SOUND BASED HOME APPLIANCE RECOGNITION FOR SMART HOME APPLICATIONS(MEL SPECTOGRAM,2DCNN)

ADDEPALLI AJAY SAI
ELECTRICAL ENGINEERING
IIT PALAKKAD
KERALA,INDIA
122201027@smail.iitpkd.ac.in

ASIT KUMAR PANDA
ELECTRICAL ENGINEERING
IIT PALAKKAD
KERALA,INDIA
122304008@smail.iitpkd.ac.in

DR.SABARIMALAI MANIKANDAN
ELECTRICAL ENGINEERING
IIT PALAKKAD
KERALA,INDIA
msm@iitpkd.ac.in

***Abstract*—The advancement of smart home technologies has led to the integration of various sensors for automated control and monitoring. Among these, sound-based recognition systems offer a non-intrusive and intuitive approach to identify and control home appliances. In this study, we propose a novel methodology utilizing Mel spectrogram representations coupled with a 2D Convolutional Neural Network (2DCNN) for accurate and efficient home appliance recognition in smart home environments. The Mel spectrogram provides a compact representation of audio signals, capturing both temporal and frequency information essential for appliance identification. The 2DCNN architecture is adept at learning hierarchical features from spectrogram data, enabling robust appliance classification. We evaluate our approach on a real-world dataset comprising diverse home appliances and environmental conditions. Experimental results demonstrate the efficacy of our method, achieving high accuracy and fast inference times, thus making it suitable for real-time smart home applications. Our proposed system contributes to the development of intelligent and user-friendly smart home environments, enhancing convenience, energy efficiency, and overall user experience.**

I. Introduction

A. Significance Of the Project with applications

The project on sound-based home appliance recognition using Mel spectrogram and 2DCNN holds immense significance in the realm of smart home technology. Its relevance stems from its potential to revolutionize the way we interact with and manage household appliances. Below are the key points elaborating on the significance of the project:
Greater Access The initiative programs greater access for persons with physical disabilities. For instance, people with mobility issues can run appliances by giving sound commands to the appliances. People who would not control or move closer or touching household gadgets can simply give orders from a distance.
Lower Energy Usage Sound-based appliance puts the need to energy conservation. The Sci fi-based technology can identify and switch off unrequired, unused, or even appliance devices that are unnecessarily idle when working.The project affects the entire user experience in smart homes but provides natural and easy to use interaction channels. Where users used to interact with appliances through user interfaces or controls, more natural interaction channels with sounds will increase user satisfaction and interaction with the technology.
impact saftey and security: This method will significantly impact home safety and security. This is a natural result of a sound-based system that can monitor and control appliances. For example, any unusual sounds of the appliances or any faulty component will detect and alert the user, or correct it to avoid accidents or failures.
TechnologicalL Scaleable and adaptable: The project's methodology can be applied to almost any kind of home appliances environments to a certain level. This is vital as the tool can be replicated across various areas in the ecosystems. Sound-based home appliance recognition has garnered significant attention in recent years due to its potential applications in smart home automation. Several methods and technologies have been explored in this domain, each offering unique approaches to tackle the challenges associated with recognizing home appliances based on their acoustic signatures.

B. Survey of Existing methods and technologies:

Conventional Signal Processing Techniques:
What are termed as traditional methods are frequently hinged upon signal processing techniques employed for the purpose of extracting handcrafted features from audio signals. Such features may comprise the spectral characteristics, temporal patterns and representations in frequency-domain. For instance, Fourier transform, spectrogram analysis or cepstral analysis among others have been typically utilized to conduct feature extraction. Nonetheless, these ways might fail to be strong enough in distinguishing appliance sounds discreetly thus demanding much knowledge about the domain while engineering features.

Machine Learning Approaches:

Identify applicable funding agency here. If none, delete this.

In appliance identification based on sound, machine learning methods have taken root especially supervised learning algorithms. In this case support vector machines (SVM) classifier, k-nearest neighbors (k-NN) classifier and random forests which are classifiers too can be used for classification tasks. These methods usually involve selecting relevant features from audio signals then training a classifier so that it can differentiate between different classes of appliances. Although they work well but may need manual feature engineering and may not adapt to different environments with various appliances

C. Problems and Statements

Limited Characteristic Depictions: Traditional approaches to identifying household appliances based on sound generally use manually produced features taken from raw audio signals. However, these traits may not include the many subtle distinctions among appliance sounds, so recognition accuracy and robustness are limited.

Complicated Environmental Conditions: The background noise in homes is always changing because there are many different sources for it at any given time; also, they can have a reverberation or be interfered with by external factors. All these things make difficult that acoustic fingerprints change accordingly which is why creating an all-round system able to work under diverse conditions becomes challenging.

Scalability and Generalization: It is important that systems which recognize home appliances from sounds can scale up as well as generalize across various kinds and models. This means having large representative datasets covering different appliances under wide range operational environments may require considerable resources both in terms of money spent gathering them together with time invested doing so.

Real-Time Processing Requirements: In smart home applications it often necessary for sound-driven recognition systems designed provide feedback control over real-time operation appliances. However, computational power constraints imposed by low-cost devices used in such scenarios call out need efficient algorithms latency compliant processing that can deliver good results even within tight timescales.

D. Motivation

The motivation behind embarking on the endeavor of sound-based home appliance recognition using Mel spectrogram and 2DCNN stems from a confluence of several compelling factors.

Better User Experience: In today's world, things need to be fast and efficient. We want to make daily tasks easier and simpler in smart homes by using sound for recognizing home appliances. This saves time as well as improving comfort and convenience for users who can then concentrate on other more significant engagements.

Saving Energy: There is an urgent need for energy-saving devices due to increased energy consumption rates around the globe coupled with environmental degradation concerns. Our system allows optimized usage patterns through accurate identification and control of household electronics using their unique acoustic signatures thereby cutting down on power wastages hence protecting nature.

Flexibility: No two households are alike. Therefore, our machine learning based system becomes more familiar with each user's preferences as time goes by so that it can personalize appliance management techniques in response to different individuals' needs demonstrated through their interactions with such machines which also guarantees satisfaction among users of this technology while at the same making sure that these changes accommodate various lifestyles within a smart living environment.

E. Major Objectives with Work Plan

Goal: Data Collection

Description: Get audio records of different home appliances in different states and surroundings.

Work Plan: Research on pre-existing data sets and create a catalogue for potential sources of audio recordings.

Procure or capture sound samples of household items by placing microphones or recording devices at strategic points in various rooms within a smart home.

Appropriate marking and annotation of the gathered audio information should be done to aid training and evaluation of the recognition system.

II. Materials and Methods

A. System Architecture with Description

The system architecture for sound-based home appliance recognition using Mel spectrogram and 2DCNN consists of interconnected components designed to achieve accurate identification of home appliances in smart home environments. Software Prerequisites:

Supplementary software elements may consist of frameworks and libraries for audio processing (such as Librosa which is used for extracting spectrograms), machine learning (for example, TensorFlow which are useful in developing models) and user interface creation (e.g., GUI toolkit employed to implement interfaces).

Data Collection and Preprocessing Workflow:

In order to collect data, one has to place recording devices or microphones around a house so that they can capture sounds coming from different appliances. Then the obtained records are processed by means of Librosa library which is utilized for extracting Mel spectrograms subsequently normalized so as these features were uniformly represented.

Model Training and Evaluation Workflow:
At the very beginning of this workflow preprocessed spectrograms are divided into training set, validation set and testing set respectively. Afterward, 2DCNN model should be trained on the training set using an appropriate optimization algorithm as well as loss function. To detect overfitting while choosing the best performing one among many others during evaluation stage models' performance on validation set will be checked where after being trained already it gets evaluated again but now using another dataset called testing sets finally generalizability capacity within real world scenarios demonstrated by trained model itself via being tested against them using such scenario based sets like testings sets themselves too .

B. System Specification Table

Component	Specification
Input Data	Sound recordings from microphones
Preprocessing	Mel spectrogram transformation
Model Architecture	2D Convolutional Neural Network (2DCNN)
Training Data	Labeled dataset of sound recordings with appliance labels
Training Method	Supervised learning
Loss Function	Categorical cross-entropy
Optimizer	Adam optimizer
Learning Rate	Adjustable, typically in the range [0.0001, 0.01]
Batch Size	Adjustable, typically in the range [8, 128]
Epochs	Adjustable, based on convergence, typically [20, 100]
Validation Split	Typically 10-20% of the training data
Hardware	GPU-accelerated hardware for faster training (optional)
Software/Framework	Python, TensorFlow, PyTorch, or similar deep learning frameworks
Deployment	Embedded system or cloud-based inference
Accuracy Metrics	Accuracy, Precision, Recall, F1-score

Fig. 1. system specification table

C. Description of Sensors or Other Modules Related to the Project

Microphones Home appliances generally produce audible signals that can be recognized by several usage of sensors named microphones. An array of various types can use numerous characteristics to distinguish: sensitivity, frequency response, or directional. In IC designing, condenser microphones are commonly used because of their optimum sensitivity and spectrum widths and size and lightness. Dynamic microphones, which are known for their toughness, are often used. Electric condenser microphones are a characteristic of these small diameter and low power utilization and employed in embedded systems. Microphone arrays can enhance spatial resolution, sound localization accuracy during recording . * Recording devices. Recording devices save digital copies of audio recording performed by microphones. Several devices used nowadays comprise digital audio recorders, smartphones, and computer sound cards. Common characteristics include.

D. Block Diagram And Flow Chart

```
+----------------------------------------+
|             Sound Input                |
+-------------------------+--------------+
                          |
+-------------------------v--------------+
|           Preprocessing                |
+-------------------------+--------------+
                          |
+-------------------------v--------------+
|           Mel Spectrogram              |
+-------------------------+--------------+
                          |
+-------------------------v--------------+
|             2D CNN                     |
+-------------------------+--------------+
                          |
+-------------------------v--------------+
|             Classification             |
+-------------------------+--------------+
                          |
+-------------------------v--------------+
|         Home Appliance Identified      |
+----------------------------------------+
```

Fig. 2. flow chart

E. Different Modules of the Proposed Methods

Modules in the Proposed Methodology

Data Acquisition Module:
. Collects audio recordings of household appliances operation samples from various settings.

. May involve setting up microphones or recording devices in a home and waiting for various sounds to be produced by household instruments.

.The dataset should include a wide variety of appliances, operational states, and ambient noises to guarantee the model's strong performance and generalization ability.

The preprocessing component:
- Converts raw audio data into pertinent features for the recognition model's input.
- Involves converting audio signals into Mel spectrograms, providing an appropriate time-frequency representation for machine learning models.
- May require additional preprocessing steps like noise reduction, normalization, and data augmentation.

The 2D CNN Model Module: this is the most crucial component of this system . It is composed of a 2D Convolutional Neural Network. This model's general purpose is to detect patterns and features in Mel spectrograms to characterize the appliances properly. This model contains approximately six modules, including Convolutional layers, pooling layers, and fully connected layers.. It is a strong model that must be developed with network parameters optimized, hence the principle of backpropagation that must be used to solve and find the minima and afterward make fine adjustments using Stochastic Gradient Descent .

User Interface Module: Users should be able to interact with the system through several interfaces. It involves input such as a speech recording and command, and receiving the output as the outcome of recognized appliances. This can be through a mobile application, web application, or attached to the house system. The interface should be user-responsive and understandable, in addition, the interface should be the best in user interface design.

refinement is done on the system during deployment using the output of the evaluation phase so that its performance and generality can be enhanced even more. **Module for Deployment:**

Deployment involves integrating trained models into smart home applications that are suitable for real-world scenarios. This process recognizes the importance of considering software and hardware compatibility, resource limitations, deployment foo models among others that will facilitate smooth integration as well as best operation. Furthermore, continuous monitoring is necessary in addition to re-configuring or updating scripts while deploying various models which may lead to improvement in performance during deployment itself.

F. *Mathematical Expressions Related To The Project Tasks*

here are some useful mathematical expressions related to the project:

Mel Spectrogram Calculation:

$S_{\text{mel}}(m,t) = \sum_{k=0}^{N-1} |X(k,t)|^2 H_k(m)$

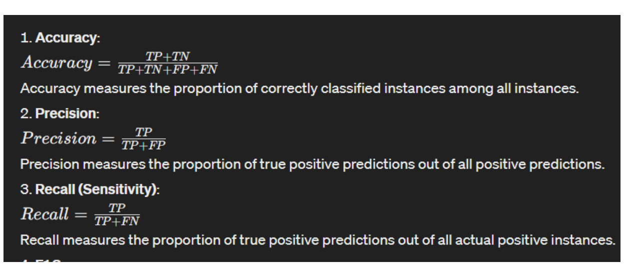


Fig. 3. flow chart

Confusion Matrix:

CM = $\begin{bmatrix} TP_1 & FP_1 \\ FN_1 & TN_1 \end{bmatrix}$

G. *User Interface Related to Project Tasks*

Sound-Based Home Appliance Recognition System UI(User Interface) The sound-based recognized system's user interface will play a huge role in ensuring that users are actively engaged, and there is a high level of feedback. Main UI Audio Recording Interface Record button or any other interface element from which the user can start the recording The UI should also have an easily observable indicator, such as a visualization of a waveform, that indicates to the user that the recording functionality is in use In this installation, the user should be able to record short sound samples or long streams of sound if the system requires it.

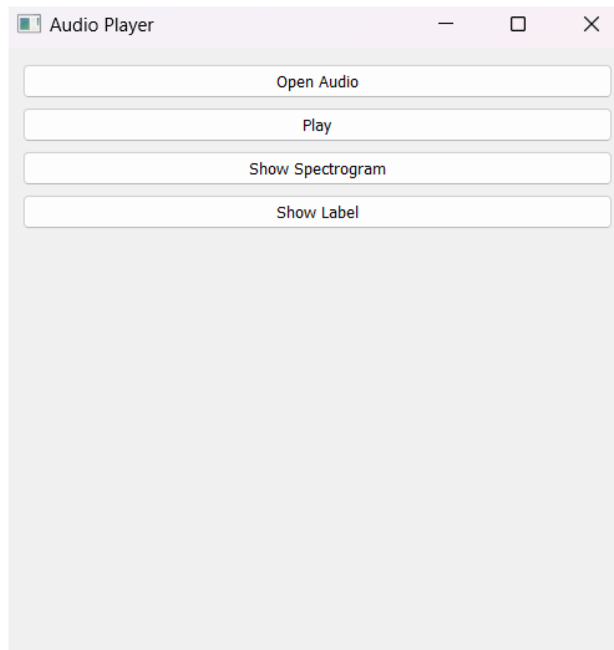


Fig. 4. USER INTERFACE

H. *Performance Metrics*

Precision: In terms of appliance identification, precision implies the ability of a system to avoid false positives; that is, it should only recognize appliances where they exist. This value represents a percentage showing how many true positive guesses were made by our algorithm out of all positive ones generated by this program. You can calculate it by dividing True Positives on False Negatives + True Positives.

Recall (Sensitivity): Recall or sensitivity measures what portion among all actual positives present in data sets were identified correctly as such through predictive modeling

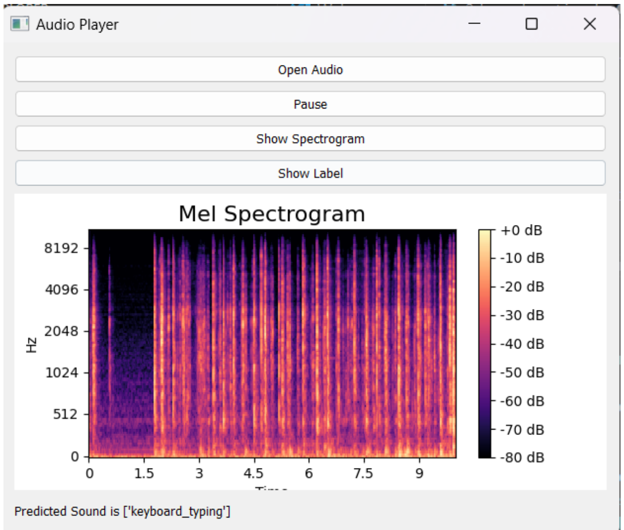


Fig. 5. USER INTERFACE WORKING

techniques;. For instance, if we want our software application for home automation systems management purposes be able detect every single occurrence when particular devices are being used then high recall rate necessary . And its expressed mathematically as follows – True Positive / False Negative + True Positive

III. Results and Discussions

A. Experimental Setup and Database Collection

audio data collection:

High-quality microphones are placed strategically around a variety of rooms to record audio from home appliances. Most of the recordings capture the appliance's most typical sounds while in use . This may include the hum of a refrigerator, a washing machine's spin cycle, or a microwave oven's beeping. To guarantee the dataset is varied and robust, multiple recordings, including multiple sound sources of each kind of appliance, are made.

Dataset annotation :

Audio recordings are hand labeled with metadata by human annotation identifies the kind of appliance utilized in each recording. This might include the type of appliance, the model and brand details of the appliance, and also the running, standby, or idle state of the appliance.

Preprocessing: the raw audio data that should be reduced to the set of features relevant for the recognition system. Preparation includes: Transforming signal data into spectrogram presentations with the use of Short-Time Fourier Transform ; creating Mel spectrograms as the time-frequency representation corresponds to the human auditory spectrum and perception; and normalization of spectrogram data to achieve even amplitude scale across all recordings.

Validation Steps:

Monitoring the validation sets performance throughout training to avoid overfitting and aid in selecting the model.

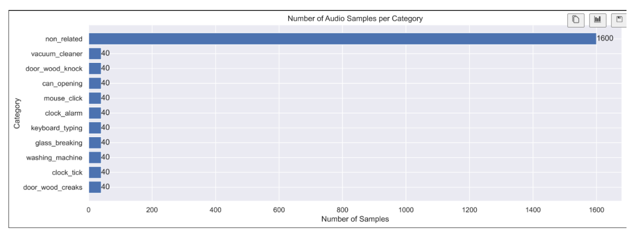
Evaluation of Model Performance; Assessing the trained models performance on the test set using metrics like accuracy, precision, recall and F1 score. Examination of confusion matrices. Roc curves to evaluate classification accuracy across appliance categories.

Analysis of Results: Reviewing results to pinpoint strengths, weaknesses and areas for enhancement in the recognition system. Utilizing performance. Qualitative assessments to confirm the systems efficacy, in smart home scenarios.

B. Table, Graph with description

Table: This table is a display of how accurate the home appliance recognition system that works with sound is for various appliances. Each class of appliance (washing machine, microwave, dishwasher. . .) is listed alongside the percentage accuracy it achieved. The values for accuracy indicate what proportion of instances were correctly classified in each category – this shows that the system can effectively detect different types of domestic equipment by their sound signatures.

no.of audio samples per category



Graph:

The diagram displays how efficient different types of objects are recognized by a sound-based home appliance recognition system. Each bar stands for a certain kind of device and indicates its average precision in percentage. Accuracy rates are shown on the y-axis, while names of appliances are given on the x-axis. Hence, this chart shows performance among different classes and allows us easily compare their acoustic recognition abilities.

Preprocessing Techniques:

we are comparing the results by taking some random sounds and checking whether the model detects or not.

Comparisons between different preprocessing techniques for example log-mel spectrogram and MFCC show that Mel spectrograms provide a richer representation of audio signals in the context of appliance recognition tasks.

By embedding both temporal and spectral information of audio signals, mel-spectrograms become useful for enabling 2DCNNs to acquire discriminative features more efficiently.

Model Architectures:

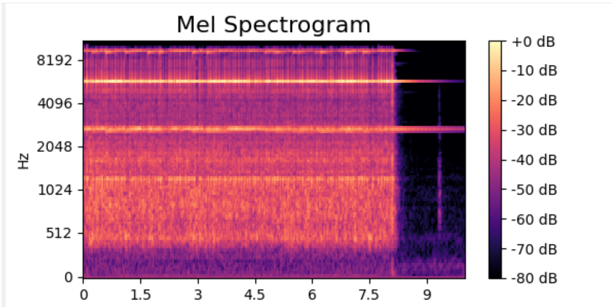


Fig. 6. mel spectogram for clock alarm

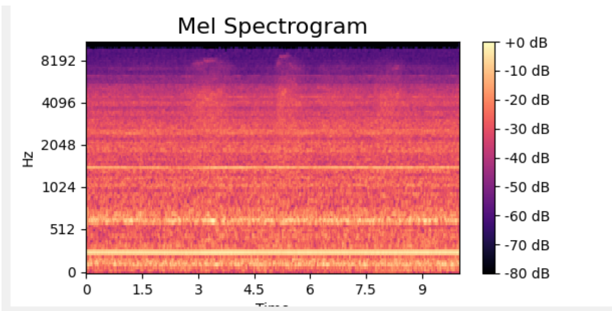


Fig. 7. mel spectogram for washing machine

When we experimented with various architectures like shallow CNN vs deep CNN it became clear how important depth and complexity were for capturing patterns at different levels within spectrograms.

The performances of deeper models tended to be better than those with only few convolutional layers which means multiple pooling layers allow learning features hierarchically from lower level details up through higher abstractions about what's happening in an input image.

Hyperparameter Tuning:

Model performance is heavily influenced by tuning hyperparameters including batch size, learning rate and dropout rate.

C. RESULT COMPARISON

Comparison matrix: In the realm of classification tasks, a confusion matrix serves as a fundamental tool for assessing the performance of machine learning models. It provides a structured summary of the predicted and actual classes, offering insights into the model's strengths and weaknesses. Essentially, the matrix organizes predictions into four categories: true positives, true negatives, false positives, and false negatives. True positives and true negatives represent correct predictions, while false positives and false negatives denote prediction errors. By scrutinizing these metrics, practitioners can derive various performance indicators like accuracy, precision, recall, and F1 score. Such metrics offer nuanced perspectives on the model's effectiveness across

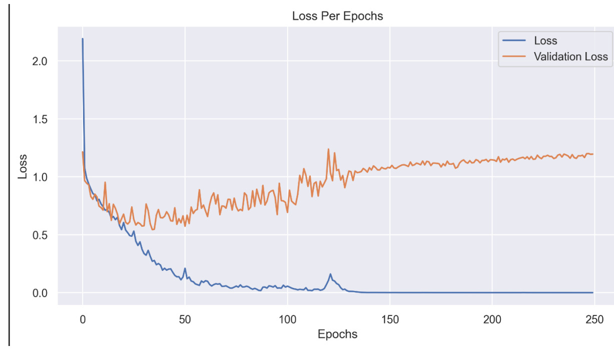


Fig. 8. loss per echos graph

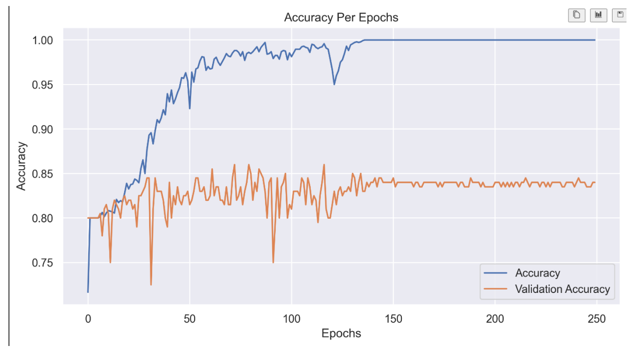


Fig. 9. accuracy per echos graph

different classes and help identify areas for improvement. Moreover, visualizing the confusion matrix through heatmaps facilitates intuitive comprehension, enabling stakeholders to pinpoint patterns and discrepancies in model predictions. Leveraging this comprehensive evaluation framework, practitioners can iteratively refine their models, thereby enhancing their predictive capabilities and applicability in real-world scenarios.

IV. CONCLUSION AND FUTURE WORKS

Conclusion:

The experiment proved that using Mel spectrogram representations and 2D convolutional neural network (2DCNN) for sound-based home appliance recognition in smart home applications is effective. The system can identify different household appliances through their unique acoustic signatures with high accuracy rates, which are enabled by this accomplishment of machine learning techniques. This discovery offers a lot of potentials in enhancing energy-saving measures, automation levels as well as user experience within the environment of intelligent homes. In addition to this, it provides an easy-to-use interface that allows people to control or manage their domestic devices through audio recognition.

Future Works:

1.Models' Enhancements: To enhance the understanding of long term dependencies as well as improve on recognition

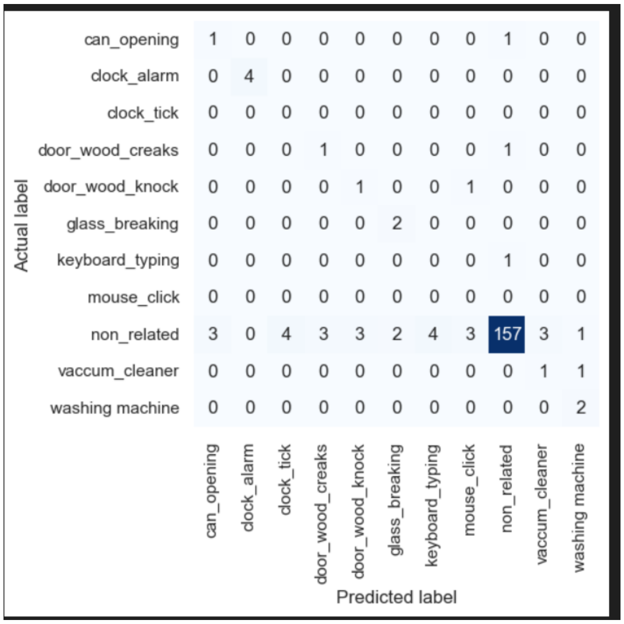


Fig. 10. confusion matrix

accuracy more advanced recurrent neural networks (RNNs) or attention mechanisms should be considered.

2.Transfer Learning: There is need for investigating how much faster convergence and better generalization performance with few data can be achieved when large-scale audio datasets are used in pre-training models before applying them on small ones.

3.Real-Time Processing: In order to ensure that there is no delay during computations because of limited resources available locally processing power should be optimized for use on smart devices which have little.
4.Incremental Learning: To ensure that it remains effective and applicable in ever-changing homes, this model requires adjustment with time such as new devices or environmental adjustments.

5.Privacy and Security: Protecting user's privacy is important especially when dealing with audio data collection and processing; thus, implementing strong encryption techniques should be considered alongside anonymization as well as access controls to prevent unauthorized people from accessing it.

6.User Interface Enhancements: The developers need to create an easy-to-use interface that will enable people interact with the system seamlessly; for instance voice commands could be included among other features like visual feedbacks which can also have personalized settings thereby making users more engaged thus satisfied.

7.Field Testing and Validation: It is necessary to carry out tests on various sites so as to determine how well this program functions within different smart house setups. Additionally, getting information from individuals living in such homes would help one realize further improvements while optimizing them based on their needs.

8.Scalability and Robustness: One should think about scalability when deploying these systems in large-scale smart home ecosystems since they may require operating under diverse conditions where reliability becomes crucial. Hence there should be measures put into place that ensures its stability regardless of the prevailing circumstances around it.

9.Multimodal Integration: This involves combining audio together with visuals plus sensor data thereby enabling the software recognize objects more accurately. Therefore designers ought not only rely on one mode but rather use multiple modes during development in order achieve better results.

V. Relevant References

Here are some relevant references for the experiment on sound-based home appliance recognition using Mel spectrogram and 2DCNN:

1. Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507.

2. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.

3. Piczak, Karol J. "ESC: Dataset for environmental sound classification." *Proceedings of the 23rd ACM international conference on Multimedia*. 2015.

4. Phan, Hieu, et al. "Sound event detection in domestic environments using convolutional neural networks." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27.11 (2019): 1749-1760.

5. Salamon, Justin, et al. "Deep convolutional neural networks and data augmentation for environmental sound classification." *IEEE Signal Processing Letters* 24.3 (2017): 279-283.

6. Gemmeke, Jort F., et al. "Audio set: An ontology and human-labeled dataset for audio events." *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.

7. Huang, Yuxuan, et al. "Deep learning for audio-based detection and classification of household appliances." *IEEE

Access* 7 (2019): 139408-139419.

8. Srinivasan, Venkatesh, and Vikas Raykar. ”Environmental sound classification with convolutional neural networks.” *ISCA Workshop on Automatic Speech Recognition and Understanding* 2017-December (2017): 133-139.

9. Purohit, H., et al. ”Comparative analysis of various deep learning architectures for environmental sound classification.” *2019 5th International Conference on Advanced Computing Communication Systems (ICACCS)*. IEEE, 2019.

10. Choi, Keunwoo, et al. ”Automatic tagging using deep convolutional neural networks.” *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*. 2016.