

## CHAPTER V

### CONCLUSION

#### 5.1 Conclusion

Based on the results of implementation, testing, and analysis of the system Case-Based Reasoning (CBR) integrated with the MultiSURF feature weighting algorithm for early detection of diabetes using Pima Indians Diabetes dataset Database (PIDDD), several conclusions were obtained as follows:

1. The preprocessing stage , which includes Missing Value Handling using in-Fold based median imputation, Min-Max Scaling normalization, and class balancing using SMOTE-ENN, was successfully implemented without causing data leakage in the validation process. This approach ensures that the model's performance evaluation reflects the true generalization capabilities of data that have never been seen before.
2. The MultiSURF algorithm successfully identified the level of relevance of each clinical feature on the diabetes dataset. The Glucose feature gained the highest weight with a value of 1.0000, followed by BMI and Age, while BloodPressure and SkinThickness gained the lowest weights. These results show that not all features have the same contribution to the early detection process of diabetes.
3. The integration of the weight of the MultiSURF feature into the calculation of Weighted Euclidean Distance at the CBR retrieval stage has been proven to be able to improve the quality of similarity measurements between cases. Feature weighting makes the process of searching for similar cases more focused on attributes that have high diagnostic relevance, resulting in better retrieval quality compared to the no-weighted approach.
4. The results of the evaluation of 20 combinations of test scenarios showed that the variation of the  $n\_Neighbors$  value in the SMOTE-ENN method and the  $K$  value in the Case-Based Reasoning method affected the performance of the resulting classification. Based on the overall test results, the best configuration was obtained at  $n\_Neighbors = 3$  and  $K = 9$ , with an Accuracy score of 76.03%, Precision of 65.87%, Recall of 64.54%, and an F1-Score of 64.95%. The configuration was chosen because it provides a good balance of performance across all evaluation metrics.

5. A comparison between the Baseline CBR model and the MultiSURF-based Weighted CBR model using the best configuration shows that the application of feature weighting improves classification performance. Weighted CBR increased Accuracy by 1.70%, Precision by 3.06%, and F1-Score by 0.19% compared to the Baseline model, despite a decrease in Recall by 2.62%. These results show that feature weighting using MultiSURF is able to improve the overall accuracy of predictions and produce a more effective retrieval process through emphasis on features that have higher relevance.

Overall, this study proves that the integration of the MultiSURF algorithm as a feature weighting method in Case-Based Reasoning able to improve process quality Retrieval and provides improved classification performance compared to unweighted CBR. Although the increase obtained is relatively small, the results show that giving weight based on the level of feature importance can result in a more representative case search process so that the Weighted Case-Based Reasoning MultiSURF-based can be an alternative in supporting the early detection of Type 2 Diabetes Mellitus based on clinical data.

## **5.2 Suggestions**

Based on the limitations and findings obtained during the study, the following suggestions are put forward as recommendations for further development and research:

1. Further research can explore more adaptive methods of handling MissingValues, such as KNN Imputation or Multiple Imputation, especially in Insulin features that have a fairly high proportion of lost data, so that the information contained in these features can be better retained.
2. Comparison with weighting methods or other feature selection, such as ReliefF, SURF, mRMR, or modern Machine Learning-based approaches, needs to be done to find out the most effective method in improving the quality of Retrieval in the CBR system.
3. The development of a more adaptive similarity function can be considered to reduce the number of False Negative cases, especially in Diabetes patients with clinical characteristics that do not follow the majority pattern. The instance-

based feature weighting approach may be one of the alternatives worth researching.

4. Testing on a larger and diverse diabetes dataset, including medical record data from different populations, is needed to evaluate the model's generalization capabilities before it is applied to a broader clinical setting.

This page has been left blank on purpose