European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 02, 2020

An Emotion Recognition Based Fitness Application for Fitness Blenders

M. Suresh Anand[1], Nisha G Mathur[2], Angappan Kumaresan[3], J.K. Periasamy[4]

[1,2,4]*Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India.*

[3]*Department of Information Technology, School of Computing Sciences, HITS, Chennai, India.*

Abstract: The primary objective of the proposed work is to recognize emotions through speech using Recurrent-Neural-Networks and suggest physical exercises accordingly for the beneficiary of people who prefer to choose a balanced life style in an efficient way. Recurrent-Neural-Network (RNN) is a classifier technique which is utilized to classify various emotions which are happiness, disgust, sadness, fear, surprise and anger. Culmination of multiple features on similar databases are compared and clearly discussed. The outcome gives good result. At the outset the experimental results would shows that combining feature will yield higher accuracy rate on emotional databases using RNN classifier. Bayes Classifier, Hidden Markov Model (HMM), Support Vector Machine (SVM), Kernel deterioration and K Nearest Neighbors approach (KNN), Gaussian Mixture Model (GMM), Naïve-Bayes-classifier(NBC) were classifiers used by many researchers for human emotion recognition and translation. The main disadvantage is that only single input and corresponding output is given and combinations of output is a tough task and the algorithm gets costly. Our proposed system is used to suggest the physical exercise from speech. Energy, format, pitch, few spectrum features Mel Frequency Ceptral Coefficients (MFCC) and spectral Modulation features are the various common features extracted and used in modern research. In this work to extract emotional features the modulation spectral features used. To analyze, recognize and classify the emotions from audio speech samples the Standard English emotional database are used. This system will be helpful to people who need guidance in recognizing their emotion and to get a better physical exercise suggestion.

Keywords: Emotion Recognition, Fitness, Blenders, Recurrent-Neural-Networks.

1. INTRODUCTION

The proposed system's main objective is to recognize emotions through speech using Recurrent-Neural-Networks and suggest physical exercises accordingly for the beneficiary of people who prefer to choose a balanced life style in an efficient way. RNN i.e. Recurrent Neural Networks classifier is used to classify the emotions like sadness, happiness, disgust, fear, surprise and anger.

The results for products mix of the highlights and on comparative information bases are looked at and clarified. The general exploratory outcomes would uncover that the component mix has the most noteworthy precision rate on enthusiastic information bases utilizing RNN classifier. Numerous analysts have utilized various classifiers for human feeling recognition and classification from speech audio such as Bayes Classifier, Hidden Markov Model (HMM), Support Vector Machine (SVM), Kernel deterioration and K Nearest Neighbors approach (KNN), Gaussian Mixture Model (GMM), Naïve-Bayes-Classifier (NBC), etc. The main disadvantage is that only single input and corresponding output is given and combinations of output is a tough task and the algorithm gets costly. Our proposed system is used to suggest the

physical exercise from speech. In late exploration, numerous regular highlights are extricated, for example, pitch, configuration, energy and some range highlights, for example, Mel Frequency Ceptral Coefficients (MFCC) and unearthly Modulation highlights. In this work, Modulation phantom highlights will be utilized, to remove the enthusiastic highlights. Discourse tests from Standard English enthusiastic information base are utilized in breaking down, recognition and grouping of feelings from sound examples. This system will be helpful to people who need guidance in recognizing their emotion and to get a better physical exercise suggestion.

2. RELATED WORK

Alexis martinez, Shichuan alluded that in psychological science and neuroscience, [1] there have been two driving models portraying how people see and order outward appearances of feeling—the nonstop and the downright model. The ceaseless model characterizes every outward appearance of feeling as a component vector in a face space. This model clarifies, for instance, how articulations of feeling can be seen at various forces. Interestingly, the unmitigated model comprises of C classifiers, each tuned to a particular feeling classification. This model clarifies, among different discoveries, why the pictures in a transforming arrangement between an upbeat and an unexpected face are seen as either glad or shock however not something in the middle. While the constant model has a more troublesome time legitimizing this last finding, the absolute model isn't as acceptable with regards to clarifying how articulations are perceived at various forces or modes. In particular, the two models have issues clarifying how one can perceive mixes of feeling classifications, for example, cheerfully astounded versus irately amazed versus shock. To determine these issues, in the previous quite a long while, we have dealt with a re-examined model that legitimizes the outcomes detailed in the psychological science and neuroscience writing. This model comprises of C particular constant spaces. Numerous (compound) feeling classifications can be perceived by directly consolidating these C face spaces. The components of these spaces are demonstrated to be generally configurable. As indicated by this model, the significant errand for the arrangement of outward appearances of feeling is exact, definite location of facial milestones instead of acknowledgment. We give a review of the writing supporting the model, show how the subsequent model can be utilized to construct calculations for the acknowledgment of outward appearance of feeling, and propose research bearings in AI and PC vision specialists to continue pushing the best in class in these zones. We likewise examine how the model can help in investigations of human discernment, social communications and issues.

Advantage Alexis Martinez, Shichuan system could be Fuzzy logic and neural network based approach has improved overall efficiency. Same way disadvantage may be perception is used that may not be appropriate always.

Nithya Roopa has said that outward appearance recognition is the piece of Facial recognition [2] which is increasing more significance and requirement for it increments hugely. Despite the fact that there are strategies to distinguish expressions utilizing AI and Artificial Intelligence procedures, this work endeavors to utilize profound learning and picture arrangement strategy to perceive expressions and characterize the expressions as indicated by the pictures. Diverse datasets are inspected and examined for planning articulation recognition model are explained in this paper. Origin-Net is used for demeanor recognition with Karolinska Directed Emotional Faces and Kaggle-Facial Expression Recognition Challenge datasets. Last precision of this articulation recognition model using Inception Net v3 Model is 35%(~). Document Terms—Facial recognition, demeanor recognition significant learning, picture recognition,

Facial development, signal dealing with, picture portrayal. Advantage Nithya Roopa work is proved efficient in many ways and Disadvantage is Quite Expensive method.

People share a widespread and central arrangement of feelings which are displayed through steady outward appearances said James Pao[3]. A calculation that performs location, extraction, and assessment of these outward appearances will take into consideration programmed acknowledgment of human feeling in pictures and recordings. Introduced here is a crossover highlight extraction and outward appearance acknowledgment technique that uses Viola-Jones course object finders and Harris corner key-focuses to extricate countenances and facial highlights from pictures and uses head part investigation, direct discriminant examination, histogram-of-oriented gradients (HOG) include extraction, and support vector machines (SVM) to prepare a multi-class indicator for characterizing the seven essential human outward appearances. The half breed approach takes into account brisk starting grouping by means of projection of a testing picture onto a determined eigenvector, of a premise that has been explicitly determined to underline the detachment of a particular feeling from others. This underlying advance functions admirably for five of the seven feelings which are simpler to recognize. In the event that further forecast is required, at that point the computationally more slow HOG highlight extraction is performed and a class expectation is made with a prepared SVM. Sensible exactness is accomplished with the indicator, subject to the testing set and test feelings. Precision is 81% with hatred, a hard to-recognize feeling, included as an objective feeling and the run-season of the half breed approach is 20% quicker than utilizing the HOG approach solely.

Monika Dubey, Prof. Lokesh Singh inspected that Emotional angles have tremendous effect on Social insight like correspondence understanding, dynamic and furthermore helps in understanding conduct part of human. Emotion assume crucial function during correspondence. Feeling acknowledgment is done in assorted manner, it might be verbal or non-verbal .Voice (Audible) is verbal kind of correspondence and Facial appearance, activity, body positions and development is non-verbal sort of correspondence. [1] When conceding just 7% impact of message is contributes by verbal part generally speaking, 38% by vocal part and 55% impact of the speaker's message is contributed by outward appearance. Subsequently robotized and predictable outward appearance would acknowledge significant part in human and machine affiliation. Outward appearance confirmation would be helpful from human working environments to clinical practices. Evaluation of outward appearance acknowledge huge capacities with respect to applications which depend after feeling certification like Human Computer Interaction, Social Robot, Animation, Alert System and Pain looking for patients. This paper presents brief presentation of outward appearance in segment I. Region II depicts six expansive outward appearances and highlights. Area III gives brief detail on near assessment of eminent procedures proposed before for Automatic Facial Emotion Recognition System. Area IV joins seasons of Automatic Facial Emotion Recognition System. Zone V joins Applications of Facial Emotion Recognition System.

Programmed facial emotion recognition is an effectively developing exploration in Emotion Recognition. Liliana expands the Deep CNN i.e. Convolution-Neural-Network approach to manage facial attitude affirmation task. This endeavor is done by recognizing the occasion of FACS-Facial Action Coding System is inherited from Action Units which addresses human inclination. In the CNN totally related layers we use a regularization methodology called "dropout" that wind up being very amazing to diminish over fitting. This assessment uses the sweeping Cohn-Kanade dataset that is accumulated for facial recognition affirmation investigate. The framework execution increment typical accuracy movement of 92.81%. The framework had been adequately assembled eight basic inclination classes. Likewise, they proposed strategy is shown to be suitable for feeling affirmation.

The imperative research part of the emotion recognition is the analysis of emotional state in the facial expression by Jothimani and et al. The subject matter of this study is to aid the human – computer interaction more empathetic with the help of automatic emotion recognition system which will be a great step forward in the robotic field. This study proposes a novel method for the emotion recognition utilized of Face detection applied Haar highlight based course classifiers, saliency planning and CNN design patterns are executed. The outward appearance of the people from the informational index is taken care of into the Saliency utilizing hyper-complex Fourier Transform (SHFT). The resulting saliency map has the extracted feature which is given as input to the CNN to perform feature modeling and output the emotional state of the human. We also exhibit that the proposed saliency model can emphasize on both minor and major salient regions more accurately than the other saliency models.

The goal of emotion recognition is recognizing emotions of a human. The emotion can be caught either from verbal correspondence face or from face by Jyostna Devi Bodapati, N. Veeranjaneyulu. In this work we concentrate in distinguishing human emotion from outward appearances. Facial emotion recognition is one of the valuable assignment and can be utilized as a base for some ongoing applications. It tends to be utilized as a piece of many intriguing and helpful applications like Monitoring security, treating patients in clinical field, advertising research, E-learning and so forth;. We people can undoubtedly recognize the emotion of different people with no exertion. Programmed location of emotion of a human face is significant because of its utilization continuously applications. The ongoing development in GPU has taken numerous applications like face recognition, manually written digit recognition and article recognition to the following level. Particularly the pre-prepared CNN based highlights better speak to the pictures.

Pre-prepared CNN highlights speak to the most discriminative highlights and subsequently considers better execution. Highlight portrayal assumes a significant part on the exhibition of any AI calculation. Subsequent to watching unimaginable execution with deep learning models, we propose to utilize the deep convolution highlights to all the more likely speak to the given picture as opposed to utilizing the customary carefully assembled highlights. The drawback of the deep learning models is that they require huge datasets to acquire better execution. To use the utilization of deep learning models without the necessity of enormous datasets is to utilize pre-prepared models. For highlight extraction pre-prepared Convolution-Neural-Networks model (VGG16) is utilized and the idea of Deep Neural Networks model is utilized for order. To show the exhibition of the proposed model, Extended Cohn-Kanade (CK+) benchmark dataset is utilized for the test contemplates. In light of the test results we guarantee that these unaided highlights better speak to the images thought about the handmade highlights.

Wisal Hashim Abdulsalam and et al had alluded that it is remarkable that understanding human outward appearances is a key part in getting feelings and finds far reaching applications in the field of human-PC communication, has been a long-standing issue. In this paper, we shed light on the utilization of a DCNN can be expanded as deep-convolution-neural-networks for facial feeling recognition from accounts using the TensorFlow AI library from Google. This work had tested with to 10 feelings by the organization called Amsterdam Dynamic Facial Expression Set Bath Intensity Variations (ADFES-BIV) dataset and tested by using 2 datasets.

3. PROPOSED SYSTEM

The Objective of the proposed system is to recognize emotions through speech using Recurrent-Neural-Networks and suggest physical exercises accordingly for the beneficiary of people who prefer to choose a balanced life style in an efficient way. Recurrent Neural Network

classifier is applied for the classification to classify the emotions such as happiness, fear, sadness, disgust, anger and surprise. Outcome for multiples culmination of the features and on similar databases were compared. Outset experimental results would shows that combining feature will yield higher accuracy rate on emotional databases using RNN classifier. Our proposed system is used to suggest the physical exercise from speech**.** In late exploration, numerous regular highlights are extricated, for example, pitch, configuration, energy and some range highlights, for example, Mel Frequency Ceptral Coefficients (MFCC) and unearthly Modulation highlights. Modulation spectral features will be used, to extract the emotional features.

Speech samples from Standard English emotional database are applied in analyze, recognize and classify emotions from data sample.

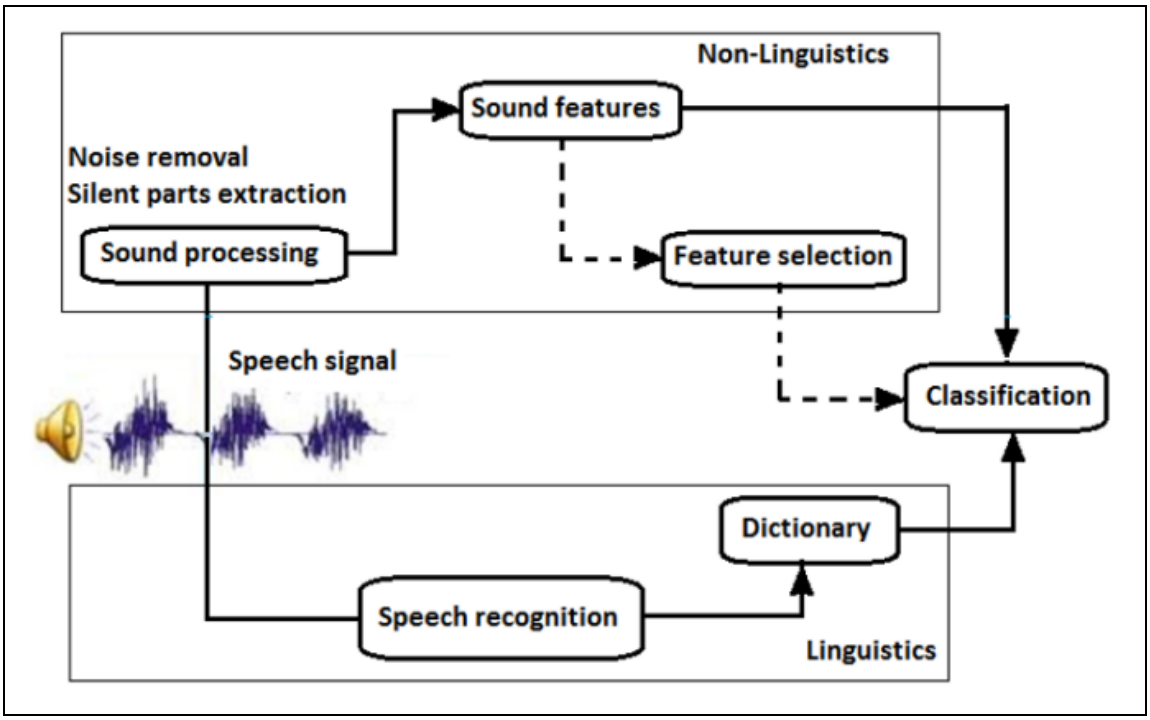

Figure 1 Architecture Diagram

Advantages

- Sound processing is done to filter the noises from the input that is in the form of speech.
- Speech Recognition is where the input speech is compared with the existing emotion database.
- Feature Comparison and Selection is the step in which the existing features provided as the sample data are compared with the current input, in order to select the category of emotion.
- Classification is where the exact emotion category is fixed and the relevant data is searched for.
- The ‘Exercise Suggestion List’ is generated by enlisting all the data fetched with respect to the current emotion recognized.

Experimental Approach, Result and Discussion

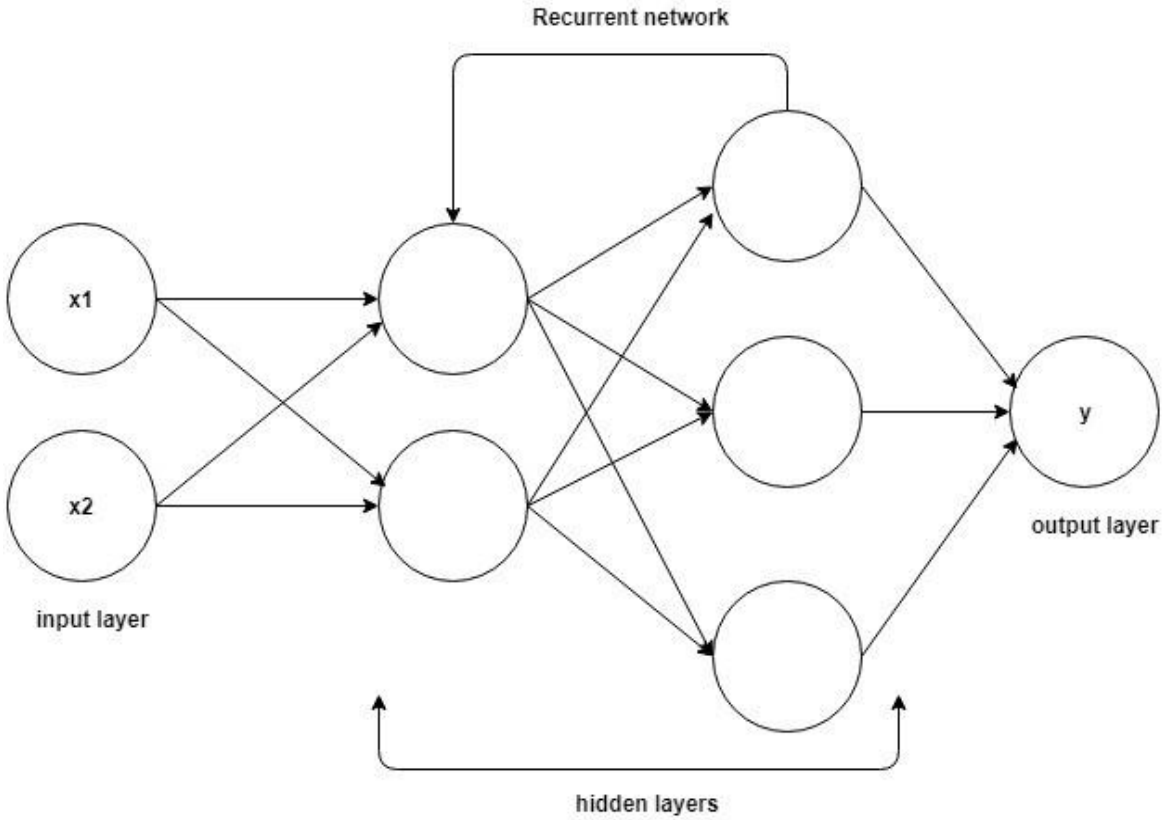

Figure 2 Recurrent-Neural-Networks Approach

Initially, the entirety of the information sound are tended to two minutes and changed over to a uniform arrangement which is mono channel, wav documents Then the ground truth is set as the normal of these abstract tests. By along these lines, preparing informational index is accessible by gathering the abstract tests and removed sound highlights for every sound example. MIR tool kit can without much of a stretch be introduced in MATLAB. This tool stash is very simple to utilize and lessens calculation intricacy. The repeat of a sound wave is what ears fathom as pitch. A higher repeat sound has a higher pitch and a lower repeat sound has a lower pitch. The pitch repeat can be dictated by using auto relationship procedure in the toolbox. The pitch seasons of a given music record is prepared by finding the postpone thinks about to the second greatest top from the central apex of autocorrelation gathering. By then pitch repeat is evaluated from the pitch time spans. The sound sign is initial rotted into hear-capable channels using bank of channels.

The envelope of each channel is eliminated. By then the envelopes are isolated, half wave changed, before being finally included again. This gives a definite portrayal of the assortment of essentialness conveyed by each event from the various hear-capable channels. The range is changed over from the recurrence area to the pitch space by applying a log-recurrence change. The circulation of the along the pitches is known as the Chroma gram. The Chroma gram is wrapped by intertwining the pitches having a place with a similar pitch classes. The wrapped Chroma gram shows a circulation of the vitality concerning the twelve potential pitch classes. The most predominant resonance is viewed as the resonance up-and-comer with highest-correlation.

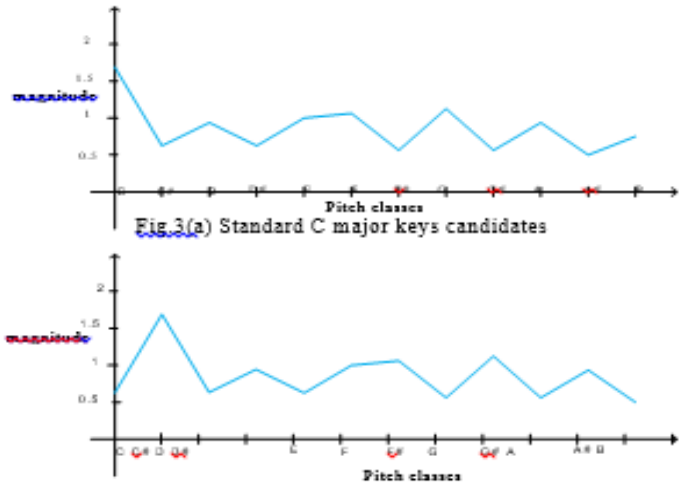

Fig 3(a) Standard C major keys candidates

Figure 3 Rotate Right-shift of the Standard C Major Key

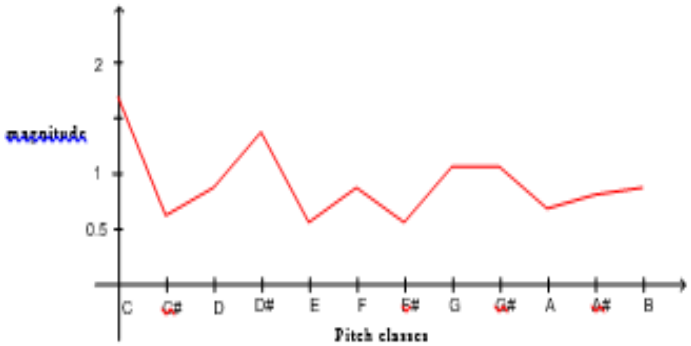

Figure 4 Graph of Line x=y

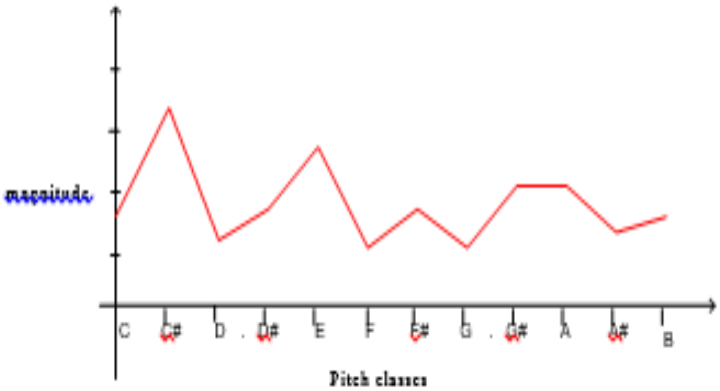

Figure 5 Standard C Minor Key Candidates

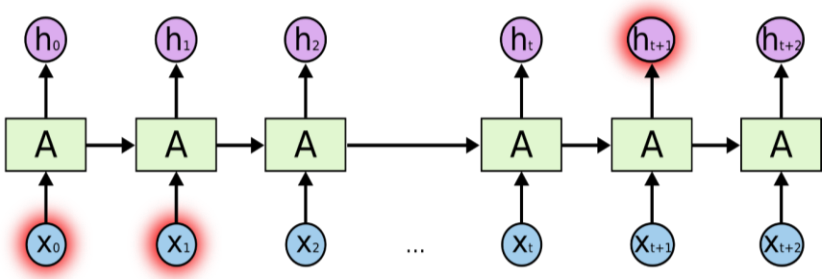

Figure 6 RNN Classification

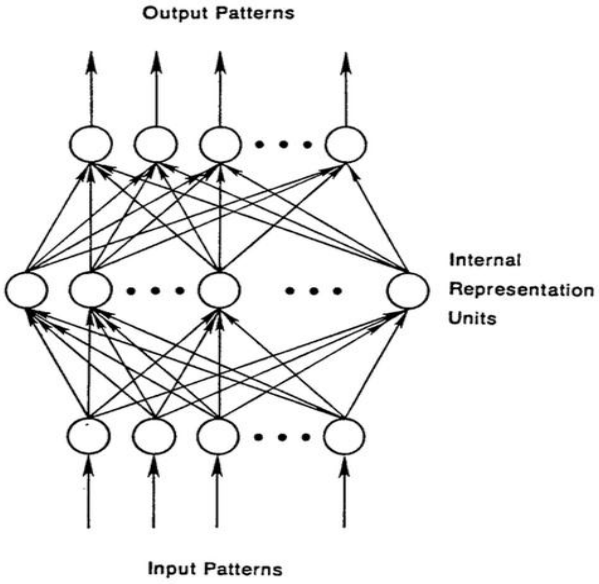

Figure 7 Input and Output Combinations in RNN

With reference to, we will give contribution to the shrouded layer at each progression. An intermittent neuron presently stores all the past advance info and unions that data with the current advance information. Consequently it additionally catches some data with respect to the connection between's present information step and the past advances. The choice at a time step t-1 influences the choice taken at time t. This is a lot of like how we as people take choices throughout our life. We join the current information with late past to accept an approach a specific issue nearby. This model is exorbitantly simple however on a basic level it lines up with our dynamic capacity.

Table 1: Extracted Audio Features

Features	**No. of Features**	**General description of audio Features**
Pitch	1	A term used to describe high or low of a note played by musical Instruments
Tempo	1	The pulse of varying strength described in terms of tempo
Tonality	1	The relation between the notes of a scale or key
Dynamic	1	The amplitude of a sound in rms energy
Total	4	

4. CONCLUSION

Now-a-days most of the people are becoming health conscious and need proper guidance to achieve good health. The proposed system would be highly useful in keeping users both physically and mentally fit. To achieve this we use the RNN algorithm in order to classify the emotions based on the audio sample provided as the input. The Mel-frequency database that is provided as training data set will be used to compare the frequencies of training set of audio samples and the user's audio sample in order to recognize the current emotional state the particular user is in. Further, for the effective detection, noise would be removed from the input audio signals using the least mean squares filter. Hence the features of the audio sample would be extracted that serves as the input parameters for the RNN algorithm to classify emotions

accordingly and finally generates a list of exercise routine to follow that would help the user balance his mental and physical state. The output has approximately shown results up to 98% accuracy. This serves to be the highest accuracy level of emotion detection using audio features among other existing systems.

5. REFERENCES

1. Kumaresan. A, Mohan Kumar. N, Suresh Anand. M and Suganya. "An Effective Recognition of Partial Speech Using Non Audible Murmur (Nam) For Speech Impaired Children". *International Journal of Control Theory and Applications,* 9(28) 2016, pp. 165-174
2. Mathavan Suresh Anand, Nagarajan Mohan Kumar and Angappan Kumaresan. "Recognition of Nam Speech-Indian English Alphabets Using Self-Designed Nam Microphone with Hmm-Viterbi". *International Journal of Control Theory and Applications,* 9(28) 2016, pp. 157-164.
3. G. Ramprabu, S. Nagarajan. "Design and Analysis of Novel Modified Cross Layer Controller for WMSN". *Indian Journal of Science and Technology,* Volume 8, Issue 4, March 2015, pp.438-444.