The null hypothesis (H_0) indicates that there is no difference in classification accuracy between the two models, while the alternative hypothesis (H_1) asserted that GRMobiNet performs significantly better than MobileNetV2. The test was performed using IBM SPSS Statistics software (version 26), where accuracy scores were input as paired variables across the 10 evaluation instances. Besides the mean and standard deviation, the standard error of the mean (SEM) and confidence intervals were calculated to assess result variability and robustness. A statistically significant result would validate the performance gain of GRMobiNet as non-random and repeatable, providing stronger evidence than a single accuracy metric typically used in DNN benchmarking. The statistical test also provided valuable insights into prediction stability, as the variability (standard deviation and SEM) of the accuracy scores across runs was evaluated. Lower variability in GRMobiNet's predictions, in combination with higher mean accuracy, would further affirm its robustness for deployment in real-time scenarios.

Results and Discussion

Results

To evaluate the inference accuracy of both models, GRMobiNet and MobileNetV2, experiments were conducted over 100 classification instances (10 runs across 10 image categories). The resulting confusion matrices are presented in Figure 2.



(a) GRMobiNet (b) MobileNetV2

Figure 2: Confusion matrices of the experimental result of (a) GRMobiNet and (b) MobileNetV2

Table 2: Prediction accuracy over ten runs for each image category using GRMobiNet and MobileNetV2

Runs	Image Category ID	Correctly Predicted Images (out of 10)	
		GRMobiNet	MobileNetV2
1	0	7	8
2	1	10	9
3	2	8	9
4	3	6	3
5	4	10	3
6	5	8	4
7	6	3	1
8	7	10	9
9	8	10	8
10	9	8	3

In these matrices, the diagonal entries represent the number of true positive (TP) predictions, indicating correctly classified images out of ten per category, while the off-diagonal entries correspond to false positive (FP) predictions. The values from the confusion matrices were systematically organized into Table 2 to facilitate statistical analysis using SPSS, where a paired samples t-test was subsequently performed. SPSS statistical software with a 95% confidence interval was used to perform the analysis [40,46]. The following two tables, Paired Samples Statistics (Table 3) and Paired Samples Test result (Table 4), provide an overview of the findings of the analysis.

Paired samples statistics

The mean accuracy scores in Table 3 indicate that GRMobiNet outperforms MobileNetV2 in classification performance across repeated test runs.

Specifically, GRMobiNet achieved an average of 8.0 correct predictions out of 10, translating to 80% accuracy, while MobileNetV2 achieved an average of 5.7 correct predictions, or 57% accuracy. This 22.5% increase in accuracy suggests that the architectural improvements made in GRMobiNet such as compound scaling, enhanced feature extraction, and quantization contribute to more effective learning and generalization, even under the computational constraints for which the model is designed. In terms of reliability and consistency, GRMobiNet also performed more stably across the ten independent test runs.

This is evidenced by its lower standard deviation of 2.26, compared to MobileNetV2's higher standard deviation of 3.16. A lower standard deviation means that GRMobiNet's performance was more consistent and less variable between runs, while MobileNetV2 showed more fluctuation in its results. This consistency is crucial for deployment in real-world scenarios, particularly in RCEs, where predictability and reliability are as important as overall accuracy.

Additionally, the standard error of the mean (SEM) which measures how accurately the sample mean estimates the population mean, was lower for GRMobiNet (0.71) than for MobileNetV2 (1.00). A smaller SEM indicates that the observed mean accuracy for GRMobiNet is more precise and trustworthy. In essence, this reinforces the confidence in GRMobiNet's superior performance and confirms that the model not only performs better on average but does so with greater stability and reliability.

Paired samples t-test

To determine whether the observed accuracy differences between GRMobiNet and MobileNetV2 were statistically significant, a paired samples t-test was conducted.

Table 4 indicates the t-test result generated from SPSS 26. The mean difference between the two models' accuracies across ten test runs was 2.3, indicating that GRMobiNet correctly classified, on average, 2.3 more images than MobileNetV2 per run.

Table 3: Paired samples statistics

Model	Mean	N	Std. Deviation	Std. Error Mean
GRMobiNet	8.0000	10	2.26078	0.71492
MobileNetV2	5.7000	10	3.16403	1.00056

Table 4: Paired samples test result

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
GRMobiNet - MobileNetV2	2.30000	2.54078	0.80346	0.48244	4.11756	2.863	9	0.019

This performance gap is substantial given the small size of the evaluation set (10 classes with 10 images each per run), and it provides a strong indication that GRMobiNet offers real, practical improvements over its predecessor.

The t-value of 2.863 measures how far the observed difference is from the null hypothesis (which assumes no difference), relative to the variability of the sample differences. A higher t-value suggests a more substantial and reliable difference. In this case, with 9 degrees of freedom (df = number of pairs − 1), the calculated p-value was 0.019. A p-value of 0.019 means there is only a 1.9% probability that the observed difference or a more extreme one, could occur due to random chance if there were truly no difference between the models. Since this p-value is less than the commonly accepted threshold of 0.05 (*i.e.* $p < 0.05$), we can reject the null hypothesis with 95% confidence. This statistical significance confirms that the superior performance of GRMobiNet is not incidental, but rather a reliable and reproducible improvement. In practical terms, the t-test results provide strong evidence that GRMobiNet's enhancements have a meaningful impact on classification performance. This statistical validation is particularly important for research focused on real-world applications, where consistency, robustness, and provable gains are essential for model adoption in production environments.

Discussion

The experimental results presented in this study provide compelling evidence that GRMobiNet offers statistically significant and practically meaningful improvements over MobileNetV2 in image classification tasks, particularly in the context of resource-constrained environments (RCEs). The observed mean increase of 2.3 correctly classified images per run, corresponding to a 23% relative improvement, affirms the effectiveness of the architectural enhancements introduced in GRMobiNet. While this may seem like a modest numerical gain, it is substantial when considering the fixed test set size and the critical role of classification accuracy in real-time edge applications, where even small improvements can have outsized impacts—such as in medical diagnosis, precision agriculture, or autonomous sensing systems. From a performance stability perspective, the results are equally noteworthy. GRMobiNet consistently demonstrated lower standard deviation (2.26) and lower standard error of the mean (0.71) across 10 independent test runs, compared to MobileNetV2's standard deviation of 3.16 and SEM of 1.00. This reduction in variability indicates that GRMobiNet produces more predictable and reliable outputs, a feature that is crucial for deployment in environments where input conditions may vary and hardware constraints limit retraining or extensive tuning. The consistent accuracy across runs implies that GRMobiNet has better internal feature representations and generalization capability, likely stemming from its use of compound scaling and quantization-aware training. These improvements suggest that GRMobiNet is not only a better model statistically, but also a more deployment-ready model in practical terms. In resource-constrained scenarios, model stability can be as important as raw accuracy. Fluctuations in model performance, as seen in MobileNetV2, can result in unpredictable behavior that undermines user trust or system effectiveness especially in mission-critical tasks. In contrast, GRMobiNet's consistent performance across runs positions it as a more robust solution, capable of operating under real-world uncertainties with minimal risk of failure.

Another important implication of these findings is that enhancing model architecture does not necessarily require increasing model complexity or computational burden. GRMobiNet maintains the lightweight footprint of MobileNetV2 while delivering improved performance, demonstrating that architectural efficiency and predictive power are not mutually exclusive. This is particularly significant for edge AI deployments, where models must fit within strict memory, latency, and energy constraints. By carefully integrating compound scaling and quantization at the design level rather than as post-hoc optimizations GRMobiNet achieves a strong balance between performance and efficiency.

However, while the findings are promising, it is also important to consider the limitations of the current evaluation framework. The test dataset, ImageNet10, was constructed from a subset of ImageNet categories, and although it was designed to reflect variability in object types and image conditions, it may not fully capture the diversity and complexity of real-world visual data. Future evaluations should incorporate larger, more diverse datasets, and test the model's robustness to adversarial inputs, occlusion, or domain shifts (*e.g.,* different lighting, background clutter). Furthermore, the evaluation focused on accuracy as the primary performance metric. While accuracy is critical, other metrics—such as precision, recall, F1-score, and inference latency—may also offer valuable insights, particularly in application domains where false positives or negatives carry different risks. Measuring energy efficiency, memory usage, and model throughput on different hardware platforms (*e.g.,* smartphones, Raspberry Pi, and micro-controllers) would also strengthen the real-world relevance of GRMobiNet's deployment potential.

Finally, the statistical analysis using the paired samples t-test has confirmed that the observed performance gain is unlikely to be due to random chance. However, further studies involving larger sample sizes and cross-validation techniques would reinforce the reliability of these conclusions. Additionally, testing the model across multiple independent datasets and tasks, including fine-grained classification and object detection, would help generalize the findings and solidify GRMobiNet's value across broader vision applications.

Conclusion

This study presents GRMobiNet, a lightweight convolutional neural network designed to address the dual challenges of efficiency and accuracy in image classification within RCEs. Building upon the architectural foundations of MobileNet, GRMobiNet integrates key enhancements, namely compound scaling and quantization to deliver higher predictive performance without increasing computational complexity. The results from repeated experiments demonstrate that GRMobiNet achieves a mean classification accuracy of 80%, significantly outperforming MobileNetV2's 57%, while also exhibiting lower variability and higher prediction consistency across runs. Critically, the application of a paired samples t-test confirms that the observed performance gains are statistically significant ($p = 0.019$), thereby validating that GRMobiNet's improvements are not due to random variation but represent a reliable advancement in lightweight deep learning model design. This finding is particularly important for RCE-based deployments where robust, repeatable performance under hardware and power constraints is essential. GRMobiNet provides a meaningful step forward in the development of lightweight neural networks, providing not only theoretical and architectural innovation, but also empirical and statistically validated improvements that are immediately relevant to practical AI deployment in low-resource settings.

Orcid

Raafi Careem: 0009-0001-8632-4019
Md Gapar MD Johar: *0000-0001-7986-3834*
Ali Khatibi: *0000-0002-2531-7720*

Reference

[1]. H.-I. Liu, M. Galindo, H. Xie, L.-K. Wong, H.-H. Shuai, Y.-H. Li, W.-H. Cheng, Lightweight deep learning for resource-constrained environments: A survey, *ACM Computing Surveys*, **2024**, *56*, 1-42. [Crossref], [Google Scholar], [Publisher]

[2]. Y. Zhong, Z. Teng, M. Tong, LightMixer: A novel lightweight convolutional neural network for tomato disease detection, *Frontiers in plant science*, **2023**, *14*, 1166296. [Crossref], [Google Scholar], [Publisher]

[3]. H.K. Kondaveeti, V.K. Vatsavayi, S. Mangapathi, R.M. Yasaswini, Lightweight deep learning: introduction, advancements, and applications, advancement in business analytics tools for higher financial performance, *IGI Global Scientific Publishing*, **2023**, 242-259. [Google Scholar], [Publisher]

[4]. L. Zhao, L. Wang, Y. Jia, Y. Cui, A lightweight deep neural network with higher accuracy, *Plos one*, **2022**, *17*, e0271225. [Crossref], [Google Scholar], [Publisher]

[5]. Y. Yang, H. Wang, R. Jiang, X. Guo, J. Cheng, Y. Chen, A review of IoT-enabled mobile healthcare: technologies, challenges, and future trends, *IEEE Internet of Things Journal*, **2022**, *9*, 9478-9502. [Crossref], [Google Scholar], [Publisher]

[6]. A. Singh, S.C. Satapathy, A. Roy, A. Gutub, Ai-based mobile edge computing for iot: Applications, challenges, and future scope, *Arabian Journal for Science and Engineering*, **2022**, *47*, 9801-9831. [Crossref], [Google Scholar], [Publisher]

[7]. H. Hussain, P. Tamizharasan, C. Rahul, Design possibilities and challenges of DNN models: a review on the perspective of end devices, *Artificial Intelligence Review*, **2022**, *55*, 5109-5167. [Crossref], [Google Scholar], [Publisher]

[8]. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *Journal of big Data*, **2021**, *8*, 53. [Crossref], [Google Scholar], [Publisher]

[9]. F. Jauro, H. Chiroma, A.Y. Gital, M. Almutairi, S.i.M. Abdulhamid, J.H. Abawajy, Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend, *Applied Soft Computing*, **2020**, *96*, 106582. [Crossref], [Google Scholar], [Publisher]

[10]. R. Baron, H. Haick, Mobile diagnostic clinics, *ACS sensors*, **2024**, *9*, 2777-2792. [Crossref], [Google Scholar], [Publisher]

[11]. E. Osei, T.P. Mashamba-Thompson, Mobile health applications for disease screening and treatment support in low-and middle-income countries: A narrative review, *Heliyon*, **2021**, *7*. [Google Scholar], [Publisher]

[12]. J. Mendoza-Bernal, A. González-Vidal, A.F. Skarmeta, A convolutional neural network approach for image-based anomaly detection in smart agriculture, *Expert Systems with Applications*, **2024**, *247*, 123210. [Crossref], [Google Scholar], [Publisher]

[13]. A.M. Mishra, S. Harnal, V. Gautam, R. Tiwari, S. Upadhyay, Weed density estimation in soya bean crop using deep convolutional neural networks in smart agriculture, *Journal of Plant Diseases and Protection*, **2022**, *129*, 593-604. [Crossref], [Google Scholar], [Publisher]

[14]. V. Dhanya, A. Subeesh, N. Kushwaha, D.K. Vishwakarma, T.N. Kumar, G. Ritika, A. Singh, Deep learning based computer vision approaches for smart agricultural applications, *Artificial Intelligence in Agriculture*, **2022**, *6*, 211-229. [Crossref], [Google Scholar], [Publisher]

[15]. M. Altalak, M. Ammad uddin, A. Alajmi, A. Rizg, Smart agriculture applications using deep learning technologies: A survey, *Applied Sciences*, **2022**, *12*, 5919. [Crossref], [Google Scholar], [Publisher]

[16]. J. Alfred Daniel, C. Chandru Vignesh, B.A. Muthu, R. Senthil Kumar, C. Sivaparthipan, C.E.M. Marin, Fully convolutional neural networks for LIDAR–camera fusion for pedestrian detection in autonomous vehicle, *Multimedia Tools and Applications*, **2023**, *82*, 25107-25130. [Crossref], [Google Scholar], [Publisher]

[17]. F. Martinez, H. Montiel, F. Martinez, Comparative study of optimization algorithms on convolutional network for autonomous driving, *International Journal of Electrical and Computer Engineering*, **2022**, *12*, 6363. [Crossref], [Google Scholar], [Publisher]

[18]. L. Chen, S. Lin, X. Lu, D. Cao, H. Wu, C. Guo, C. Liu, F.-Y. Wang, Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey, *IEEE Transactions on Intelligent Transportation Systems*, **2021**, *22*, 3234-3246. [Crossref], [Google Scholar], [Publisher]

[19]. S. Grigorescu, B. Trasnea, T. Cocias, G. Macesanu, A survey of deep learning techniques for autonomous driving, *Journal of field robotics*, **2020**, *37*, 362-386. [Crossref], [Google Scholar], [Publisher]

[20]. A. Shabbir, N. Arshad, S. Rahman, M.A. Sayem, F. Chowdhury, Analyzing surveillance videos in real-time using AI-powered deep learning techniques, *International Journal on Recent and Innovation Trends in Computing and Communication*, **2024**, *12*, 950-960. [Google Scholar], [Publisher]

[21]. C. Fontes, E. Hohma, C.C. Corrigan, C. Lütge, AI-powered public surveillance systems: why we (might) need them and how we want them, *Technology in Society*, **2022**, *71*, 102137. [Crossref], [Google Scholar], [Publisher]
[22]. M.J. Iqbal, M.M. Iqbal, I. Ahmad, M.O. Alassafi, A.S. Alfakeeh, A. Alhomoud, Real-time surveillance using deep learning, *Security and Communication Networks*, **2021**, *2021*, 6184756. [Crossref], [Google Scholar], [Publisher]
[23]. M.P. Kumar, D. Hasmitha, B. Usha, B. Jyothsna, D. Sravya, Brain tumor classification using mobilenet, 2024 international conference on integrated circuits and communication systems (ICICACS), *IEEE*, **2024**, 1-7. [Crossref], [Google Scholar], [Publisher]
[24]. T.H. Arfan, M. Hayaty, A. Hadinegoro, Classification of brain tumours types based on MRI images using mobilenet, 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech), *IEEE*, **2021**, 69-73. [Crossref], [Google Scholar], [Publisher]
[25]. M.A.B. Abbass, Y. Ban, MobileNet-based architecture for distracted human driver detection of autonomous cars, *Electronics*, **2024**, 13, 2 , 365. [Crossref], [Google Scholar], [Publisher]
[26]. D. Sinha, M. El-Sharkawy, Thin mobilenet: An enhanced mobilenet architecture, 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON), *IEEE*, **2019**, 0280-0285. [Crossref], [Google Scholar], [Publisher]
[27]. A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint arXiv:1704.04861*, **2017**. [Crossref], [Google Scholar], [Publisher]
[28]. T. Magadza, S. Viriri, Brain tumor segmentation using partial depthwise separable convolutions, *IEEE Access*, **2022**, *10*, 124206-124216. [Crossref], [Google Scholar], [Publisher]
[29]. S.-F. Hsiao, B.-C. Tsai, Efficient computation of depthwise separable convolution in MoblieNet deep neural network models, 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), *IEEE*, **2021**, 1-2. [Crossref], [Google Scholar], [Publisher]
[30]. F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE conference on computer vision and Pattern Recognition*, **2017**, 1251-1258. [Google Scholar], [Publisher]
[31]. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **2018**, 4510-4520. [Google Scholar], [Publisher]
[32]. C. Lin, P. Yang, Q. Wang, Z. Qiu, W. Lv, Z. Wang, Efficient and accurate compound scaling for convolutional neural networks, *Neural Networks*, **2023**, *167*, 787-797. [Crossref], [Google Scholar], [Publisher]
[33]. S.J. Colaco, D.S. Han, Deep learning-based facial landmarks localization using compound scaling, *IEEE Access*, **2022**, *10*, 7653-7663. [Crossref], [Google Scholar], [Publisher]
[34]. M.F. Hashmi, S. Katiyar, A.W. Hashmi, A.G. Keskar, Pneumonia detection in chest X-ray images using compound scaled deep learning model, *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, **2021**, *62*, 397-406. [Crossref], [Google Scholar], [Publisher]
[35]. T. Liang, J. Glossner, L. Wang, S. Shi, X. Zhang, Pruning and quantization for deep neural network acceleration: A survey, *Neurocomputing*, **2021**, *461*, 370-403. [Crossref], [Google Scholar], [Publisher]
[36]. H.F. Langroudi, V. Karia, Z. Carmichael, A. Zyarah, T. Pandit, J.L. Gustafson, D. Kudithipudi, Alps: Adaptive quantization of deep neural networks with generalized posits, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, **2021**, 3100-3109. [Google Scholar], [Publisher]
[37]. Y. Zhou, S.-M. Moosavi-Dezfooli, N.-M. Cheung, P. Frossard, Adaptive quantization for deep neural network, Proceedings of the AAAI Conference on Artificial Intelligence, **2018**. [Crossref], [Google Scholar], [Publisher]
[38]. A. Polino, R. Pascanu, D. Alistarh, Model compression via distillation and quantization, *arXiv preprint arXiv:1802.05668*, **2018**. [Crossref], [Google Scholar], [Publisher]

[39]. O.G. Atanda, W. Ismaila, A.O. Afolabi, O.A. Awodoye, A.S. Falohun, J.P. Oguntoye, Statistical Analysis of a deep learning based trimodal biometric system using paired sampling T-Test, 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), *IEEE*, **2023**, 1-10. [Crossref], [Google Scholar], [Publisher]

[40]. S. Afifah, A. Mudzakir, A.B.D. Nandiyanto, How to calculate paired sample t-test using SPSS software: From step-by-step processing for users to the practical examples in the analysis of the effect of application anti-fire bamboo teaching materials on student learning outcomes, *Indonesian Journal of Teaching in Science*, **2022**, *2*, 81-92. [Crossref], [Google Scholar], [Publisher]

[41]. R. Careem, G. Md Johar, A. Khatibi, Analyzing baseline models for optimizing deep neural networks in resource-constrained environments, *Journal of Engineering in Industrial Research*, **2024**, 5, 2, 81–89. [Crossref], [Google Scholar], [Publisher]

[42]. R. Careem, M. Gapar Md Johar, A. Khatibi, Baseline model for deep neural networks in resource-constrained environments: an empirical investigation, *International Journal of Computing and Digital Systems*, **2024**, *16*, 1147-1161. [Crossref], [Google Scholar], [Publisher]

[43]. R. Careem, G. Johar, A. Khatibi, Deep neural networks optimization for resource-constrained environments: techniques and models, *Indonesian Journal of Electrical Engineering and Computer Science*, **2024**, *33*, 1843-1854. [Crossref], [Google Scholar], [Publisher]

[44]. A. Jayasimhan, P. Pabitha, A comparison between CPU and GPU for image classification using Convolutional Neural Networks, 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), *IEEE*, **2022**, 1-4. [Crossref], [Google Scholar], [Publisher]

[45]. H. Kimm, I. Paik, H. Kimm, Performance comparison of TPU, GPU, CPU on Google Colaboratory over distributed deep learning, *Proceedings of the 2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, **2021**, 312–319. [Crossref], [Google Scholar], [Publisher]

[46]. U.K.B. Dubey, D.P. Kothari, *Research Methodology: Techniques and Trends* (1st ed.), Chapman and Hall/CRC, **2022**. [Crossref], [Google Scholar], [Publisher]

[47]. H. Chen, G. Zhou, W. He, X. Duan, H. Jiang, Classification and identification of agricultural products based on improved MobileNetV2, *Scientific Reports*, **2024**, *14*, 3454. [Crossref], [Google Scholar], [Publisher]

[48]. L. Dang, P. Pang, J. Lee, Depth-wise separable convolution neural network with residual connection for hyperspectral image classification, *Remote Sensing*, **2020**, *12*, 3408. [Crossref], [Google Scholar], [Publisher]

[49]. S. Benkrama, N.E.H. Hemdani, Deep Learning with EfficientNetB1 for detecting brain tumors in MRI images, 2023 International Conference on Advances in Electronics, Control and Communication Systems (ICAECCS), *IEEE*, **2023**, 1-6. [Crossref], [Google Scholar], [Publisher]

[50]. B. Espejo-Garcia, I. Malounas, N. Mylonas, A. Kasimati, S. Fountas, Using EfficientNet and transfer learning for image-based diagnosis of nutrient deficiencies, *Computers and Electronics in Agriculture*, **2022**, *196*, 106868. [Crossref], [Google Scholar], [Publisher]

[51]. M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, *International Conference on Machine Learning, PMLR*, **2021**, 10096-10106. [Google Scholar], [Publisher]

[52]. V.T. Hoang, K.-H. Jo, Practical analysis on architecture of EfficientNet, 2021 14th International Conference on Human System Interaction (HSI), *IEEE*, **2021**, 1-4. [Crossref], [Google Scholar], [Publisher]

[53]. G. Habib, S. Qureshi, GAPCNN with HyPar: Global Average Pooling convolutional neural network with novel NNLU activation function and HYBRID parallelism, *Frontiers in Computational Neuroscience*, **2022**, *16*, 1004988. [Crossref], [Google Scholar], [Publisher]

[54]. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, **2012**, *25*. [Google Scholar], [Publisher]

[55]. H.A. Al-Jubouri, S.M. Mahmmod, A comparative analysis of automatic deep neural networks for image retrieval, *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, **2021**, *19*, 858-871. [Crossref], [Google Scholar], [Publisher]

[56]. L. Xu, S. Qiu, B. Yuan, J. Jiang, C. Renggli, S. Gan, K. Kara, G. Li, J. Liu, W. Wu, Stochastic gradient descent without full data shuffle: with applications to in-database machine learning and deep learning systems, *The VLDB Journal*, **2024**, *33*, 1231-1255. [Crossref], [Google Scholar], [Publisher]