DRIVER ASSISTANT DECISION-MAKING OVER HUMAN CURRENT STATE-OF-MIND

Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted By

Aryan Bhargav

IEC2019026

Under the supervision of

Dr. Surya Prakash

ELECTRONICS AND COMMUNICATION ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD, PRAYAGRAJ - 211012

MAY,2023

CANDIDATE DECLARATION

I hereby declare that the work presented in this report entitled "DRIVER ASSISTANT DECISION-MAKING OVER HUMAN CURRENT STATE-OF-MIND", submitted towards the fulfillment of BACHELOR'S THESIS report of Electronics and Communication Engineering at the Indian Institute of Information Technology Allahabad, is an authenticated record of our original work carried out under the guidance of Dr. Surya Prakash. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Aryan Bhargav
IEC2019026
Department of ECE

CERTIFICATE FROM SUPERVISOR

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief. The project titled "DRIVER ASSISTANT DECISION-MAKING OVER HUMAN CURRENT STATE-OF-MIND" is a record of the candidates' work carried out by him under my guidance and supervision. I do hereby recommend that it should be accepted in the fulfillment of the requirements of the Bachelor's thesis at IIIT Allahabad.

Dr. Surya Prakash

CERTIFICATE OF APPROVAL

The forgoing thesis is hereby approved as a creditable study carried out in the area of Information Technology and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose for which it is submitted.

Committee on the final examination for the evaluation of the thesis:

Dr. Surya Prakash (Coordinator) .....

Dr. Ramesh Kumar Bhukya (Chairman) ……..

Dr. Rahul Meshram (Committee Member) ……..

Dr. Shanti Chandra (Committee Member) ……..

Dean(A & R)

ACKNOWLEDGEMENTS

I wish to happily acknowledge the generous cooperation and the blessings of many in the completion of this thesis.

I would like to express my deep sense of gratitude to my supervisor Dr. Surya Prakash for his valuable guidance in the preparation of this thesis and for providing me with this theme of research in the area of emotion detection using facial, speech, and text. His expertise was invaluable in formulating the methodology, which pushed me to bring my work to a higher level.

I also extend my gratitude to our Hon'ble Director, Prof. (Dr.) Prof MS Sutaone, Dr. Sajay Singh, Head of the Department (ECE), IIIT Allahabad, and Dr. Manish Goswami, Associate Dean (Admission, Assessment, and Award- AAA) for providing outstanding lab facilities and an excellent environment for research. I am thankful to all the faculty members of IIIT Allahabad for their interactive involvement, valuable suggestions, and guidance during the progress seminars.

I would like to acknowledge my colleagues for their wonderful collaboration. A special thanks to my encouraging and inspiring senior colleagues at IIIT Allahabad, who have been a source of motivation and inspiration throughout my time at the institute.

Last but certainly not least, I owe a great deal to my family for molding me into the person I am today. My parents' unwavering support for my studies throughout has enabled me to obtain my Bachelor's today.

Plag Report

Plag Report

ORIGINALITY REPORT

11%	7%	6%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.freepatentsonline.com Internet Source	1%
2	"Artificial Neural Networks and Machine Learning – ICANN 2018", Springer Science and Business Media LLC, 2018 Publication	1%
3	Submitted to University of East London Student Paper	1%
4	Submitted to Indian Institute of Information Technology, Allahabad Student Paper	1%
5	Submitted to Cranfield University Student Paper	1%
6	Submitted to University of Warwick Student Paper	1%
7	www.ncbi.nlm.nih.gov Internet Source	<1%
8	tiptekno.net Internet Source	<1%

9	www.mdpi.com Internet Source	<1 %
10	Submitted to University of Auckland Student Paper	<1 %
11	Submitted to Savitribai Phule Pune University Student Paper	<1 %
12	cs.lnu.se Internet Source	<1 %
13	kipdf.com Internet Source	<1 %
14	Submitted to RMIT University Student Paper	<1 %
15	www.cs.ccu.edu.tw Internet Source	<1 %
16	Submitted to United Colleges Group - UCG Student Paper	<1 %
17	link.springer.com Internet Source	<1 %
18	"Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications", Springer Science and Business Media LLC, 2022 Publication	<1 %

19	Submitted to University of Liverpool Student Paper	<1 %
20	Himaja Avula, Ranjith R, Anju S Pillai. "CNN based Recognition of Emotion and Speech from Gestures and Facial Expressions", 2022 6th International Conference on Electronics, Communication and Aerospace Technology, 2022 Publication	<1 %
21	Qirong Mao, Ming Dong, Zhengwei Huang, Yongzhao Zhan. "Learning Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks", IEEE Transactions on Multimedia, 2014 Publication	<1 %
22	Shuo Jia, Pingzhao Hu. "ChrNet: A re-trainable chromosome-based 1D convolutional neural network for predicting immune cell types", Genomics, 2021 Publication	<1 %
23	Submitted to The University of Wolverhampton Student Paper	<1 %
24	Submitted to University of Central Florida Student Paper	<1 %
25	Submitted to University of Sheffield Student Paper	<1 %

ABSTRACT

Road accidents are a major cause of mortality and morbidity in today's world. One of the main reasons for this is the driver's emotional state, which can distract them from responding promptly to sudden changes in the environment, leading to delayed or inadequate control of the vehicle, ultimately resulting in a collision. To mitigate this risk, an autonomous driving system that can detect the driver's emotional state and intervene accordingly has been proposed. This system, known as the Triple Verified Driver Assistant (TVDA), utilizes three primary indicators, namely facial expressions, voice pitch, and social media text analysis, to accurately determine the driver's emotional state. The system verifies these 3 indicators against each other to provide a more precise assessment of the driver's mental state and improve the overall safety of the driving experience.

The facial emotion detection model was trained using Convolutional Neural Network (CNN) with a batch size of 32 for 21 epochs and achieved a training accuracy of 86.14% and a validation accuracy of 81.24%. The speech emotion recognition model achieved an accuracy of 81.33%, which is relatively high compared to other existing models. The text analysis model, which utilized Logistics Regression, achieved an accuracy of 83%.

The proposed approach has the potential to significantly enhance driver safety by allowing for the development of advanced driver assistance systems that can detect the emotional state of the driver. This could provide valuable insights into the driver's mental state and alert the system to potential distractions, fatigue, or other issues that could affect the driver's ability to safely operate the vehicle.

In addition, the ability to switch control from manual to autonomous drive modes based on the driver's emotional state could further improve safety by reducing the risk of accidents caused by driver error or fatigue. Overall, the proposed approach has the potential to revolutionize the field of driver assistance systems and significantly enhance the safety and comfort of drivers and passengers alike.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER - 1

INTRODUCTION

1.1 Emerging Technologies: Real-Time Mental State Detection

In recent years, there has been a growing interest in developing technologies that can accurately determine a person's mental state in real-time. The potential applications of such technologies are vast, ranging from personalized healthcare to the development of advanced driver assistance systems.

Various methodologies have been explored to determine a person's mental state, including analyzing facial expressions to detect emotions, evaluating social media posts to infer personality traits, and measuring voice pitch and amplitude to control the operation of intelligent vehicles. While each of these approaches has shown promise individually, accurate evaluation of a person's mental state remains elusive.

In response to this challenge, the current study aims to determine a person's mental state through a combination of emotion, voice tone, and social media analysis, which will be used to control the movement of a vehicle. This approach builds upon and extends previous research that has focused on these areas individually.

1.2 Facial Expression Analysis

Facial expression analysis has long been recognized as an important tool for detecting emotions. Advances in computer vision and machine learning techniques have made it possible to accurately recognize emotions from facial expressions in real-time. However, facial expressions alone do not provide a complete picture of a person's mental state.

1.3 Speech Emotion Analysis

Voice tone analysis is another promising approach for detecting a person's mental state. Research has shown that changes in voice tone, such as increased pitch and amplitude, can be indicative of emotions such as anger, stress, and anxiety. By analyzing voice tone in real-time, it may be possible to detect a person's mental state with a high degree of accuracy.

1.4 Social Media Analysis

Social media analysis is another area of research that has shown promise for detecting a person's mental state. By analyzing a person's social media posts, it may be possible to infer their personality traits, emotional state, and even their risk of developing certain mental health conditions. However, social media analysis is not without its challenges, such as the difficulty of accurately interpreting text-based data .
To achieve this goal, the study will use a combination of machine learning and signal processing techniques. The facial expression and voice tone data will be processed using deep learning algorithms to extract features that are indicative of the driver's emotional state. The social media data will be analyzed using natural language processing techniques to infer the driver's personality traits and emotional state.

The current study aims to combine these three approaches to develop a more accurate and reliable method for detecting a person's mental state. Specifically, the study will analyze facial expressions, voice tone, and social media data to infer the emotional state of the driver. This information will then be used to control the movement of a vehicle, with the goal of improving driver safety and reducing the risk of accidents caused by distracted or fatigued drivers.
Another challenge is the need to ensure the accuracy and reliability of the model in real-world settings. The study will need to be conducted in a real-world driving environment, with all the attendant challenges that come with that, such as changes in lighting conditions, weather, and traffic patterns.

CHAPTER - 2

LITERATURE REVIEW

There are three primary research domains: Facial Emotion Detection, Speech Emotion Recognition and Social Media Text Analysis.

2.1 Facial emotion recognition

H. Avula and A. S. Pillai present a study on utilizing convolutional neural networks (CNN) for recognizing emotions and speech from gestures and facial expressions[9]. The research was conducted as part of the 6th International Conference on Electronics, Communication, and Aerospace Technology in 2022. The authors' approach involved training a CNN model to extract relevant features from gesture and facial expression data and classify them into specific emotions and speech patterns. The results of the study demonstrated the effectiveness of CNN-based recognition in accurately identifying emotions and speech based on gestures and facial expressions

The paper titled "Facial Emotion Recognition Using Deep Convolutional Neural Network" by E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya[10] presents a study on utilizing deep convolutional neural networks (CNN) for facial emotion recognition. The research was conducted in the 6th International Conference on Advanced Computing and Communication Systems in 2020. The authors propose a CNN-based approach to extract features from facial images and classify them into different emotions. The study demonstrates the effectiveness of deep CNNs in accurately recognizing facial emotions. The results show promising performance, highlighting the potential of deep learning techniques for facial emotion recognition. This research contributes to the field of emotion recognition and provides insights into the application of deep CNNs for analyzing facial expressions.

2.2 Speech Emotion Recognition

Huang, Z., Dong, M., Mao, Q., & Zhan, Y proposes a semi-CNN approach for Speech Emotion Recognition (SER)[4] to learn affect-salient features. The results demonstrate improved recognition performance on benchmark datasets, especially in complex scenes with distortion. The approach outperforms existing SER features and provides stable and robust performance. This research highlights the effectiveness of utilizing deep learning techniques for emotion recognition in speech and contributes to the advancement of SER systems

Luengo, E. Navas, and I. Hernández present a study on analyzing and evaluating features for automatic emotion identification in speech [5] investigate various features to determine their effectiveness in accurately identifying emotions from speech signals and explores the impact of features such as pitch, intensity, and spectral information, and evaluates their performance using machine learning algorithms. The results provide insights into the discriminative power of different features for emotion recognition in speech. This research contributes to the field by identifying key acoustic and prosodic features and their relevance in developing robust automatic emotion identification systems.

They used a Context-Dependent Deep Neural-Network Hidden Markov model, which was able to handle large amounts of speech-to-text transcription data sets. The system used tied-state triphones and pre-training of deep-belief networks to improve accuracy.

2.3 Text Emotion Recognition

Rout et al. (2017)[11] presents a model for sentiment and emotion analysis of unstructured social media text. The model combines sentiment analysis, emotion analysis, and text analysis techniques to extract meaningful insights from social media posts. It aims to enhance the understanding of user sentiments and emotions on social media platforms, providing valuable information for various applications.

Agarwal et al. (2011)[2] focus on sentiment analysis of Twitter data. It explores techniques to analyze sentiments expressed in short text messages on Twitter and evaluates different

methods for classifying tweets into positive, negative, or neutral sentiments. The study demonstrates the effectiveness of their proposed approach in accurately identifying sentiment in Twitter data. This research contributes to the field of sentiment analysis by addressing the unique characteristics of Twitter messages and providing insights into the sentiment expressed on the platform.

2.4 Autonomous Vehicle Safety

Sabaliauskaite et al. (2018) [6]proposed an approach to integrate autonomous vehicle safety and security processes in compliance with international standards. The approach considers the three functions that determine autonomous driving function: perception, decision, and control. Perception is performed using sensors, a fusion sensor, a world model, semantic understanding, and localization. Decision-making is performed by an onboard computer with cognitive driving intelligence, a vehicle platform, a communication system, and external sensors including LIDAR, cameras, and ultrasonic sensors. The final component is a vehicle platform with actuators and controllers (ECU) that implement the desired movement. The STPA method is used to ensure compliance with international standards such as SAE J3016, SAE J3061, and ISO 26262. Currently, vehicle security systems do not take into account autonomous vehicle systems.

Barrett et al. (2009) created a system and method for processing a safety signal in a self-driving car[12]. The vehicle uses its onboard local area sensors and perceptual context software to detect the presence of unsafe conditions while operating unmanned. In addition, the vehicle receives and responds to the operator's emergency signals. In either case, processing of the detected or signaled data leads to manipulation of vehicle input devices in some embodiments to ensure proper response to the detected or signaled data.

CHAPTER - 3

METHODOLOGY AND DISCUSSION

3. l Dataset

There are 6 kinds of datasets used for the TVDA model which are summarized in Table 1.

Dataset	Sub-System	Size	Features
1. FER 2013	Expression to Emotion	54Mb	35.9k image files,48x48 pixels, seven emotional categories
2. RAVDESS	Speech to Emotion	24.8Gb	This dataset contains a total of 7,400 audio files and is categorized into 24 different actors
3. TESS	Speech to Emotion	4.7Gb	It consists of a total of 2,960 audio files. The dataset features two actresses in it.
4. CREMA-D	Speech to Emotion	1.65Gb	It contains a total of 7,442 audio files. Each audio file corresponds to a specific emotional expression performed by one of the actors.
5. SAVEE	Speech to Emotion	259Mb	It contains a total of 480 audio files. These audio files are performed by four male actors and cover seven different emotions.
6. SENTIMENT140	Text to Sentiment	305Mb	Sentiment140 dataset, which consists of 1.6 million tweets with sentiment labels.

Table 1: Dataset features for the TVDA model

The proposed system comprises three subsystems to accurately estimate the driver's state of mind. The first subsystem involves detecting facial expression to predict the driver's emotion, which is verified using speech expressions in the second subsystem followed by third social media analysis. Each subsystem uses different datasets and corresponding processing algorithms, which are elaborated on below.

3.2 Facial recognition model.

3.2.1 Data Preprocessing and Preparation.

We are performing several operations on image data and emotion labels. It starts by defining a dictionary that maps numeric labels to different emotions.

Then, we processes the image data. It converts the pixel values, which are initially stored as strings separated by spaces, into numerical arrays. These arrays are reshaped to have dimensions of 48x48x1, representing grayscale images. The data type of the arrays is changed to float32. All the resulting image arrays are then combined into a single multidimensional array.

Next, we encode the emotion labels using a technique that assigns a unique numeric value to each emotion category. This encoded representation is further converted into a one-hot encoded form, which is useful for classification tasks.

Finally, a mapping is created between the unique emotion classes and their corresponding encoded values. This mapping allows for easier interpretation of the numeric labels assigned to each emotion category.

3.2.2 Proposed Modeling

This is a deep convolutional neural network (DCNN) model using the Keras Sequential API. The model architecture consists of total 14 layers.

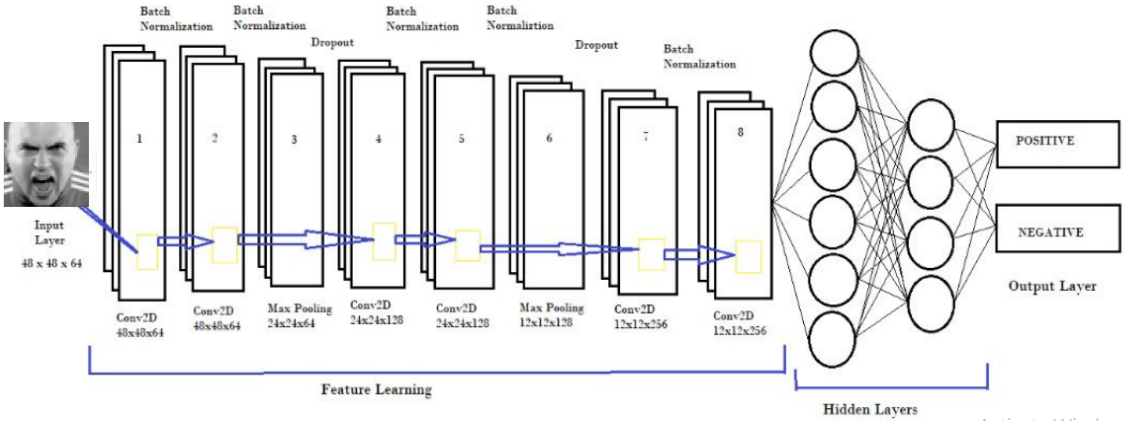

Figure 2 : CNN Architecture for Facial Emotion Recognition

The input layer is a 2D convolutional layer (Conv2D) with 64 filters, a kernel size of (5,5), with an activation function called Exponential Linear Unit (ELU). It takes an input shape of 48x48x64 which represents the dimensions of the input images. The padding is set to 'same' to preserve the spatial dimensions, and the kernel weights are initialized using the He normal initialization method. Batch normalization is applied after this layer to regularize the network.

Following the first convolutional layer, there is another identical Conv2D layer with batch normalization.

Next, a max-pooling layer with a pool size of (2,2) is added, reducing the spatial dimensions of the feature maps. A dropout layer with a rate of 0.4 is applied to reduce overfitting.

This pattern of two Conv2D layers, batch normalization, max pooling, and dropout is repeated twice more, gradually increasing the number of filters to 128 and then 256.

After the third set of convolutional layers, there is a Flatten layer that converts the 2D feature maps into a 1D vector.

A fully connected layer (Dense) with 128 units and ELU activation is added, followed by batch normalization and dropout.

Finally, the output layer (Dense) with the number of units equal to num_classes = 3 and softmax activation is added to perform multi-class classification.

The model is compiled with the categorical cross-entropy loss function, with ‘ADAM’ optimizer , and accuracy as the evaluation metric.

3.2.3 Training the CNN Model

During the training process of the model, two callbacks were utilized: early stopping and ReduceLROnPlateau. Early stopping is employed to prevent overfitting. It monitors a specified validation metric, such as validation loss or accuracy, and halts the training process if the metric does not show any improvement. This helps to avoid training the model for an excessive number of epochs, which can lead to overfitting on the training data and poor generalization to new, unseen data.

ReduceLROnPlateau is another callback used to dynamically adjust the learning rate during training. If the monitored validation metric fails to improve for a certain number of epochs, the learning rate is reduced. This can be beneficial for fine-tuning the model's convergence and ensuring progress in optimizing the training process.
The Adam optimizer was chosen for training the model. Adam is an efficient optimization algorithm that combines elements of both gradient descent and momentum methods. It adapts the learning rate based on the gradients of the model's parameters, resulting in faster convergence and better optimization.

3.2.4 Results and Predictions of Facial Emotion Recognition

The model got 86.14 % accuracy on training set and 81.24 % on validation when trained over 50 epochs in batch size of 32.

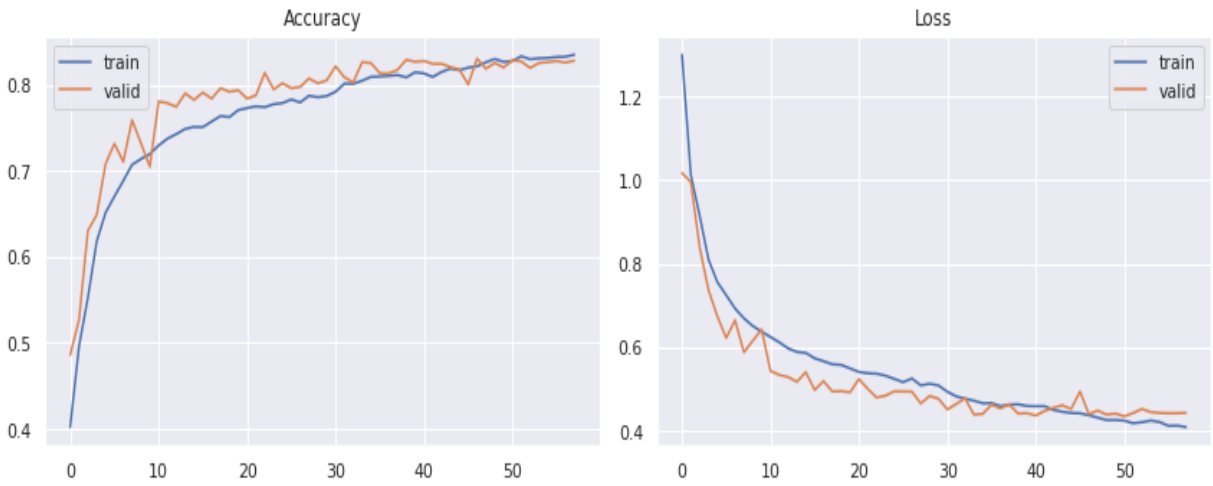

Figure 3 : Accuracy and Loss Graph for Facial Model.

Figure 4 : Actual and Predicted Emotions.

3.3 Speech Emotion Recognition

First of all, the 4 datasets (Ravdess, Tess, Savee, Crema-D) are combined to create a single bigger dataset using dataframes in pandas. Then the following operations are performed on

them. These 4 datasets when combined consists of 12,162 samples of Positive (Happy, Neutral, Surprise) and Negative (Sad, Angry, Disgust, Fear) emotions.

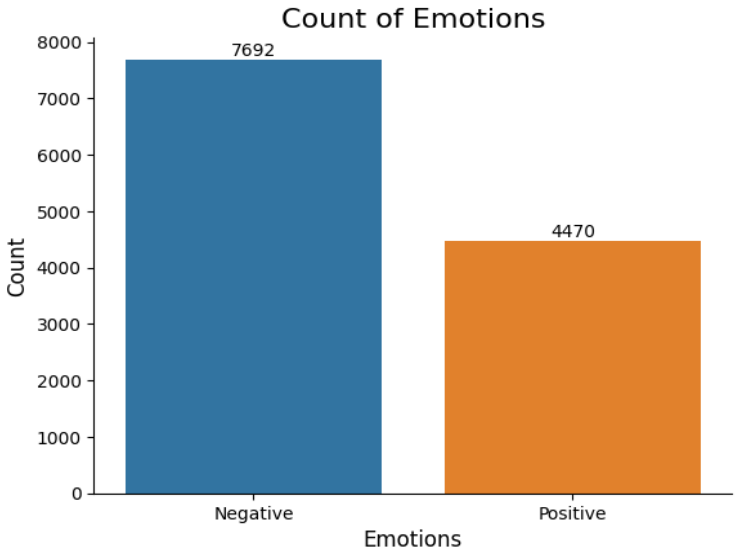

Figure 5 : Count of Emotions in Concatenated Dataset.

3.3.1 Data Augmentation

Data augmentation techniques are used to artificially increase the diversity of the training data by applying various transformations or modifications to the existing samples. In the context of speech signals, four data augmentation techniques are mentioned: noise injection, stretching, shifting, and pitching.

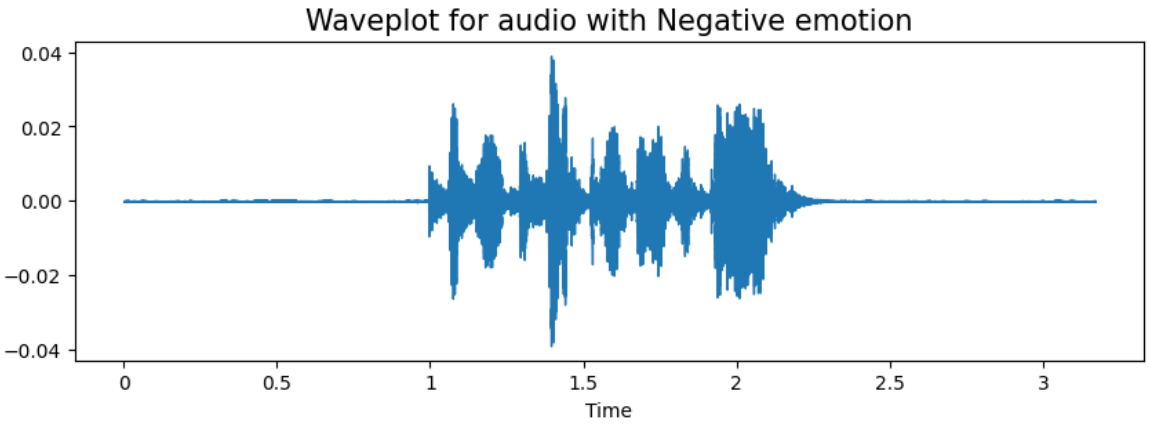

Figure 6 : Waveplot for Audio with Negative Emotion

1. Noise Injection: This technique introduces random noise to the speech signal. It

calculates the maximum amplitude of the signal, generates random noise within a specific range, and multiplies it with a standard normal distribution. By adding noise, the model becomes more robust to noisy environments and variations in input quality.

2. Stretching: This technique modifies the speech signal by compressing or expanding its time axis. It randomly applies a stretching factor between 0.8 and 1.2 to the signal, using the time_stretch() function from the librosa library. This manipulation changes the temporal characteristics of the signal, allowing the model to learn variations in speech duration and tempo. It helps to enhance the model's robustness to different speaking rates and improves its ability to handle time-related variations in the input data.
3. Shifting: This technique shifts the speech signal by a random number of samples between -5000 and 5000. It uses the roll function from numpy to perform the signal shifting. Shifting the signal in time helps the model learn to handle temporal misalignments or variations in the starting point of speech segments.
4. Pitching: This technique alters the pitch of the speech signal by a random number of semitones between -2 and 2. It utilizes the pitch_shift function from the librosa library. By changing the pitch, the model can adapt to variations in the fundamental frequency of the speech, improving its ability to handle different vocal characteristics or variations in pitch across speakers.

By applying these data augmentation techniques, the training data is expanded with modified versions of the original speech signals, allowing the model to generalize better and perform well on unseen data with different noise levels, durations, temporal alignments, and pitch variations.

3.3.2 Feature Extraction

We extract the following features with the help of librosa library so that our data is compatible to be trained on.

1. Zero Crossing Rate (ZCR): The zero crossing rate represents the rate at which a signal changes from positive to negative or vice versa. It provides information about

the frequency of the waveform and the amount of signal change over time. Signals with higher ZCR values tend to have more high-frequency content or rapid changes, while signals with lower ZCR values are smoother or less noisy.

2. Chroma STFT: Chroma refers to the 12 different pitch classes of the musical octave. The Chroma STFT feature captures the presence and distribution of these pitch classes in the audio signal. It is calculated using the Short Time Fourier Transform (STFT), which converts the audio signal into the frequency domain.
3. MFCC (Mel-Frequency Cepstral Coefficients): MFCCs are widely used features in speech recognition and audio processing. They capture the spectral envelope of the speech signal and represent the power spectrum in the mel frequency scale. The calculation of MFCC involves dividing the audio signal into short overlapping frames, applying the Fourier Transform to each frame, mapping the resulting spectrum onto the mel scale, and finally computing the cepstral coefficients.
4. Root Mean Square Value (RMS): The RMS value is calculated as the square root of the average of the squared values of the speech signal. It provides information about the amplitude or energy of the signal. Higher RMS values indicate a higher overall loudness or intensity of the signal, while lower RMS values indicate a softer or quieter signal.
5. Mel Spectrogram: The Mel spectrogram represents the spectral content of the speech signal in the mel frequency scale. It is calculated using the STFT, which converts the audio signal into the time-frequency domain. The mel spectrogram captures the distribution of energy across different frequencies over time and is useful for analyzing the spectral characteristics of the signal.

These audio features provide valuable information about the frequency content, pitch, energy, and spectral characteristics of the speech signal. They help to capture important aspects of the audio that can be used as input for our speech recognition model.

3.3.3 Building Model of Speech Emotion System

The model architecture consists of total 14 layers where four Conv1D layers followed by four max pooling layers, which are used to downsample the output of the previous layer

and extract hierarchical features. Two Dropout layers are included to mitigate overfitting by randomly deactivating some neurons during training. The first Conv1D layer has 256 filters, a kernel size of 5, and a stride of 1. The first 11 layers are responsible for feature learning, extracting relevant patterns and representations from the input data, while the last three layers are classification layers.

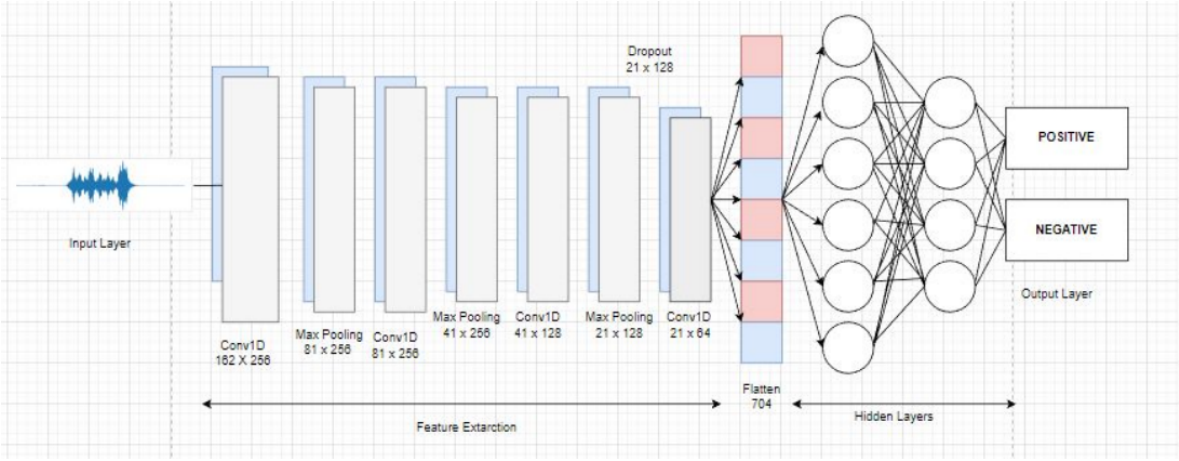

Figure 7 : Network Architecture for Speech Emotion Recognition Model.

The final layer is a Dense layer with two units and a softmax activation function, producing a probability distribution over the two classes. The model is compiled with the 'adam' optimizer, 'categorical_crossentropy' loss function, and 'accuracy' metric, indicating that it aims to minimize the difference between predicted and true labels by adjusting the model's weights and biases using gradient-based optimization.

3.3.4 Results

This model got accuracy of 81.33 % when trained over 31 epochs with batch size of 32 and 571 steps in each epoch.

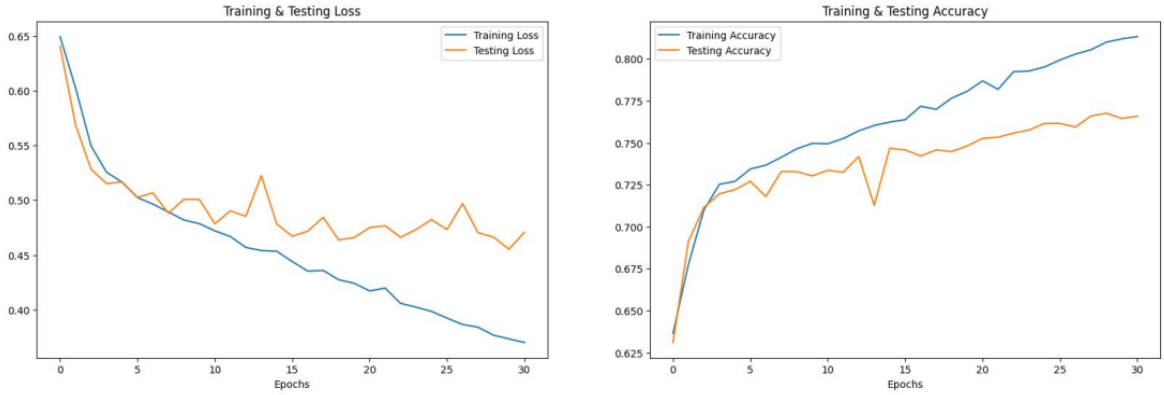

Figure 8 : Accuracy and Loss Graph for Speech Model.

	precision	recall	f1-score	support
Negative	0.78	0.87	0.82	3783
Positive	0.74	0.59	0.65	2298
accuracy			0.77	6081
macro avg	0.76	0.73	0.74	6081
weighted avg	0.76	0.77	0.76	6081

Figure 9 : Precision, Recall and F1-Score for Speech Model.

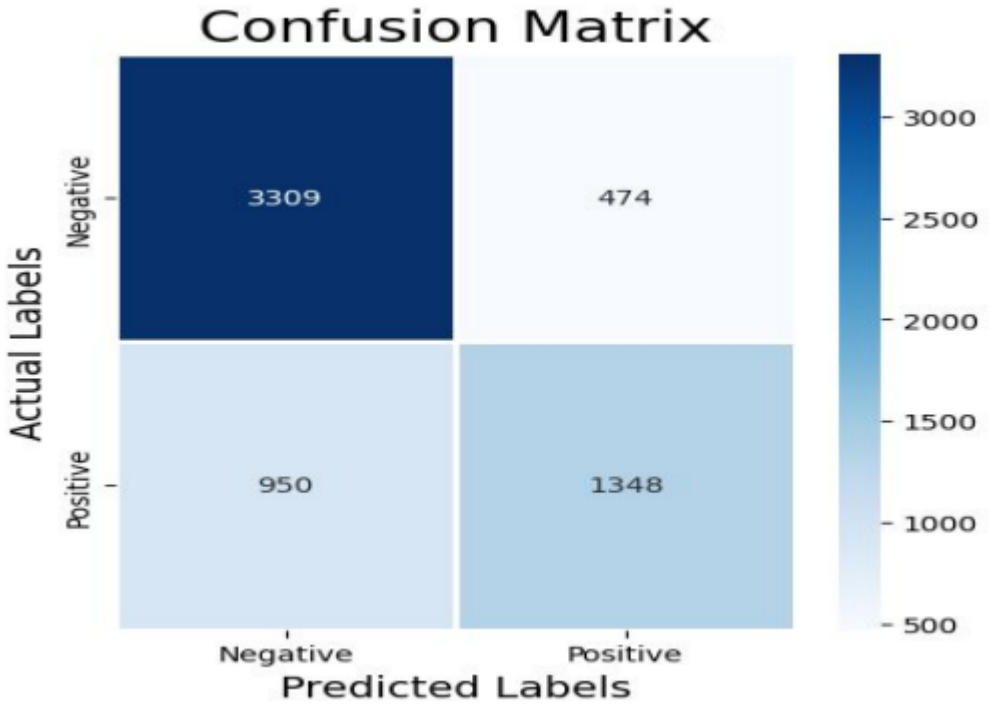

Figure 10 : Confusion Matrix for Speech Model.

3.4 Text Emotion Analysis

The dataset used here is Sentiment140 which consists of 1.6 million tweets with sentiment labels but are need to be processed first before feeding to a machine learning model.

3.4.1 Data Preprocessing

The given preprocessing steps aim to clean and standardize text data for further analysis. The steps are as follows:

1. Converting the text to lowercase: This step ensures that all letters in the text are converted to lowercase. It helps in achieving case insensitivity and avoids multiple representations of the same word.
2. Replacing URLs: Any URLs present in the tweet are replaced with the string 'URL'. This step removes URLs from the text, which are not typically informative for text analysis.
3. Replacing emojis: Emojis in the tweet are replaced with the string 'EMOJI'

concatenated with a unique identifier. This step allows treating emojis as separate entities rather than considering them as part of the text, as they often convey emotional or expressive content.

4. Replacing usernames: Usernames in the social media site are replaced with the string 'USER'. This step anonymizes usernames and treats them as a uniform entity rather than individual usernames, which may not contribute much to the analysis.
5. Replacing non-alphanumeric characters: Any non-alphanumeric characters (e.g., punctuation marks, symbols) in the tweet are replaced with a space. This step helps remove noise and special characters that may not add much meaning to the text analysis.
6. Replacing consecutive letters: Sequences of three or more consecutive letters in the tweet are reduced to two letters. This step reduces redundancy and helps normalize the text by removing elongated words or repeated letters.
7. Lemmatizing words: The words in the tweet are lemmatized using the WordNetLemmatizer. Lemmatization reduces words to their base or dictionary form, capturing their essential meaning and allowing better text analysis by grouping similar word forms together.
8. Appending lemmatized words: The lemmatized words are appended to the tweetwords string. This step aggregates the preprocessed words into a single string that can be used for further analysis or modeling tasks.

By performing these preprocessing steps, the text data is cleaned, standardized, and made more suitable for sentiment analysis.

3.4.2 Logistics Regression

The logistic regression model is a popular statistical model that excels in binary classification problems. Its main objective is to predict the probability that an instance belongs to a specific class. It achieves this by employing the logistic function, also known as the sigmoid function, to model the relationship between independent variables and the probability of binary outcomes. The logistic regression model uses a cost function to minimize the error between the predicted output and the actual binary labels. In this

context, a smaller value of C is chosen, which corresponds to stronger regularization. Regularization helps prevent overfitting by penalizing large coefficients, promoting a simpler and more generalized model. By carefully tuning the regularization parameter, logistic regression can effectively handle binary classification tasks by providing accurate probability estimates and distinguishing between two classes of Positive and Negative states based on the input features.

3.4.3 Results of Text Analysis

The logistic regression gives accuracy of 83%.

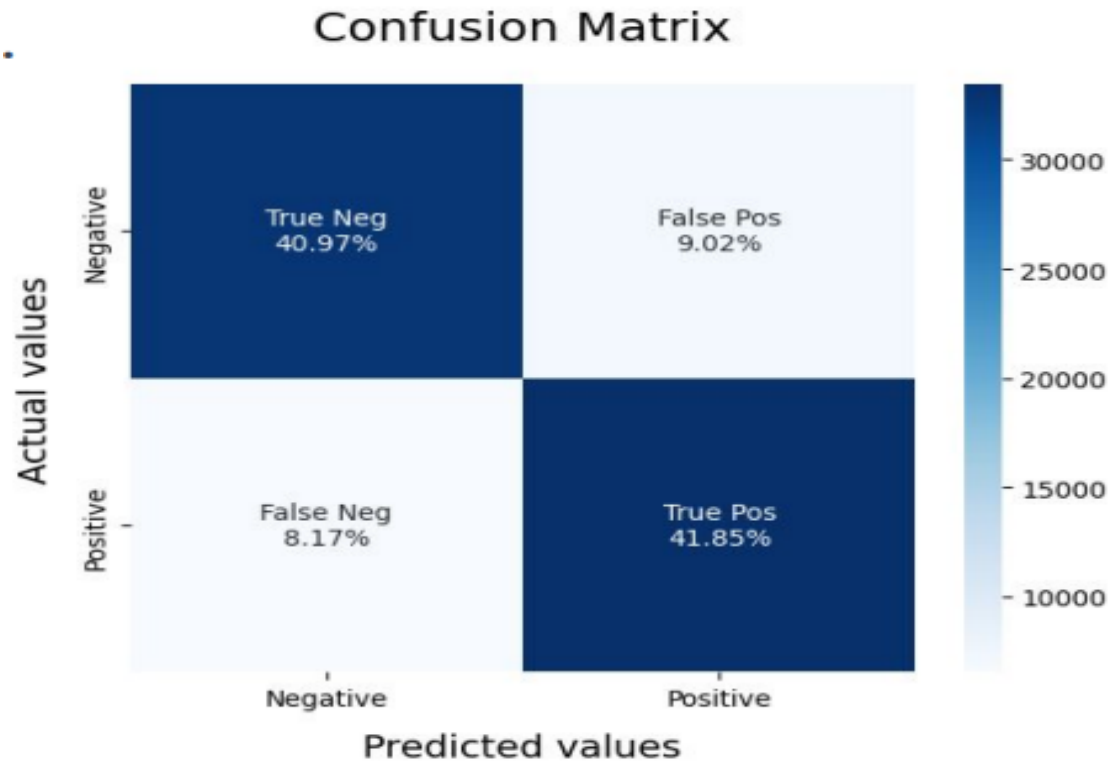

Figure 11 : Confusion matrix for Text Model.

	precision	recall	f1-score	support
0	0.83	0.82	0.83	39989
1	0.82	0.84	0.83	40011
accuracy			0.83	80000
macro avg	0.83	0.83	0.83	80000
weighted avg	0.83	0.83	0.83	80000

Figure 12 : Precision, Recall and F1- Score of Text Model.

3.5 Outcomes

3.5.1 Manual Mode:

Utilizing a combination of facial expression recognition, voice recognition, and social media history analysis, the vehicle will remain in manual mode if two of three analyses produce positive results, potentially improving the accuracy and reliability of driver state monitoring for accident prevention.

3.5.2 Auto Mode:

The proposed driver state monitoring system utilizes a combination of facial expression recognition, voice recognition, and social media history analysis to switch the vehicle from manual to auto mode if two out of three analyses produce negative results, improving the accuracy and reliability of the system for accident prevention.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 Conclusion

In conclusion, the Triple Verified Driver Assistant (TVDA) presents a promising approach to accurately determine the driver's emotional state through facial expressions, voice pitch, and social media text analysis. By leveraging these three primary indicators and cross-verifying them, the system achieves a more precise assessment of the driver's mental state.

The facial emotion detection model, trained using Convolutional Neural Network (CNN), demonstrates high accuracy with a training accuracy of 86.14% and a validation accuracy of 81.24%. The speech emotion recognition model also performs well, achieving an accuracy of 81.33%, surpassing existing models. Furthermore, the text analysis model, employing Logistics Regression, achieves an accuracy of 83%.

The proposed approach holds significant potential for enhancing driver safety by enabling the development of advanced driver assistance systems. By accurately detecting the driver's emotional state, these systems can gain valuable insights into potential distractions, fatigue, or other factors that may impact the driver's ability to operate the vehicle safely.

The ability to switch control from manual to autonomous driving modes based on the driver's emotional state further enhances safety by reducing the risk of accidents caused by driver error or fatigue. This integration of emotion analysis into driver assistance systems has the potential to revolutionize the field, enhancing the safety and comfort of both drivers and passengers.

However, further research and development are required to refine and validate the proposed approach. This includes exploring additional indicators, refining the machine learning models, and addressing privacy and ethical considerations associated with the collection and analysis of personal data.

Overall, the TVDA system presents a promising future for the field of driver assistance systems, offering a more comprehensive understanding of the driver's mental state and improving safety on the roads.

4.2 Future Scope of the Work

In the future, the integration of emotion and sentiment analysis into vehicles has the potential to enhance driver safety and personalize the driving experience. One area of application is the enhancement of Advanced Driver Assistance Systems (ADAS). By analyzing the driver's emotional state, the vehicle could adapt its behavior to provide calming cues during moments of stress or adjust the interior environment to create a more comfortable atmosphere.

Another potential application is in personalized ride-sharing services. Emotion analysis could be used to match passengers with drivers who have similar emotional profiles, offering a more tailored and comfortable experience. Additionally, sentiment analysis of social media history could be employed to understand passengers' preferences and customize the in-vehicle environment accordingly.

Vehicles equipped with emotion analysis capabilities could also provide mental health support while driving. By detecting signs of emotional distress or anxiety, the vehicle could offer calming techniques, play relaxing music, or even initiate a call with a helpline service.

Moreover, integrating emotion and sentiment analysis into vehicles could generate valuable data for research and behavioral analysis. This data could lead to a deeper understanding of driver behavior, preferences, and emotional responses, ultimately resulting in the

development of improved driving environments and interventions.

CHAPTER-5

REFERENCES

1. Bakshi, Rushlene Kaur, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. “Opinion mining and sentiment analysis.” In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 452-455. IEEE, 2016.
2. SA2 Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J. Passonneau. “Sentiment analysis of Twitter data.” In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38. 2011.
3. Carlini, Nicholas, and David Wagner. “Audio adversarial examples: Targeted attacks on speech-to-text.” In 2018 IEEE Security and Privacy Workshops (SPW), pp. 1-7. IEEE, 2018.
4. Huang, Z., Dong, M., Mao, Q., & Zhan, Y. (2014). Speech Emotion Recognition Using CNN. In Proceedings of the 22nd ACM international conference on Multimedia, [Orlando Florida USA] (pp. [801-804]).
5. E. Luengo, A. Navas, and I. Hernández, "Analyzing and evaluating features for automatic emotion identification in speech," in IEEE Transactions on Multimedia, vol. 12, no. 6, pp. 490-501, Oct. 2010
6. Sabaliauskaite, Giedre, Lin Shen Liew, and Jin Cui. “Integrating autonomous vehicle safety and security analysis using stpa method and the six-step model.” International Journal on Advances in Security 11, no. 1&2 (2018): 160-169.
7. EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo 2020.
8. Emotion Recognition from Physiological Signal Analysis: A Review Egger Maria, Ley Matthias, Hanke Sten 2019.
9. H. Avula, R. R and A. S Pillai, "CNN-based Recognition of Emotion and Speech

from Gestures and Facial Expressions," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1360-1365.

10. E. Pranav, S. Kamal, C. Satheesh Chandran, and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 317-320.
11. J. K. Rout, K.-K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," Springer Science+Business Media New York, 2017.
12. Barrett, David S., James Allard, Misha Filippov, Robert Todd Pack, and Selma Svendsen. “System and method for processing safety signals in an autonomous vehicle.” U.S. Patent 7,499,774, issued March 3, 2009.
13. Koopman, Philip, and Michael Wagner. “Autonomous vehicle safety: An interdisciplinary challenge.” IEEE Intelligent Transportation Systems Magazine 9, no. 1 (2017): 90-96.