



CNN-Based Smoking Image Classification

Muhammad Abraar Md Khir¹, Nurulhuda Ismail^{1*}, Nabihah binti Ahmad¹, Munirah Ab. Rahman¹, Zarina Tukiran², Noraisah Sudin², and Eddy Irwan Shah Saadon³

¹Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, Johor, Malaysia.

²Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, Johor, Malaysia.

³Center for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Pagoh, 84600, Johor, Malaysia.

KEYWORDS

Smoking Image Classification
Convolution Neural Network
Binary Classifier
Data Augmentation

ABSTRACT

Cigarette smoke exposure is a major public health concern that can harm nearly every organ of the body and cause many diseases. To better regulate smoking in public areas, this project presents a binary image classification system for smoking versus non-smoking images using Convolutional Neural Networks (CNN), designed for robust and scalable environmental monitoring. The system is trained on a dataset of annotated images of smoking and non-smoking individuals. Image preprocessing and data augmentation techniques were applied to enhance model performance and mitigate overfitting. A baseline CNN model was first developed and evaluated, followed by the implementation of transfer learning using three pre-trained models: ResNet50V2, Efficient NetB0, and MobileNetV2. Among these, ResNet50V2 combined with data augmentation achieved the highest accuracy of 92.5%. Comparative analysis showed that data augmentation significantly improved model robustness, while transfer learning optimized efficiency. These findings demonstrate the efficacy of CNNs in identifying smoking-related features within static images, providing a strong foundation for deploying the trained model into real-world, automated smoking detection systems.

ARTICLE HISTORY

Received 23 July 2026
Accepted 6 October 2026
Available online 20 October 2026

© 2026 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Cigarette smoking remains one of the leading factors of global health challenges and the primary mainstay of preventable deaths globally [1]. It is a harmful activity that introduces over 7,000 toxic chemicals into the human body, many of which trigger oxidative stress, DNA damage, and chronic diseases [2]. One of the most prevalent consequences of long-term smoking is lung cancer, primarily non-small cell lung cancer (NSCLC), which accounts for nearly 85% of all lung cancer cases, with squamous cell carcinoma being a prominent subtype linked directly to tobacco exposure. According to the World Health Organization (WHO), lung cancer caused an estimated 1.8 million deaths in 2020, with cigarette smoking being the primary risk factor [3]. Therefore,

substantial strategies to monitor and reduce smoking habits, especially in public smoking environments, are required.

In general, cigarette smoke detection and monitoring rely on chemical sensors or air-quality control systems, which often require costly development and complex infrastructure [4]. Recent advancements in the Internet of Things (IoT) and embedded digital systems have demonstrated significant potential to enable fast, low-cost, and highly viable identification systems. This study focuses on static image classification, in which the system examines a single image and identifies whether smoking activity is present. This technique is different from real-time object detection, that requires localized bounding boxes (e.g., pinpointing small objects such as cigarette butts or smoke plumes) or temporal video analysis. While this study focuses on global image classification as a baseline for developing a robust model [6], future work is

*Corresponding author:

E-mail address: Nurulhuda binti Ismail <nhuda@uthm.edu.my>.

<https://doi.org/10.56532/mjsat.v6iS2.807>

2785-8901/ © 2026 The Authors. Published by Penteract Technology.

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

required to extend this technique to dynamic, real-time object detection that can be applied to real-world public surveillance and monitoring applications [5][7].

Exposure to particulate matter from cigarette smoking induces oxidative stress, primarily by disrupting the balance between oxidants and antioxidants in the lungs. Long-term exposure to particulate matter can produce multiple reactive oxygen species (ROS) that can harm pulmonary cells and tissues and impair cellular functions [7].

Commercial smoking detectors are very expensive and cover only enclosed spaces, indicating they are not suitable for broader environmental applications [8]. Deep Learning remains relevant in image processing for grid-like data applications, while preserving its specialised ability to learn image classification from data such as edges, patterns, and labels [9]. One limitation of traditional machine learning algorithms, such as logistic regression, is the need for manual feature extraction, which requires careful selection to identify meaningful characteristics. For example, considering countless attributes of our images and selecting the most meaningful ones can make this process very time-consuming. However, deep neural networks, which use deep learning, can handle large image datasets, as seen with Convolutional Neural Networks (CNNs) in applications such as medical diagnosis, autonomous vehicles, surveillance, and image analysis with convolutional filters. It is able to identify regions of interest in images by early layers recognize simple patterns, while deeper layers identify complex structures like faces and objects [10]. Generally, common target-detection algorithms are not ideal for small targets, and detection speed could be improved. The difficulty of determining the position relationship between human and cigarette butts using the algorithm, as cigarette butts are small, makes it inaccurate in identifying positioning [11].

Thus, this research proposes a smoking image classification system based on CNNs and deep learning to classify images as smoking or non-smoking. It is a scalable, software-based classifier that provides a strong baseline model for a future automated smoking detection system. It has the potential to be improved or extended in the future to monitor smoking activities and contribute to lung cancer prevention efforts.

2. METHODOLOGY

This section explains the development of a binary image classifier using a CNN. This study begins with the dataset collection and preprocessing, followed by CNN model design, training, and evaluation. The system uses the 'Smoking – CNN' dataset. This dataset consists of 224×224 RGB pixel images, divided into smoking and non-smoking classes. The study utilizes a total of 2,010 original unique images across all partitions. For the training set, the system initially utilizes 1,610 unique images (805 smoking and 805 non-smoking). Once data augmentation is applied, this training dataset size is doubled to 3,220 total images (1,610 per class), so that the model generalization can be improved. The validation set and testing set comprise 160 images (80 per class) and 240 images (120 per class), respectively.

The custom-built CNN model serves as a baseline, while a pre-trained architecture using transfer learning is applied to improve reputation and enhance prediction accuracy. Selecting a CNN as the core model is due to its ability to extract features for image-based classification. It consists of multiple layers: input, hidden, and output.

The development of the baseline model began with data collection and organization, in which labelled images of 'smoking' and 'non-smoking' classes were partitioned into training, validation, and testing. Images were preprocessed to improve clarity and ensure a suitable format. Then, data augmentation was used to help the model generalize better. By applying transformations such as rotations and colour adjustments, the size of the training dataset increased, significantly improving the model's performance.

Transfer learning was then implemented using pre-trained models, including ResNet50V2, EfficientNetB0, and MobileNetV2. These architectures were retained for their deep feature-extraction capabilities and compared using validation accuracy and loss metrics. The results from both the CNN and transfer learning will be analyzed to determine the model performance. The whole process is illustrated in Figure 1.

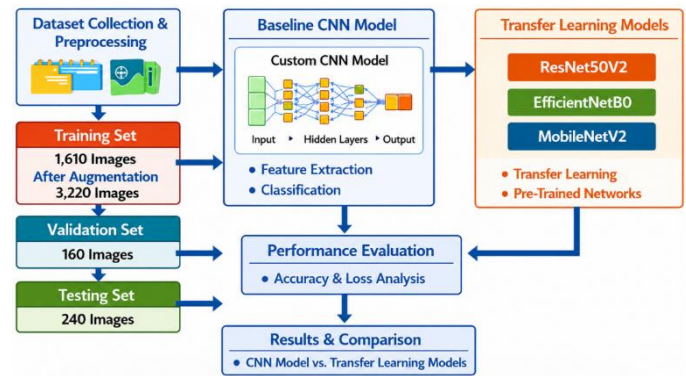


Fig. 1. The flowchart of Smoking Image Classification based on CNN

2.1 Data Collection and Preprocessing

The dataset used in this project was obtained from the publicly available GitHub repository "BesherTaye/Smoking-CNN", which is adapted from a published study that addresses smoking behavior classification through visual data analysis. The dataset comprises labeled images categorized into two classes: smoking and non-smoking. It is partitioned into three main subsets: training, validation, and testing.

In this study, the dataset used 2,010 labelled images, which were partitioned into training, validation, and testing sets. The original training subnet consisted of 1,610 unique images (805 smoking and 805 non-smoking). Images for the validation and test sets were drawn from an additional pool of 400 images. These images were divided into a validation set of 160 images (80 smoking and 80 non-smoking), while the remaining 240 images (120 smoking and 120 non-smoking) were manually reviewed and verified for the test set. The original 400 unlabeled test images in the repository were completely

excluded from this study due to the absence of verified ground-truth labels.

All input images were resized to 224×224 pixels to match the expected input dimensions of most CNN architectures. Furthermore, rescaling was performed by normalizing pixel values to the range of, which is essential for stabilizing the learning process during training.

In addition, data augmentation techniques were employed on the training dataset using TensorFlow's ImageDataGenerator. The augmentation involved operations such as horizontal flipping, rotation, zooming, and shearing to artificially increase the diversity of the training samples. This process effectively helped the model learn more robust features and adapt better to unseen test images. Table 1 summarizes the distribution of the dataset across the training, validation, and testing subsets, highlighting the increase in training samples from 1,610 to 3,220 images following the application of data augmentation techniques. The specific transformation configurations and ranges used during data augmentation, including rotation, shearing, and zooming, are detailed in Table 2.

Table 1. Datasets before and after augmentation

Subnet	Before Augmentation	After Augmentation
Training Set	1,610 images (805 per class)	3,220 images (1,610 per class)
Validation Set	160 images (80 per class)	160 images (80 images per class)
Test Set	240 images (120 per class)	240 images (120 per class)

Table 2. Data augmentation parameter

Augmentation class	Parameter
Rotation	Range ~ 20 degrees
Shear range	0.2
Zoom range	0.2
Width shift range	0.2
Height shift range	0.2
Horizontal flip	True
Vertical flip	True
Re-scale	1/255

2.2 CNN Architecture

The CNN architecture developed in this study is designed to classify image data into two categories: smoking and non-smoking. CNNs are a class of deep learning algorithms that have demonstrated exceptional performance in visual pattern recognition tasks by automatically learning spatial hierarchies of features from input images. The architecture implemented in this project comprises a sequence of layers designed to progressively extract relevant features from the input images.

The data fed into the CNN model consists of RGB images resized to a fixed dimension of 224×224 pixels. This size is chosen to ensure compatibility with pre-trained feature extraction models and to balance computational efficiency.

The initial layers of the model consist of multiple convolutional layers, each equipped with a small number of filters. These filters scan across the image and detect local

features such as edges, corners, and textures. Rectified Linear Unit (ReLU) is used as the activation function to introduce non-linearity into the model and to allow for the learning of complex patterns

The objective of max pooling layers is to downsample the feature maps. Pooling reduces the spatial dimensions of the data, helping to decrease overfitting while maintaining the dimension. The outputs of the convolutional and pooling operations are flattened into a one-dimensional vector, which is then fed into fully connected layers for final classification. The fully connected (dense) layers process the flattened vector and learn complex relationships between features. A final dense layer with a sigmoid activation function produces a binary classification between smoking and non-smoking.

Using a binary classifier, smoking and non-smoking input images are fed through the same CNN feature-extraction pipeline. These two behaviours are distinguished entirely by the final Sigmoid output layer, which maps the extracted features to either a smoking or non-smoking classification. The process starts with the input data, which is fed into the CNN model. RGB images are resized to a fixed dimension of 224×224 pixels. This size is chosen to ensure compatibility with pre-trained feature extraction models and to balance computational efficiency.

2.3 Transfer Learning Approach

Transfer learning was adopted in this study to enhance the model's ability to classify images as smoking or non-smoking, particularly with limited datasets. This approach leverages pre-trained convolutional base models trained on large datasets (ImageNet) to extract generalized features and builds custom classification layers on top.

To enhance performance and reduce training time, the project also explores transfer learning using pre-trained models from TensorFlow Hub, including ResNet50V2, EfficientNetB0, and MobileNetV2. Each transfer learning model was integrated by using its convolutional base as a frozen feature extractor (with `include_top=False`) and adding custom fully connected layers for binary classification. The model's architecture is defined using TensorFlow's Functional API.

The fully connected (dense) layers process the flattened vector and learn complex relationships between features. A final dense layer with a sigmoid activation function produces a binary classification between smoking and non-smoking. To enhance performance and reduce training time, the project also explores transfer learning using pre-trained models from TensorFlow Hub, including ResNet50V2, EfficientNetB0, and MobileNetV2. These models are used as feature extractors, can overcome the limitations of a relatively small dataset, and are trained on large-scale datasets such as ImageNet to improve classification accuracy on the smoking dataset. Table 3 summarizes the architectural layout of the baseline CNN model following data augmentation, detailing the specific filter configurations, output shapes, and the total number of trainable parameters for each layer.

Table 3. Summary of Baseline CNN Model with augmentation

Layer	No of filters	Filter size	Pool size	Output shape	Trainable parameters
1 st convolutional layer	10	3x3	-	(222, 222, 10)	280
1 st Max_pool with ReLU activation	-	-	2x2	(111, 111, 10)	0
2 nd convolutional layer	10	3x3	-	(109, 109, 10)	910
2 nd Max_pool with ReLU activation	-	-	2x2	(54, 54, 10)	0
3 rd convolutional layer	10	3x3	-	(52, 52, 10)	910
3 rd Max_pool with ReLU activation	-	-	2x2	(26, 26, 10)	0
Flatten	-	-	-	(6760)	0
Dense	-	-	-	(1)	6,761
Total params: 26,585 (103.85 KB)					
Trainable params: 8,861 (34.61 KB)					
Non-trainable params: 0 (0.00 B)					

3. RESULTS AND DISCUSSIONS

This section discusses the results of the cigarette-smoking classification system developed using CNN and transfer learning. The results are evaluated based on model performance metrics, including accuracy, precision, recall, F1-score, and specificity. Both baseline CNNs and transfer learning models, before and after data augmentation, are compared to determine the most effective model configuration for cigarette smoke classification.

3.1 CNN Baseline Model Performance

The first model developed was a custom CNN without data augmentation. This model served as a baseline to understand the potential of a manually designed network. However, the model experienced unstable training behavior with moderate performance, resulting in a validation accuracy of around 65%. To enhance the model's generalization ability, data augmentation was applied, effectively doubling the number of training samples and exposing the model to various transformations. The augmented custom CNN achieved a validation accuracy of 62.08%. The accuracy slightly drops, indicating a performance anomaly due to a few factors, as discussed below.

As detailed in Table 3, the custom baseline CNN model is very lightweight, containing only three conventional layers (10 filters each) and 8,861 trainable parameters. In addition, data

augmentation increases geometric complexity and introduces photometric changes (rotation, shearing, zooming, shifting, flipping). These changes are too complex for the custom CNN model, which causes underfitting as it has a limited learning capacity. The results suggest that while the baseline CNN struggled initially, data augmentation significantly enhanced its ability to classify images more reliably.

3.2 Transfer Learning Model Performance

Transfer learning was then explored using three pre-trained models: ResNet50V2, Efficient NetB0, and MobileNet V2. All models were first evaluated on the original dataset before being evaluated on the augmented dataset.

3.2.1 Without Data Augmentation

Among the three models, ResNet50V2 outperformed the others with consistent validation accuracy 87.90%, and a high F1-score (0.879) demonstrating good generalization. It can effectively differentiate between 'smoking' and 'non-smoking' categories. MobileNet V2 achieved reasonable results with 85% accuracy, while Efficient NetB0 struggled to converge, maintaining around 50% accuracy, indicating it was not learning effectively. The result is illustrated in Table 4. The performance of a classification model is evaluated by comparing actual values against predicted values using a 2x2 confusion matrix, as shown in Figure 2. The results are divided into 4 categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). ResNet50V2 outperforms, as it shows a clear diagonal line (high TP and TN), indicating that it can identify both classes (smoking and non-smoking) very well.

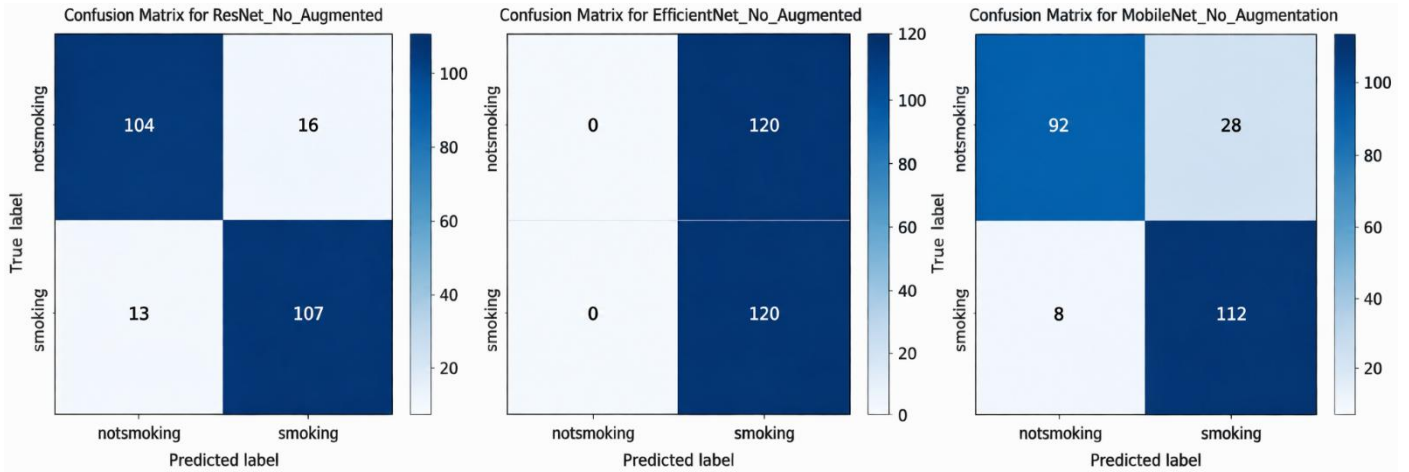
3.2.2 With Data Augmentation

With data augmentation, ResNet50V2 shows a dramatic performance increase, peaking at 92.5% validation accuracy, indicating excellent generalization across smoking and non-smoking images, as shown in Table 5. MobileNet V2 remained stable and accurate, which also benefited from the augmented dataset. However, Efficient NetB0 again failed to improve, suggesting that it may require more advanced fine-tuning or that it is not well-suited to this dataset. These results show that, with data augmentation, ResNet50V2 is the most suitable model for classifying smoking versus non-smoking images. It provides the best balance of accuracy, reliability, and training stability.

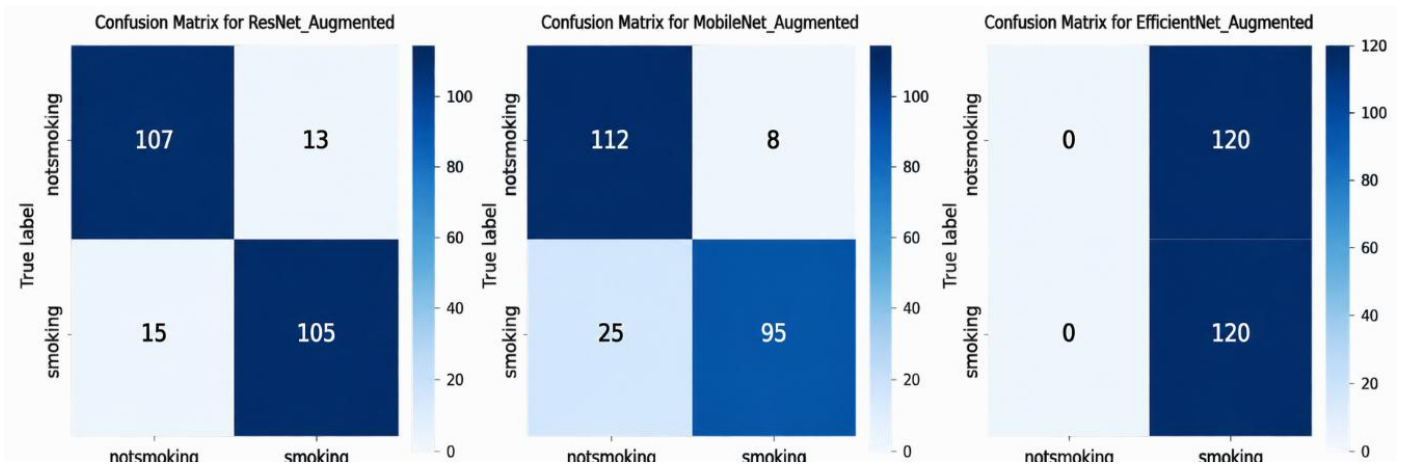
In general, data augmentation improved the performance of machine learning models. By increasing the variety and amount of training data, augmentation improved model accuracy, precision, recall, and F1 scores across the board. This technique also led to better generalization, as evidenced by the reduction in false negative rates (FNR) and false positive rates (FPR). Figure 3 illustrates the confusion matrix of CNN models with data augmentation. ResNet50V2 correctly identified 107 as non-smoking and 105 as smoking. On the other side, Efficient NetB0's matrix is entirely blank as it predicted 'smoking' for every single sample, resulting in 120 FP and 0 TN.

Table 4. Comparison between CNN Models without Data Augmentation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score	Specificity	False Positive Rate (FPR)	False Negative Rate (FNR)
ResNet50V2	87.9	87.9	87.9	0.87914	0.86667	0.13333	0.10833
Efficient NetB0	50	25	50	0.3	0	1	0
MobileNet V2	85	86	85	0.84895	0.76667	0.23333	0.06667

**Fig. 2.** Evaluation matrices of CNN models without data augmentation**Table 5.** Comparison between CNN Models with Data Augmentation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score	Specificity	False Positive Rate (FPR)	False Negative Rate (FNR)
ResNet50V2	92.5	92.5	92.5	0.925	0.925	0.10833	0.12500
Efficient NetB0	50	25	50	0.3	0	1	0
MobileNet V2	86.25	86.99	86.25	0.86180	0.93333	0.06667	0.20833

**Fig. 3.** Evaluation matrices of CNN models with augmentation

3.2.3 Performance Comparison (Without and With Data Augmentation)

Based on the results in Tables 4 and 5, ResNetV2 performs better than the other three transfer learning models, achieving consistent validation accuracies of 87.90% (without augmentation) and 92.50% (with augmentation), illustrating excellent generalization. In addition, MobileNetV2 also shows high performance, achieving 85.00% accuracy (without augmentation) and 86.25% (with augmentation).

In contrast, EfficientNetB0 underperformed in both conditions, maintaining a flat accuracy of approximately 50.00%. It predicts every image as the 'smoking' class, resulting in 120 false positives and 0 true negatives, as illustrated in Figure 3. One of the three primary reasons for this failure is a scaling mismatch in the input preprocessing. EfficientNetB0 is designed to automatically scale the pixels of raw images before processing. However, in this study, the images were already manually scaled during preprocessing. As a result, the images were double-scaled, making the pixel values become too small and approach zero. The amount of useful information that is extracted by the model is reduced, and the learning process is less effective.

EfficientNetB0 is more sensitive to the learning rate compared to other CNN models. The learning rate used in this study is suitable for ResNet50V2 and MobileNetV2, but not for EfficientNetB0, because it was too high. Consequently, the model's pre-trained knowledge was disrupted early in training, preventing it from learning the new task effectively.

4. CONCLUSION

This study successfully demonstrates the implementation of a binary image classifier for smoking and non-smoking images using a custom CNN architecture and transfer learning. The development process includes data collection, image preprocessing, augmentation, CNN model architecture, transfer learning implementation, and evaluation using several performance metrics. The custom CNN model was developed and trained using the original dataset, achieving moderate performance due to limited data diversity. To address this limitation, data augmentation was applied, resulting in significant improvements in classification accuracy and generalization capability. The augmented CNN model outperformed the non-augmented version, achieving higher validation accuracy and a better F1-score, thereby highlighting the importance of image diversity in enhancing deep learning model performance.

To further enhance the model's performance, transfer learning was applied using three pre-trained architectures: ResNet50V2, Efficient NetB0, and MobileNet V2. Results revealed that ResNet50V2 was the most effective model, particularly with data augmentation, achieving a validation accuracy of 92.5%. In contrast, Efficient NetB0 underperformed across both augmented and non-augmented scenarios, suggesting that its architecture may not align well with the dataset's nature.

In conclusion, the combination of ResNet50V2 and data augmentation was determined to be the optimal approach for cigarette smoke image classification. This work lays the

foundation for future development of real-time smoke detection systems and potential deployment in scalable public surveillance and monitoring applications.

REFERENCES

- [1] Y. He, G. T. Fong, K. M. Cummings, A. Hyland, R. Freitas-lemos, and C. Shang, "The economic factors that determine smoking abstinence : a survival and mediation analysis," *Int. J. Drug Policy*, vol. 142, no. May, p. 104852, 2025, doi: <https://doi.org/10.1016/j.drugpo.2025.104852>.
- [2] Y.-S. Seo, J.-M. Park, J.-H. Kim, and M.-Y. Lee, "Cigarette Smoke-Induced Reactive Oxygen Species Formation: A Concise Review," *Antioxidants*, vol. 12, no. 9, Art. no. 1732, Sep. 2023, doi: <https://doi.org/10.3390/antiox12091732>.
- [3] W. Ye, W. Lu, C. Xu, and C. Ji, "Long-Term Changes in Premature Death of Lung Cancer in a Developed Region of China : A Population-Based Study From 1973 to 2019," pp. 1–18, 2019, doi: <https://doi.org/10.21203/rs.3.rs-509107/v1>.
- [4] Z. Wu, H. Zhang, W. Sun, N. Lu, M. Yan, Y. Wu, Z. Hua, and S. Fan, "Development of a Low-Cost Portable Electronic Nose for Cigarette Brands Identification," *Sensors*, vol. 20, no. 15, Art. no. 4239, Jul. 2020, doi: <https://doi.org/10.3390/s20154239>.
- [5] M. M. Engelhard, J. A. Oliver, R. Henao, M. Hallyburton, and L. E. Carin, "Identifying Smoking Environments From Images of Daily Life With Deep Learning," vol. 2, no. 8, pp. 1–13, 2019, doi: <https://doi.org/10.1001/jamanetworkopen.2019.7939>.
- [6] Ali Khan, Somaiya Khan, Bilal Hassan, Zhonglong Zheng, "CCN-Based Smoker Classification and Detection in Smart City Application," *Sensors*, 22(3), 892, 2022, doi: <https://doi.org/10.3390/s22030892>.
- [7] J. Yoo, Y. Lee, Y. Park, J. Lee, J. Y. Choi, H. Lee, and J. U. Lim, "Update in Association between Lung Cancer and Air Pollution," *Tuberc. Respir. Dis.*, vol. 88, no. 2, pp. 228–236, Apr. 2025, doi: <https://doi.org/10.4046/trd.2024.0092>.
- [8] A. Pant, J. Chen, A. Sivasubramanian, A. Tulsian, and R. Mourad, "Low-Cost Vaping Detectors to Mitigate Teenage Vaping," 2020 IEEE MIT Undergrad. Res. Technol. Conf., no. February, pp. 1–4, 2020, doi: <https://doi.org/10.1109/URTC51696.2020.9668868>.
- [9] A. I. Khan and S. Al-habsi, "Machine Learning in Computer Vision," *Procedia Computer Science*, vol. 167, no. 2020, pp. 1444–1451, 2020, doi: <https://doi.org/10.1016/j.procs.2020.03.355>.
- [10] A. Khan, S. Khan, B. Hassan, and Z. Zheng, "CNN-Based Smoker Classification and Detection in Smart City Application," *Sensors*, vol. 22, no. 3, p. 892, Jan. 2022, doi: <https://doi.org/10.3390/s22030892>.
- [11] C. Jiang, "An Intelligent Detection Approach for Smoking Behavior," *Int. J. Cogn. Inform. Nat. Intell.*, vol. 17, no. 1, pp. 1–18, Jun. 2023, doi: <https://doi.org/10.4018/IJCINI.324115>.