ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

SMART VOICE LOCK USING MACHINE LEARNING

Aman Panchal, Aakash Yadav, Ashish Meena, Ajay Joshi, Dheeren divya, Prachi Goyal
E-Mail Id: pra05chi@gmail.com
Department of Computer Science & Engineering, AIET, Jaipur, Rajasthan

Abstract- In the fast-growing world, the AI companies are mainly focusing on more contactless after covid. Earlier to lock and unlock any lock system was totally manual. We have to always carry keys which were very uncomfortable to carry everywhere and there were chances of getting lost or duplicated. But as of technology advancement the new alternative is coming for user identification. Such as: fingerprint, eye scanner, faces recognition, etc. And sound is also one of them because every person has different voice. This is also more secure and more reliable. We present novel audio dataset which we can use for classification task on particular user. And perform various Machine Learning and Deep Learning algorithms for audio classification and speech recognition.
Keywords: AI, Machine Learning, Deep Learning, Audio Classification, Speech Recognition.

1. INTRODUCTION

This new emerging world of AI and Machine Learning, which has potential to globally change the human lie in positive way. Many applications around the world have been implementing different operation, based on the background of ML, and also offer many services for smart services. ML has become an efficient tool to build smart cities. As we can see everywhere increase graph of crime and the technique used during crime is also increases, we should also have to modify our systems to increase or security. So, we start to build a device which uses person audio to lock and unlock a lock system.

This device used user specific voice to lock and unlock. Speech is the most natural way of communicating for human beings. Advancement in digital signal processing technology has been the use of speech in many different areas of application such
as compression of speech enhancement, synthesis and recognition. In this, using speech processing voice
lock is accomplished through the development of a system of voice recognition using MFCC approach.

2. OBJECTIVE AND RELATED WORK

The objective of the project is to designs an IOT based system of smart voice lock with the model of speech recognition using MFCC extraction. This device lock/unlock on user voice.

Related work:

- Many electronics companies used this type of system for controlling home appliances. Such as: lights, washing machines, etc.
- This type of technology used in smart cars which interact with user and perform many tasks.
- Mobile companies use this technique in their devices so their customer control and operate their phones with their speech.

3. DATASET AND METHODS

3.1 Audio dataset

The audio dataset is a small dataset consist of 200 audio files of English sentences of 25 lines with 4 repetitions of each line for each of two different speakers. Audio average .5 seconds of length. Each of this audio is labeled with two labels person1 and person2. And the dataset is labeled so we can easily perform supervised machine learning algorithms.

Table-3.1 Example labels from the audio dataset

filename	**class**
WhatsApp Audio 2021-06-09 at 7.33.	akash
22 PM (1)	
WhatsApp Audio 2021-06-09 at 7.02	akash
.59 PM (1)	
…	…
Recording (99)	aman
Recording (100)	aman

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

3.2 Methodology

The step-by-step design approach of the proposed system and workflow of complete system have been mentioned below:

- Record audio to create audio dataset in order to train our machine learning algorithms.
- Convert audio into values using mfcc.
- Comparisons of various algorithms on accuracy and performance.
- Selection of best algorithm from performance characteristics.
- A full fledged tentative design Arduino with well connected small LED and microphone.
- Now user can input her voice and result will be showed on LED.

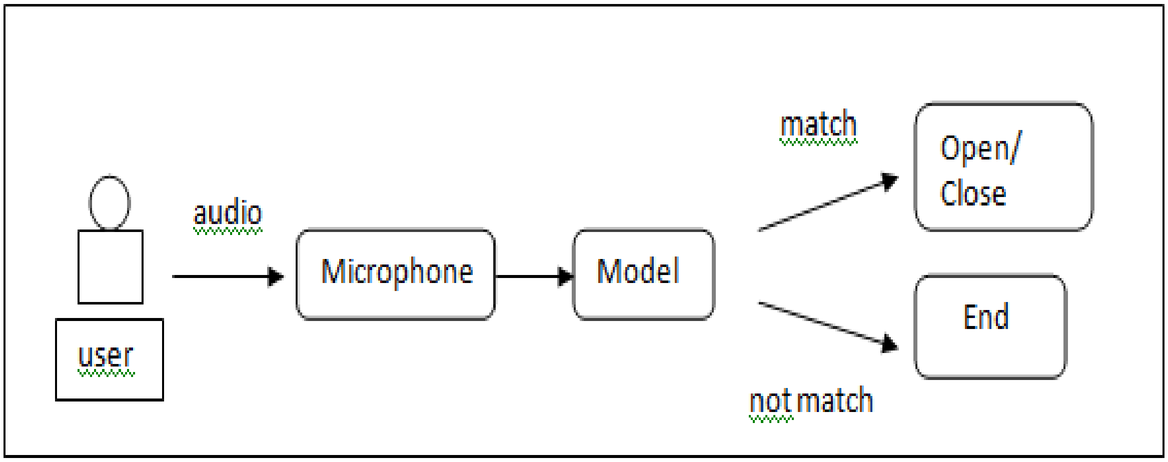

Fig. 3.1 Working of device

4. EXPERIMENTAL FRAMEWORK

4.1 Training

First, we take balanced audio dataset of 200 audio files equally 100 of each two different person. And a csv file containing audio files name and another column of class. Then created augmentation to upscale the dataset. Now use mfcc library. And by using this library we convert these audio files into numerical values and extract features.
Then divide this into two for training and testing into 80:20 ratios. Training part is used for different machine learning algorithm.

4.2 Architecture

4.2.1 Logistic Regression

In this we use l2 regularization, tolerance level .0001,
lbfgs as an optimization, random state to 32 and maximum iteration of 32.

4.2.2 SVM

In this we use kernel as rbf, gamma value is 0.1, value of C to 1.0, degree equal to 3 and random state of 32.

4.2.3 Random Forest

In this the parameter used are number of estimators are 1, maximum depth of a tree should be 2, criterion used was gini, minimum sample leaf was 1 and the random state was 32.

4.2.4 DNN

We use deep neural network and use sequential model. In first layer use 100 starting node and RELU activation with dropout of 0.5. And 2 hidden layer one with 200 nodes and another with 100 node and RELU activation function with 0.5 dropouts. And final layer 2 node with softmax functions. Categorical cross entropy was a loss function used and optimizer was Adam.

5. RESULT

Model performance is summarized in Table 5.1 in terms of training accuracy and testing accuracy. All models have different accuracy due to different –different ways they perform and parameters used. Data can be showed in below table.

ISSN Number: 2454-2024(Online)

International Journal of Technical Research & Science

Table-5.1 All model accuracy

Architecture	Training Accuracy	Testing Accuracy
Logistic Regression	85.01	85.53
SVM	100	71.38
Random Forest	97.66	97.43
DNN	85.09	98.07

CONCLUSION

In this work, we learn how to use multiple audio formats and perform various machine learning algorithms. The result, how dataset is handle and features extracted from audio. Then dataset is divided in training and testing. Architecture and different parameters used in various machine learning algorithms.

REFERENCES

[1] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe,K. Hansen, and K.-R. Muller, “How to explain individual clas-¨sification decisions,” Journal of Machine Learning Research,vol. 11, no. Jun, pp. 1803–1831, 2010.

[2] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, “Unsupervisedfeature learning for audio classification using convolutional deep belief networks,” in Advances in Neural Information Processing Systems (NIPS), 2009, pp. 1096–1104.

[3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed,N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath et al. , “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[4] M. Anusuya and S. K. Katti, “Speech recognition by machine;a review,” International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 181–205, 2009.

[5] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, Software available from tensorflow.org.

[6] R. Grosse, R. Raina, H. Kwong, and A.Y. Ng. Shift-invariant sparse coding for audio classification. In UAI, 2007.

[7] Vittal Kumar Mittal, Manish Mukhija, “Cryptosystem based on modified Vigenere Cipher using Encryption Technique”, International Journal of Trend in Scientific Research and development, Vol.3, Issue 5, pp. 1936-1939, ISSN: 2456-6470, August 2019.

[8] Swati Bhargava, Manish Mukhija, “Hide Image and Text Using LSB, DWT and RSA based on Image Steganography”, ICTACT Journal on Image and Video Processing, Vol.09, Issue 3, pp. 1940-1946, February 2019, ISSN: 0976-9102, DOI:10.21917/ijivp.2019.0275.

[9] Chen L, Mao X, Xue Y-L, Cheng LL (2012) Speech emotion recognition: features and classification models. Digital Signal Process 22(6):1154–1160

[10] Goodfellow IJ, Bengio Y, Courville AC (2016) Deep learning (Adaptive Computation and Machine Learning series). MIT Press.

[11] Vishal Pratap Singh, Manish Kumar, Himanshu Arora, “Enhanced image security technique with combination of ARNOLD transformation and RSA algorithm”, International Journal of TEST engineering and management, Vol.83, pp. 30550-30560, May/June, 2020, ISSN: 0193-4120 (Scopus).