CHAPTER V
CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Based on the results of research on the application of the hybrid EfficientNet-B0 and Prototypical Network model for multi-class diabetic retinopathy classification using the *APTOS 2019 dataset*, the following conclusions are obtained:

1. The hybrid model integrating EfficientNet-B0 as a feature extractor and Prototypical Network as a classifier was successfully applied for multi-class diabetic retinopathy classification with five severity levels. Implementation was performed *end-to-end* where both components were trained simultaneously through backpropagation, enabling joint optimization of feature extraction and class prototype formation. The proposed model without preprocessing achieved the best performance with 82.26% accuracy, macro precision of 0.6696, macro recall of 0.6345, macro F1-score of 0.6454, and Quadratic Weighted Kappa (QWK) of 0.8807. This QWK value falls within the almost perfect agreement category, indicating a high level of agreement between model predictions and *ground truth* labels in the context of ordinal classification on an imbalanced dataset.
2. Prototypical Network integration as a classifier provides performance improvements compared to the baseline EfficientNet-B0 model with softmax classifier. With identical feature extractors, the proposed model recorded a QWK improvement of 0.0278 points (from 0.8529 to 0.8807) or 3.26%, and an accuracy improvement of 1.91% (from 80.35% to 82.26%). The greater QWK improvement compared to accuracy indicates that the Prototypical Network is more effective in understanding the ordinal classification structure, not just improving general prediction accuracy. Confusion matrix analysis confirms the proposed model produces error patterns more consistent with the ordinal structure, where prediction errors tend to occur in adjacent classes and rarely in distant classes. t-SNE visualization also shows the proposed model forms a more structured

embedding space with better inter-class separation compared to baselines, reflecting the model's ability to capture ordinal relationships between diabetic retinopathy severity levels.

5.2 Recommendations

Based on the research results and limitations that have been identified, several suggestions for further research are as follows:

1. Conduct testing on independent datasets such as Messidor, EyePACS, IDRiD, or local institution datasets to evaluate model generalization capability under different imaging conditions and populations.
2. Explore methods such as class-weighted loss, focal loss, or resampling techniques to improve performance on minority classes, particularly Severe DR.
3. Test other *backbone* architectures such as larger EfficientNet variants, ResNet, DenseNet, or Vision Transformer to obtain the optimal combination between performance and computational efficiency.
4. Conduct systematic hyperparameter tuning using approaches such as Bayesian optimization or random search to obtain optimal configurations.
5. Develop a prototype that can be integrated with electronic medical record systems to support clinical practice implementation.
6. Explore ensemble strategies and model development for multi-retinal disease detection to improve system utility as a comprehensive screening tool.