CHAPTER V
CONCLUSION

This chapter presents the conclusions drawn from the results discussed in the previous chapters, as well as suggestions for future research.

5.1 Conclusion

Based on the experimental results, performance evaluations, and feature visualization analyses conducted to compare Adaptive Minimal Ensemble, EfficientNetV2-S, and EfficientNetV2-L in the vehicle class classification task, the following conclusions can be drawn.

1. The Adaptive Minimal Ensemble (AME) and EfficientNet V2-L (End-to-End) models demonstrated the best and most optimal performance with perfect metric results, namely 100% accuracy, 100% precision, 100% recall, and an F1-score of 100%. These results prove that combining models through the AME ensemble approach successfully achieves results on par with large-scale single models trained end-to- end. Meanwhile, the single EfficientNet V2-S architecture also recorded very solid performance with an accuracy of 96% and an F1-score of 96.04%. The most significant drop in performance occurred when applying the fine-tuning method to EfficientNet V2-L, which yielded an accuracy of only 67%.
2. The AME architecture offers highly optimized computational efficiency by balancing high accuracy with moderate hardware load. During the training phase, AME took 267 seconds and used 1,477.5 MB of VRAM. These figures are far more efficient than those of the End-to-End V2-L model, which recorded the longest training time of 1,710 seconds and consumed 8,053.2 MB of VRAM. During the testing phase, AME was able to keep VRAM usage at 1,375.8 MB with a power consumption of 63.9 W, demonstrating the energy efficiency of this ensemble architecture, with AME's inference time recorded at 0.0228 seconds per sample.
3. Based on heatmap visualization extracted using the84 HiResCAM, the architecture AME and EfficientNet V2-L (End-to-End)85 demonstrate the most logical feature localization process. For most classes, both models consistently focus on crucial structural features that determine vehicle class,

such as the rear wheel arrangement and chassis. The EfficientNet V2-S model, despite logically focusing on important features, still experiences classification errors in some images. In contrast, the worst feature localization is demonstrated by EfficientNet V2-L (Fine- Tune), which often fails to capture important features such as the wheel axles and instead exhibits a bias toward the background. This indicates that the model's accuracy performance is directly proportional to the precision of the feature localization captured by the model.

5.2 Recommendations

Based on the problem limitations and implementation during the research process, several suggestions for future research are as follows.

1. Exploration of a more diverse dataset to improve the model's generalization ability; future research is recommended to expand the training dataset by including variations of extreme environmental conditions such as nighttime, rainy weather, or low light. Simulating these real-world conditions is crucial so that the arterial road monitoring system can operate stably under various actual lighting situations.
2. Testing in an edge computing environment: Given that the AME architecture has proven to be highly efficient in terms of power consumption and VRAM usage during the inference process, direct performance testing on edge computing devices such as single-board computers integrated with highway CCTV cameras is necessary to validate real-time latency in the field.
3. Exploration of weak learner architectural combinations: The base model for AME in this study is limited to the EfficientNet V2-S variant. Future developments could involve designing combination layers that integrate features from fundamentally different architectures to evaluate whether such diversity in feature extraction can further enhance prediction accuracy.