Lung Cancer Classification Based on Ensembling EfficientNet Using Histopathology Images

1st Akif Rachmat Hidayah
, *School of Computing*
Telkom University
Bandung, Indonesia
akifrh@student.telkomuniversity.ac.id

2nd Untari Novia Wisesty*
School of Computing
Telkom University
Bandung, Indonesia
untarinw@telkomuniversity.ac.id
*corresponding author

***Abstract*—Lung cancer is a leading cause of cancer-related deaths, and accurate, early diagnosis is critical for effective treatment. Histopathological analysis is a standard diagnostic approach but requires significant expertise and time. This study aims to improve lung cancer classification through an ensemble of EfficientNetV2 models (B0-B3) applied to histopathological images. EfficientNetV2 was chosen for its scalability and strong performance in image classification tasks. Data augmentation was used to enhance robustness, simulating variability in histopathological slides, while transfer learning from ImageNet pre-trained models enabled faster convergence with limited data. The models were trained on the LC25000 dataset, containing augmented images, and evaluated individually and in ensemble configurations. Grad-CAM provided interpretability, generating heatmaps that highlight model focus, aiding in understanding decision-making. Results showed that individual EfficientNetV2 models achieved near-perfect accuracy, with the ensemble approach further improving performance. Ensemble models, particularly those using hard voting, achieved up to 100% accuracy, precision, and recall, underscoring the effectiveness of combined predictions. However, the high accuracy may be partially due to the dataset's limited unique images, as repeated patterns in augmented data might inflate performance. Future work will test the ensemble on larger, more diverse datasets to validate generalizability. These findings demonstrate the potential of EfficientNetV2 ensembles in lung cancer diagnostics, paving the way for reliable, interpretable AI-based pathology tools in clinical settings.**

Keywords—Lung cancer classification, EfficientNetV2, histopathological images, ensemble learning, Grad-CAM

I. INTRODUCTION

Cancer is an uncontrolled proliferation disease caused by mutated cells evolving under natural selection [1]. These cells can originate in almost any part of the human body, composed of trillions of cells. Lung cancer arises when such abnormal cell growth occurs in the lungs [2]. Given its high prevalence and mortality rate, early and accurate detection of lung cancer is critical to improving patient outcomes and reducing the overall burden on healthcare systems.

Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a transformative technology in medical image analysis, automating complex diagnostic tasks and reducing the workload on healthcare providers. CNNs have shown remarkable success in various cancer detection applications by analyzing histopathology images, which are essential for understanding the cellular structure of cancerous tissues. Among CNN architectures, EfficientNet, introduced by Tan and Le (2019) [3], is recognized for its balance between high accuracy and computational efficiency, achieved through compound scaling that optimally adjusts model depth, width, and resolution. This efficiency makes EfficientNet a suitable choice for medical imaging tasks where computational resources may be limited.

Researchers have employed deep learning and machine learning techniques in previous studies to analyze medical data, particularly histopathology images, for cancer detection. Hatuwal et al. [4] used a CNN architecture with three convolutional layers and one fully connected layer for lung cancer detection based on histopathology images. While this approach achieved high accuracy, the model's simplicity could limit its ability to capture complex patterns in cancerous tissues. Anjum et al. [5] expanded this research by comparing various EfficientNet variants for the classification of lung cancer using histopathology images, ultimately identifying EfficientNetB2 as the most effective model. In a separate study not limited to lung cancer, Kallipolitis et al. [6] explored an ensemble model based on EfficientNet for cancer image classification, incorporating Grad-CAM to improve the interpretability of the ensemble's classification decisions. However, this study applied limited image augmentation, which may have restricted the model's generalizability on diverse histopathology images. Abioye et al. [7] evaluated EfficientNetV2 models on benign breast cancer images, with EfficientNetV2L showing the highest accuracy. However, their study was limited to benign cases at 40x magnification, which may limit generalizability to malignant cases or other magnifications. Additionally, the augmentation was limited, potentially restricting the model's adaptability to diverse histopathology images.

This study addresses the gaps identified in previous research by leveraging an ensemble of EfficientNetV2 variants for lung cancer classification using histopathology images. By combining EfficientNetV2 variants and applying comprehensive image augmentation, this study seeks to enhance model robustness, accuracy, and generalization specifically for lung cancer detection. Additionally, Grad-CAM will improve interpretability, providing insights into the model's decision-making process. This research aims to establish a robust lung cancer diagnosis framework through these techniques, enhancing histopathological image analysis's accuracy and reliability and ultimately contributing to improved diagnostic workflows and patient care strategies.

II. RELATED WORK

Lung cancer histopathology image classification using artificial intelligence has garnered significant research interest in recent years, with deep learning techniques showing potential to enhance diagnostic accuracy and efficiency. In 2020, Hatuwal and Thapa [4] developed a CNN model with

979-8-3315-0649-0/24/$31.00 ©2024 IEEE

three convolutional layers and one fully connected layer, specifically for lung cancer classification on histopathology images. Data preprocessing steps included resizing all images to a uniform aspect ratio (1:1) with a final resolution of 180x180 pixels, pixel normalization to a (0,1) range, and data augmentation techniques such as horizontal and vertical flipping and zooming within a 0.2 range. The model achieved a training accuracy of 96.11% and a validation accuracy of 97.2%, though the relatively simple architecture may limit its ability to capture complex features in lung cancer histopathology.

In 2023, Anjum et al. [5] advanced lung cancer classification by developing and comparing various EfficientNet architectures for histopathology image analysis. Before training, data augmentation was applied using three techniques: random rotation between 25 degrees to the right and 25 degrees to the left, the addition of random noise, and horizontal flipping. Image dimensions were adjusted according to the specific EfficientNet variant used. Classification accuracies for models EfficientNet-B0 through B7 were recorded as follows: B0 (95.87%), B1 (96.26%), B2 (97.24%), B3 (95.63%), B4 (96.83%), B5 (94.31%), B6 (93.76%), and B7 (95.59%), with EfficientNet-B2 achieving the highest accuracy.

In 2021, Kallipolitis et al. [6] employed an ensemble model based on EfficientNet variants for cancer image classification. Two ensemble groups were selected: EfficientNet-B0, B1, and B2, and EfficientNet-B1, B2, and B3. Grad-CAM interpretability techniques highlighted visual patterns responsible for each class prediction, enhancing model transparency. The classification accuracies achieved by the ensemble models were 99.25% for breast cancer and 99.46% for colon cancer with the EfficientNet-B0 to B2 ensemble, and 98.55% for breast cancer and 98.56% for colon cancer with the EfficientNet-B1 to B3 ensemble.

In 2024, Abioye et al. [7] evaluated the performance of various EfficientNetV2 models for classifying benign breast cancer histopathology images. Using the BreaKHis dataset at 40x magnification, they compared EfficientNetV2 variants, including B0, B1, B2, B3, S, M, and L. Among these, EfficientNetV2L achieved the highest performance, with an accuracy of 97% and precision, recall, and F1-score all reaching 0.97. The study highlighted the robustness of EfficientNetV2L in classifying benign cases; however, the authors noted that the model's generalizability might be limited due to the focus on benign images and the use of only a single magnification level. Data augmentation was applied to mitigate overfitting, though the scope of augmentation techniques was somewhat limited, potentially restricting adaptability to more diverse histopathology images.

Although various studies have applied deep learning and ensemble techniques to cancer classification, there remains a significant gap in research focused on ensembling EfficientNet models specifically for lung cancer histopathology images. Existing studies have either employed individual EfficientNet variants or ensemble methods for other types of cancer, but none have fully explored the benefits of an EfficientNetV2 ensemble tailored for the unique challenges of lung cancer detection. This research aims to bridge this gap by developing an ensemble of EfficientNetV2 variants optimized for lung cancer classification. Through data augmentation and the use of Grad-CAM for enhanced interpretability, this study seeks to improve model robustness, and transparency, providing a valuable tool for pathologists in diagnosing lung cancer and ultimately contributing to better patient care.

III. Materials And Method

This study employs a structured methodology to classify lung cancer subtypes from histopathological images. Key steps include dataset selection, data augmentation to enhance sample diversity, modification and fine-tuning of EfficientNetV2 models, and robust training strategies. Explainability is addressed through Grad-CAM, and an ensemble approach using soft, hard, and weighted voting further strengthens classification accuracy.

A. Dataset

The dataset used in this study is the Lung and Colon Cancer Histopathological Image LC25000 image dataset [8] containing 25,000 images across five classes, each with 5,000 images. The images are in JPEG format and have a resolution of 768 x 768 pixels. Since the study focuses on lung cancer, only three classes are used: Lung benign tissue, Lung adenocarcinoma, and Lung squamous cell carcinoma, totaling 15,000 images.

However, upon inspection, it was identified that the dataset contains some duplicate images, which could potentially affect the performance and generalization of the model [9]. These duplicates were removed to ensure a more reliable dataset for training and evaluation. After removing duplicates, the images were split into training (60%), validation (20%), and test sets (20%). TABLE I shows the detailed breakdown of the data split.

TABLE I. DISTRIBUTION OF DATASET ACROSS TRAINING, VALIDATION, AND TEST SETS

Class	Training	Validation	Test
Lung adenocarcinoma	2,836	945	946
Lung benign tissue	2,846	949	949
Lung squamous cell carcinoma	2,834	945	945

Fig. 1 shows histopathology images of lung tissues are displayed to illustrate the three classes involved in this study: (a) lung adenocarcinoma, characterized by glandular structures; (b) benign lung tissue, showing normal cell morphology; and (c) lung squamous cell carcinoma, distinguished by the presence of abnormal squamous cells [5]. These images serve as a visual reference for the classification task undertaken in this research

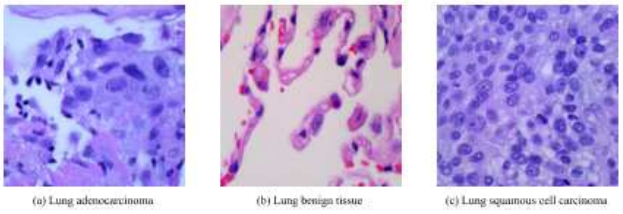

Fig. 1. Histopathology images of lung tissue showing different classifications. (a) Lung adenocarcinoma, (b) Lung benign tissue, and (c) Lung squamous cell carcinoma.

B. Data Augmentation

Data augmentation is a crucial technique used in training deep learning models to enhance the diversity of the training dataset and improve model generalization [10]. By applying transformations such as rotation, width shift, shear, zoom, vertical flipping, and brightness adjustment, the model is exposed to a wider range of variations, which helps it learn more robust and invariant features. These transformations prevent the model from overfitting to specific patterns in the training data, thereby enhancing its ability to generalize to new, unseen data [11], [12]. Random noise is added to increase variability and resilience to minor perturbations further.

For this study, all images are resized to a standardized size of 180x180 pixels to ensure consistency in model input dimensions. This resizing facilitates efficient processing while balancing computational load and image quality. The detailed parameters for each augmentation technique are provided in TABLE II.

TABLE II. DATA AUGMENTATION TECHNIQUES

Technique	Values
Rotation	25°
Width Shift	0.2
Shear	0.2
Zoom	0.2
Vertical Flip	True
Fill Mode	nearest
Brightness Adjustment	[0.8, 1.2]
Preprocessing Function	Add_random_noise

C. Model Architecture

EfficientNetV2 [13] variants, specifically EfficientNetV2-B0 through EfficientNetV2-B3, are used as the primary models for classification in this study. EfficientNetV2 employs a compound scaling approach to balance model depth, width, and input resolution, achieving high accuracy with reduced computational requirements. This architecture also integrates fused MBConv layers, combining depthwise separable and standard convolutions to enhance training speed and efficiency. Pre-trained on the ImageNet [14] dataset, each EfficientNetV2 variant provides robust feature extraction capabilities, which are fine-tuned for this specific classification task through transfer learning. Custom layers, as shown in TABLE III, are added on top of the EfficientNetV2 base to adapt the model for multi-class classification. These include dense and dropout layers to enhance learning capacity while reducing overfitting. The GlobalAveragePooling2D layer condenses spatial dimensions, and a final softmax layer provides a probability distribution for the three classes.

TABLE III. LAYERS ADDED ON TOP OF EFFICIENTNETV2 MODEL FOR MULTI-CLASS CLASSIFICATION

Layer (type)	Unit / Paramters	Activation	Output Shape
GlobalAveragePooling2D	-	-	(None, 1280)
Dense	64	RelU	(None, 64)
Dropout (20%)	-	-	(None, 64)
Dense	128	RelU	(None, 128)
Dropout (30%)	-	-	(None, 128)
Dense	3	Softmax	(None, 3)

D. Fine Tuning

In fine-tuning, the pre-trained layers of the EfficientNetV2 model are unfrozen, allowing them to update during training. This process enables the model to adapt the general features learned from the ImageNet [14] dataset to the specific characteristics of the new dataset. By unfreezing these layers, the model can refine low-level features, like edges and textures, and higher-level patterns, making them more relevant to the new classification task. A reduced learning rate is typically used to maintain stability and preserve valuable pre-trained knowledge, allowing for small, precise updates to the model's weights. This careful adjustment improves performance on the new data without losing the foundational features from pre-training.

E. Model Training

The EfficientNetV2 variants (B0 to B3) were trained using the Adam optimizer with a learning rate of 0.001, a batch size of 128, and 10 epochs. The choice of 10 epochs was made to balance effective learning with computational efficiency, as the model was observed to achieve stable convergence and high performance within this range, with minimal improvements beyond 10 epochs. Early stopping was employed to monitor validation performance, terminating training if the model's accuracy or loss showed signs of stabilizing. This technique prevented unnecessary epochs, reduced computational cost, and minimized overfitting risks by ensuring that training stopped once the model had sufficiently learned from the data.

F. Explainability Using Grad-CAM

This study uses Grad-CAM (Gradient-weighted Class Activation Mapping) to enhance interpretability and gain insights into the model's decision-making process. Grad-CAM generates heatmaps highlighting the regions of input images most influential to the model's predictions, providing a visual explanation for each classification. By applying Grad-CAM to test set images, this study assesses whether the model's focus aligns with meaningful features, such as patterns, colors, or textures indicative of disease in lung tissue. Consistent focus on relevant regions suggests that the model has learned meaningful features rather than relying on irrelevant patterns. Additionally, Grad-CAM helps identify potential misclassifications by revealing cases where the model's attention may be on misleading areas. This approach enhances transparency and serves as a qualitative check, supporting the model's reliability and validity in lung cancer classification.

G. Ensemble Voting Techniques

The ensemble approach in this study combines EfficientNetV2 variants (B0 to B3) using soft, hard, and weighted voting to enhance accuracy and robustness. Each model provides unique feature extraction capabilities, contributing to a more generalized prediction. In soft voting, the models' probability outputs are averaged, and the class with the highest probability is selected, allowing for nuanced decision-making based on model confidence. Hard voting involves each model making an independent prediction, with

the final class determined by a majority vote, providing robustness by reducing the impact of individual model errors. Weighted voting further refines the ensemble by assigning weights to each model based on their performance metrics, ensuring models with higher accuracy have a greater influence on the final prediction. Together, these voting methods leverage the strengths of each EfficientNetV2 variant, resulting in a balanced, adaptive, and effective classification system.

H. Evaluation Metric

The lung cancer classification model's performance is evaluated using accuracy, precision, recall, and F1-score standard metrics for multi-class tasks. Accuracy measures the overall correctness, while precision and recall assess the model's ability to avoid false positives and false negatives, respectively. The F1-score balances precision and recall, especially useful in cases of uneven class distribution. Confusion matrices were also used to identify specific misclassifications across classes. These metrics collectively offer a thorough assessment of the model's effectiveness in classifying lung cancer from histopathology images.

IV. Result And Discussion

The results demonstrate the performance of EfficientNetV2 variants and the ensemble approach in classifying lung cancer subtypes. Key observations include the baseline accuracy, the impact of data augmentation, and the benefits of combining models through ensemble voting. Grad-CAM visualizations provide further insights into model focus, highlighting relevant regions for classification.

A. Baseline Performance of EfficientNetV2B0

The baseline performance of the EfficientNetV2B0 model, evaluated without data augmentation, achieved a high accuracy of 99.82%. This accuracy is supported by the confusion matrix in Fig. 2, which shows a clear distribution of correct and incorrect classifications across classes, indicating that the model effectively learned distinguishing features for each class with minimal overfitting. The model's high baseline accuracy demonstrates its inherent capability to classify lung cancer subtypes from histopathology images without requiring data augmentation to achieve strong generalization.

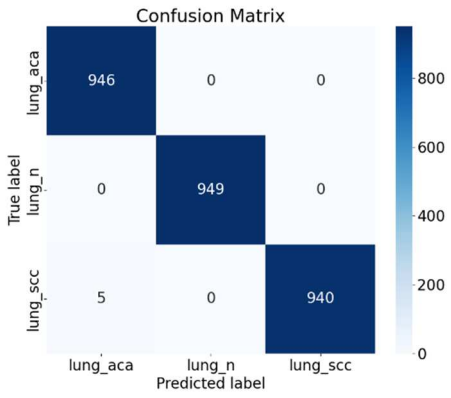

Fig. 2. Confusion matrix for EfficientNetV2B0 without data augmentation, illustrating the model's high accuracy and minimal misclassifications across lung cancer subtypes.

TABLE IV shows that the model achieved near-perfect precision, recall, and F1-scores for all classes, with only slight variance. These results confirm that EfficientNetV2B0 is well-suited for this classification task, providing a solid reference point for evaluating the impact of additional techniques such as data augmentation and ensembling.

TABLE IV. PERFORMANCE METRICS FOR EFFICIENTNETV2B0 ON THE TEST DATA WITHOUT AUGMENTATION

Model	Label	Accuracy	Precision	Recall	F1-Score
Efficient NetV2B0	Lung_aca	0.9982	0.9947	1	0.9974
	Lung_n		1	1	1
	Lung_scc		1	0.9946	0.9973

[a.] 'Lung_aca' represents Lung adenocarcinoma

[b.] 'Lung_n' represents Lung benign tissue

[c.] 'Lung_scc' represents Lung squamous cell carcinoma

B. Impact of Image Augmentation on Model Performance

As shown in TABLE V, the model trained with data augmentation achieved high precision, recall, and F1-scores across all classes, with only slight variance in the lung_scc category. The confusion matrix in Fig. 3 illustrates that the augmented EfficientNetV2B0 model accurately classified most instances across the three lung cancer subtypes with minimal misclassifications.

TABLE V. PERFORMANCE METRICS FOR EFFICIENTNETV2B0 ON THE TEST DATA WITH AUGMENTATION

Model	Label	Accuracy	Precision	Recall	F1-Score
Efficient NetV2B0	Lung_aca	0.9975	0.9927	1	0.9963
	Lung_n		1	1	1
	Lung_scc		1	0.9926	0.9963

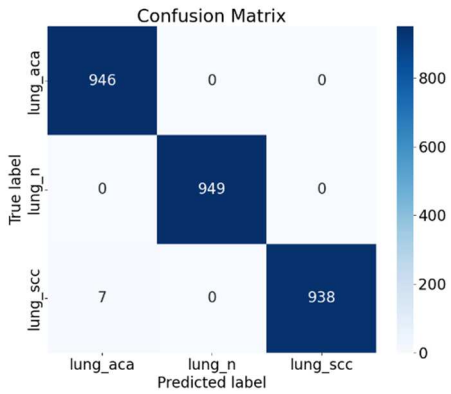

Fig. 3. Confusion matrix for EfficientNetV2B0 with data augmentation, showing accurate classification across lung cancer subtypes with minimal misclassifications.

The accuracy and loss curves for the EfficientNetV2B0 model are shown in Fig. 4 and Fig. 5, illustrating how data augmentation impacts robustness. The model without augmentation achieved rapid convergence, with training and validation accuracy stabilizing around 99.82% within the first

few epochs, suggesting effective learning but potentially limited robustness due to a lack of variability in the training data. In contrast, the model with augmentation exhibited a more gradual convergence, with training accuracy steadily increasing and validation accuracy stabilizing at 99.75%. This smoother, more controlled progression indicates enhanced generalization, as data augmentation introduced variability that helped the model learn more adaptable features. While the model with augmentation reached slightly lower peak accuracy, its stable validation performance and reduced sensitivity to fluctuations suggest improved robustness, making it more reliable for real-world applications where data variability is expected.

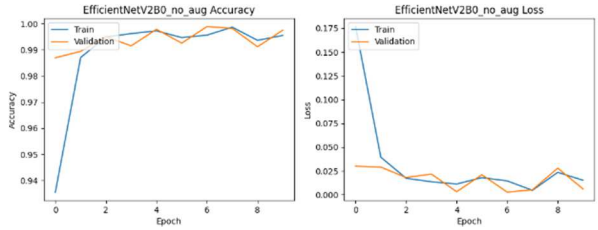

Fig. 4. Training and validation accuracy and loss curves for EfficientNetV2B0 without data augmentation, showing close convergence and minimal loss fluctuations.

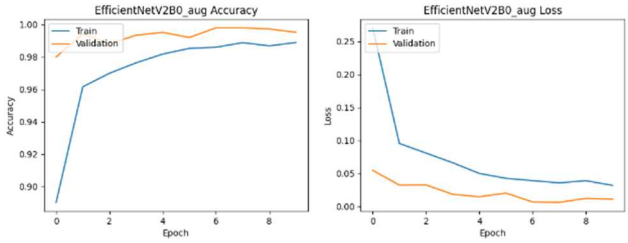

Fig. 5. Training and validation accuracy and loss curves for EfficientNetV2B0 with data augmentation, showing gradual convergence and stabilized validation performance across epochs.

C. Performance Comparison of EfficientNetV2 Variants

The performance of various EfficientNetV2 model variants (B0 through B3) is summarized in TABLE VI, illustrating their accuracy, precision, recall, and F1-scores. Models were trained on augmented data. Each model demonstrates high performance, with EfficientNetV2-B3 achieving the highest accuracy at 99.96%. The macro averages of precision, recall, and F1-score consistently reflect strong classification capabilities across all variants, indicating that increasing model complexity slightly improves performance. This table clearly compares each variant's effectiveness, highlighting the incremental gains achieved by more advanced versions.

TABLE VI. PERFORMANCE METRICS OF EFFICIENTNETV2 VARIANTS EVALUATED USING MACRO AVERAGES.

Model Variant	Accuracy	Precision (Macro Average)	Recall (Macro Average)	F1-Score (Macro Average)
EfficientNetV2-B0	0.9975	0.9976	0.9975	0.9975
EfficientNetV2-B1	0.9993	0.9993	0.9993	0.9993
EfficientNetV2-B2	0.9961	0.9961	0.9961	0.9961
EfficientNetV2-B3	0.9996	0.9996	0.9996	0.9996

D. Explainability Using Grad-CAM

The Grad-CAM heatmaps in Fig. 6, Fig. 7, and Fig. 8 reveal that each EfficientNetV2 model variant (B0 through B3) highlights regions associated with key diagnostic features for lung cancer subtypes. The highlighted regions correspond to medically relevant features, such as glandular structures in lung adenocarcinoma and abnormal squamous cells in lung squamous cell carcinoma. This alignment indicates that the models base their predictions on meaningful features consistent with each subtype's known characteristics. The consistent focus on diagnostically relevant areas across different variants underscores the reliability of the models' decision-making processes. These findings suggest that Grad-CAM could assist pathologists by providing visual explanations for AI predictions, helping to identify key areas of interest within histopathology images, and potentially accelerating the diagnostic process.

The models' focus on relevant features is enhanced when combined into an ensemble. Hard voting ensures robustness by relying on the majority consensus, even if some models have minor variations in focus. Soft voting averages the probabilities from all models, balancing predictions to reduce the influence of outliers and provide a more nuanced decision. Weighted voting prioritizes models with stronger performance metrics, giving their predictions more influence and improving overall accuracy. These voting techniques effectively combine individual models' strengths, as Grad-CAM demonstrated, which highlights diagnostically meaningful regions. This approach results in a reliable classification system that leverages each model's unique and shared focus areas while mitigating the risks of relying on any single model's perspective.

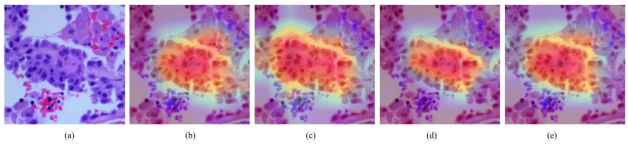

Fig. 6. Grad-CAM heatmap overlays for the Lung adenocarcinoma class: (a) Original image, (b) EfficientNetV2-B0, (c) EfficientNetV2-B1, (d) EfficientNetV2-B2, (e) EfficientNetV2-B3.

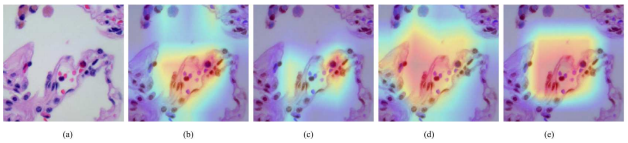

Fig. 7. Grad-CAM heatmap overlays for the Lung benign class: (a) Original image, (b) EfficientNetV2-B0, (c) EfficientNetV2-B1, (d) EfficientNetV2-B2, (e) EfficientNetV2-B3.

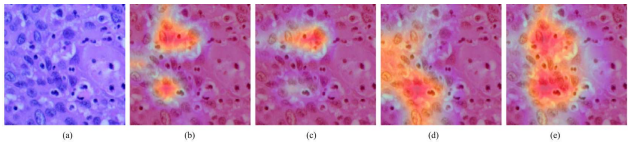

Fig. 8. Grad-CAM heatmap overlays for the Lung squamous cell carcinoma class: (a) Original image, (b) EfficientNetV2-B0, (c) EfficientNetV2-B1, (d) EfficientNetV2-B2, (e) EfficientNetV2-B3.

E. Analysis of Ensembling

As shown in TABLE VII, the ensemble configurations of EfficientNetV2 models (B0, B1, B2, and B3) achieve consistently high accuracy, demonstrating the effectiveness of hard, soft, and weighted voting strategies. Certain

configurations, specifically [v2B0, v2B1, v2B3] and [v2B0, v2B2, v2B3], reach perfect accuracy (1.0000) across all three methods, highlighting their ability to capture complementary strengths of individual models. By assigning higher weights to more reliable models based on performance metrics, weighted voting ensures that stronger models like EfficientNetV2-B1 and B3 contribute more to the final prediction. This high performance across configurations reflects the robustness of the ensemble method, as it combines multiple perspectives from individual models, each focusing on different relevant features identified in Grad-CAM analysis, thereby enhancing classification accuracy and generalizability.

TABLE VII. ACCURACY COMPARISON OF VARIOUS EFFICIENTNETV2 ENSEMBLE CONFIGURATIONS USING HARD, SOFT, AND WEIGHTED VOTING.

Models Included in Ensembling	Accuracy		
	Hard Voting	Soft Voting	Weighted Voting
[v2B0, v2B1, v2B2]	0.9996	0.9996	1.0000
[v2B0, v2B1, v2B3]	1.0000	1.0000	1.0000
[v2B0, v2B2, v2B3]	1.0000	1.0000	1.0000
[v2B1, v2B2, v2B3]	0.9996	0.9996	1.0000

V. CONCLUSION AND FUTURE WORK

This study used an ensemble of EfficientNetV2 model variants (B0 through B3) to classify lung cancer subtypes from histopathology images, achieving exceptionally high accuracy. By leveraging hard, soft, and weighted voting strategies, the ensemble approach demonstrated robust performance across various model combinations, achieving perfect accuracy in some configurations. Grad-CAM visualizations further validated that each model variant focused on relevant features, confirming that the ensemble effectively combined diverse yet meaningful perspectives, enhancing classification robustness and interpretability.

However, it is important to note that the dataset used in this study contains a limited number of unique images, with many samples generated through augmentation. This limitation may have contributed to the high accuracy observed, as the augmented dataset may inadvertently make the classification task "easier" by providing repetitive patterns that are more readily learned by the models. Additionally, no similar datasets are currently available to test the model's performance on external data, limiting the ability to evaluate its generalizability beyond this dataset.

Future work should focus on testing the ensemble on larger and more diverse datasets when they become available, ideally with more unique, non-augmented images, to assess its generalizability to real-world data better. Additionally, testing on varying-quality images would offer insights into the model's robustness in real-world conditions, ensuring its reliability across different settings. Exploring other model architectures or advanced ensembling techniques, such as stacking, could optimize performance by combining predictions from multiple models through a meta-learner.

The EfficientNetV2 ensemble, combined with Grad-CAM for interpretability, can provide accurate classifications and visual explanations as a decision-support tool. Integrated into digital pathology systems, these tools could streamline workflows, reduce workloads, and improve diagnostic consistency. Future research should prioritize the development of user-friendly interfaces to facilitate seamless clinical adoption. Through these efforts, this research can continue advancing the accuracy and real-world applicability of deep learning for histopathology-based lung cancer classification.

REFERENCES

[1] J. S. Brown et al., "Updating the definition of cancer," Molecular Cancer Research, vol. 21, no. 11, pp. 1142–1147, Nov. 2023, doi: 10.1158/1541-7786.MCR-23-0411.

[2] "What is Cancer?" National Cancer Institute, Oct. 11, 2021. [Online]. Available: https://www.cancer.gov/about-cancer/understanding/what-is-cancer. Accessed: September 3, 2024.

[3] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," arXiv, Sep. 11, 2020. [Online]. Available: http://arxiv.org/abs/1905.11946. Accessed: September 3, 2024.

[4] B. K. Hatuwal and H. C. Thapa, "Lung cancer detection using convolutional neural network on histopathological images," IJCTT, vol. 68, no. 10, pp. 21–24, Oct. 2020, doi: 10.14445/22312803/IJCTT-V68I10P104.

[5] S. Anjum et al., "Lung cancer classification in histopathology images using multiresolution efficient nets," Computational Intelligence and Neuroscience, vol. 2023, pp. 1–12, Oct. 2023, doi: 10.1155/2023/7282944.

[6] A. Kallipolitis, K. Revelos, and I. Maglogiannis, "Ensembling EfficientNets for the classification and interpretation of histopathology images," Algorithms, vol. 14, no. 10, p. 278, Oct. 2021, doi: 10.3390/a14100278.

[7] O. A. Abioye, A. E. Evwiekpaefe, and A. J. Olalekan, "Performance Evaluation of EfficientNetV2 Models on the Classification of Histopathological Benign Breast Cancer Images," Science Journal of the University of Zakho, vol. 12, no. 2, pp. 208–214, Apr.-Jun. 2024.

[8] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and Colon Cancer Histopathological Image Dataset (LC25000)," arXiv preprint arXiv:1912.12142v1 [eess.IV], 2019. doi: 10.48550/arXiv.1912.12142.

[9] L. Manjusha and V. Suryanarayana, "Detect and Remove Duplicate Images from a Dataset for Deep Learning," Journal of Positive School Psychology, vol. 6, no. 2, pp. 606-609, 2022

[10] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The Impact of Class Imbalance in Classification Performance Metrics Based on the Binary Confusion Matrix," Pattern Recognition, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.

[11] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[12] F. Shahidi, S. M. Daud, H. Abas, N. A. Ahmad, and N. Maarop, "Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study," IEEE Access, vol. 8, pp. 187531-187543, 2020, doi: 10.1109/ACCESS.2020.3029881.

[13] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," arXiv preprint arXiv:2104.00298, 2021.

[14] J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 2009, pp. 248–255.