CHAPTER 5   CONCLUSION AND RECOMMENDATIONS

5.1   Conclusion

Based on the results and analysis obtained in this study, it can be concluded that:

1. The system device has been created and can function properly in detecting images. The detection results are sent to the database via an API and displayed on an LCD screen. The best model for detection is the CNN model with 100 epochs. The accuracy achieved for SIBI is 95.91% and for BISINDO is 92.64%.

2. Detection works well on images with varying light saturation, both dark and light.

3. Communication from non-disabled people to disabled people is done through STT, which converts speech to text and displays it on an LCD or website. This allows users with disabilities to read the message directly.

4. This system is already capable of two-way communication, equipped with the implementation of gesture recognition and speech-to-text (STT) and text-to-speech (TTS) functionality, which are functioning properly.

5.2   Recommendations

The detection system design using computer vision and speech-to-text (STT) in this study still needs to be further developed to help people with speech and hearing disabilities. The following are suggestions for further research:

1. Develop a more sophisticated computer vision model by combining additional machine learning and deep learning-based architectures to improve the accuracy and reliability of the system.

2. Expanding the dataset with video-based training, as sign language involves emotions, continuous movements, and dynamic contexts that are better captured in video format than static images.

3. Improving the system's output presentation, for example by integrating a larger screen or graphical display in addition to the 16x2 LCD, so that full sentences and more complex messages can be displayed more effectively.

4. Optimizing real-time interaction with enhanced STT and TTS features, so that voice input can be transcribed more accurately and developed into gestures, and gesture detection results can be directly converted into voice to support smoother two-way communication.