CHAPTER V
CONCLUSION AND RECOMMENDATIONS

This chapter presents the conclusions and recommendations derived from the conducted research. The conclusions are formulated based on the architecture design, hyperparameter testing scenarios, model performance evaluation, and implementation described in the previous chapters, thereby directly addressing the research questions. The recommendations are provided as considerations for future research to further improve the deepfake detection system.

5.1. Conclusion

Based on the entire research process, including architecture design, hyperparameter testing, model performance evaluation, and Android application prototype implementation, several conclusions are obtained. These conclusions summarize the main findings related to model development, performance, generalization, and mobile implementation. The conclusions are as follows:

1. This study successfully designed and implemented a deepfake detection model integrating MobileViT, FFT, and Modified CBAM within a single face-image-based classification pipeline. The pipeline begins with face detection and cropping using YOLOv11n, followed by spatial representation using RGB images and frequency representation using FFT, which are combined through Fusion Convolution before being processed by MobileViT and enhanced using Modified CBAM. The resulting features are then classified using Global Average Pooling, Dropout, and a Fully Connected Layer, demonstrating that all components can be implemented and operate according to the proposed design.
2. The proposed model with the best configuration demonstrates improved performance on secondary data and better generalization on primary data compared with the MobileViT baseline. On the secondary data, the proposed model achieves an accuracy of 95,30%, precision of 0,9512, recall of 0,9550, and F1-score of 0,9531, while on the primary data it achieves an accuracy of 85,80%, precision of 0,8594, recall of 0,8560, and F1-score of 0,8577. These

results indicate that integrating FFT and Modified CBAM improves secondary-data performance and provides stronger generalization to primary data than the MobileViT baseline.

3. The best-performing model was successfully implemented into the Flutter-based Telisik Android application prototype. YOLOv11n and the MobileViT + FFT + Modified CBAM model were deployed in TensorFlow Lite format to enable on-device execution. The application analyzes a sequence of video frames and aggregates the classification probability results to reach a video-level decision. Furthermore, the system demonstrated the ability to identify and differentiate non-human inputs, such as animals or inanimate objects, through an initial face detection verification stage, thereby preventing inference on irrelevant data. This implementation demonstrates that the proposed model can be effectively executed within an Android application prototype for deepfake detection.

5.2. Recommendations

Based on the research results and limitations, several recommendations are proposed for future studies. These recommendations focus on expanding the system's detection capabilities and evaluating its efficiency on mobile devices. The recommendations are as follows:

1. Future research is recommended to extend the system so that it can detect deepfakes in videos containing multiple faces within a single frame. This capability can be developed by incorporating tracking and classification mechanisms for each detected face. Such an extension would allow the system to operate beyond videos containing only one dominant face.
2. Future research is also recommended to evaluate model efficiency directly across various Android devices. The evaluation may include inference time, memory usage, power consumption, and application stability, while additional model optimization can be performed to improve detection speed and computational efficiency. Such testing would provide a clearer understanding of the model's practical performance across devices with different computational capabilities.