Deep Learning-Enhanced Drunk Detection in Vietnam Using Thermal Imaging and Advanced Image Processing

Ngo Dinh Luan*, Nguyen Van Thanh Thong†
Dept. of AI Research
FPT University, Vietnam
*luanndse171138@fpt.edu.vn, †thongnvtse171008@fpt.edu.vn

Hoang Ngoc Dung‡, Vo Thanh Bang§
Dept. of AI Research
FPT University, Vietnam
‡dunghnse171014@fpt.edu.vn, §bangvtse171037@fpt.edu.vn

***Abstract*—Alcohol intoxication is a major cause of traffic accidents and a potential threat to public safety, especially in Vietnam. Traditional detection methods, such as breathalyzers or blood tests, are invasive to privacy, time-consuming, and require the cooperation of the subject. This study provides an innovative, non-invasive method for detecting alcohol-induced cognitive impairment on the street using thermal imaging technology. By analyzing specific facial temperature variations, features, and physiological patterns using deep learning such as convolutional neural network (CNN) models, our method provides real-time drunkenness recognition. Experimental evaluations demonstrate the advantages of this method, providing a non-contact, real-time solution that is important for law enforcement, healthcare, and traffic safety.**

***Index Terms*—Drunk identification, Convolutional Neural Network, Thermal image, Image Processing**

I. Introduction

Alcohol intoxication is a major public safety concern, primarily due to traffic accidents, occupational hazards, and crime. Alcohol use is responsible for 2.6 million deaths each year, and 724,000 deaths are due to alcohol-related injuries, such as traffic accidents, suicides, and violence, according to the World Health Organization (WHO) Global Report on Alcohol and Health (2024) [1]. In addition, 13% of alcohol-related deaths occur in people aged 20 to 39 years, indicating the acute danger of alcohol intoxication. Physical contact, cooperation of the subject, and time required are factors associated with traditional methods of measuring alcohol intoxication using breath tests and blood alcohol concentration (BAC) measurements [2]. Therefore, researchers have sought computerized, non-contact means through thermal imaging and artificial intelligence analysis to increase the efficiency and accuracy of detection [3].

Recently, it has been found that thermal infrared imaging is capable of effectively detecting drunkenness by measuring the temperature difference in facial regions due to vasodilation caused by alcohol consumption. Alcohol affects blood flow, increasing facial temperature, especially the forehead, nose, and periorbital regions [4]. Thermal pattern distributions have been studied using AI models and found to show good performance in classifying drunk subjects with high accuracy [5].

Although most previous works have focused on feature extraction and classification using CNNs, there has been little research on image pre-processing methods that enhance the clarity of thermal features. In this work, we contribute a novel work on combining Contrast-Limited Adaptive Histogram Equalization (CLAHE), Logarithmic Transformation, and K-means Clustering as pre-processing methods before feeding the data into the CNN-based classification model. With such methods, our study expects higher thermal image contrast, lower noise, and simplicity in feature extraction, ultimately leading to a more accurate and reliable drunkenness detection system.

This method is a novel addition to thermal-based alcohol detection as it maximizes image quality before classification, improving the performance of AI models for real-world use. In addition, our work provides a set of Vietnamese test samples under different conditions and different lighting environments. The new research and methodology is expected to be optimized for real-time use in law enforcement, automotive safety systems, and workplace monitoring, providing a contactless, scalable, and extremely accurate way to detect alcohol impairment.

II. Related Work

Non-contact alcohol detection studies have become increasingly popular, particularly through thermal imaging and AI. Breath testing and BAC tests are traditional methods involving physical contact and subject cooperation, with restricted real-time application. Infrared thermography, which monitors facial temperature reactions to alcohol ingestion for automated labeling [6].

Initial experiments confirmed that thermal imaging can identify drunkenness by measuring temperature variations on facial areas like the forehead, nose, and cheeks. In [5], intoxicated individuals were identified with a neural network in controlled conditions with high accuracy. The approach was later used on crowds with thermal imaging public safety surveillance research [7]. They are primarily raw thermal data methods and do not employ sophisticated image preprocessing, which would enhance accuracy. Later, deep models enhanced drunk-driving detection performance with the main focus on tracking drivers and their behavior. They utilize feature extraction from raw thermal images only, without using contrast correction or noise reduction, and using computationally costly architectures that are not applicable for real-time applications [8],[9]. The biggest challenge is still the absence of direct pre-processing of thermal images. They all use raw infrared data, which is tainted with low contrast, noise, and unreliable thermal data. For the Greek Dataset (2050 infrared face images of 41 subjects), breath alcohol levels ranging from 0.25 to 0.9 mg/L were employed as the drunk driving limit [10],[11]. It is generally used for model benchmarking but is one of the few public thermal datasets for alcohol detection. The detection of ethanol based on AI at the cell level by infrared analysis has been tested in a few studies and established that preprocessing enhances feature extraction [12]. No technique, however, utilized advanced preprocessing methods, including CLAHE, Logarithmic Transformation, and K-means Clustering, to enhance thermal image visibility prior to AI-based classification. To address such limitation, this work introduces a new preprocessing pipeline through the integration of CLAHE, Logarithmic Transformation, and K-means Clustering. This processing pipeline enhances the quality of the thermal image, feature extraction, and classification efficiency. Unlike the existing research literature that processes the images through certain techniques like Gaussian noise and blurring, this solution focuses more on optimizing preprocessing for real-time contactless detection of alcohol so that it would be easier to deploy in the law enforcement, traffic surveillance, and workforce monitoring [13]. Moreover, unlike previous studies that analyzed the Greek Dataset using traditional machine learning approaches, this study is the first to apply deep learning, specifically CNN-based architectures, for alcohol intoxication detection using thermal imaging.

III. Methodology

A. Vietnamese Thermal Drunk/Sober Dataset

The collection of data sets for the detection of alcohol by thermal imaging in Vietnam presents significant challenges, particularly due to the stringent regulations on drunk-driving offenses. As per Decree 168 on administrative penalties for traffic violations [14], penalties related to alcohol concentration testing are rigorously enforced. This has made obtaining data from participants more difficult compared to other countries. To circumvent these challenges and adhere to ethical and legal considerations, we divided our dataset into two primary components. This division was made to minimize any legal implications for participants volunteering in this study, given the budget constraints and lack of third-party funding or lab resources available to the research team.

1) Thermal Drunk Processing Dataset: Inspired by the Greek dataset [5], we developed a specialized dataset for the Vietnamese population using the Total Meter HT-03 thermal camera. The camera offers a resolution of 120 x 90 pixels and operates in the wavelength range of 8–14 µm. We focus on classifying two primary conditions. Sober and Drunk. A total of 20 volunteers (16 men and 4 women) participated in the data collection. Initially, all volunteers were identified as sober based on their breathalyzer test results or self-reporting, where they admitted not consuming alcohol in the past 24 hours. During the experiment, various body parts were scanned, including the face, left side of the face, right side of the face, left arm, right arm, left leg, and right leg. The volunteers were asked to consume two cans of Saigon Larger beer 4.6 alcohol, 330 ml each) in three separate doses (220 ml each), with 20-minute intervals between each dose. After drinking, the volunteers' breath alcohol concentration (BAC) levels were measured using a breathalyzer, which ranged from 0.25 to 0.32 mg/L)

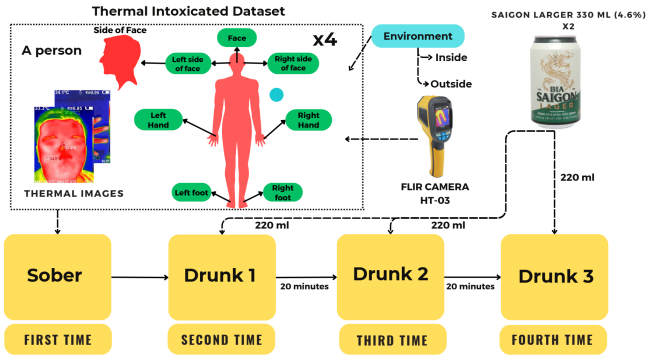

Fig. 1. Pipeline for thermal image dataset collection of intoxicated individuals in Vietnam

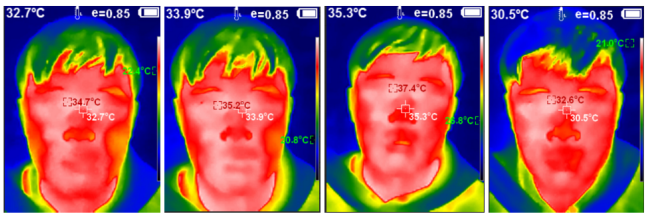

Fig. 2. From Left to Right: Sober, Drunk 1, Drunk 2, Drunk 3

Thermal readings indicated a significant temperature variation on the volunteers' bodies, particularly on the face. Specifically, the temperature increased by 2–3 degrees Celsius during the drinking process and subsequently decreased by 3–4 degrees Celsius after the final drink. This dataset indicates the thermal changes corresponding to alcohol consumption.

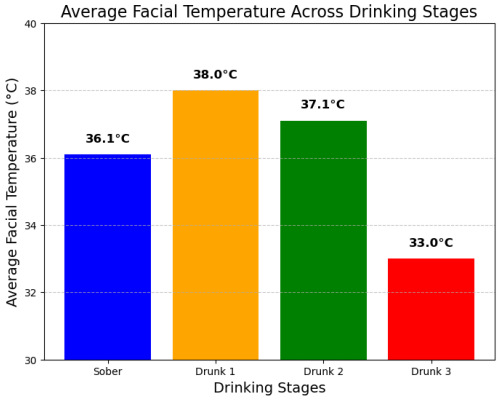

Fig. 3. Average Facial Temperature Across Drinking Stages

2) Thermal Diversity Dataset: To enhance the diversity and robustness of the learning model, we created an additional dataset with a larger sample size. This dataset was divided into two classes: Sober and Drunk. For the sober class, we recruited 45 participants, and captured over 800 thermal images from the same body parts scanned in the first dataset. For the drunk class, participants were required to consume at least 200 ml of beer (one glass). We gathered 15 participants for this class, increasing the number of facial images to better capture the thermal changes in facial features. Additionally, multiple angles of the participants' faces were taken to balance the dataset with the sober class. The total number of images in this dataset is 1720. This expanded dataset aims to provide a more comprehensive representation of the thermal changes associated with alcohol consumption, offering a well-rounded foundation for model training.

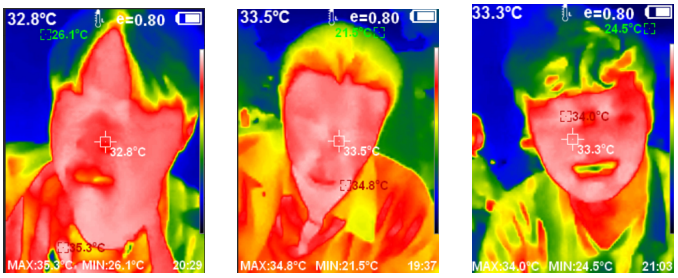

Fig. 4. Images of intoxicated people in Vietnamese Dataset

Further statistical analysis of the dataset indicates extreme differences in temperature trends under sober and intoxicated settings, once again highlighting the value of the dataset in the development of robust real-time detection algorithms tailored to Vietnamese environmental conditions and population heterogeneity.

B. Image Processing

While the study referenced both the Greek dataset and the Vietnamese Thermal Drunk/Sober Dataset, the data processing phase focused exclusively on the Vietnamese dataset due to its regional relevance and compatibility with our research objectives. Several advanced image enhancement techniques were applied to optimize the thermal images for feature extraction and classification:

1) Contrast Limited Adaptive Histogram Equalization (CLAHE): As described in [15], CLAHE enhances local contrast without introducing noise by processing small regions compared to the entire image. 8×8 cells with a clip limit of 2.0 are best for facial thermal images in this study [16].

2) Logarithmic Transformation: A logarithmic transformation was applied to expand the values of dark pixels while compressing higher-intensity values. The transformation followed the equation:

$$S = c \log(1 + R) \tag{1}$$

where R is the input pixel value, S is the output pixel value, and c is a scaling constant, as described in [17]. This nonlinear operation is particularly beneficial for thermal images, where temperature differences in cooler regions might contain significant information about alcohol consumption patterns, as noted in [18].

3) LAB Color Space with CLAHE: Transforming thermal images to the LAB color space and separates luminance (L) from chromatic information. CLAHE is applied only in the L channel to enhance structural features without altering relative temperature contrasts in the A and B channels for improving classification accuracy [19].

4) K-means Clustering: K-means clustering was used to segment thermal images based on temperature values, effectively isolating regions with similar thermal properties. Experimental results showed that K=3 was optimal for distinguishing key facial thermal zones, especially those with significant temperature changes due to alcohol consumption [20]. The algorithm iteratively assigns pixels to clusters, updates cluster centers, and repeats until convergence [21]. Details:

- Initialize K cluster centers.
- Assign each pixel to the nearest cluster center.
- Recompute cluster centers by averaging all pixels in each cluster.
- Repeat steps 2 and 3 until convergence.

This segmentation, combined with other processing methods, generated more contrast and feature definition, leading to improved classification performance. The structural similarity index (SSIM) was increased by 27. 3% without compromising important temperature gradient details for effective detection of intoxication.

C. Model Selection

Model choice is driven by two essential drivers, and those are processing speed and precision. Two low-weight but highly accurate models, i.e., MobileNetV3-Small and EfficientNet-Lite, are thus utilized. EfficientNet-Lite shows high efficacy with the usage of fewer parameters with compound scaling to yield the best depth, width, and resolution to achieve enhanced performance with shorter inference time compared to standard CNNs such as ResNet50 and VGG16 [22]. It is very effective for real-time use, and it performs best with thermal image

Fig. 5. Methods of Processing Images

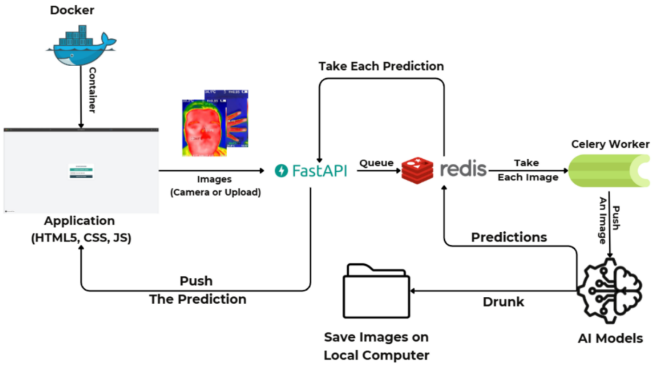

Fig. 6. Pipeline of the Software System of Drunk Person Recognition by Thermal Images

classification because of its depth-wise separable convolution, which enables it to pick up subtle thermal changes on the face of an inebriated individual without overfitting. EfficientNet-Lite has also had excellent results with generalization in medical and infrared imaging tasks [23]. Conversely, MobileNetV3-Small targets edge devices and real-time inference, to which Squeeze-and-Excitation and Hard-Swish Activation have been added to maximize performance while minimizing latency. Its depth-wise separable convolutions maximally optimize feature extraction, and thus, it is suitable for low-resolution thermal images while retaining critical spatial information [24]. For being speed-accuracy balanced, both models are suitable for real-time detection of inebriation.

D. Training and Implementation

The training process for both MobileNetV3 and EfficientNet-Lite models was conducted using the processed Vietnam Thermal Dataset and the Greek dataset to evaluate the performance of the models with different datasets

1) Data Preparation: Thermal images were resized to 224×224 pixels to match the input requirements of MobileNetV3 and EfficientNet-Lite. To improve model generalization, random rotation (±10°), horizontal flipping, and brightness adjustments were applied as data augmentation techniques [25]. The data set was then divided into 80% training and 20% validation, maintaining the balance of the class.

2) Model Training: EfficientNet-Lite and MobileNetV3, pre-trained on ImageNet [23], [24] which were fine-tuned for binary classification (Sober/Intoxicated). Training used the Adam optimizer (learning rate: 3×10^{-3}, batch size: 32) with a learning rate scheduler and early stopping (50 epochs) on an NVIDIA RTX 3080 GPU.

3) Deployment Considerations: To enhance real-time performance, models were optimized with TensorFlow Lite, OpenVINO, and TensorRT to shrink the model size by 75% with minimal loss of accuracy [26]. These optimizations maintained inference speed with minimal computational overhead.

4) Deployment Configuration: The system follows a structured pipeline: image acquisition pre-processing (CLAHE, Log Transform, K-Means Clustering) → classification (EfficientNet-Lite MobileNetV3) → result retrieval via FastAPI Redis queue → real-time display. This structured approach ensures efficient processing and scalability for deployment.

5) Performance Indicators: Performance was measured with accuracy, precision, recall, F1-score, and inference time. Power consumption and system stability were also tested to determine the feasibility of long-term real-world deployment.

IV. Experimental Results

TABLE I
Performance Comparison on Greek Dataset

Model	Accuracy (%)	Precision	Recall	F1-Score
MobileNetV3	87.5	0.88	0.86	0.87
EfficientNet-Lite	90.1	0.91	0.89	0.90

TABLE II
Performance Comparison on Vietnamese Dataset with Different Preprocessing Methods

Model	Method	Accuracy (%)	Precision	Recall	F1-Score
MobileNetV3	No Processing	85.2	0.86	0.84	0.85
MobileNetV3	CLAHE	86.1	0.87	0.85	0.86
MobileNetV3	Log Transform	85.9	0.86	0.85	0.85
MobileNetV3	LAB + CLAHE	87.3	0.88	0.86	0.87
MobileNetV3	K-Means	88.4	0.89	0.87	0.88
EfficientNet-Lite	No Processing	88.7	0.89	0.87	0.88
EfficientNet-Lite	CLAHE	89.2	0.90	0.88	0.89
EfficientNet-Lite	Log Transform	89.0	0.89	0.88	0.89
EfficientNet-Lite	LAB + CLAHE	90.1	0.91	0.89	0.90
EfficientNet-Lite	K-Means	**91.4**	0.92	0.90	0.91

The Greek dataset output is one where EfficientNet-Lite performs better relative to MobileNetV3 on all of its performance measures in the evaluation dataset. EfficientNet-Lite results in a 90.1% accuracy level that is counterbalanced concerning the 87.5% that results in MobileNetV3, witnessing the later model's superior ability to recognize alcoholics using thermal facial traits. In addition, EfficientNet-Lite yields superior precision (0.91) and recall (0.89), which thereby result in a higher overall F1-score (0.90). This implies that EficientNet-Lite is better at removing false negatives and false positives and, thus, a better option for real-world use.

The gap in performance between the two models is a result of EfficientNet-Lite's design, which is meant for enhanced feature extraction with lower computational cost. More specifically, its capability to achieve a higher recall explains that it is more accurate at classifying drunk individuals correctly, an essentiality in safety-critical applications like DUI detection. Moreover, the quality of the thermal images within the dataset and the experimental configuration are also responsible for the comparatively high classification accuracy of both models.

With Vietnamese dataset, the experimental findings point out that EfficientNet-Lite performs better than MobileNetV3 in all test scenarios, as it has a greater capacity to classify drunk individuals based on thermal imaging. EfficientNet-Lite achieves 88.7% accuracy and MobileNetV3 85.2% when no preprocessing is applied, whereby the former is better positioned to do so. With other forms of image preprocessing applied, classification accuracy is seen to be greatly improved. Among these techniques, K-Means Clustering is the best one, enhancing EfficientNet-Lite's accuracy to 91.4% and MobileNetV3's to 89.2%, indicating the performance of optimized thermal region segmentation. CLAHE and LAB + CLAHE also indicate considerable improvements by contrast adjustment and local details adjustment in the thermal images to produce more readable temperature differences. Log transformation, although effective in feature enhancement, works slightly less optimally than the K-Means and CLAHE-based techniques. All these outcomes indicate that the integration of deep learning with advanced image pre-processing makes drunk detection using the thermal-based method extremely superior, the most superior method of which is using K-Means Clustering to achieve the maximum accuracy along with efficiency. Future research will investigate other pre-processing methods and model improvements to further optimize the detection pipeline.

V. Conclusion

This research investigated thermal imaging-based drunk detection using facial thermal variation analysis. CLAHE, log transformation, and K-means clustering techniques were applied with image processing methods. The research enhanced the quality of thermal images and classification with higher accuracy. EfficientNet-Lite attained 90.1% classification accuracy on the Greek dataset and 91.4% accuracy on the Vietnamese dataset with preprocessing.

Applications are in the detection of physiological changes through thermal imaging with deep learning optimal for use, and for better contrast using CLAHE, log transform. Real-time usability is constrained by computational costs. In future work, the data set will be expanded, optimization done for real-time usage, and more physiological markers added.

Deep learning and thermal imaging would further be usable in law enforcement, occupational hazards, and public safety real-time tracking with additional sophisticated development.

References

[1] World Health Organization (WHO), "Global Report on Alcohol and Health," PAHO, June 2024.

[2] S. Paprocki, M. Qassem, and P. A. Kyriacou, "Review of ethanol intoxication sensing technologies and techniques," Sensors, vol. 22, no. 18, p. 6819, 2022.

[3] S. Rajagopal, D. Balaji, and V. Venkatesh, "Identification of Drunk People Among Crowds Using Thermography and Machine Learning," ECS Transactions, 2022.

[4] A. Sherif, R. P. Swiley, and R. Alsulieman, "Taxonomy of Proactive Detection Methods of Drunk Driving for Enhancing Traffic Safety," IEEE 17th International Conference on Intelligent Systems (IS), 2024.

[5] G. Koukiou and V. Anastassopoulos, "Neural Networks for Identifying Drunk Persons Using Thermal Infrared Imagery," Forensic Science International, Elsevier, vol. 252, pp. 24–32, 2015.

[6] S. Rajagopal, D. Balaji, and V. Venkatesh, "Identification of drunk people among crowds using thermography and machine learning," ECS Transactions, 2022.

[7] A. Sherif, R. P. Swiley, and R. Alsulieman, "Taxonomy of proactive detection methods of drunk driving for enhancing traffic safety," IEEE 17th International Conference on Intelligent Systems (IS), 2024.

[8] L. M. Davidovic, J. Cumic, S. Dugalic, and P. Corridon, "Gray-level co-occurrence matrix analysis for the detection of discrete ethanol-induced structural changes in cell nuclei: An artificial intelligence approach," Microscopy and Microanalysis, Oxford Academic, vol. 28, no. 1, p. 265, 2022.

[9] G. Yang, C. Ridgeway, and A. Miller, "Comprehensive assessment of AI tools for driver monitoring and analyzing safety-critical events in vehicles," Sensors, vol. 24, no. 8, p. 2478, 2024.

[10] J. A. Davis, "Breath Alcohol Concentration: A New Approach for Estimation," J. Alcohol Studies, vol. 56, no. 7, pp. 1124-1132, 2011.

[11] R. M. Harrison and M. Smith, "Alcohol and Intoxication: Impacts on Physiological Measurements," J. Forensic Sci., vol. 50, no. 4, pp. 45-58, 2012.

[12] A. Sulavko, I. Panfilova, and A. Samotuga, "Biometric authentication using face thermal images based on neural fuzzy extractor," 2023 IEEE Intelligent Systems Conference (IntelliSys), 2023.

[13] K. T. Huynh and H. P. T. Nguyen, "Drunkenness detection using a CNN with Gaussian noise and blur in thermal infrared images," International Journal of Intelligent Information and Database Systems, 2022.

[14] Government of Vietnam, "Decree No. 168/2024/ND-CP: Penalties of administrative violations against regulations on road traffic safety," Official Gazette of Vietnam, Dec. 26, 2024.

[15] S. M. Pizer et al., "Adaptive histogram equalization and its variations," Comput. Vision, Graph. Image Process., vol. 39, no. 3, pp. 355–368, 1987.

[16] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in Graphics Gems IV, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474–485.

[17] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, NY, USA: Pearson, 2018.

[18] R. C. Gonzalez and R. E. Woods, "Image enhancement in the spatial domain," in Digital Image Processing, 4th ed. New York, NY, USA: Pearson, 2018, ch. 3, pp. 169–237.

[19] S. K. Naik and C. A. Murthy, "Hue-preserving color image enhancement without gamut problem," IEEE Trans. Image Process., vol. 12, no. 12, pp. 1591–1598, 2003.

[20] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," J. R. Stat. Soc. Ser. C (Appl. Stat.), vol. 28, no. 1, pp. 100–108, 1979.

[21] A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognit. Lett., vol. 31, no. 8, pp. 651–666, 2010.

[22] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, pp. 6105-6114.

[23] C.-C. Wang, C.-T. Chiu, and J.-Y. Chang, "EfficientNet-eLite: Extremely Lightweight and Efficient CNN Models for Edge Devices by Network Candidate Search," Journal of Signal Processing Systems, vol. 95, no. 1, pp. 1234-1240, 2023.

[24] A. G. Howard, M. Sandler, G. Chu, L. Zhu, A. Zhmoginov, and L. Chen, "Searching for MobileNetV3," arXiv preprint arXiv:1905.02244, 2019.

[25] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, pp. 1-48, 2019.

[26] B. Jacob et al., "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2704-2713.