CHAPTER V
CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

Based on a series of comparative experiments utilizing the DenseNet121 architecture across four scenarios encompassing two levels of data imbalance, this research has successfully answered all the established research questions:

- First, Cross-Entropy Loss is proven to exhibit a bias toward the majority class under imbalanced dataset conditions. The S-1 model achieved an overall accuracy of 95.81%, yet it concealed a distinct disparity: the Gray Leaf Spot class only achieved a Recall of 0.8421 with 18 undetected images under moderate conditions, and drastically deteriorated to a Recall of 0.4615 under extreme conditions (S-3), wherein more than half of the GLS images were missed and entirely misclassified as Blight. This phenomenon serves as empirical confirmation of the Accuracy Paradox high global accuracy while sensitivity toward the most critical minority class remains low.
- Second, Focal Loss is proven to improve the detection sensitivity of the minority Gray Leaf Spot class, albeit accompanied by a trade-off of decreased precision and overall accuracy. In the moderate scenario (S-2), GLS Recall increased from 0.8421 to 0.8684 (+0.0263), while precision dropped from 0.8889 to 0.7674, resulting in a decline in the GLS F1-Score from 0.8649 to 0.8148. In the extreme scenario (S-4), the increase in Recall was far more dramatic, rising from 0.4615 to 0.6923 (+0.2308, or 7.7 times larger than the moderate scenario), proving that the effectiveness of Focal Loss is proportional to the degree of data imbalance. However, the trade-off also strengthened proportionally: GLS precision dropped drastically to 0.3913 in the extreme scenario.
- Third, the effectiveness of Focal Loss is highly dependent on the degree of data imbalance. At a moderate Imbalance Ratio (IR) of

1:2.3, Cross-Entropy Loss still provides a better overall precision-recall balance; the GLS F1-Score of the CE model (0.8649) is higher than that of the FL model (0.8148). At an extreme IR (1:10), Focal Loss becomes a crucial intervention because CE fails to detect more than half of the minority class instances. This finding indicates that there is an imbalance ratio threshold at which the benefits of the Focal Loss mechanism begin to outweigh the detriments of its precision trade-off.

- Fourth, DenseNet121 coupled with a transfer learning approach has proven to be a highly suitable architecture for this task. The CE baseline model already achieved an accuracy of 95.81% on a dataset of 4,188 images using standard augmentation without complex techniques, confirming that the parameter efficiency of DenseNet121 (~8 million parameters) and its dense connection mechanism are capable of effectively learning the discriminative features of corn leaf diseases on a medium-sized dataset. The selection of the loss function must be predicated on a cost-benefit analysis based on application priorities: if avoiding the omission of any genuine GLS cases is more critical, Focal Loss is recommended; however, if ensuring the reliability of every GLS prediction prior to taking action is the priority, Cross-Entropy is more optimal.

5.2. Recommendations

Based on the research findings and limitations, several recommendations are proposed for future work:

- First, exploring the hyperparameters of Focal Loss via a grid search across combinations of gamma values (0.5, 1.0, 1.5, 2.0) and various alpha schemes is necessary to determine the optimal balance point between improving Recall and maintaining Precision, particularly under moderate imbalance conditions.
- Second, incorporating synthetic augmentation techniques, such as Mosaic or CutMix, specifically for the GLS class could be combined with Focal Loss to achieve a more potent synergistic effect.

- Third, a comparative architectural analysis using models such as EfficientNet or ResNet50 should be conducted to determine whether the observed trade-off pattern is specific to DenseNet121 or represents a more generalized phenomenon.
- Fourth, integrating per-class AUC-ROC (Area Under the Receiver Operating Characteristic Curve) analysis and the Matthews Correlation Coefficient (MCC) would provide a more comprehensive evaluation framework for imbalanced dataset conditions.
- Fifth, testing on real-world field datasets with highly variable lighting and background conditions is required to validate the model's generalization capability in actual agricultural environments, considering that the PlantVillage dataset utilized in this study was collected under controlled laboratory conditions.