Jurnal Kejuruteraan 37(6) 2025: 3035-3043
https://doi.org/10.17576/jkukm-2025-37(6)-37

Emergency Sound Recognition and Direction Indication Using Machine Learning for Individuals with Hearing Loss

Vern Yee Lim[a,b], Huang Shen Chua[b], Salina Mohmad[b], Kim Boon Lau[b,c] & Sin Jin Tan[b*]

[a]*Texas Instruments Electronics Malaysia, Free Trade Zone, Batu Berendam,*

Taman Perindustrian Batu Berendam, 75350 Melaka, Malaysia

[b]*School of Engineering,*

University of Wollongong, Malaysia, 40150 Shah Alam, Malaysia

[c]*School of Engineering, UOW Malaysia KDU Penang University College,*

Jalan Anson, George Town, 10400, Pulau Pinang, Malaysia

[*]*Corresponding author: sj.tan@uow.edu.my*

Received 4 December 2024, Received in revised form 2 May 2025
Accepted 2 June 2025, Available online 30 September 2025

ABSTRACT

Sound is a vital means of expression and communication. However, individuals with hearing loss face considerable challenges, particularly in emergency situations. During such events, alerts and critical information are typically conveyed through sirens or verbal commands, which are ineffective for those with hearing loss. To overcome this limitation, an attachable or wearable assistive device is necessary. The proposed wearable system in this work consists of (i) microphone arrays embedded in a headband, and (ii) a wristband with a motor vibrator and display attached to it. The neural network algorithm, specifically the multi-layer perceptron model, is used for selective emergency sound (dog bark, car honk and siren) recognition and classification. Once an emergency sound is detected by the microphone arrays and classified by the algorithm, a vibration signal is transmitted to the motor, while the direction of the sound is simultaneously indicated on the wristband display. The trained model and real-time implementation achieved an accuracy of 90% and 77%, respectively.

Keywords: Machine learning; neural network; assistive technology; emegency sound; hearing loss

INTRODUCTION

Hearing loss is a condition that impairs a person's ability to hear sounds. Hearing loss ranges from minor to severe and can be caused by many factors, like genetic factors, chronic diseases, ageing, accidents, and nutritional deficiencies. More than 5% of the world's population, which is approximately 430 million people, requires rehabilitation to address the hearing loss issue (Bisgaard et al. 2022). By 2050, it is estimated that over 700 million people, or over 0.07% of the projected world population (9.8 billion), will need hearing rehabilitation and 2.5 billion people (over 0.26%) are estimated to have some extent of hearing loss (World Health Organisation, 2023; Graydon et al. 2019). Beyond its prevalence, research has also explored the underlying mechanisms of noise-induced hearing loss and its association with related conditions such as tinnitus and vestibular dysfunction, which can impair balance and increase the risk of accidents (Le et al. 2017).

Hearing is important for humans, especially in emergency situations. In times of emergency, alerts and information are usually delivered in the form of sirens and commands, as well as reports of hazards, harm, and damage. But these measures are not helpful for the deaf. A study stated that being able to be notified on the spot of what kind of emergency it is and the location of the emergency matters a lot to people with hearing disabilities (Tannenbaum-Baruchi, 2014). For individuals who are deaf or have hearing loss, receiving emergency alerts without relying on third-party assistance requires an attachable or

wearable device that can inform them through alternative means (Yağanoğlu, 2021; Mohd Ramzi et al. 2024). Various approaches, including assistive technologies and wearable devices, have been developed to support individuals with hearing loss, particularly in enhancing road safety. Mobile assistive devices based on smart phones are also introduced. For instance, Mielke and Brueck presented a smart phone that was paired with a watch via a Bluetooth connection application to provide alerts (Mielke & Brueck, 2015). When a sound was recognised, the source of the sound was shown both on the smart phone and the watch. Participants whom they interviewed commented on the importance of indication on the direction of sound, but this feature was not integrated and tested in their work.

Lately, machine learning technologies are often incorporated into these assistive devices for better sound classification accuracy (Saifan et al., 2018; Adeyanju et al. 2021; Shahira et al. 2022; Mulfari et al., 2021). In one work demonstrated by Otoom and Aloufee, direction instructions from a global positioning system (GPS)-based navigation system were analysed and classified using six classifiers, with the aim of guiding the deaf drivers in real time by means of various vibrotactile stimulation patterns (Otoom & Aloufee, 2022). In another work, deep learning-based siren identification system was designed to assist hearing-impaired individuals in noisy urban environment (Ramirez et al. 2022). Accuracy of 98% was achieved using training data and accuracy of 91% was attained on real street recordings. Building on such advancements, haptic and tactile technology have also gained prominence in enhancing road safety for individuals with hearing difficulties (Spiers et al. 2016; Mirzaei et al. 2020).

The focus of the work is to develop alternative, low-cost assistive devices specifically for pedestrians with hearing loss. With the aid of machine learning integration, the developed prototype incorporates the functionality to identify the source of emergency sounds and their directions, thereby providing valuable assistance while users are on the road.

METHODOLOGY

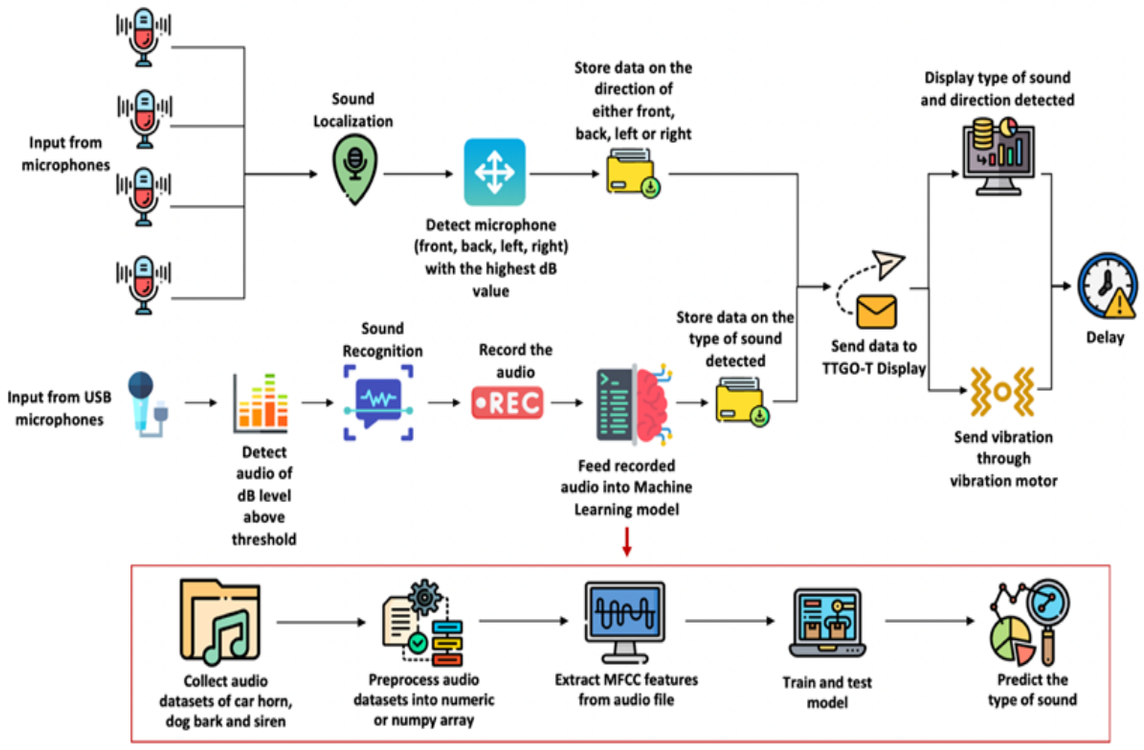
EXPERIMENTAL SETUP

Figure 1 ilustrates the overview of the developed system architecture and communications links established among components of prototype. Figure 1(a) shows the overall workflow, from the initial data acquisition stage through signal processing and classification, to the final output presentation. It receives signals from the four microphones (microphone array MAX4466) and a USB microphone. A USB microphone is used to detect the type of sound, while the microphone array is responsible for sound detection and localisation. The received signal is sampled at a frequency of 44100 Hz. Sound recognition only occurs when the detected audio is above the threshold value of 80 dB. The sound is recorded for 5 seconds and fed into the machine learning model. As for the input from the microphone array, the direction of the sound is obtained through the detection of the microphone with the highest dB value. Both data (type of sound and direction) are stored and displayed on the TTGO-T display. Every time the TTGO-T display receives data, it sends a vibration signal through the vibration motor. Figure 1(b) shows the connection of the ADC chip, ADS1115, and the four microphone amplifiers, MAX4466, to the Raspberry Pi, as well as the connection of the vibration motor to the TTGO-T display.

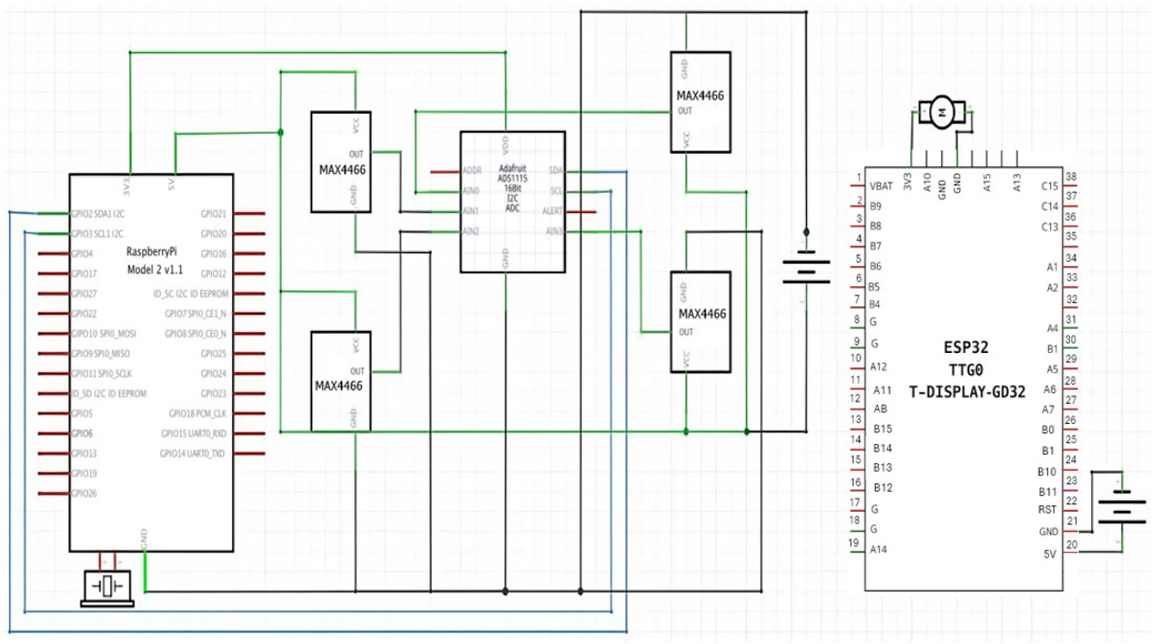
MODEL TRAINING

Training a sound recognition and classification model involve several stages. It starts by collecting audio dataset of car honks, dog barks, and sirens. The dataset is obtained from Freesound, which is a free source on the web that offers different kinds of voice recordings. Class labels are used to designate the audio files. The audio dataset is then be pre-processed to extract relevant features. In this case, the mel-frequency cepstral coefficients (MFCC) features of every audio are extracted before proceeding to the training process. The model architecture is then defined. The machine learning algorithm used in training the model is a neural network, specifically a multi-layer perceptron (MLP). MLP has several layers of interconnected nodes. Every node performs a linear transformation of input, followed by a non-linear activation function. The input layer receives the data, while the output layer generates the model's final predictions. The intermediate layers, also known as the hidden layers, transform the input data by extracting relevant features and learning representations that support accurate predictions. MLP learns to model complex non-linear relationships between the input and output data by stacking multiple hidden layers. An optimisation algorithm minimises the difference between the predicted and actual target outputs during the training process. This is done by adjusting the weights and biases of the neurones, or nodes. This process is repeated over many iterations until the model achieves reliable accuracy in predicting new data. The dataset is then divided into training and testing sets, typically 70% for training and 30% for testing. During training, the model learns to recognize patterns in the audio data that correspond to different sound classes. This process involves feeding the model with audio samples that are labelled and having the

parameters of the model adjusted to minimise the difference between the predicted and actual labels. After the training, the model is evaluated on the 30% testing set to measure its accuracy in correctly classifying the audio data. In this work, Google Colab is used to train the model.



(a)



(b)

FIGURE 1. Proposed project work for signal detection, classification and alert for individual with hearing loss (a) overview of project (b) connections between components

SOUND RECOGNITION

After training the model, the trained model is used to classify audio signals in real-time. The sound recognition process starts by receiving input from the USB microphone. Sound is recorded every 5 seconds and sampled at the sampling frequency of 44100 Hz. If the recorded audio exceeds 80 dB, it is fed into the machine learning model, which then predicts the type of sound. The threshold is set to 80 dB because the average dog barks range from 80 to 90 dB, while the car honks and sirens generally fall between 100 dB to 120 dB. However, their sound level may decrease depending on the ambient noise level.

SOUND LOCALIZATION

To determine the direction of the sound, four MAX4466 microphones are positioned at the front, back, left, and right of the headband. These microphones capture the audio input and compute sound intensity based on the recorded voltage values. The higher the voltage value, the higher the intensity of the sound. The microphone that records the highest voltage value is used to determine the direction of the sound.

DISPLAY OF RESULTS

The results from the sound recognition and sound localisation is displayed on the TTGO T-Display wirelessly via Wi-Fi connection. The communication between the Raspberry Pi and the TTGO T-Display is made through Message Queuing Telemetry Transport (MQTT). MQTT is a lightweight messaging protocol that sends messages using the publish or subscribe model between networked devices. The flow of this protocol is that when the publisher, which in this case is the Raspberry Pi, generates a message, it publishes the message along with a topic to the MQTT broker. The MQTT broker receives and stores the messages accordingly. A device that subscribes to that particular topic is called a subscriber. The TTGO T-Display, which is the subscriber, connects to the broker and every message that is published on the topic is sent to the subscriber. This protocol enables the publisher and subscriber to send and receive messages via the broker. In this programme, the Raspberry Pi publishes the results of sound recognition and sound localisation on a topic to the MQTT broker. The TTGO T-Display, which is the subscriber to the topic, receives the results, have them displayed, and send vibrations at the same time to notify the user that there is an incoming message.

RESULTS AND DISCUSSION

DEVELOPED PROTOTYPE

Figure 2 shows the complete prototype of head band, wrist band and arm band. The microphone array is mounted on the headband. The microphone array consists of four MAX4466s which are connected to ADS1115. The ADS1115 is connected to the Raspberry Pi which is placed in the armband. The USB microphone connecting to the Raspberry Pi is also placed at the armband. As for the wristband, the TTGO T-Display is connected to it. To make both the Raspberry Pi and TTGO T- Display portable, they are powered up using power bank with the output of 5V, 3A and LIPO battery of 3.7V with battery shield. LIPO battery of 3.7V is chosen to power up the TTGO T-Display because it is rechargeable which does not require user to keep changing the battery. A battery shield is also used to prevent overcurrent flow to the microcontroller board which could cause damage to it.

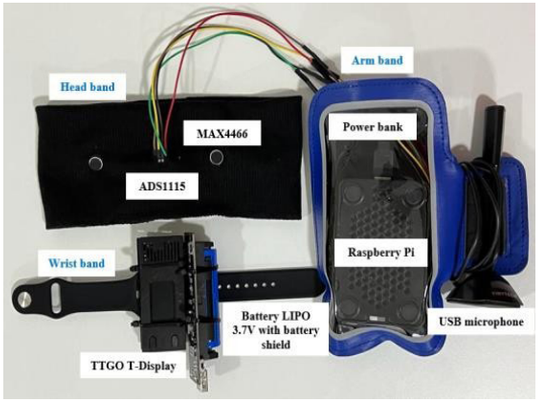


FIGURE 2. Developed prototype

MODEL ACCURACY AND MODEL LOSS

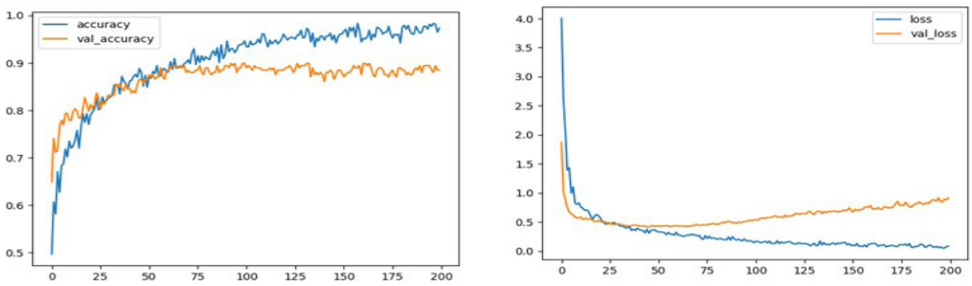
The model is trained using a total of 465 audio samples, consisting of 151 car honks, 169 dog barks and 145 sirens. The experiment is conducted over 200 epochs with a batch size of 32. Figure 3 shows the performance metrics of the training model. In the accuracy model illustrated in Figure 3(a), the training accuracy shows a sudden increase at the first few epochs from 0.5 to 0.7 and a gradual increase for the remaining epochs from 0.7 to approximately 1.0. The validation accuracy also shows a sudden spike at the first few epochs, from 0.65 to 0.75. It then slowly increases to 0.85 and maintains its range within 0.85 to 0.90. Similar to the first experiment, the initial validation accuracy is higher than the training accuracy for the first 30 epochs but lower for the remaining epochs. For the remaining

epochs, it can be seen that the training accuracy increases gradually but the validation accuracy shows a slow decline after the hundredth epoch. As for the loss model depicted in Figure 3(b), the training loss shows a drastic decrement from 4.0 to 1.0 for the first few epochs and decreases gradually for the remaining epochs, reaching almost 0.0 at the end of the epochs. The validation loss starts at 1.9 and decreases to 0.5 at the first few epochs. At the first 30 epochs, the loss is lower than the training loss. Upon reaching the thirtieth epoch, the validation loss starts to increase gradually. The validation loss ranges from 0.5 to 2.0.

CONFUSION MATRIX AND CLASSIFICATION METRICS

Figure 4 describes the confusion and classification metrics for model's performance on the test dataset. A confusion matrix describes the performance of a classification model based on a set of test data where the true values can be determined. The confusion matrix table in Figure 4(a) shows the combinations of predicted and actual values. The dark-coloured squares show the number of predictions that are correct for each class. The result shows that dog bark has the most correct predictions, followed by the car honk and, lastly, the siren.

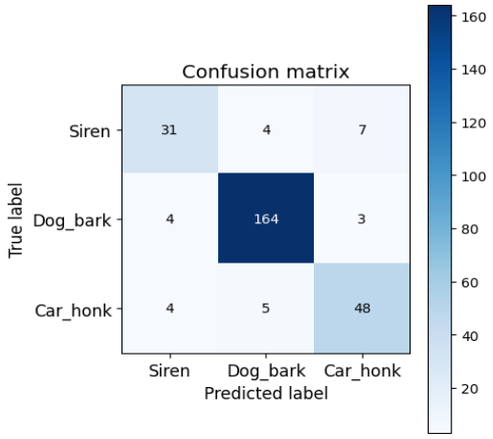
In classification metrics, precision shows the number of correctly predicted cases that turned out to be positive. Recall shows the number of actual positive cases that can be predicted correctly with the model. The F1 score is a weighted average of precision and recall. The macro average is the average of the metrics calculated for each individual class. The weighted average takes the imbalance in class sizes into account by weighing the average based on the number of instances in each class. From Figure 4(b), dog bark has the highest value for all three scores among the three classes and siren has the lowest value.



(a) (b)

FIGURE 3. Performance metrics of trained model (a) model accuracy (b) model loss

	precision	recall	f1-score	support
Siren	0.79	0.74	0.77	42
Dog_bark	0.95	0.96	0.95	171
Car_honk	0.83	0.84	0.83	57
accuracy			0.90	270
macro avg	0.86	0.85	0.85	270
weighted avg	0.90	0.90	0.90	270



(a) (b)

FIGURE 4. Confusion matrix and classification metrics for test data (a) confusion matrix (b) classification metrics

MULTICLASS RECEIVER OPERATING CHARACTERISTICS (ROC) CURVE

The ROC curve measures the performance and efficiency of a machine learning model. The higher the ROC score, the better the performance of the model in differentiating the positive and negative classes. A ROC score of 1 indicates that the model can differentiate between the positive and negative class points perfectly. Figure 5 shows that the ROC scores for the classes dog bark, car honk, and siren are all above 0.90, with values of 0.97, 0.95, and 0.92, respectively. Among the three classes, dog bark gives the highest accuracy, car honk gives the second highest and siren has the lowest accuracy.

SOUND RECOGNITION AND SOUND LOCALIZATION

Figure 6 shows the output when the audio detected does not exceed or exceeds the threshold value of 80 dB and the voltage values from all four MAX4466s. In Figure 6(a), when the audio dB level does not exceed 80 dB, the sound recognition process will not take place and hence, no output will be displayed at the TTGO T-Display. When the audio detected has a dB level above 80 dB, the sound recognition process will take place, as shown in Figure 6(b). The type of sound and direction detected is sent to the TTGO T-Display to be displayed.

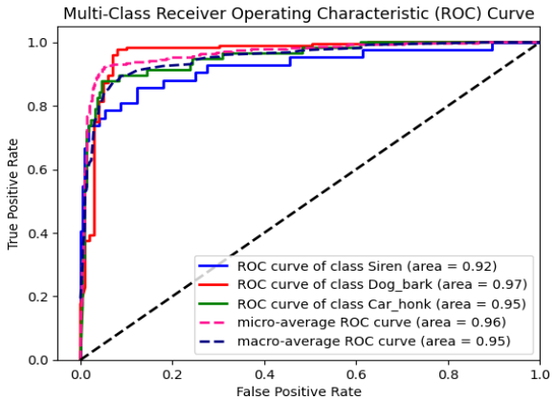
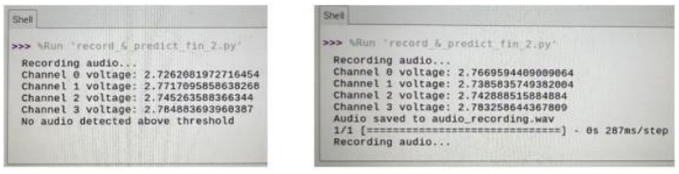


FIGURE 5. Multiclass ROC curve



(a) (b)

FIGURE 6. Output of voltage values from four MAX4466s and when audio level detected is (a) less than 80 dB (b) more than 80 dB

DISPLAY OF RESULTS

To allow communication between the Raspberry Pi and TTGO T-Display, the TTGO T-Display subscribes to the MQTT topic where the Raspberry Pi is sending messages. Subscription to the topic allows the MQTT broker to route messages between the two devices. When the TTGO T-Display receives a message from the subscribed topic, it extracts the information, specifically the sound type and its direction, and presents it on the display. When the message is received, the vibration motor, which is connected to the TTGO T-Display, vibrates for 2 seconds. Figure 7 illustrates the direction of sound shown at the TTGO T-Display. Figure 7(a) indicates siren is detected in the front direction. Figure 7(b) and Figure 7(c) show the dog bark, and car honk are detected in the front direction as well.

PROTOTYPE TESTING IN DIFFERENT ENVIRONMENT

Testing is carried out in two environments: on the laptop and in real-time. Testing on a laptop is where the audio is inputted directly into the programme to make predictions on what class the sound falls to. As for real-time testing, audio is obtained from the USB microphone in both indoor

and outdoor environments. A total of 50 audios are used in testing the model. The following figure shows the accuracy of the model in detecting the three types of sounds in three different environments.The chart in Figure 8 shows that tests conducted on the laptop give the highest accuracy for all three types of sounds, followed by indoor and, lastly, outdoor environments. Among the three types of sounds, dog bark gives the highest accuracy and siren gives the lowest accuracy since the utilised classification sound dataset has more samples for dog barking as compared to siren, which made the model struggle with sounds like siren.

(a) (b) (c)

FiGURE 7. TTGO T-Display when sound is detected coming from the front direction (a) siren (b) dog bark (c) car honk

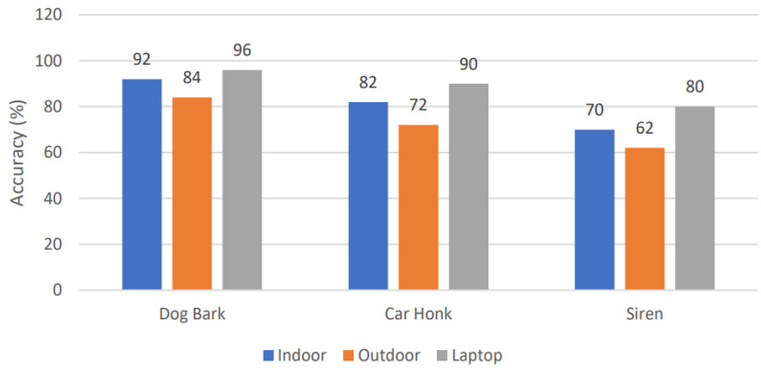


FIGURE 8. Sound recognition accuracy recorded at laptop, indoor and outdoor

DISCUSSION

These wearable devices have the capability of detecting sound in real-time and can be worn over a long period of time with the use of a rechargeable battery for the TTGO T-Display and a power bank for the Raspberry Pi. With the high computing power and memory from the Raspberry Pi and Wi-Fi connectivity on both the Raspberry Pi and TTGO T-Display, it can process data at a very fast rate. The system achieved a success rate of approximately 77% in real-time applications, which is lower than the accuracy obtained during analysis on a laptop using direct audio input. This is attibuted to some contributing factors. The first factor is due to the lack of dataset used in training the model. As can be seen in the accuracy and loss models in Figure 3, the curves are overfitting. This indicates that the model starts memorising the data, including the noise, making it fit way too closely to the training set and unable to generalise new unseen data. A model trained with a larger dataset generally achieves higher accuracy, as it can learn from a broader range of samples. This exposure enables the model to capture more complex relationships and patterns within the data. Training a model with more data increases its ability to identify consistent patterns and relationships across different subsets of the dataset. This improves generalisation and helps the model perform well on new data. Larger dataset expose the model to more diverse scenarios and help the model handle different types of inputs and variations in data. It increases the robustness of the model, making it less prone to overfitting or underfitting. A larger dataset also helps to reduce bias in the model because it could help prevent the model from making predictions based on a limited set of samples and improve its ability to handle inputs of different variations. It can be deduced that training a model with larger dataset makes the model more versatile and provides better ability to handle a variety of tasks. Other factors that may affect performance include the distance of the sound source or the presence of background noise. In this system, communication between the device and the wearer is achieved through vibration feedback. A vibration motor placed on the wrist band allows the user to feel vibration

and get notified every time an emergency sound is detected. The display on the wristband showing the type of emergency and the direction of the sound would be a great help since the hearing loss individuals have been having trouble being notified and locating the direction of the emergency sound.

CONCLUSION

The proposed prototype successfully detected and classified emergency sounds such as a siren, a dog barking, and a car honking. In the event of an emergency, the user will be notified by means of vibration and the direction of the emergency sound will be displayed on the wristband. In conclusion, the trained model achieved an accuracy of 90%, while real-time testing of the prototype in both indoor and outdoor environments produced an average accuracy of 77%. This work can be further improved by using the large dataset for training the model and noise reduction techniques to minimise the effects of ambient noise. For future work, it is recommended to train the model with larger dataset in order to improve its accuracy. Ambient noise is also another factor causing performance in outdoor environments to decline. A better noise-filtering system is suggested to reduce ambient noise.

ACKNOWLEDGEMENT

The authors express their appreciation for the financial support provided by PGRC of University of Wollongong Malaysia.

DECLARATION OF COMPETING INTEREST

None.

REFERENCES

Abdulhamied, R. M., Nasr, M. M., & Abdulkader, S. N., 2023. Real-time recognition of American sign language using long-short term memory neural network and hand detection. *Indonesian Journal of Electrical Engineering and Computer Science* 30(1): 545-556.

Adeyanju, I. A., Bello, O. O., & Adegboye, M. A., 2021. Machine learning methods for sign language recognition: A critical review and analysis. *Intelligent Systems with Applications* 12: 200056.

Bisgaard, N., Zimmer, S., Laureyns, M. & Groth, J., 2022. A model for estimating hearing aid coverage world-wide using historical data on hearing aid sales. *International Journal of Audiology* 61(10): 841-849.

Graydon, K., Waterworth, C., Miller, H., & Gunasekera, H., 2019. Global burden of hearing impairment and ear disease. *The Journal of Laryngology & Otology* 133(1): 18-25.

Le, T. N., Straatman, L. V., Lea, J., & Westerberg, B. 2017. Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. *Journal of Otolaryngology-Head & Neck Surgery* 46(1): 41.

Mielke, M., & Brueck, R., 2015. Design and evaluation of a smartphone application for non-speech sound awareness for people with hearing loss. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 5008-5011.

Mirzaei, M., Kán, P., & Kaufmann, H., 2020. EarVR: Using ear haptics in virtual reality for deaf and Hard-of-Hearing people. *IEEE Transactions on Visualization and Computer Graphics* 26(5): 2084-2093.

Mohd Ramzi, I. A. K., Mohd Saufi, M. S. R., Md. Zain, M. Z., Ab Wahid, M., Ab Talib, M. H., Waziralilah, N. F., & Hassan, K. A. 2024. Malaysian Sign Language Detection with Convolutional Neural Network for Disabilities. *Journal of Advanced Research in Applied Sciences and Engineering Technology* 58(1): 33–41.

Mulfari, D., Meoni, G., Marini, M., & Fanucci, L., 2021. Machine learning assistive application for users with speech disorders. *Applied Soft Computing* 103: 107147.

Otoom, M., Alzubaidi, M. A., & Aloufee, R., 2022. Novel navigation assistive device for deaf drivers. *Assistive Technology* 34(2): 129-139.

Ramirez, A. E., Donati, E., & Chousidis, C., 2022. A siren identification system using deep learning to aid hearing-impaired people. *Engineering Applications of Artificial Intelligence* 114: 105000.

Saifan, R. R., Dweik, W., & Abdel□Majeed, M., 2018. A machine learning based deaf assistance digital system. *Computer Applications in Engineering Education* 26(4): 1008-1019.

Shahira, K. C., Sruthi, C. J., & Lijiya, A., 2022. Assistive technologies for visual, hearing, and speech impairments: Machine learning and deep learning solutions. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications* 397-423.

Spiers, A. J., & Dollar, A. M., 2016. Design and evaluation of shape-changing haptic interfaces for pedestrian navigation assistance. *IEEE Transactions on Haptics* 10(1): 17-28.

Tannenbaum-Baruchi, C., Feder-Bubis, P., Adini, B., & Aharonson-Daniel, L. 2014. Emergency situations and deaf people in Israel: Communication obstacles and recommendations. *Disaster Health 2*(2): 106-111.

Yağanoğlu, M. 2021. Real time wearable speech recognition system for deaf persons. *Computers & Electrical Engineering* 91: 107026.