WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

RTP VOICE PREDICTOR

Barapatre Utkarsh Ramesh [*a] , Gursal Aniket Rangnath[*b] , Jadhav Nupoor Hanmat [*c] , Lengare Vaishnavi Suresh [*d] , Prof. Rupali Adhau[*e]

[*a,b,c,d]Final Year Students,Department of Computer Engineering,Indira College of Engineering and Management,Pune,Maharashtra,India.

A B S T R A C T :

In the processes of team recruitment, formation, instruction, assessment, and re-instruction, automatic workload and performance prediction is essential. The purpose of this study was to determine whether team workload (TW) and team performance could be predicted using speech analysis of team-based communication (TP). The labels for the TW and TP categories were objective. A computer-based command-and-control simulation that requires the dissemination of task-specific information to each team member was given to ten teams of three participants. Convolution Neural Network (CNN) models for each team was trained individually using recordings of each participant's voice communications. The TP prediction networks were trained to predict TW in the first experiment, and vice versa, the TW prediction networks were tuned to predict TP. The TP or TW prediction based on the combination of interconnected TP and TW classier was put to the test in the second experiment. The theory was supported by both experiments. It has been demonstrated that task-related prerequisite knowledge included into neural network models decreased training time and enhanced performance without growing the size of the training data set. The baseline strategy utilizing a single CNN model trained to predict either TW or TP alone beat predictions based on combined TW and TP classification outcomes (using either separate or interconnected TW or TP classifiers). The accuracy of classification was commensurate with previously reported predictions of cognitive strain based on objective measurements.

Keywords: Team Performance, Data analysis, Machine Learning.

1. INTRODUCTION

In human factors research, assessing cognitive workload (CL) is crucial, particularly in safety-critical fields like air traffic control. Cognitive workload refers to the mental effort required to complete a task, and excessive CL can hinder performance, decision-making, and learning. The link between CL and task performance has been explored extensively, leading to key theoretical advancements. With recent advancements in numerical techniques, there is renewed interest in modeling and predicting CL. Computational analysis of biosignals and audio-visual data now allows objective predictions of mental states.

This method provides valuable insights into how CL impacts human performance, with studies using both objective and arbitrary metrics to evaluate the relationship. Additionally, the connection between CL and factors like familiarity and trust can be further explored. In this context, our study introduces two new machine learning methods: a two-channel decision-making system and a pre-trained Convolutional Neural Network (CNN), both aimed at predicting team workload (TW) and team performance (TP).

2. LITERATURE SURVEY

[1] To predict team Luz santamaria-granados et al.: In this paper,Despite being designed for object recognition in images, convolutional networks showed superior performance in the detection of emotions in physiological signals when compared to the standard machine learning techniques. It was possible to identify morphological features appropriate for the prediction of affective state by the preprocessing of the ECG and GSR signal peaks as an entry vector to the CNN. The trial outcomes supported the suggested approaches and enhanced the Dataset AMIGOS's performance in emotion classification.

[2] Martin Gjoreski et al.: The participants in this study were subjected to varied levels of cognitive stress while two datasets of multimodal data were presented. To the best of our knowledge, these are the first datasets with such comprehensive sensor data combined with participant personality trait data. The experimental design used to gather the datasets included a number of cognitive activities carried out on a PC and a smartphone.

[3] Russell Li et al: In this paper,they showed that deep neural networks can be used to create reliable, continuous, and non-invasive techniques for stress detection and emotion classification, with the ultimate goal of enhancing quality of life.

[4] Catherine Sandoval et al.:They have looked into auditory speech qualities as markers of familiarity and trust between people. The ability to distinguish between various classes of trust, team members' familiarity with one another, and trust change following a mission involving team-based problem-solving was trained into a number of CNN models.

[5] Xiang Li et al.: This study tested the mental workload prediction model using three methodologies and four indices. The experimental results outlined in Section 4 show that the model's predictions of changes in mental workload were highly correlated with the actual outcomes obtained using various mental workload measuring techniques, validating the model's claims.

[6] Valentina Emilia Balas et al.: In this paper, In order to give each user with individualised assistance and increase learning effectiveness, this study seeks to apply the findings to the performance prediction module in an effort to uncover the information that is ignored by the user during the learning process.

[7] AurélienAppriou et al.: In order to assess the EEG-based workload level classification performances using both user-specific and user-independent calibration, they offered a comparison of 4 contemporary machine learning methods. Our findings revealed that CNN can categorise two workload levels (low vs. high) better than CSP approaches, for both user-specific and user-independent trials. In order to execute more reliable neuroergonomics and neuroadaptive HCI, CNN could be used. The CNN is especially interesting for designing calibration-free neuroadaptive systems in the future because it can achieve satisfactory performances in a user-independent calibration using only 2 sec of EEG data and only 21 users for calibration. Investigating whether this CNN can accurately predict other cognitive states, such as curiosity, attentiveness, etc., might also be intriguing.

[8] Johann Benerradi et al.: The categorization of mental workload from brain data is still mainly an unsolved topic, despite the fact that other sensor data solutions, such as step recognition from gyroscope data, are already rather sophisticated.SVM examples have been used in relevant research, but little has been done to assess the various machine learning techniques that will be most effective for the task, particularly when it comes to tailoring the features used to best suit the unique characteristics of each model.

[9] Maria Vukovic et al.:The results of this study support the use of a portion of test speaker utterances in training sets for PSD (partially speaker dependent) models of cognitive load estimate. Additionally, as demonstrated in this study, using speech data that has been subjectively classified and dividing class boundaries using a sigmoidal curve that reflects subjective workload assessment may increase the accuracy of affect categorization.

[10] Vidhyasaharan Sethu et al.: SVMs have been used to explore the automatic estimate of cognitive load from speech cues. The SVM classifiers used in this paper's investigations were created utilising acoustic utterance-level information. Additionally used and researched were SSL approaches and feature selection for redundancy reduction.The speech used in this study was captured during a dynamic military simulation exercise where the participants did complex tasks outside of a lab environment while also self- reporting their cognitive load levels

3. METHODOLOGY

To predict team workload (TW) and team performance (TP), a system was developed to handle large-scale voice communication data. This system needed to be accessible and efficient, with minimal hardware and software requirements. The decision was made to build the system using machine learning models, focusing on Convolutional Neural Networks (CNN) for voice analysis. These models were trained on communication data collected from teams during task simulations. Various technologies were used to create this system.

Technologies Used:

Front End:

Python (for data preprocessing)

Backend:

TensorFlow (for CNN model training and deployment)

Keras (for model building)

Fig. 1 – Use case Diagram: A Use Case Diagram illustrates the various functions (use cases) that a system performs, from the perspective of an external observer. These functions are represented as ovals, and the actors (users or other systems) interacting with these functions are represented as stick figures.

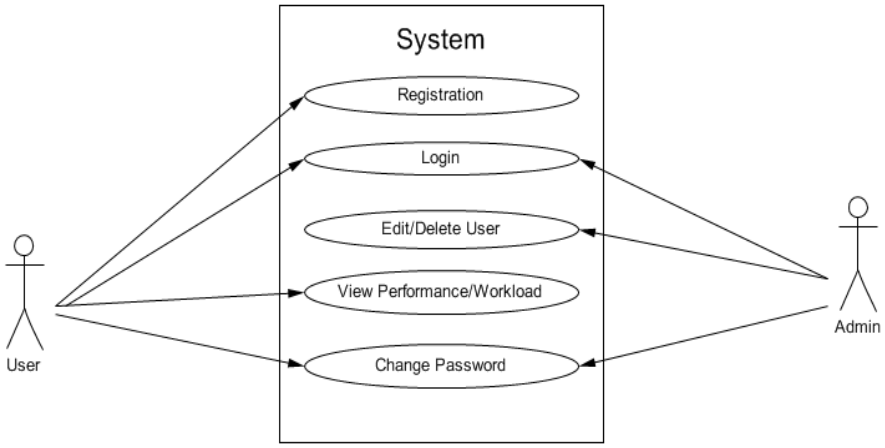
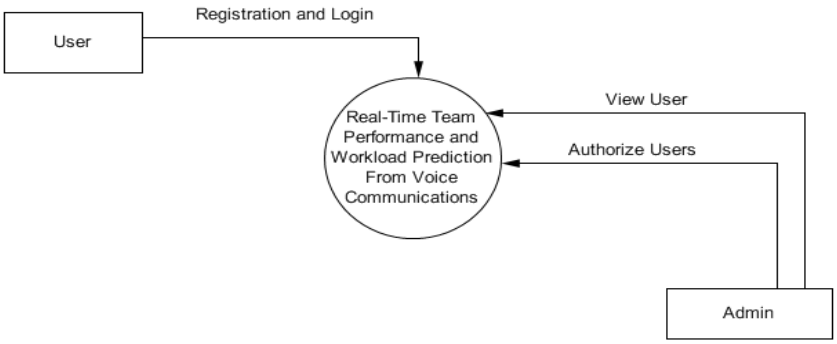


Fig. 2- DFD Level 0: A DFD represents the flow of data within a system, illustrating how data is processed by the system and how it moves between different components.



4. PROJECT TRACKING

This application aims to connect the teacher and parent. The objective of the project is to build an application that will allow fixing an online appointment with the teacher. The online system for booking an appointment responds to the current needs of the institute. The purpose of the application is to create a software product that will help as many users as possible.

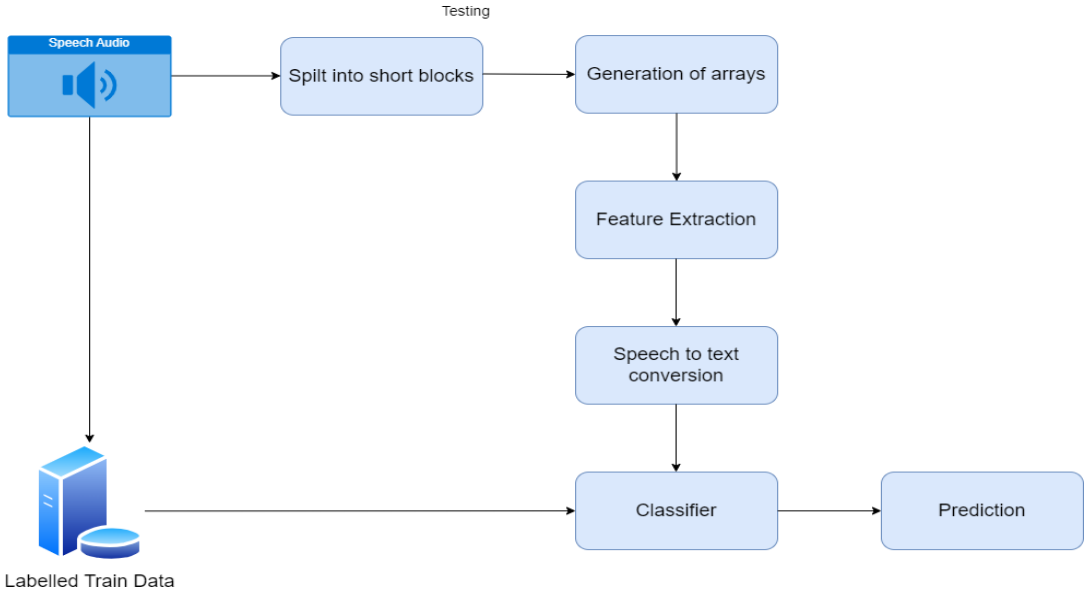
Proposed System

The proposed system for voice prediction utilizes a structured approach, leveraging deep learning techniques to analyze and classify voice characteristics in real-time. The system begins with the collection of speech recordings, which are divided into a training subset (for model learning) and a testing subset (for validation). Each RTP voice stream is initially segmented into short, manageable time blocks. This segmentation captures subtle changes in voice characteristics, making the model more robust in identifying patterns in transient audio data.

In the training phase, each segmented block undergoes transformation into a spectrogram, a time-frequency representation that visualizes the frequency components of audio signals over time. These spectrograms are then converted into RGB images, providing a rich, multi-dimensional input that enhances feature extraction for the deep learning model. These RGB images are subsequently fed into a classifier, such as a Convolutional Neural Network (CNN), which is trained on labeled data to recognize distinct voice patterns and characteristics. This classifier learns complex voice patterns, allowing it to distinguish among various speaker identities or voice features.

During the testing phase, the unseen data blocks from the testing subset are processed through the same pipeline, generating spectrograms and RGB images as input for the trained classifier. The classifier applies its learned parameters to predict the labels of the testing data accurately. This approach ensures consistency between the training and testing processes, optimizing the model's accuracy in predicting real-time voice characteristics in RTP streams.

The output is a set of predicted labels for each audio block, which could indicate speaker identity, emotional tone, or other voice attributes. By employing spectrograms and image-based classification, this system provides a reliable, scalable, and efficient framework for real-time RTP voice prediction, with potential applications in telecommunication, security, and interactive voice systems.



System Analysis

The baseline method (as shown in Fig. 1) involves training a single classifier to identify either team workload (TW classifier) or team performance (TP classifier). In both cases, the model uses a Convolutional Neural Network (CNN) that has been pre-trained to categorize RGB images generated from speech recordings. This approach allows the classifier to analyze image-based representations of speech data, applying neural network techniques to distinguish between different levels of workload or performance based on the input data.
The system for RTP voice prediction is structured to classify and predict voice characteristics in real-time, focusing on capturing, transforming, and analyzing audio data through a deep learning-based approach. The process begins with data collection, where RTP audio streams are recorded and divided into two subsets: one for training the model and the other for testing. Each audio stream is then broken down into short time blocks. This segmentation is essential, as it ensures that the system can handle real-time data by processing smaller, manageable chunks.

In the training phase, each audio block is transformed into a spectrogram—a visual representation of the audio frequencies over time. Spectrograms provide rich information about speech patterns, which is crucial for accurate classification. To make use of advanced computer vision techniques, these spectrograms are then converted into RGB images, adding depth to the input data by leveraging color channels. These RGB images are fed into a deep learning classifier, such as a Convolutional Neural Network (CNN), which is trained on labeled voice data. This model learns the distinct patterns associated with different voice characteristics, such as tone, pitch, and rhythm, enabling it to differentiate between various speakers or emotional tones.

In the testing phase, the system follows the same process for the testing subset, converting audio blocks into spectrograms and then into RGB images. These images are passed through the trained classifier, which generates predictions based on the patterns it learned during training. The output of this phase is a set of predicted labels, which could represent specific characteristics like speaker identity, emotion, or other voice traits.

This system is optimized for real-time performance, enabling immediate predictions on RTP audio streams. By processing audio data in segmented blocks and converting it into visual patterns, the system achieves high accuracy and scalability. The integration of deep learning and image processing in this approach ensures that the system can handle complex voice patterns efficiently, making it suitable for applications in telecommunications, security, and real-time voice analytics.

Requirement Analysis

- ✓ **Hardware requirements –**
  Processor: 2.2 GHz CPU
  RAM: 8 GB
  Device: Android Phone
- ✓ **Software Requirements –**
  Development Tool: Visual Code Studio version 10
  Xampp Server

4. TOOLS AND TECHNOLOGIES

A. Hardware used laptop with
   1. System type – 64-bit OS
   2. Installed memory - 4 GB and more ram

B. Platform
   1. Desktop PC or Android Phone

C. Software Tool Used
   1. Microsoft windows 7 or above.
   2. Visual Code Studio
   3. XAMPP Server
   4. Browser

5. CONCLUSION

The tasks of TP and TW prediction based on objective speech labelling were the focus of this work. Without having to expand the size of the training data or the complexity of the classification model, it was expected that adding relevant pre-requisite knowledge to the prediction model could enhance classification results for the task at hand .By combining the TP and TW data into a single CNN model by double transfer learning or into a multichannel decision-making system of parallel TW and TP classifiers, experiments were done to evaluate the research hypothesis.The idea was supported in both instances by an increase in prediction accuracy for the TP and TW stages

6. ACKNOWLEDGEMNTS

Perseverance, Inspiration & Motivation have always played a key role in the success of any venture. At this level of understanding it is difficult to understand the wide spectrum of knowledge without proper guidance and advice, hence we take this opportunity to express our sincere gratitude to our respected Project Guide an example appendix

7.REFERENCES

[1] Catherine Sandoval ,Real-Time Team Performance and Workload Prediction from Voice Communication.29 July 2022.

[2] Khushkirat Singh, Sunil Kumar Chawla, Gurpreet Singh, Punit Soni. "Stress Detection using Machine Learning Techniques: A Review" , December 2023

[3] Phie Chyan, Andani Achmad, Ingrid Nurtanio, Intan Sari , " Hybrid Deep Learning Approach for Stress Detection Model Through Speech Signal" , December 2023

[4] Mamadou Dia, Ghazaleh Khodabandelou, Alice Othmani , "A Novel Stochastic Transformer-Based Approach for PostTraumatic Stress Disorder Detection Using Audio Recording of Clinical Interviews" , March 2024

[5] Zihan Wu, Neil Scheidwasser-Clow, Karl El Hajal, Milos Cernak, " Speaker Embeddings as Individuality Proxy for Voice Stress Detection.", June 2023

[6 Maria Vukovic , VidhyasaharanSethu , Jessica Parker ,Lawrence Cavedon , Margaret Lech , John Thangarajah, "Estimating cognitive load from speech gathered in a complex real-life training exercise", Article 2018

[7] Maria Vukovic , VidhyasaharanSethu , Jessica Parker ,Lawrence Cavedon , Margaret Lech , John Thangarajah, "Estimating cognitive load from speech gathered in a complex real-life training exercise", Article 2018

[8] Russell Li1 and Zhandong Liu2,3," Stress detection using deep neural networks", The International Conference on Intelligent Biology and Medicine (ICIBM) 2020 Virtual. 9-10 August 2020..

[9] Martin Gjoreski, Tine Kolenik, TimotejKnez, MitjaLuštrek, MatjažGams,HristijanGjoreskiand VeljkoPejovi´c," Datasets for Cognitive Load Inference Using Wearable Sensors and Psychological Traits", Article Appl. Sci. 2020, 10, 3843.