CHAPTER V
CONCLUSION AND SUGGESTIONS

5.1. Conclusion

From the overall system testing, the results of this study can be concluded as follows:

1. Identification of five Javanese dialect variations (Arekan, Pantura Banten, Banyumas, Cirebon, and Solo-Yogya) was successfully carried out automatically using the Log-Mel Spectrogram feature classified through a CNN-TCN hybrid architecture. This integration proved effective, CNN excels in extracting low-level spatial patterns, while TCN successfully models long-term temporal dependencies. In the optimal configuration (Scenario 3: batch size 16 and learning rate 0.001), the application of Soft Voting prediction aggregation to a 10-second audio segment resulted in an accuracy rate of 90.68% and an F1-Score of 0.9032. This achievement significantly outperformed the comparison single models, namely CNN (83.53%) and TCN (86.37%), and proved to be the most stable convergence point compared to the scenario of reducing the learning rate to 0.0001 (Scenario 5) which only achieved an accuracy of 88.49%.
2. Based on classification tests on 6,471 segments from 139 unique audio files, this hybrid algorithm successfully overcomes the challenge of cross-dialect confusion fairly and without bias towards the majority class. The model's performance stands out with a near-perfect F1-Score of 0.96 for the Pantura (Banten) dialect. Furthermore, the model proved robust in predicting the Arekan dialect, a previous weakness of the single model, by recording an F1-Score increase to 0.85 thanks to a balanced Precision and Recall ratio. This entire best-in-class system has been comprehensively implemented into a web interface tool, allowing users to directly predict dialect classes based on the highest probability (confidence score).

5.2. Suggestions

Based on the results of the analysis and limitations in the research that has been conducted, here are some suggestions that can be considered for further system development:

1. Model testing indicates that the Arekan dialect is often the most challenging class and prone to cross-confusion. Given the limited availability of public digital resources for regional languages, additional, more balanced data collection, especially for minority classes, is highly recommended to improve the recall value and generalizability of the model.
2. The current system focuses on recognizing five main dialect variations (Pantura-Banten, Cirebon, Banyumas, Solo-Yogya, and Arekan). To support more comprehensive digital preservation efforts, the dataset's scope could be expanded to recognize other specific sub-dialect variations on the island of Java.
3. This system uses a purely phonetic approach (acoustic and prosodic characteristics) and does not rely on word transcription. To achieve a higher level of Dialect Identification (DID) precision, further development is highly recommended to combine acoustic features (CNN-TCN) with text-based linguistic representations (such as the Transformer language model). This multimodal approach has been proven to effectively boost system performance because it is able to understand dialects not only from the “way of pronunciation” but also from the “choice of vocabulary.”