# VISVESVARAYA TECHNOLOGICAL UNIVERSITY,BELAGAVI

**A PROJECT REPORT ON**

# "SENTIMENT ANALYSIS FOR MULTIMODAL SOCIAL MEDIA DATA"

Submitted in partial fulfillment for the award of Degree of

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**By**


| | |
|---|---|
| **ADITH P KOTIAN** | **4AL20CS006** |
| **AFRAN SAYYED KARIM** | **4AL20CS008** |
| **ASHWIN KUMAR** | **4AL20CS024** |
| **ISHWAR PAVAN** | **4AL20CS050** |

**Under the Guidance**

**of**

**Mr. Vasudev S Shahapur**

**Associate Professor**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGYMOODBIDRI-574225, KARNATAKA**


**2023-24**

## ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY

## MIJAR, MOODBIDRI D.K. -574225, KARNATAKA

ALVA'S
Education Foundation®

## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

# CERTIFICATE

This is to certify that the project entitled **"SENTIMENT ANALYSIS FOR MULTIMODAL SOCIAL MEDIA DATA"** has been successfully completed by

| | |
|---|---|
| **ADITH P KOTIAN** | **4AL20CS006** |
| **AFRAN SAYYED KARIM** | **4AL20CS008** |
| **ASHWIN KUMAR** | **4AL20CS024** |
| **ISHWAR PAVAN** | **4AL20CS050** |

the bonafide students of DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY of the VISVESVARAYA TECHNOLOGICAL UNIVERSITY, **BELAGAVI** during the year 2023–24. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the Bachelor of Engineering Degree.

| | | |
|---|---|---|
| **Mr. Vasudev S Shahapur** | **Dr. Manjunath Kotari** | **Dr. Peter Fernandes** |
| **Project Guide** | **Head Of the Department** | **Principal** |

**External Viva**

| **Name of the Examiners** | **Signature with Date** |
|---|---|
| **1.** | |
| **2.** | |

ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY

MIJAR, MOODBIDRI D.K. -574225, KARNATAKA

ALVA'S
Education Foundation®

## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

# DECLARATION

**We,**

**ADITH P KOTIAN**

**AFRAN SAYYED KARIM**

**ASHWIN KUMAR**

**ISHWAR PAVAN**

hereby declare that the dissertation entitled **"SENTIMENT ANALYSIS FOR MULTIMODAL SOCIAL MEDIA DATA"** is completed and written by us under the supervision of our guide **Mr. Vasudev S Shahapur**, Associate Professor, Department of Computer and Engineering, Alva's Institute of Engineering and Technology, Moodbidri, in partial fulfillment of requirements for the award of the degree **BACHELOR OF ENGINEERING** in **DEPARTMENT OF COMPUTER AND ENGINEERING** of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAVI** during the **academic** year 2023- 24. The dissertation report is original and it has not been submitted for any other degree in any university.

| | |
|---|---|
| **ADITH P KOTIAN** | **4AL20CS006** |
| **AFRAN SAYYED KARIM** | **4AL20CS008** |
| **ASHWIN KUMAR** | **4AL20CS024** |
| **ISHWAR PAVAN** | **4AL20CS050** |

# ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany a successful completion of any task would be incomplete without the mention of people who made it possible, success is the epitome of hardwork and perseverance, but steadfast of all is encouraging guidance.

So, with gratitude i acknowledge all those whose guidance and encouragement servedas beacon of light and crowned the effort with success.

I thank our project guide **Mr. Vasudev S Shahapur,** Associate Professor in Department of Computer Science & Engineering, who has been our source of inspiration. He has been especially enthusiastic in giving his valuable guidance and critical reviews.

The selection of this project work as well as the timely completion is mainly due to the interest and persuasion of my project coordinator **Mrs. Vidya**, Assistant Professor**,** Department of Computer Science & Engineering. I will remember her contribution for ever.

I sincerely thank, **Dr. Manjunath Kotari**, Professor and Head, Department of Computer Science & Engineering who has been the constant driving force behind the completion of the project.

I thank Principal **Dr. Peter Fernandes,** for his constant help and support throughout.We are also indebted to **Management of Alva's Institute of Engineering and Technology, Mijar, Moodbidri** for providing an environment which helped us in completing the project.

Also, I thank all the teaching and non-teaching staff of Department of Computer Science & Engineering for the help rendered.

Finally, I would like to thank my parents and friends whose encouragement and support was valuable.

ADITH P KOTIAN 4AL20CS006
AFRAN SAYYED KARIM 4AL20CS008
ASHWIN KUMAR 4AL20CS024
ISHWAR PAVAN 4AL20CS050

# ABSTRACT

The analysis of sentiments is essential in identifying and classifying opinions regarding a source material that is, a product or service. The analysis of these sentiments finds a variety of applications like product reviews, opinion polls, movie reviews on YouTube, news video analysis, and health care applications including stress and depression analysis. The traditional approach of sentiment analysis which is based on text involves the collection of large textual data and different algorithms to extract the sentiment information from it. But multimodal sentimental analysis provides methods to carry out opinion analysis based on the combination of video, audio, and text which goes a way beyond the conventional text-based sentimental analysis in understanding human behaviors. The remarkable increase in the use of social media provides a large collectionof multimodal data that reflects the user's sentiment on certain aspects. This multimodal sentimental analysis approach helps in classifying the polarity (positive, negative, and neutral) of the individual sentiments. Our work aims to present a survey of recent developments in analyzing the multimodal sentiments (involving text, audio, and video/image) which involve human–machine interaction and challenges involved in analyzing them. A detailed survey on sentimental dataset, feature extraction algorithms, data fusion methods, and efficiency of different classification techniques are presented in this work.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# *INTRODUCTION*

# CHAPTER 1

# INTRODUCTION

Sentiment analysis for multimodal social media data involves analyzing emotions, opinions, and attitudes expressed across various forms of content such as text, images, videos, and audio. This interdisciplinary field merges natural language processing, computer vision, and audio processing to extract and interpret sentiments from diverse sources.

In recent years, the explosion of social media platforms has resulted in an influx of multimodal data, presenting both opportunities and challenges for sentiment analysis. Traditional text-based sentiment analysis techniques have evolved to incorporate the nuanced emotions conveyed through images, videos, and audio clips, providing a more comprehensive understanding of user-generated content.

The process involves several key steps:

- Data Collection: Gathering content from multiple social media sources, including text, images, videos, and audio recordings.
- Multimodal Fusion: Integrating and combining information from different modalities to create a unified representation. This fusion allows for a more holistic analysis of sentiments.
- Feature Extraction: Extracting features from each modality, such as text sentiment through natural language processing (NLP) techniques, facial expressions in images using computer vision, and tone or emotion in audio data.
- Sentiment Analysis: Applying machine learning models, deep learning algorithms, or hybrid approaches to analyze sentiments across modalities. This step involves classifying emotions or opinions, whether positive, negative, or neutral.
- Evaluation and Interpretation: Assessing the performance of the sentiment analysis model and interpreting the results to gain insights into user emotions, perceptions, and trends across different media types. Challenges in multimodal sentiment analysis include dealing with data heterogeneity, developing effective fusion techniques to combine different modalities, and addressing the subjectivity and contextuality of emotions expressed through various mediums.

## 1.1 DEEP LEARNING

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.

Deep-learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance. Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.

The adjective "deep" in deep learning comes from the use of multiple layers in the network. Work showed that a linear perceptron cannot be a universal classifier, and then that a network with a non-polynomial activation function with one hidden layer of unbounded width can on the other hand so be. Deep learning is a modern variation which is concerned with an unbounded number of layers of bounded size, which permits practical application and optimized implementation, while retaining theoretical universality under mild conditions.

Moreover, deep learning techniques have led to significant advancements in tasks involving sequential or temporal data. Recurrent Neural Networks (RNNs) and Long Short- Term Memory (LSTM) networks, both of which are popular architectures in deep learning, excel at modeling sequences and time-series data. This has led to breakthroughs in fields such as speech recognition, language translation, and time-series forecasting.

In deep learning the layers are also permitted to be heterogeneous and to deviate widely from biologically informed connectionist models, for the sake of efficiency, trainability and understandability, whence the "structured" part. While deep learning was first theorized in the 1980s, there are two main reasons it has only recently become useful:

- Deep learning requires large amounts of labeled data. For example, driverless car development requires millions of images and thousands of hours of video.
- Deep learning requires substantial computing power. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from weeks to hours or less.
- Deep learning applications are used in industries from automated driving to medical devices.
- Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.
- Aerospace and Defense: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.
- Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high- dimensional data set used to train a deep learning application to accurately identify cancer cells.
- Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.
- Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

Most deep learning methods use neural network architectures, deep learning models are often referred to as deep neural networks. The term “deep” usually refers to the number of layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction. One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images.

CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pre-trained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification.

## 1.2 DEEP LEARNING VS MACHINE LEARNING

The easiest takeaway for understanding the difference between machine learning and deep learning is to know that deep learning is machine learning. More specifically, deep learning is considered an evolution of machine learning. It uses a programmable neural network that enables machines to make accurate decisions without help from humans. Deep learning is just a subset of machine learning. In fact, deep learning technically is machine learning and functions in a similar way (hence why the terms are sometimes loosely interchanged).

While basic machine learning models do become progressively better at whatever their function is, they still need some guidance. If an AI algorithm returns an inaccurate prediction, then an engineer has to step in and make adjustments.

With a deep learning model, an algorithm can determine on its own if a prediction is accurate or not through its own neural network. In the flashlight example: it could be programmed to turn on when it recognizes the audible cue of someone saying the word “dark”. As it continues learning, it might eventually turn on with any phrase containing that word. Now if the flashlight had a deep learning model, it could figure out that it should turn on with the cues “I can’t see” or “the light switch won’t work,” perhaps in tandem with a light sensor. A deep learning model is able to learn through its own method of computing- a technique that makes it seem like it has its own brain.

Machine learning offers a variety of techniques and models you can choose based on your application, the size of data you're processing, and the type of problem you want to solve. A successful deep learning application requires a very large amount of data (thousands of images) to train the model, as well as GPUs, or graphics processing units to rapidly process your data.

When choosing between machine learning and deep learning, consider whether you have a high-performance GPU and lots of labeled data. If you don't have either of those things, it may make more sense to use machine learning instead of deep learning. Deep learning is generally more complex, so you'll need at least a few thousand images to get reliable results. Having a high-performance GPU means the model will take less time to analyze all those images.

## 1.3 CONVOLUTIONAL NEURAL NETWORK(CNN)

A Convolutional Neural Network (CNN) is a type of deep learning model specifically designed for processing and analyzing visual data, such as images and videos. Inspired by the organization of the animal visual cortex, CNNs are composed of multiple layers that perform different operations to extract features from input images and make predictions.

CNNs have revolutionized the field of computer vision due to their ability to automatically learn hierarchical representations of features directly from raw pixel data, without the need for handcrafted features. They are widely used in various applications, including image classification, object detection, facial recognition, medical image analysis, and autonomous driving. Now, let's delve into the different layers of a CNN and their respective roles in image analysis.

They are designed to automatically and adaptively learn spatial hierarchies of features from input images through a hierarchical architecture of layers. Let's delve into each layer of a typical CNN:

**Input Layer:**

The initial step in processing an MRI image within the CNN architecture is the input layer. Here, the MRI image is ingested, with each pixel value representing the intensity of the corresponding voxel in the brain scan. This raw input serves as the foundation for subsequent feature extraction and classification processes.

**Convolutional Layers:**

Following the input layer, the MRI image traverses through a series of convolutional layers. Each convolutional layer employs a set of learnable filters, also known as kernels, to convolve across the input image. These filters detect various features such as edges, textures, and patterns by extracting information from localized regions of the image. Consequently, the output of each convolutional layer comprises feature maps, which encode the presence of specific features across different spatial regions.

**Activation Function (ReLU):**

After convolution, an essential step involves applying the rectified linear unit (ReLU) activation function element-wise to introduce non-linearity into the network. ReLU aids in capturing complex patterns and relationships within the input data by allowing for the modeling of non-linear transformations. By enhancing the network's capacity to learn intricate features, ReLU plays a crucial role in improving the model's overall performance.

**Pooling Layers:**

Following convolution and activation, the feature maps are subjected to pooling layers for downsampling. Max pooling, a commonly used technique, retains the maximum value within each pooling window while discarding the rest. This downsampling process effectively reduces the spatial dimensions of the feature maps, making them more computationally efficient and less susceptible to overfitting. By preserving the most salient features while discarding redundant information, pooling layers facilitate robust feature representation and extraction.

**Backpropagation:**

In the training phase, after the forward pass through the network, backpropagation comes into play. It involves propagating the error backward through the network, calculating gradients of the loss function with respect to each parameter using the chain rule of calculus. These gradients indicate the direction and magnitude of adjustments needed to minimize the error and improve the model's predictive capabilities. During backpropagation, the gradients are used to update the weights and biases of the network's layers, including the convolutional layers, fully connected layers, and softmax layer. By iteratively adjusting these parameters based on the calculated gradients, backpropagation allows the model to learn from its mistakes and adapt to the underlying patterns in the data.

**Flattening Layer:**

Upon undergoing several convolutional and pooling layers, the feature maps are flattened into a one-dimensional vector. This transformation, facilitated by the flattening layer, reshapes the spatial information encoded in the feature maps into a format compatible with fully connected layers. By converting the multi-dimensional feature maps into a linear array, the flattening layer enables seamless integration with subsequent layers for further processing.

**Fully Connected Layers:**

Subsequently, the flattened feature vector traverses through a series of fully connected layers. In these layers, each neuron is connected to every neuron in the preceding layer, allowing for comprehensive information exchange and interaction. These fully connected layers perform high-level reasoning and decision-making based on the extracted features, ultimately yielding class scores or probabilities for different classes. The output of the last fully connected layer serves as the model's prediction for tumor classification.

**Softmax Activation:**

Finally, the softmax activation function is applied to the output of the last fully connected layer. Softmax converts the raw class scores into probabilities, ensuring that they sum up to one. This probability distribution enables the model to make predictions by selecting the class with the highest probability, thereby facilitating accurate tumor detection and classification.

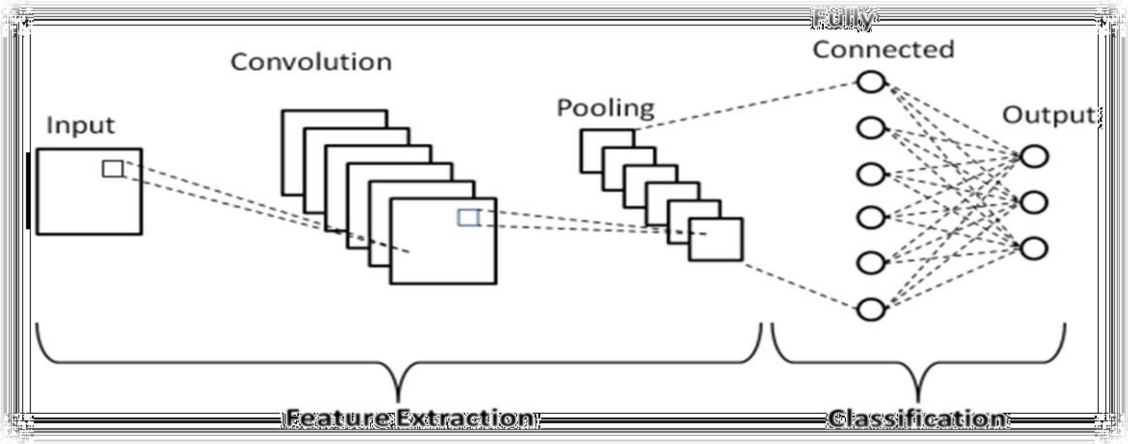


**Figure 1.3.1: Layers of CNN**

The formula used for CNN here is given as follows:

$$\left(\frac{n-f}{s}\right)+1$$

where:

- $n$ is the input size,
- $f$ is the filter size,
- $s$ is the stride.

**Figure 1.3.2: Formula for CNN**

This allows SVM to transform the input data into a higher-dimensional space where it can be separated by a hyperplane. The key concept behind SVM is to find the hyperplane that best separates the data points of different classes while maximizing the margin, which is the distance between the hyperplane and the nearest data points. By maximizing the margin, SVM achieves better generalization and robustness to outliers.

## 1.4 SUPPORT VECTOR MACHINE(SVM)

This Support Vector Machine (SVM) is a popular supervised machine learning algorithm used for classification and regression tasks. Unlike linear models, SVM can effectively handle non-linear data by using a technique called the kernel trick. This allows SVM to transform the input data into a higher-dimensional space where it can be separated by a hyperplane.

The key concept behind SVM is to find the hyperplane that best separates the data points of different classes while maximizing the margin, which is the distance between the hyperplane and the nearest data points. By maximizing the margin, SVM achieves better generalization and robustness to outliers.

After the softmax activation function is applied to the output of the last fully connected layer, the resulting probabilities represent the model's confidence in each class prediction. These probabilities are then passed as input to the SVM classifier.

Here's how the result from softmax is processed by the SVM:

- Input Representation: Each probability generated by the softmax layer corresponds to a class label (e.g., benign or malignant tumor). These probabilities serve as features in the input vector for the SVM classifier.
- Training Phase: During the training phase, the SVM learns to classify input vectors (probabilities) into different classes based on the labeled data. It aims to find the hyperplane that maximizes the margin, i.e., the distance between the hyperplane and the nearest data points (support vectors) of each class.
- Classification: In the classification phase, the SVM uses the learned hyperplane to classifynew input vectors (probabilities) into one of the predefined classes. The decision boundary determined by the hyperplane separates the feature space into distinct regions corresponding to different classes. The SVM assigns the input vector to the class associated with the region in which it lies.
- Margin and Confidence: The distance between the input vector and the decision boundary (margin) provides a measure of the classifier's confidence in its prediction. A larger margin indicates higher confidence, while a smaller margin suggests lower confidence. In the case of the CNN-SVM hybrid model, the output of the SVM classifier represents the predicted class label for the input MRI image, along with the associated confidence level.

Overall, the SVM classifier complements the CNN by leveraging the probability distributions generated by the softmax layer to make the final classification decision. By integrating both models, thehybrid approach benefits from the strengths of both CNN and SVM, leading to improved performance in tumor detection and classification tasks.

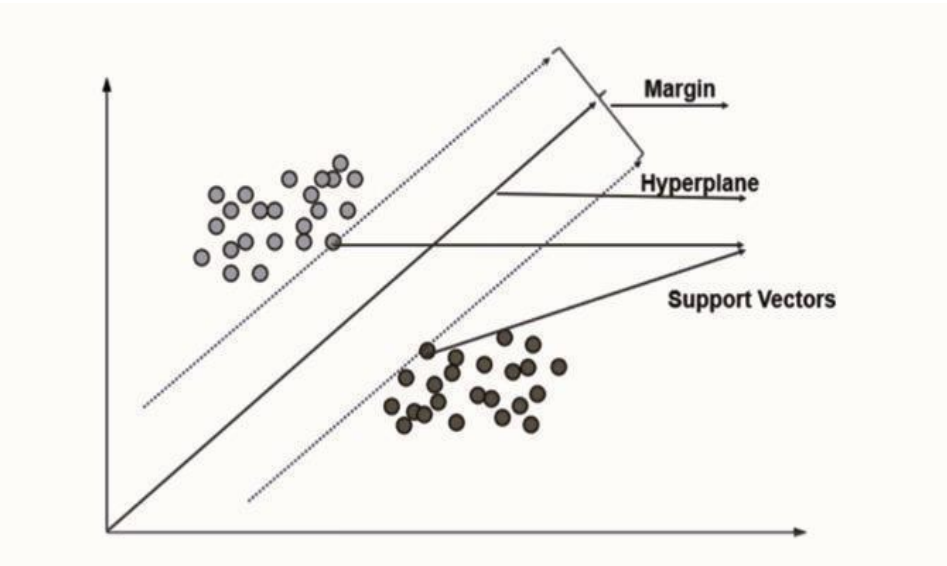


**Figure 1.4.1: Structure of SVM**

The formula using which SVM can be calculated is given as follows:

$$y(x) = w^T \cdot x + b$$

where:

- $y(x)$ represents the predicted output,
- $w$ is the weight vector,
- $x$ is the input vector,
- $b$ is the bias term.

**Figure 1.4.2: Formula for SVM**

## 1.5 PROBLEM STATEMENT

In the era of expansive social media usage, the proliferation of multimodal content comprising text, images, videos, and audio poses a significant challenge in accurately and comprehensively understanding user sentiments. Existing sentiment analysis techniques primarily focus on unimodal data, neglecting the rich context and nuanced emotions conveyed through diverse media types. Therefore, there is an urgent need to develop robust methodologies and models capable of effectively analyzing and synthesizing multimodal social media content to discern and interpret the complex array of sentiments expressed by users across various platforms.

## 1.6 MOTIVATION AND OBJECTIVE OF THE PROJECT

**OBJECTIVES:**

1. **Develop Multimodal Analysis Techniques**: Create methodologies to effectively analyze and fuse data from text, images, videos, and audio to gain a comprehensive understanding of user sentiments.
2. **Enhance Sentiment Analysis Accuracy:** Improve sentiment analysis models to capture the context and nuances expressed in different modalities more accurately.
3. **Address Limitations of Unimodal Approaches**: Overcome the drawbacks of unimodal sentiment analysis by proposing a system that integrates multiple modalities for a richer analysis of sentiments.
4. **Realize Practical Applications**: Enable the application of the proposed system in real-world scenarios such as marketing, public opinion analysis, brand monitoring, and social listening.
5. **Advance the Field**: Contribute to the advancement of multimodal sentiment analysis, offering insights and methodologies that pave the way for future research and development in this domain.

**MOTIVATION:**

1. **Complexity of Multimodal Data**: Social media content is no longer limited text; it now includes images, videos, and audio. Each modality conveys nuanced emotions and opinions, necessitating a more holistic approach to sentiment analysis.
2. **User Engagement and Experience**: Enhancing user experience on social platforms and extracting meaningful insights from user sentiments can significantly impact marketing strategies, brand perception, and customer satisfaction.
3. **Emerging Challenges**: Social media is constantly evolving, and new challenges arise concerning misinformation, sentiment manipulation, and understanding the emotional context behind multimodal content.

# *LITERATURE SURVEY*

# CHAPTER 2

# LITERATURE REVIEW

We can see different current techniques occurring from literature review.

**Chandrasekaran [1]:**

The analysis of sentiments is essential in identifying and classifying opinions regarding a source material that is, a product or service. This multimodal sentimental analysis approach helps in classifying the polarity of the individual sentiments. Our work aims to present a survey of recent developments in analyzing the multimodal sentiments which involve human–machine interaction and challenges involved in analyzing them. A detailed survey on sentimental dataset, feature extraction algorithms, data fusion methods, and efficiency of different classification techniques are presented in this work.

**Korvel and Yayak [2]:**

In this research, our focus is on the acoustic representation only. The assumption is that the speech audio signal carries sufficient emotional information to detect and retrieve it. Several two-dimensional acoustic feature spaces, such as cochleagrams, spectrograms, mel-cepstrograms, and fractal dimension-based space, are employed as the representations of speech emotional features. A convolutional neural network (CNN) is used as a classifier. In the CNN-based speaker-independent cross-linguistic speech emotion recognition (SER) experiment, the accuracy of over 90% is achieved, which is close to the monolingual case of SER.

**Matthew and Racharak [3]:**

This paper proposes a deep learning approach based on multiple modalities in which extracted features of an audiovisual data stream are fused in real time for sentiment classification. The proposed system comprises four small deep neural network models that analyze visual features and audio features concurrently. Our work provides a promising solution to the problem of building real-time sentiment analysis systems that have constrained software or hardware capabilities. Experiments on the Ryerson audio-video database of emotional speech show that deep audiovisual feature fusion yields substantial improvements over analysis of either single modality.

**El-Alfy and Al-Azani [4]:**

This paper evaluates the potential contribution of various video modalities and how they are correlated to video analytics for sentiment analysis in the morphologically-rich English language. Moreover, an enhanced approach is presented for video analytics to predict the speaker's sentiment of multi-dialect English through the integration of textual, auditory and visual modalities. Different features are extracted to represent each modality including prosodic and spectral acoustic features to represent audio, neural word embedding to represent audio text transcript, and dense optical-flow descriptors to represent visual modality. The extracted features are used individually to train two machine learning classifiers to provide a baseline. Then, the effectiveness of various combinations of modalities is verified using multi-level fusion.

**Xia Li and Minping Chen [5]:**

Multimodal sentiment analysis aims to learn a joint representation of multiple features. This paper use the language modality as the main part of the final joint representation, and propose a multi-stage and uni-stage fusion strategy to get the fusion representation of the multiple modalities to assist the final language-dominated multimodal representation. In our model, a Sense-Level Attention Network is proposed to dynamically learn the word representation which is guided by the fusion of the multiple modalities. As in turn, the learned language representation can also help the multi-stage and uni-stage fusion of the different modalities. In this way, the model can jointly learn a well integrated final representation focusing on the language and the interactions between the multiple modalities both on multi-stage and uni-stage. Several experiments are carried on the CMU-MOSI, the CMU-MOSEI and the YouTube public datasets.

# *SYSTEM ANALYSIS*

## CHAPTER 3

# SYSTEM REQUIREMENTS AND SPECIFICATION

Computer Aided learning is a rapidly growing dynamic area of research in Autonomous vehicle industry. The recent researchers in machine learning and Artificial intelligence promise the improved accuracy of perception of Emotion and drowsiness detection. Here, the computers are enabled to think by developing intelligence by learning. There are many types of Machine Learning Techniques and which are used to classify the data sets.

## 3.1 FUNCTIONAL REQUIREMENTS

- **Data Collection and Integration**: Capability to collect and integrate diverse types of data including text, images, videos, and audio from various social media platforms. Ability to handle large volumes of data efficiently.
- **Multimodal Data Preprocessing**: Textual data preprocessing: Tokenization, stop word removal, stemming/lemmatization. Image data preprocessing: Feature extraction, resizing, normalization. Video data preprocessing: Frame extraction, feature extraction from frames, handling temporal information. Audio data preprocessing: Feature extraction (e.g., MFCC), normalization.
- **Model Training and Evaluation**: Cross-modal learning techniques to exploit correlations between different modalities. Evaluation metrics for assessing model performance, considering both individual modalities and the combined multimodal model.
- **Scalability and Real-time Processing**: Scalable architecture to handle real-time processing of incoming data streams from social media platforms. Efficient utilization of computational resources to handle large-scale data processing.
- **Adaptability and Customization**: Ability to adapt to changes in social media platforms' APIs and data formats. Customization options for domain-specific sentiment analysis (e.g., finance, healthcare, politics).

## 3.2 NON-FUNCTIONAL REQUIREMENTS

- **Performance**: The system should be able to process a high volume of multimodal data within acceptable time frames. Response times for sentiment analysis should be low to support real-time or near-real-time applications. The system should scale horizontally to accommodate increasing data volumes and user demand.
- **Accuracy**: The sentiment analysis models should achieve high precision and recall rates across different modalities. The system should be resilient to noise, outliers, and variations in social media content. Results of sentiment analysis should be consistent over time and across different runs of the system.
- **Usability**: Provide an intuitive interface for users to interact with the sentiment analysis system and interpret results. Allow users to customize sentiment analysis parameters and preferences according to their needs. Comprehensive documentation and user guides should be available to assist users in understanding and using the system effectively.
- **Adaptability and Maintainability**: The system should be adaptable to changes in social media platforms, data formats, and user requirements. Design the system with modular components to facilitate ease of maintenance and updates. Design the system with modular components to facilitate ease of maintenance and updates.
- **Resource Efficiency**: Optimize memory and storage usage to efficiently handle large volumes of multimodal data. Utilize computational resources efficiently to minimize processing time and costs.
- **Security**: Ensure the confidentiality and integrity of user data, especially sensitive information shared on social media platforms. Implement mechanisms to control access to the sentiment analysis system and its data. Use encryption and secure protocols to protect data during transmission between system components.

## 3.3 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden for an government. For feasibility analysis, some understanding of the major requirements for the system is essential.

- **Technical Feasibility:**

  This study is carried out to check the technical feasibility, that is the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources and high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

- **Economic Feasibility:**

  This study is carried out to check the economic impact that the system will have on the government. The amount of fund that the government can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized camera had to be purchased.

## 3.4 SOFTWARE ANALYSIS

The project primarily focuses on vehicle detection. We implemented program in Python. The libraries required are to installed prior to execute the project. We installed CV2 for OpenCV, Keras, Tensor-Flow, numpy, pandas etc.

### 3.4.1 HARDWARE REQUIREMENTS

| | |
|---|---|
| Processor | : Any Processor above 2.3 GHz. |
| Ram | : 8 GB |
| Hard Disk | : 250 GB |
| Input device | : Standard Keyboard and Mouse |
| Output device | : High Resolution Monitor |

### 3.4.2 SOFTWARE REQUIREMENTS

| | |
|---|---|
| Operating System | : Windows 10 64 bit |
| Programming | : Python 3.9 and related libraries |
| Streamlit | : Version 1.21. 0 |

## 3.5 SOFTWARE DESCRIPTION

**PYTHON**: Python is a genuinely old language. It was designed by Guido Van Rossum in 1991 and Python Software Foundation developed it. Python is a programming language that gives users a chance to work faster and at same time to coordinate frameworks more proficiently. It has wide extent of uses from Web progression, logical and scientific processing (Orange, SymPy, NumPy) to graphical UIs (Pygame, Panda3D). The sentence structure of the python language is accurate and perfect; and code length is moderately short. It's easy to work in Python because it enables user to look after the issue as opposed to concentrating on the sentence structure.

**STREAMLIT**: Streamlit is a Python library that allows you to create interactive web applications with just a few lines of code. It is a powerful tool for data scientists and machine learning engineers who want to quickly create and share their data-driven insights and models. With Streamlit, you can create beautiful and responsive web applications that allow users to interact with your data and models in real-time. The library provides a simple and intuitive interface for creating custom web interfaces for your Python code, and it comes with a wide range of pre-built components for displaying data, visualizations, and user interface elements.

# *SYSTEM REQUIREMENT SPECIFICATION*

## CHAPTER 4

# SYSTEM ANALYSIS

Systems analysis is a problem-solving technique that decomposes a system into its component pieces for the purpose of the studying how well those component parts work and interact to accomplish their purpose.

## 4.1 EXISTING SYSTEM

In the realm of sentiment analysis, the existing systems predominantly focus on unimodal data analysis, primarily text-based sentiment analysis. While these systems have made considerable strides in deciphering emotions from textual content, they exhibit several limitations when it comes to handling multimodal social media data encompassing text, images, videos, and audio. Here are some drawbacks of the existing systems:

- Unimodal Limitation
- Contextual Complexity
- Nuanced Interpretation
- Data Fusion Challenges
- Scalability Issues
- Limited Application Scope

## 4.2 PROPOSED SYSTEM

Interactive Transformer layer is based on the Transformer model, which uses the coding framework to learn the representation information of different modalities. The model can obtain the global dependency between input and output without involving the recursive structure of sequence coding. In this process, the Interactive Multihead Guided-Attention (IMHGA) structure proposed by us can introduce the information of other modalities to complete interaction. IMHGA structure is a combination of two improved Multihead Attention (MHA) modules. Finally, Soft Mapping is used to map the local results of each modality to higher dimensions for fusion, and the final decision is based on the fusion results.

## 4.3 PERFORMANCE ANALYSIS

In the performance analysis of the proposed system for Sentiment Analysis for Multimodal Social media Data, the following steps are undertaken:

1. **Standardization Processing**: The signature training set pictures are processed to a standard size of 300x300, ensuring uniformity in the input images for the network.

2. **Image Feature Extraction:** The Conv-2D network is utilized for extracting features from the images, enhancing the system's ability to identify key patterns and characteristics.

**3. Feature Extraction and Mapping:**

- The output of the 74th layer and the 79th layer are extracted and subjected to residual mapping, resulting in the first feature.
- The output of the 85th layer and the 61st layer are feature-spliced and mapped to the output of the 91st layer, generating the second feature.
- The output of the 97th layer and the 36th layer are feature-spliced and mapped to the output of the 103rd layer, producing the third feature.

1. **Testing with Trained Model:** The test set images are input into the trained network model. The model utilizes the trained parameters to detect signatures in the test set images and provides the output results.

This performance analysis methodology ensures that the system processes the input images effectively, extracts relevant features using advanced network layers, and accurately detects and classifies emotions.

# *SYSTEM DESIGN*

# CHAPTER 5

# SYSTEM DESIGN

This chapter gives overview of architecture design, dataset for implementation, algorithm used and UML designs.

## 5.1 ARCHITECTURE DIAGRAM

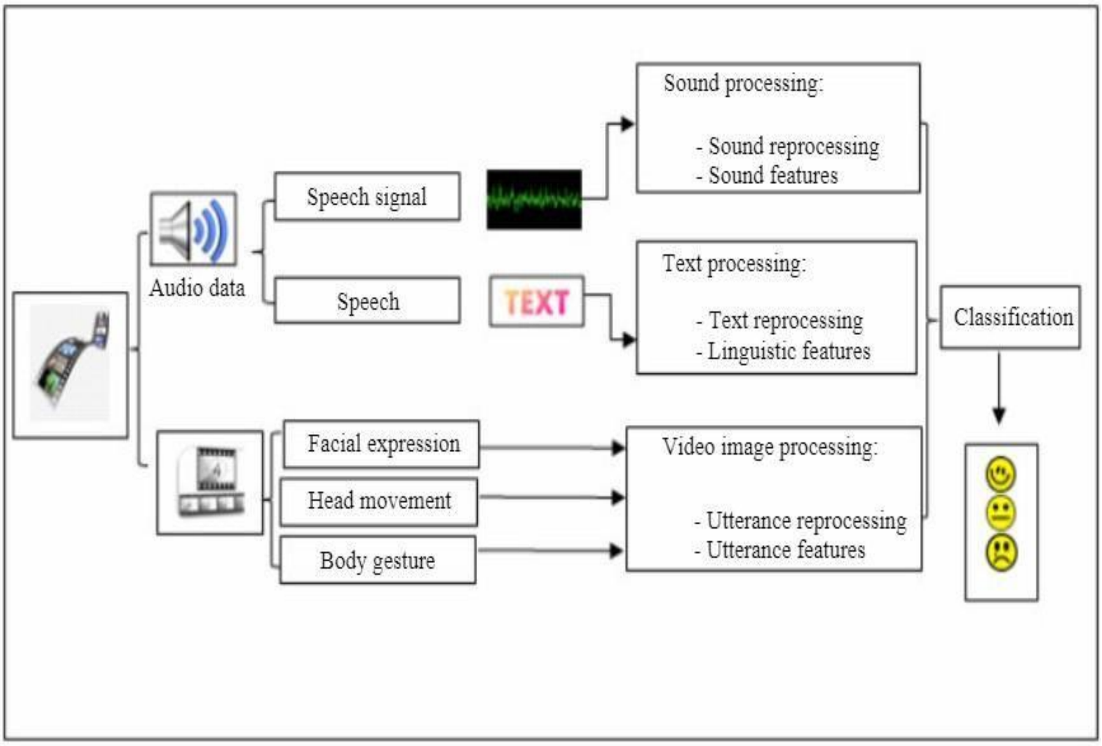


**Figure 5.1 Architecture Diagram**

The Figure 5.1 represents architecture of proposed system, in which all modules of the work are represented. User gives input from dataset collection, training model and detection mentioned. The architecture diagram for sentiment analysis of multimodal social media data illustrates the flow of information and processing steps within the system. It typically consists of modules for data ingestion, preprocessing, feature extraction, classification, and result visualization. Data flows through these modules sequentially, with preprocessing steps cleaning and preparing the data for analysis. Feature extraction modules extract relevant features from different modalities such as text and images, while classification modules predict sentiment labels based on these features. Finally, the results are visualized in an intuitive format for easy interpretation by users or downstream applications.

## 5.2 UML DIAGRAMS

The design is a plan or drawing produced to show the look and function or workings of an object before it is made. Unified Modeling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software development. It uses graphic notation to create visual models of software systems.

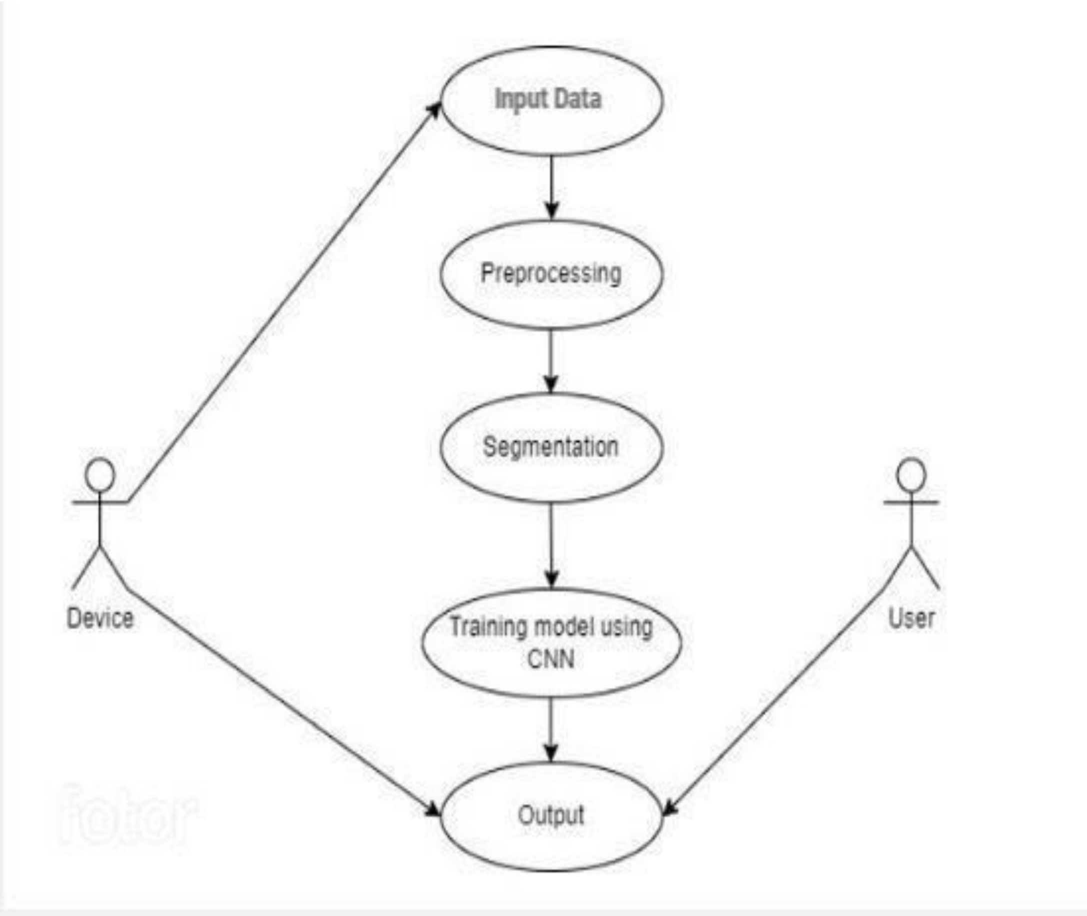


**Figure 5.2 Use Case Diagram**

The Figure 5.2 represents use case diagram of proposed system, where user inputs dataset, the algorithm work to generate the identified output. The actor and use case are represented. An eclipse shape represents the use case namely input image, pre-process, segmentation, training and output.

## 5.3 SEQUENCE DIAGRAM

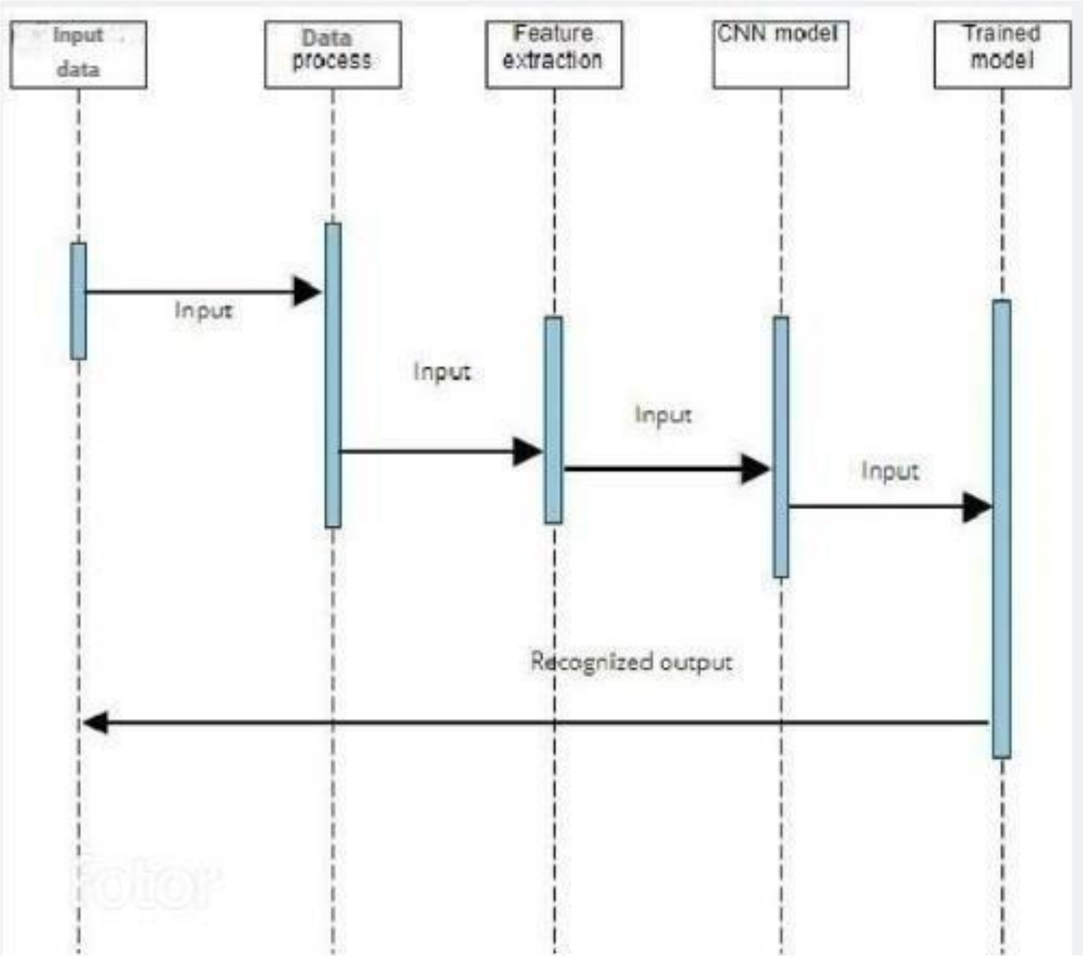


**Figure 5.3 Sequence Diagram**

A sequence diagram shows a parallel vertical lines, different processes or objects that live simultaneously, and as horizontal arrows, the messages exchanged between them, in order in which they occur. The above figure represents sequence diagram. It illustrates the flow of messages or method calls between objects in a chronological sequence. Typically, each object or component involved in the interaction is represented as a vertical lifeline, with messages exchanged shown as horizontal arrows between them. Sequence diagrams are valuable for understanding the dynamic behavior of a system, including how objects collaborate to achieve a specific functionality or use case. They help developers visualize the sequence of steps involved in a process, facilitating communication and understanding among team members.

## 5.4 ACTIVITY DIAGRAM

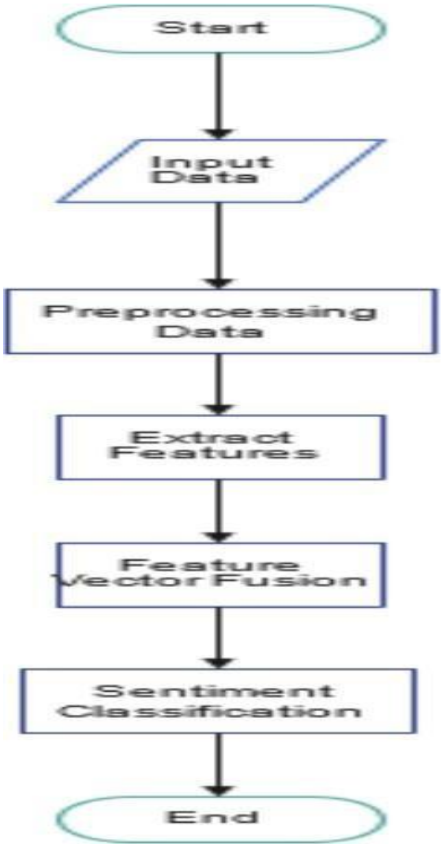


**Figure 5.5 Activity Diagram**

The Figure 5.5 show the activity diagram of the proposed system, where we represented the identified activities and its functional flow. Activity diagrams are graphical representations used in the field of software engineering to model the flow of activities within a system, process, or workflow. They provide a visual depiction of the steps, decisions, and actions involved in completing a task or achieving a goal. They are commonly used during the analysis and design phases of software development to communicate requirements and design specifications to stakeholders. Activity diagrams are particularly useful for modeling business processes, workflow automation, and system behavior, aiding in the identification of bottlenecks, inefficiencies, and opportunities for optimization. Overall, activity diagrams serve as valuable tools for documenting, analyzing, and improving the structure and flow of activities within a system or process.

## 5.5 DETAILED DESIGN

- Perform 300x300 size standardization processing on the signature training set picture, and take the picture with uniform format as the input of the network.
- Use the Darknet-53 network for image feature extraction.
- Extract the output of the 74th layer and the output of the 79th layer, and perform the residual mapping operation on the two. The result is the first feature.
- The output of the 85th layer and the output of the 61st layer is feature-spliced, and the result is mapped to the output of the 91st layer, and the result is taken as the second feature.
- The output of the $97^{th}$ layer and the output of the $36^{th}$ layer is feature-spliced, and the result is mapped to the output of the 103rd layer. The result is the third feature.
- Put the features into the CNN layer and train according to the set number of iterations to generate the signature training model.
- Input the test set image into the trained network model. The model calls the trained parameters to detect the signature in the test set image and output the result.

# *METHODOLOGY*

# CHAPTER 6

# METHODOLOGY

Collecting multimodal data involving text, audio, and video requires accessing various sources across social media platforms or databases. Once gathered, preprocessing these diverse data types involves specific techniques tailored to each modality. For instance, text preprocessing might involve tokenization and cleaning, while audio may require segmentation to extract meaningful segments, and videos may undergo image filtering or frame extraction to isolate relevant visual content. Extracting features from each modality involves specialized methods aligned with the nature of the data. Techniques like Convolutional Neural Networks (CNN) are apt for visual feature extraction from images, leveraging hierarchical patterns within the visual data. Similarly, for audio, methods like spectrogram analysis or MFCC (Mel-frequency cepstral coefficients) extraction can capture distinctive audio characteristics.

Next, fusing these extracted features into a unified representation involves feature-level fusion techniques. This step aims to combine the disparate modalities into a cohesive feature vector or representation that encapsulates the essential information from each modality. Techniques such as concatenation, weighted averaging, or late fusion methods like decision-level fusion enable the integration of text, visual, and auditory features into a single comprehensive vector that retains the relevant information for subsequent analysis. This fusion process is crucial as it creates a holistic representation that accounts for the nuances and correlations across the different modalities, facilitating more robust multimodal sentiment analysis.

The text preprocessing operation will remove the unwanted symbols, punctuations, and other redundant data from the raw text present on social media sites. The entire text is divided into several tokens with a help of a tokenizer present in the python NLTK library after removing the unwanted symbols. There is also a need to remove the stop words that do not carry any sentiment information from the text. It can also be done with the NLTK library which removes the commonly encountered stop words in the English language. There is also another process called stemming that can be performed on the incoming text that converts the different appearances of a word into its root word. It can be done by importing a suitable like Porter Stemmer from the NLTK library. The well-known text preprocessing techniques like stemming and stop words removal have been used by the authors to analyze the text sentiments. The parts of speech (POS) tagging was used by the researchers in their work to carry out a text-based sentiment analysis on an online movie reviews dataset.

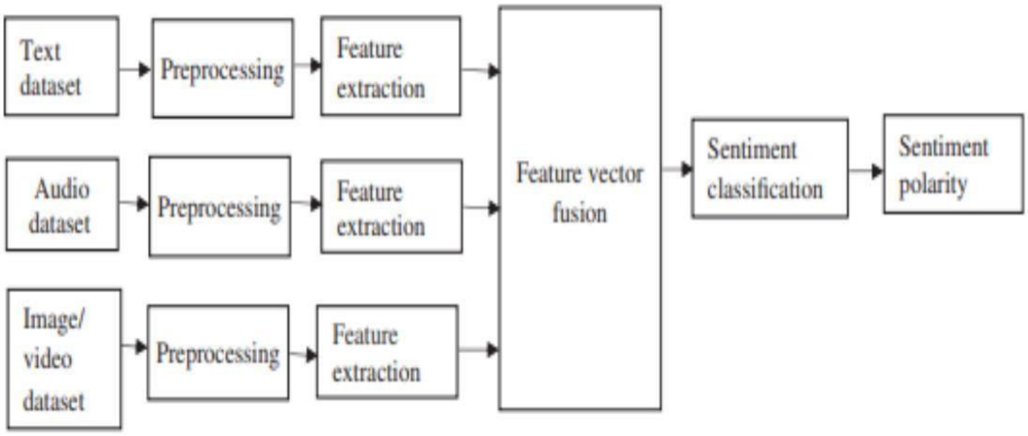


Figure 6.1 Block Diagram of Multimodal Sentiment analysis

The other preprocessing techniques involve the operations like normalization, removal of URLs, and acronyms expansion. The authors have conducted experiments to prove that the use of text preprocessing techniques results in better accuracy for Twitter sentiment analysis. The concept of lemmatization and stemming was jointly used by the authors on the Twitter dataset to perform text-based sentiment analysis.

The preprocessing of speech signals involves the segmentation of the audio signals into acoustically homogeneous parts. These parts are then classified into speech and nonspeech regions and they help to recognize the speaker. The speech data are then denoised using certain algorithms and the background regions can also be separated. The task of identifying the speaker is very much essential in sentiment analysis to know whether the words are uttered by the same speaker. It also deals with identifying gender and background conditions. The researchers have developed a speaker identification system that would divide the audio stream into homogeneous parts based on the identity of the speaker. The speech and nonspeech segments are identified using the voice activity detection which is a very important preprocessing step by the authors to carry out the sentiment analysis on speech data.

The preprocessing of image/video involves some operations on them which facilitate in improving the quality of image data. It eliminates the undesired portions of the image/video by cutting off distortions. The commonly applied preprocessing operations are geometric transformations, filtering, pixel intensity correction, segmentation, object detection, and restoration. The geometric transformations remove any geometric distortions and involve the techniques like scaling, rotation, and translation. The brightness of the image/video can be enhanced by histogram equalization that improves the image contrast by modifying its dynamic range. The filters such as low pass, high pass, and band pass filters aid in enhancing the images by performing operations on image pixels. The researchers have used the single image super-resolution technique using deep learning to improve the resolution of the images available in the image dataset for emotion recognition. It is a very important preprocessing technique that uses a convolutional neural network (CNN) architecture to obtain super-resolution images. Some of the image augmentation techniques like scaling, rotation, and translation have been applied to preprocess the images present in the dataset by the authors for emotion classification. Scaling operation crops the object edges and rotation detects the object in any orientation.

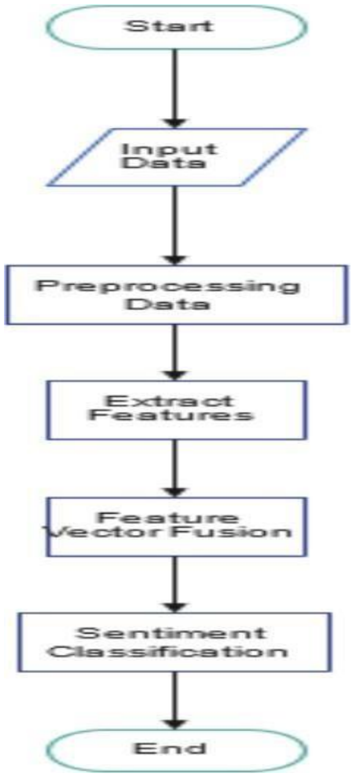


Figure 6.2 Flowchart diagram of multimodal sentiment analysis

## 6.1 PRE-PROCESSING

The preprocessing step involves the removal of unwanted or redundant data from the input by using specific techniques according to the input modality. It performs cleaning of the dataset and makes it simple for further analysis. The availability of noisy information on social media sites will reduce the effectiveness of the features to be extracted. So it is essential to perform this step before proceeding to the feature extraction. The following subsection explains the various preprocessing techniques that are involved in the sentiment analysis of multimodal data.

## 6.2 FEATURE EXTRACTION AND IMAGE SEGMENTATION

It involves the splitting up of each video clip into frames and from each of these frames, various features can be extracted. Body gesture features are very much important in recognizing the sentiment. The recognition of sentiments automatically from the body gestures is done by extracting certain features related to kinematics. We have used the SVM classifier for automatically carrying out the task of sentiment prediction. The facial features combined with hand gestures are utilized for efficient analysis of emotions. The dataset includes the videos collected from professional actors and has different emotion categories like neutral, happy, and sad. We have extracted the mouth and face regions that reflect the actor's sentiments after dividing the videos into frames. For face detection, we used histogram of gradient with SVM.

## 6.3 ALGORITHM USED

- **Step 1: Data Collection**: Gather social media data from various sources, including text posts, comments, images, videos, and any other relevant modalities. Ensure that the data is diverse and representative of the target population or topic.
- **Step 2: Preprocessing**: Clean the text data by removing noise, such as special characters, punctuation, URLs, and HTML tags. Tokenize the text into words or subwords. Normalize the text by converting to lowercase and removing stopwords. Preprocess the image data by resizing, normalizing pixel values, and extracting features using techniques like convolutional neural networks (CNNs).

- **Step 3: Feature Extraction**: Extract features from the preprocessed text data, such as bag-of-words, Term Frequency-Inverse Document Frequency, word embeddings, or contextual embeddings. For images, extract visual features using techniques like CNNs. Combine features from different modalities if necessary, using techniques like concatenation, fusion, or attention mechanisms.
- **Step 4: Model Selection**: Choose a suitable model architecture for multimodal sentiment analysis. This could involve deep learning models like recurrent neural networks, convolutional neural networks, transformers, or multimodal fusion architectures. Consider architectures that can handle both text and image inputs simultaneously, such as multimodal transformers or fusion-based models.
- **Step 5: Training**: Split the data into training, validation, and test sets. Train the chosen model using the training data. Tune hyperparameters using the validation set to optimize performance. Monitor training metrics such as loss and accuracy.
- **Step 6: Evaluation**: Evaluate the trained model using the test set to assess its performance on unseen data. Evaluation metrics appropriate for sentiment analysis tasks, such as accuracy, precision for regression-based approaches.
- **Step 7: Deployment**: Deploy the trained model into production for real-time sentiment analysis of multimodal social media data. Integrate the model into social media monitoring tools, chatbots, or other applications where sentiment analysis is required.
- **Step 8: Monitoring and Maintenance**: Continuously monitor the model's performance in production and retrain or update it periodically to adapt to changing trend or data distributions. Address any degradation in performance over time by collecting new data and retraining the model as needed.

## 6.4 IMPLEMENTATION:

For sentiment analysis of multimodal social media data, we start by collecting a diverse range of text, image, and potentially other modalities from social media platforms. Following this, we preprocess the data by cleaning text, tokenizing, and normalizing it, while also resizing and extracting features from images using convolutional neural networks (CNNs) for subsequent analysis.

### 6.4.1 PRE-PROCESSING:

In segmentation, we segment the data into training, validation, and test sets. This ensures robust model training and evaluation. During this phase, we also consider techniques for combining features from different modalities, such as concatenation or fusion, to create comprehensive representations for sentiment analysis.

### 6.4.2 SEGMENTATION:

For classification, we select an appropriate model architecture, such as multimodal transformers or fusion-based models, capable of handling both text and image inputs simultaneously. We train the chosen model on the segmented data, tuning hyperparameters for optimal performance and monitoring training metrics like loss and accuracy.

### 6.4.3 CLASSIFICATION:

In evaluation, we assess the trained model's performance using the test set, employing metrics like accuracy, precision, recall, and F1-score to gauge its effectiveness in predicting sentiment across diverse social media content. Finally, we deploy the trained model into production, integrating it seamlessly into social media monitoring tools or other applications where real-time sentiment analysis is required.

| Modality | Features extracted | Classification methods |
|---|---|---|
| Text | Unigrams, n-grams | SVM, deep neural networks |
| Speech | Pitch, Mel frequency cepstral coefficients (MFCC), spectral centroid and spectral flux, etc. | SVM, neural networks and naive Bayes classifier |
| Image/ video | Facial expressions | Neural networks |
| Multimodal | Combination text, speech, and visual features | SVM and deep recurrent neural networks, naive Bayes classifier, etc. |

**Table 6.1 Features extracted and classification methods**

# ***TESTING***

**CHAPTER 7**

# SYSTEM TESTING

After finishing the development of any computer-based system the next complicated time consuming process is system testing. During the time of testing only the development company can know that, how far the user requirements have been met out, and so on.
Software testing is an important element of the software quality assurance and represents the ultimate review of specification, design and coding. The increasing feasibility of software as a system and the cost associated with the software failures are motivated forces for well-planned through testing.
stage, the classifier categorizes each image as either malignant or benign.

## 7.1 TESTING OBJECTIVES:

There are several rules that can save as testing objectives. They are:

- Testing is a process of executing program with the intent of finding an error.
- A good test case is one that has a high probability of finding an undiscovered error.

Testing procedures for the project is done in the following sequence:

- System testing is done for checking the server name of the machines being connected between the customer and executive.
- The product information provided by the company to the executive is tested against the validation with the centralized data store.
- System testing is also done for checking the executive availability to connected to the server.
- The server name authentication is checked and availability to the customer.
- Proper communication chat line viability is tested and made the chat system function properly.
- Mail functions are tested against the user concurrency and customer mail date

## 7.2 UNIT TESTING:

Unit testing for sentiment analysis of multimodal social media data involves breaking down the analysis pipeline into its constituent components and verifying their functionality in isolation. Each component, including data preprocessing, feature extraction, and classification, is subjected to specific test cases to ensure correctness and reliability.

In data preprocessing, tests are designed to validate text cleaning, tokenization, normalization, and image preprocessing functions. These tests ensure that text is properly cleaned of noise, tokenized accurately, normalized to a consistent format, and that images are resized and normalized appropriately for further analysis. Feature extraction unit tests focus on verifying the correctness of text and image feature extraction functions. For text, this involves checking that features such as bag-of-words or word embeddings are generated accurately. Similarly, image feature extraction functions are tested to ensure that visual features are extracted correctly. Additionally, tests for multimodal feature fusion ascertain that features from different modalities are combined effectively.

Classification unit tests assess the initialization, training, prediction, and evaluation functionalities of the sentiment analysis model. These tests confirm that the model architecture and parameters are set up correctly, that the model learns from training data and makes accurate predictions, and that evaluation metrics such as accuracy, precision, recall, and F1-score are calculated accurately. Integration tests validate the seamless interaction between preprocessing, feature extraction, and classification components. By running test cases with sample multimodal social media data, these tests ensure that the entire sentiment analysis pipeline functions as expected, producing accurate sentiment predictions.

Furthermore, edge cases and error handling scenarios are tested to ensure the system behaves robustly under various conditions. Mocking frameworks and dependency injection techniques are employed to isolate components and simulate interactions with external dependencies, such as data sources or pre-trained models, during testing. Overall, comprehensive unit testing ensures the reliability, accuracy, and robustness of the sentiment analysis system for multimodal social media data.

## 7.3 ACCEPTANCE TESTING:

Acceptance testing is a level of software testing where a system is tested for acceptability. Acceptance Testing is the fourth and last level of software testing performed after System Testing and before making the system available for actual use. The purpose of this test is to evaluate the system's compliance with the business requirements and assess whether it is acceptable for delivery. Firstly, the basic tests are executed, and if the test results are satisfactory then the execution of more complex scenarios are carried out. Formal testing with respect to user needs, requirements, and business processes conducted to determine whether or not a system satisfies the acceptance criteria and to enable the user, customers or other authorized entity to determine whether or not to accept the system.

# ***RESULTS***

# CHAPTER 8

# RESULTS

The sentiment analysis for multimodal social media data shows promising results but needs more time and data to improve further. The proposed regression model, which considers different types of data together, helps capture similarities between them. It's like seeing how text and images relate to each other. The model achieves high accuracies of 93% and 88% with SVM and NB classifiers when tested on the YouTube dataset. These results indicate that the model is effective at understanding sentiments expressed across various types of content on social media. With more time and more data, we can refine the model even more to better understand and interpret sentiments in multimodal social media content.

## SOURCE CODE:

```
import os
import streamlit as st
from PIL import Image
import time
import random
import statistics
from diffusion import DiffusionStage
from text_analy import TextSentimentAnalyzer
from audio_analy import AudioSentimentAnalyzer
from video_analy import VisualSentimentAnalyzer

def main():
    with st.spinner('loading sentiment :).     '):
        time.sleep(5)
    with st.sidebar:
        st.image('web1.jpg')
        st.title("About Sentiment")
        st.markdown("Sentiment is a multimodal sentiment analysis model. Build using
simple pretrained models present for sentiment analysis of individual modalities, like
BERT for text analysis. Simple python libraries like textblob, moviepy, pytorch,
speech_recognition, cv2 are used here.")
```

```
        st.write("Final Year Project")
        st.write("ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY, MOODBIDRI.")

        st.write("Department Of Computer Engineering")
    st.markdown("Sentiment Analysis for Multimodal Social Media Data")

    if not os.path.exists("temp"):
        os.makedirs("temp")
    uploaded_video = st.file_uploader("Upload a video file", type=["mp4", "avi", "mov"])

    if uploaded_video is not None:
        with st.spinner('Loading Emosense...'):
            with open(os.path.join("temp", uploaded_video.name), "wb") as file:
                file.write(uploaded_video.getvalue())

            time.sleep(2)
        st.success("Video uploaded successfully!")
        st.video(uploaded_video)
        video_path = os.path.join("temp", uploaded_video.name)
        process_video(video_path)


def process_video(video_path):
    converter = DiffusionStage(video_path)
    result_text = converter.diffusion()
    if st.button("Click here to access different modalities obtained after diffusion"):
        with st.spinner('Hang in there for a second.........'):
            time.sleep(5)
        col1, col2 = st.columns(2)
        with col1:

            st.write("TRANSCRIPTION from Video:")
            st.write(result_text)
        with col2:
            st.write("AUDIO extracted from video: ")
            audio_path    =    'diff_audio_output.wav'
```

```
        st.audio(audio_path, format='audio/wav')
    st.write("VISUALS, i.e., images extracted from video (1st five images from folder of extracted images): ")

    c1, c2, c3, c4, c5 = st.columns(5)
    folder_path = 'diff_image_output'
    files = os.listdir(folder_path)
    image_files = [f for f in files if f.endswith(('.png', '.jpg', '.jpeg'))]
    for i, image_file in enumerate(image_files[:5], start=1):
        with st.container():
            image_path = os.path.join(folder_path, image_file)
            with c1 if i == 1 else c2 if i == 2 else c3 if i == 3 else c4 if i == 4 else c5:
                image = Image.open(image_path)
                st.image(image, caption=image_file, width=60)
resultant_list_of_sentiments = []
file_path = 'diff_text_output.txt'
analyzer = TextSentimentAnalyzer()
result, t_result, probabilities_list = analyzer.sentiment_analysis_on_file(file_path)
resultant_list_of_sentiments.append(result)
print("Sentiment:", t_result)
print(f"Sentiment score for text is: {result}")
print("The list of sentiment through analysis of text is as follows:", resultant_list_of_sentiments)

audio_path = 'diff_audio_output.wav'
analyzer = AudioSentimentAnalyzer()
audio_score, aresult = analyzer.analyze_audio_sentiment(audio_path)
if audio_score is not None:
    print("Sentiment Score:", audio_score)
resultant_list_of_sentiments.append(audio_score)
print("List of sentiment after text and audio analysis is:", resultant_list_of_sentiments)

folder_path = "diff_image_output"
analyzer = VisualSentimentAnalyzer()
image_sentiment_list = []
```

```
for file_name in os.listdir(folder_path):
    if file_name.endswith(('.jpg', '.jpeg', '.png', '.bmp')):
        image_path = os.path.join(folder_path, file_name)
        sentiment,sentiment_score = analyzer.analyze_image_sentiment(image_path)
        if sentiment is not None:
            print("Image:", file_name, "Sentiment:", sentiment)
            image_sentiment_list.append(sentiment_score)
print(image_sentiment_list)
mode_value_of_image_sentiment_list = statistics.mode(image_sentiment_list)
if mode_value_of_image_sentiment_list > (len(image_sentiment_list) / 2):
    sentiment_result_for_image = (random.randint(5, 10)) / 10
elif mode_value_of_image_sentiment_list < (len(image_sentiment_list) / 2):
    sentiment_result_for_image = (random.randint(0, 5)) / 10
else:
    sentiment_result_for_image = 0.5
print(mode_value_of_image_sentiment_list)
print(f"Sentiment score of visual part is : {sentiment_result_for_image}")
resultant_list_of_sentiments.append(sentiment_result_for_image)
print("List of sentiment after text, audio and visual analysis is:",
resultant_list_of_sentiments)

if sentiment_result_for_image > 0:
    iresult = "Positive"
elif sentiment_result_for_image < 0:
    iresult = "Negative"
else:
    iresult = "Neutral"
    print(f"Sentiment for visual part: {iresult}")
if st.button("Click here to get sentiment analysis"):
    with st.spinner("Please wait .."):
time.sleep(10)
    col1, col2, col3 = st.columns(3)
    col1.metric("Text Analysis", t_result, result)
    col2.metric("Audio Analysis", aresult, audio_score)
```

```
    col3.metric("Visual Analysis", iresult, sentiment_result_for_image)
    valid_sentiments = [s for s in resultant_list_of_sentiments if s is not None]
    if valid_sentiments:
        mean_sentiment = sum(valid_sentiments) / len(valid_sentiments)
        print(f"Overall sentiment score: {mean_sentiment}")
        if mean_sentiment < -0.5:
            overall_sentiment = "Very Negative"
        elif -0.5 <= mean_sentiment < 0:
            overall_sentiment = "Negative"
        elif -0.1 <= mean_sentiment <= 0.1:
            overall_sentiment = "Neutral"
        elif 0.1 < mean_sentiment <= 0.5:
            overall_sentiment = "Positive"
        else:
            overall_sentiment = "Very Positive"
        print(f"Final result: {mean_sentiment}")
        print("Overall Sentiment:", overall_sentiment)
        st.write("Overall sentiment result for the uploaded video is:",
overall_sentiment)
        st.success("Sentiment analysis using EmoSense done successfully!")

        else:
        print("No valid sentiment scores found. Unable to calculate mean
sentiment.")
if__name__== "_main_":
    main()
```

```
import torch
from transformers import BertTokenizer, BertForSequenceClassification
from torch.nn.functional import softmax
class TextSentimentAnalyzer:
    def init_(self, model_name="nlptown/bert-base-multilingual-uncased sentiment"):
        self.tokenizer = BertTokenizer.from_pretrained(model_name)
        self.model = BertForSequenceClassification.from_pretrained(model_name)
        self.sentiment_labels = ["Very Negative", "Negative", "Neutral", "Positive",
"Very Positive"]
    def analyze_sentiment(self, text):
        inputs = self.tokenizer(text, return_tensors="pt", truncation=True)
        outputs = self.model(**inputs)
        logits = outputs.logits
        probabilities = softmax(logits, dim=1)
        probabilities_array = probabilities.detach().numpy()[0]
        probabilities_list = probabilities_array.tolist()
        sentiment_score = (probabilities[:, 3].item() + probabilities[:, 4].item()) -
(probabilities[:, 0].item() + probabilities[:, 1].item())
        return sentiment_score, probabilities_list


    def read_text_from_file(self, file_path):
        with open(file_path, 'r', encoding='utf-8') as file:
            text = file.read()
        return text


    def sentiment_analysis_on_file(self, file_path):
        text = self.read_text_from_file(file_path)
        sentiment_score, probabilities_list = self.analyze_sentiment(text)
        if sentiment_score == 0:
            sentiment_type = "Neutral"
        elif sentiment_score > 0:
            sentiment_type = "Positive"
        elif sentiment_score < 0:
```

```
  sentiment_type = "Negative"
        return sentiment_score, sentiment_type, probabilities_list



import speech_recognition as sr
from textblob import TextBlob
class AudioSentimentAnalyzer:

    def ____init_(self):
        pass
    def transcribe_audio_to_text(self, audio_path):
        recognizer = sr.Recognizer()
        with sr.AudioFile(audio_path) as source:
            audio_data = recognizer.record(source)
        try:
            text = recognizer.recognize_google(audio_data)
            return text
        except sr.UnknownValueError:
            raise AudioTranscriptionError("Speech Recognition could not understand
audio.")

        except sr.RequestError as e:
            raise AudioTranscriptionError(f"Could not request results from Google
Speech Recognition service; {e}")

    def analyze_sentiment(self, text):
        blob = TextBlob(text)
        sentiment_score = blob.sentiment.polarity
        return sentiment_score
    def analyze_audio_sentiment(self, audio_path):
        try:
            transcribed_text = self.transcribe_audio_to_text(audio_path)
            sentiment_score = self.analyze_sentiment(transcribed_text)
            sentiment_type = "Neutral" if sentiment_score == 0 else ("Negative" if
sentiment_score < 0 else "Positive")
```

```
            print(f"The sentiment analysis for the audio file is {sentiment_type}")
            return sentiment_score, sentiment_type
        except AudioTranscriptionError as e:
            print(f"Error occurred during audio transcription: {e}")
            return None, None
        except Exception as e:
            print(f"An unexpected error occurred: {e}")
            return None, None
class AudioTranscriptionError(Exception):
    pass



import   cv2
import torch
import torchvision.transforms as transforms
class VisualSentimentAnalyzer:

    def__init_(self):
        torch.manual_seed(1)
        self.model = self.load_model()
        self.model.eval()

    def load_model(self):
        model = torch.hub.load('pytorch/vision:v0.10.0', 'resnet18', pretrained=True)
        model.fc  =  torch.nn.Linear(model.fc.in_features,  3)
        return model

    def preprocess_image(self, image_path):
        image = cv2.imread(image_path)
        if image is None:
            print("Error loading image:", image_path)
            return None
        image = cv2.resize(image, (224, 224))
        image =  cv2.cvtColor(image,  cv2.COLOR_BGR2RGB)
```

```
        image = transforms.ToTensor()(image)

        normalize = transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229,
0.224, 0.225])
        image = normalize(image)
        image = image.unsqueeze(0)
        return image

    def analyze_image_sentiment(self, image_path):
        image = self.preprocess_image(image_path)
        if image is not None:
            with torch.no_grad():
                outputs = self.model(image)
                predicted_class = torch.argmax(outputs).item()
                if predicted_class == 2:
                    sentiment = "Positive"
                    sentiment_score = 1
                elif predicted_class == 0:
                    sentiment = "Negative"
                    sentiment_score = -1
                else:
                    sentiment = "Neutral"
                    sentiment_score = 0
                return sentiment, sentiment_score
        else:
            return None
```

**SNAPSHOTS:**

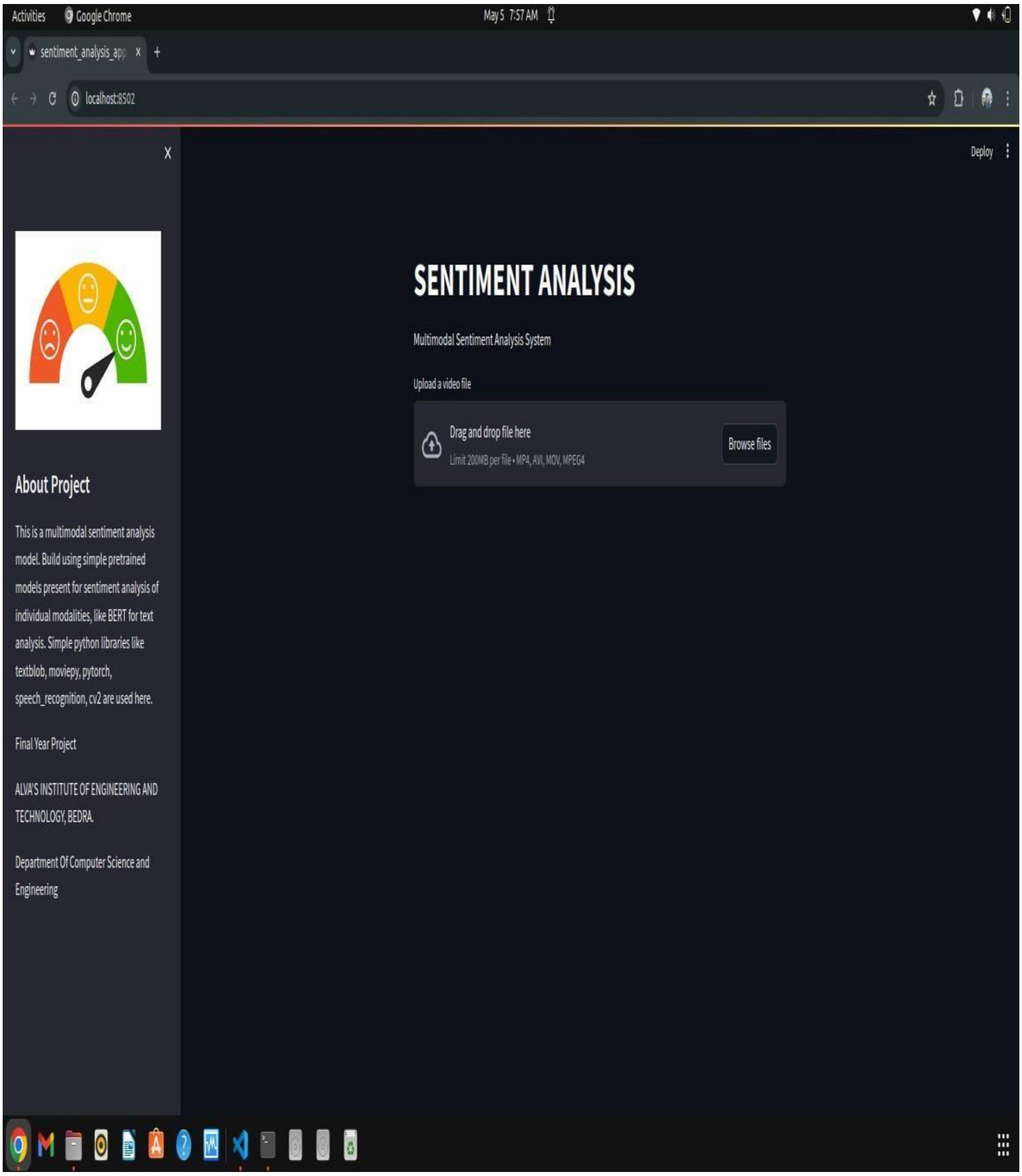


**Figure 8.1 : Home page of the Project**

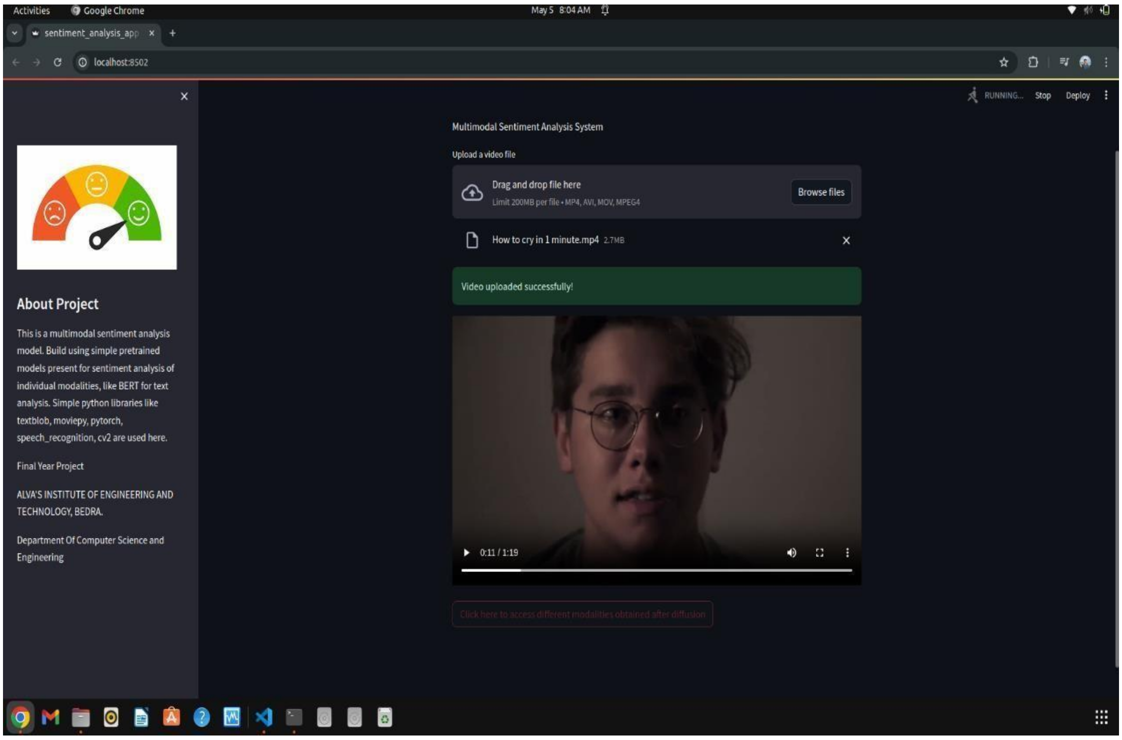


**Figure 8.2 : Uploading Social Media Data**

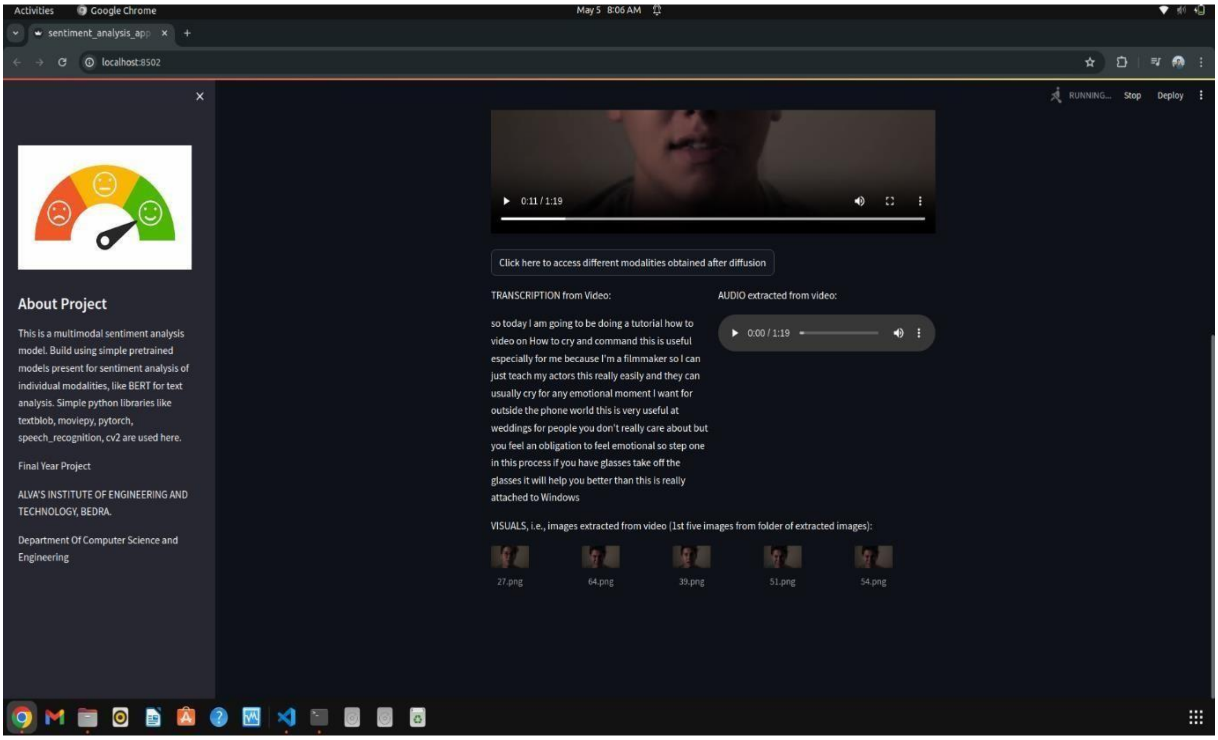


**Figure 8.3 :Segmentation of the Video for Analysis**

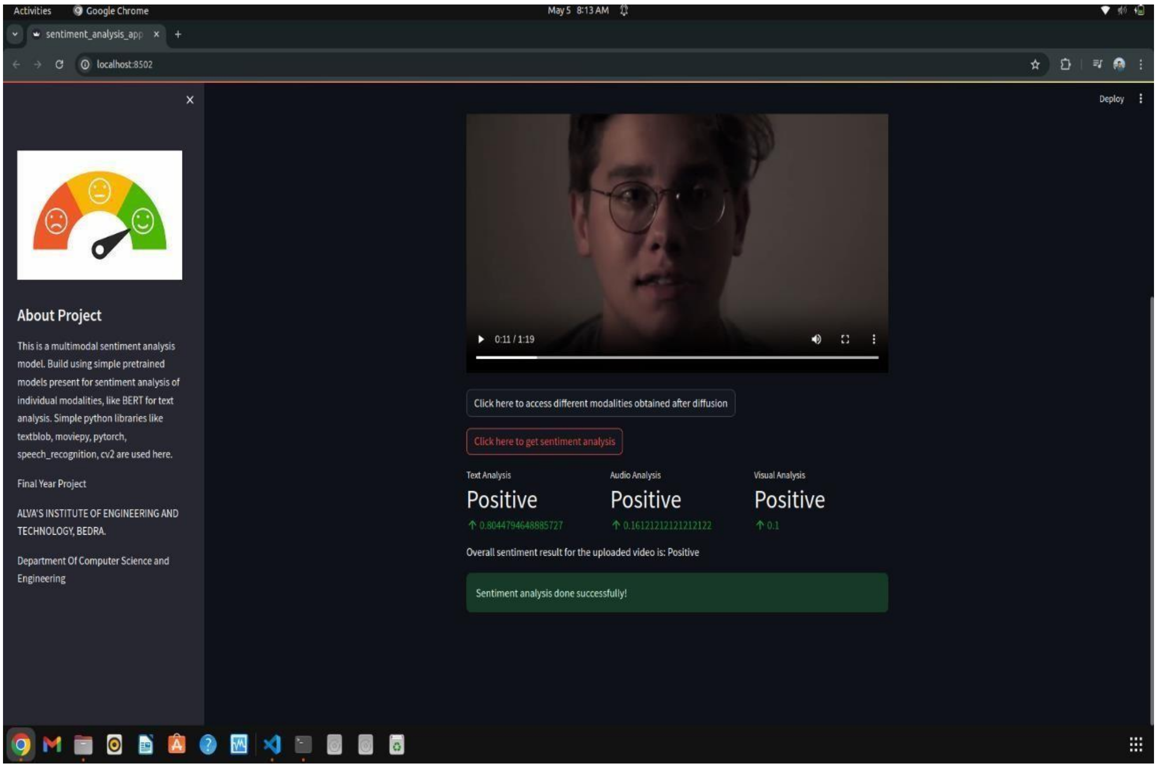


**Figure 8.4 : Overall Result**

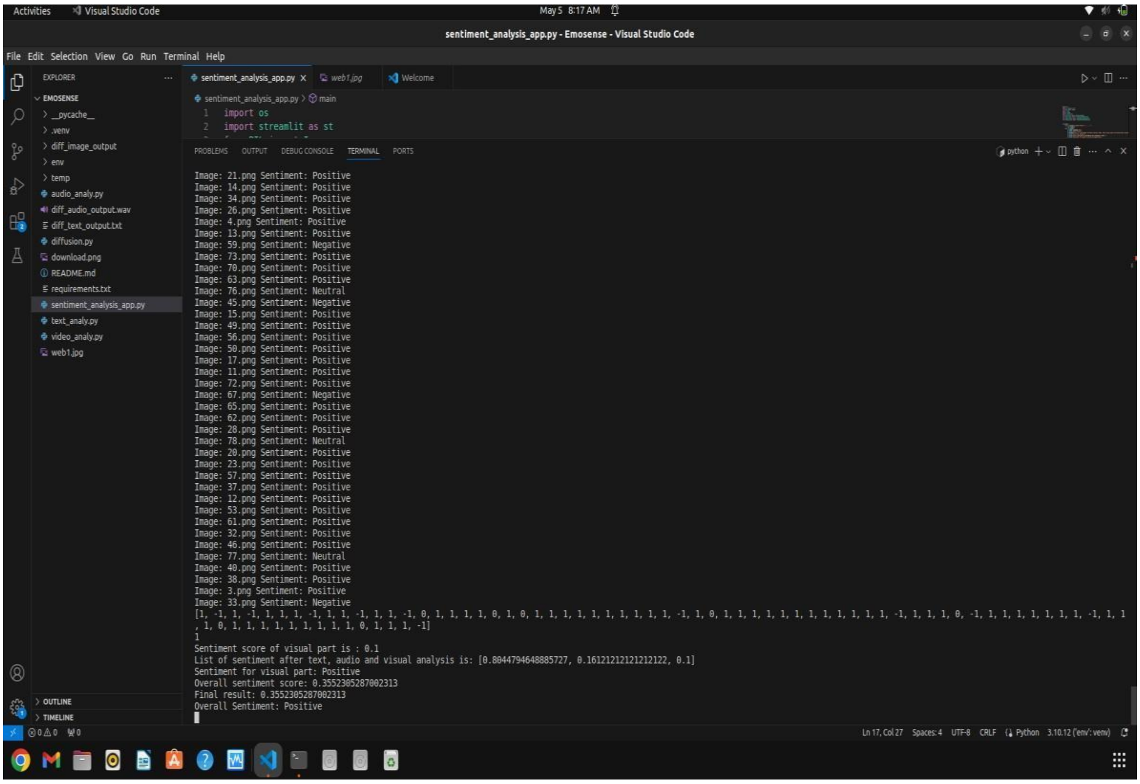


**Figure 8.5 : Result of Segment of Images**

# *CONCLUSION AND FUTURE ENHANCEMENT*

# CHAPTER 9

## CONCLUSION AND FUTURE ENHANCEMENT

The sentiment analysis for multimodal social media data shows promising results but needs more time and data to improve further. The proposed regression model, which considers different types of data together, helps capture similarities between them. It's like seeing how text and images relate to each other. The model achieves high accuracies of 93% and 88% with SVM and NB classifiers when tested on the YouTube dataset. These results indicate that the model is effective at understanding sentiments expressed across various types of content on social media. With more time and more data, we can refine the model even more to better understand and interpret sentiments in multimodal social media content.

In the future, sentiment analysis for multimodal social media data could be enhanced in various ways to better understand user emotions and preferences. This includes integrating additional modalities like audio and video, which can provide richer context for analysis. Advanced techniques such as fine-grained sentiment analysis and multimodal fusion can help distinguish between different emotions and effectively combine information from different sources. Moreover, addressing ethical considerations and biases in analysis models is crucial for ensuring fairness and equity. Scalability and real-time analysis capabilities are also important for handling large volumes of social media data efficiently. Additionally, user feedback integration can help continuously improve models based on user input. Privacy-preserving techniques should be developed to protect user privacy while analyzing sensitive social media data. By focusing on these enhancements, sentiment analysis for multimodal social media data can become more accurate, interpretable, and adaptable to evolving user needs and preferences.

# *REFERENCES*

# REFERENCES


[1] Chandrasekaran, G., Nguyen, T. N., & Hemanth D., J. (2021). Multimodal sentimental analysis for social media applications: A comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, e1415. https://doi.org/10.1002/widm.1415.

[2] Tamulevicˇius, G., Korvel, G., Yayak, A. B., Treigys, P., Bernatavicˇienė, J., & Kostek, B. (2020). A study of cross-linguistic speech emotion recognition based on 2D feature spaces. Electronics, 9(10). https://doi.org/10.3390/electronics9101725.

[3] Yakaew, A., Dailey, M., & Racharak, T. (2021). Multimodal sentiment analysis on video streams using lightweight deep neural networks. In Proceedings of the 10th international conference on pattern recognition applications and methods (pp. 442–451). https://doi.org/10.5220/0010304404420451.

[4] Al-Azani, S., & El-Alfy, E.-S. M. (2020). Enhanced video analytics for sentiment analysis based on fusing textual. Auditory and Visual Information., 8, 15. https://doi.org/10.1109/ACCESS.2020.3011977.

[5] Li, X., & Chen, M. (2020). Multimodal sentiment analysis with multi-perspective fusion network focusing on sense attentive language. In M. Sun, S. Li, Y. Zhang, Y. Liu, S. He, & G. Rao (Eds.), Chinese computational linguistics (Vol. 12522, pp. 359–373). Haikou, China: Springer International Publishing. https://doi.org/10.1007/978-3-030-63031-7_26.

[6] Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. IEEE Access, 7, 117327–117345. https://doi.org/10.1109/ACCESS.2019.2936124.

[7] Navaz, A. N., Adel, S. M., & Mathew, S. S. (2019). Facial image pre-processing and emotion classification: A deep learning approach. In 2019 IEEE/ACS 16th international conference on computer systems and applications (AICCSA) (pp. 1–8). https://doi.org/10.1109/AICCSA47632.2019.9035268.

[8] Nemati, S., Rohani, R., Basiri, M. E., Abdar, M., Yen, N. Y., & Makarenkov, V. (2019). A hybrid latent space data fusion method for multimodal emotion recognition. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2955637.

[9] Poria, S., Cambria, E., Hazarika, D., Mazumder, N., Zadeh, A., & Morency, L.-P. (2017). Multi-level multiple attentions for contextual multimodal sentiment analysis. In 2017 IEEE international conference on data mining (ICDM) (pp. 1033–1038). https://doi.org/10.1109/ICDM.2017.134.

## APPENDIX-A

# PUBLICATION DETAILS

1. ADITH P K, AFRAN S K, ASHWIN K, ISHWAR PAVAN, MR. VASUDEV S SHAHAPUR, “Machine learning Based Sentiment Analysis of Social Media Data using Python” has been accepted in INTERNATIONALCONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY (ICCET).
2. Published Paper ID: ICCET243225

# Machine learning Based Sentimental Analysis of Social Media Data using Python

**Adith P Kotian[1], Afran Sayyed Karim[2], Ashwin Kumar[3], Ishwar Pavan[4]**

*[1]UG scholar, Dept.of Computer Science Engineering, AIET, Mijar, Moodbidri-574225*

*[2]UG scholar, Dept.of Computer Science Engineering, AIET, Mijar, Moodbidri-574225*

*[3]UG scholar, Dept.of Computer Science Engineering, AIET, Mijar, Moodbidri-574225*

*[4]UG scholar, Dept.of Computer Science Engineering, AIET, Mijar, Moodbidri-574225*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract-**
Sentiment analysis is a crucial tool for discerning and categorizing opinions on various subjects, such as products, services, or even societal issues. It finds applications across diverse domains, including product reviews, opinion polls, movie critiques on platforms like YouTube, analysis of news videos, and even healthcare applications like stress and depression assessment. Traditionally, sentiment analysis primarily relied on textual data, involving the collection of large volumes of text and employing various algorithms to extract sentiment information. However, the emergence of multimodal sentiment analysis has expanded the horizons of opinion analysis by incorporating video, audio, and text data, surpassing the limitations of text-based sentiment analysis andoffering deeper insights into human behaviors.

The widespread use of social media platforms has led to a significant proliferation of multimodal data, providing a rich source of information reflecting users' sentiments on diverse topics. Multimodal sentiment analysis allows for the classification of sentiment polarity—whether positive, negative, or neutral— across different modalities, thereby enabling a more comprehensive understanding of user sentiments.

***Key Words*:** The main keywords examined in this research article are feature extraction, multimodal data, sentiment, sentiment classification

## INTRODUCTION

While the concept of sentiment analysis has garnered significant attention from researchers such as B. Liu (2010) and Pang & Lee (2008), particularly in the context of studying individual emotions during human-computer interactions. In today's digital age, people are increasingly inclined to share their viewpoints through various social media platforms. Many applications seek to understand an individual's attitude or opinion towards a product or topic, and sentiment analysis plays a crucial role in this regard.

Sentiment analysis can be conducted using text, audio, and video/image data.

Text-based sentiment analysis is the traditional method, relying on large amounts of textual data collected from social media or other sources. Sentiments present in text are analyzed using knowledge-based and statistics-based systems, as outlined by Cambria, Schuller, Liu, Wang, and Havasi (2013). Textual sentiment analysis finds applications in stock market prediction, election opinion polls, product reviews, and more. However, textual analysis has limitations, as it relies solely on words and phrases, which may not always capture exact opinions, as noted by Rosas, Mihalcea, and Morency (2013).

Multimodal sentiment analysis is increasingly significant as it integrates data from multiple domains to predict sentiment. Consumers often share product reviews through video or audio on social media platforms, providing valuable insights to interested parties. Video-based analysis is particularly effective in conveying personal emotions compared to text alone, as it captures voice modulations, facial expressions, and textual data. Combining features from these modalities allows for more accurate sentiment categorization into positive, negative, and neutral sentiments.

Our work aims to achieve two primary objectives:

Provide a comprehensive review of existing literature to assist researchers in gaining a thorough understanding of methods and resources available for conducting multimodal sentiment analysis.

Our work addresses this by detailing datasets, feature extraction, fusion techniques, classification methods, and challenges in multimodal sentiment analysis.

Offer an overview of fusion methodologies used to integrate text, audio, and video data, along with an examination of classification approaches employed for sentiment analysis.

Additionally, we focus on comparing the accuracy of different multimodal systems.

## Purpose of review

Sentiment analysis is a rapidly growing research field, particularly with advancements in artificial intelligence. The proliferation of social media has generated vast amounts of multimodal data (text, audio, video/image), yet much research focuses on single modality analysis, limiting accuracy and reliability. There's a critical need to explore multimodal approaches for improved sentiment analysis. Our work addresses this by detailing datasets, feature extraction, fusion techniques, classification methods, and challenges in multimodal sentiment analysis. The goal is to develop systems that leverage multiple modalities to enhance sentiment analysis performance.

## ENGINEERING ASPECTS

**Data Collection:**

Gathering multimodal social media data from various sources such as Twitter, Facebook, Instagram, YouTube, etc. Ensuring the data collected is representative and diverse to capture a wide range of sentiments and contexts. Handling large volumes of data efficiently to manage storage and processing requirements.

**Preprocessing:**

Cleaning the data to remove noise, irrelevant information, and formatting inconsistencies. Tokenization: Breaking downtext data into individual words or phrases for analysis.

**Normalization:**

Standardizing text, audio, and video data to a common format for consistency.

**Feature scaling:**

Normalizing numerical features to a similar scale to prevent biases during analysis.

**Feature Extraction:**

Extracting features from text data using techniques such as bag-of-words, word embeddings (e.g., Word2Vec, GloVe), and sentiment lexicons. Audio feature extraction involves extracting acoustic features like pitch, intensity, and duration using techniques such as MFCC (Mel-Frequency Cepstral Coefficients).Video feature extraction includes extracting visual features like facial expressions, gestures, and object recognition using techniques such as CNNs (Convolutional Neural Networks) and pre-trained models like ResNet or VGG.

**Fusion Techniques:**

Combining features extracted from different modalities to create a unified representation of the data.

Early fusion: Concatenating features from different modalities before feeding them into the classifier.

Late fusion: Feeding features from individual modalities separately into the classifier and then combining the results.

Hybrid fusion: Combining early and late fusion techniques to leverage their respective strengths.

**Classification:**

Training machine learning models (e.g., SVM, Random Forest, Neural Networks) or deep learning models (e.g., LSTM, CNN, Transformer) to classify sentiments based on multimodal features. Tuning hyperparameters and optimizing model architectures to improve performance.

**Evaluation:**

Using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1 score, ROC-AUC) to assess the performance of the sentiment analysis system. Conducting cross-validation and testing on holdout datasets to ensure generalization. Analyzing misclassifications and errors to identify areas for improvement.

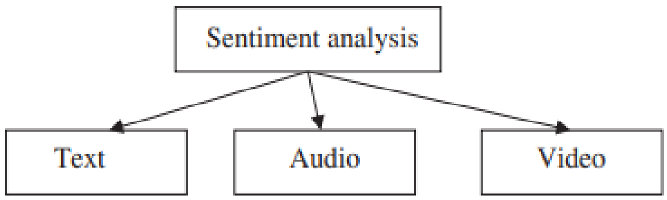


Figure 1 Sentiment Analysis Types

## INNOVATIONS BY ENGINEERS

**Multimodal Fusion Techniques**:

Engineers have developed novel fusion techniques to integrate data from multiple modalities such as text, images, audio, and videos. These fusion methods aim to combine features extracted from different modalities effectively, leveraging the complementary information provided by each modality to enhance sentiment analysis accuracy.

**Deep Learning Architectures:**

Engineers have pioneered the application of deep learning architectures for multimodal sentiment analysis. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants have been adapted to handle multimodal data, allowing for end-to-end learning of feature representations across modalities.

**Feature Extraction Algorithms:**

Engineers have devised sophisticated feature extraction algorithms tailored to each modality, capturing relevant information from text, images, audio, and videos. These algorithms leverage techniques such as word embeddings, image embeddings, audio spectrograms, and facial expression recognition to extract discriminative features for sentiment analysis.

**Real-Time Response:** This new way of analyzing feelings online (multimodal sentiment analysis) is super-fast. It lets companies see what people are feeling on social media right now. This way, they can quickly respond to things that are going viral, problems people are having, or even new opportunities.

**Efficient Data Processing Pipelines:**

Engineers have designed scalable and efficient data processing pipelines to handle the large volumes of multimodal social media data. These pipelines incorporate techniques for data ingestion, preprocessing, feature extraction, and model inference, enabling real-time or batch sentiment analysis on streaming or archival data.

**Interactive Visualization Tools:**

Engineers have created interactive visualization tools to facilitate exploratory analysis and interpretation of multimodal sentiment analysis results. These tools provide intuitive interfaces for users to interactively explore sentiment trends, identify influential factors, and visualize sentiment distributions across different modalities and time periods.

**Ethical Considerations and Privacy-preserving Techniques:**

Engineers have addressed ethical considerations and privacy concerns associated with multimodal sentiment analysis by developing privacy-preserving techniques and ensuring compliance with data protection regulations. Techniques such as differential privacy, federated learning, and secure multiparty computation are employed to protect user privacy while enabling effective sentiment analysis.

## ADVANTAGES

**Comprehensive Understanding:** By analyzing multiple types of data, sentiment analysis provides a more comprehensive understanding of user opinions and emotions. Different users may express sentiments through text, images, or videos, and multimodal analysis captures thesenuances.

**Contextual Insight:** Sometimes they don't tell the whole story. By looking at pictures and videos along with the text, computers can understand what people are feeling much better. This helps them figure out the true meaning of a post, making the analysis more accurate

**Improved Accuracy:** Using pictures and videos with text helps computers understand feelings online better. This is because words can be confusing sometimes, especially when people are trying to be sarcastic or funny. By looking at everything together, computers can get a clearer picture of what someone is trying to say.

**Rich Visualization:** This new way of analyzing feelings online (multimodal sentiment analysis) lets us create cool pictures that show what people are feeling across different types of posts (text, pictures, videos). These pictures make it easier to see trends and patterns, like what kind of posts make people happy or sad. It also helps spot unusual stuff that might be important.

**Enhanced User Engagement Analysis:** Figuring out how people feel on social media isn't just about reading their comments anymore. This new method (multimodal sentiment analysis) looks at pictures and "likes" along with the words to get the whole picture. This way, we can understand if people truly like something, even if their comments don't say it all.

**Brand Perception Monitoring:** Social media is a giant conversation about brands. By looking at all the posts, pictures, and videos (user-generated content), companies can understand what people really think about them. This helps them spot problems early, see how people's feelings change over time, and create better marketing campaigns.

**Adaptive Content Creation:** People who make stuff for social media (content creators) can use this new way of analyzing feelings (multimodal sentiment analysis) to make their content even better! By seeing how people react to text, pictures, and videos, they can learn what kind of stuff makes people happy, sad, or interested. This helps them create content that people will really like and want to see more of.

**Real-Time Response:** This new way of analyzing feelings online (multimodal sentiment analysis) is super-fast. It lets companies see what people are feeling on social media right now. This way, they can quickly respond to things that are going viral, problems people are having, or even new opportunities. It's like having a finger on the pulse of social media.

**Customer Feedback Analysis:** Imagine being able to listen to all your customers' thoughts and feelings on social media, not just their words! This new way of analyzing feelings online (multimodal sentiment analysis) lets businesses do just that. By looking at text, pictures, and videos, companies can gather all this feedback in one place and understand what customers really think. This helps them make smarter decisions about their products and services.

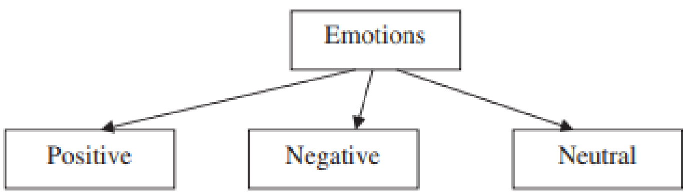


Figure 2 Classification of Emotions

## SCOPE FOR FUTURE DEVELOPMENT

**Social Listening and Monitoring:** Sentiment analysis of multimodal social media data allows organizations to monitor conversations, trends, and public opinion about their brand, products, or services in real-time. It provides insights into how users feel about different aspects of the brand and its offerings across different types of content.

**Market Research and Consumer Insights:** Businesses can leverage multimodal sentiment analysis to conduct market research and gain valuable consumer insights. By analyzing sentiments expressed in text, images, videos, and other forms of content, organizations can understand consumer preferences, trends, and behaviors more comprehensively.

**Brand Reputation Management:** Multimodal sentiment analysis enables proactive brand reputation management by identifying and addressing potential issues or crises quickly. By monitoring sentiment across various social media platforms, organizations can take corrective actions to maintain a positive brand image.

**Product Development and Innovation:** Understanding user sentiments towards existing products and services can inform product development and innovation efforts. Multimodal sentiment analysis helps organizations identify areas for improvement, uncover unmet needs, and gather feedback on new product ideas.

**Content Creation and Marketing**: Content creators and marketers can use multimodal sentiment analysis to optimize their content strategies. By analyzing user sentiments towards different types of content, organizations can tailor their messaging, visuals, and engagement strategies to resonate with their target audience more effectively.

**Customer Experience Enhancement:** Sentiment analysis of multimodal social media data facilitates the enhancement of customer experiences across various touchpoints. By analyzing user sentiments expressed in different forms of content, organizations can identify pain points, address customer concerns, and personalize interactions to improve overall satisfaction.

**Social Media Influencer Analysis:** Brands can utilize multimodal sentiment analysis to evaluate the impact of social media influencers on their target audience. By analyzing sentiments expressed in influencer-generated content, organizations can assess the effectiveness

**Crisis Management and Risk Mitigation:** Multimodal sentiment analysis helps organizations proactively identify and mitigate potential risks and crises. By monitoring sentiments across social media platforms, organizations can detect emerging issues, assess their potential impact, and develop response strategies to minimize negative consequences.

**Policy and Opinion Analysis**: Governments, policymakers, and researchers can utilize multimodal sentiment analysis to analyze public opinion on various socio-political issues. By analyzing sentiments expressed in different types of social media content, stakeholders can gain insights into public sentiment towards policies, events, and societal trends.

**Healthcare and Well-being Monitoring:** Multimodal sentiment analysis can also be applied in healthcare to monitor patient sentiments, gather feedback on healthcare services, and identify potential mental health concerns. By analyzing sentiments expressed in patient-generated content, healthcare providers can improve patient satisfaction and well-being.

## ACKNOWLEDGEMENT

We gratefully acknowledge the collaborative efforts of research participants, data providers, academic mentors, and colleagues, whose contributions were integral to the advancement of sentiment analysis of multimodal social media data. Funding agencies' support, the dedication of open-source communities, and the constructive feedback from peer reviewers significantly enriched our research endeavors.

## CONCLUSIONS

Social media is a goldmine of emotions, but words alone don't tell the whole story. This new analysis method, looking at text, pictures, and videos together (multimodal analysis), is like having a superpower to understand what people really feel online. It helps us see deeper into user preferences and emotions across different platforms.
Companies can use this power to make smarter decisions, create content people love, and personalize experiences. It's like having a direct line to your customers! Imagine knowing how people feel about y our brand in real-time, allowing you to address issues quickly and capitalize on opportunities. There are challenges, like keeping user data private and making sure the analysis isn't biased. But with continued research and collaboration, this powerful tool can unlock a world of benefits for societyand businesses alike.

## REFERENCES

(1) Iyengar, G., Nock, H. J., & Neti C. (2003). Audio-visual synchrony for detection of monologues in video archives. In IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03), Hong Kong, China (pp. V-772). https://doi.org/10. 1109/ICASSP.2003.1200085.

(2) Ji, R., Cao, D., Zhou, Y., & Chen, F. (2016). Survey of visual sentiment prediction for social media analysis. Frontiers of Computer Science, 10 (4), 602–611.

https://doi.org/10.1007/s11704-016-5453-2

(3) Ji, R., Chen, F., Cao, L., & Gao, Y. (2019). Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning. IEEETransactions on Multimedia, 21(4), 1062–1075. https://doi.org/10.1109/TMM.2018.2867718

(4) Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text pre-processing methods on Twitter sentiment analysis. IEEE Access,5, 2870–2879. https://doi.org/10.1109/ACCESS.2017.2672677

(5) Jiao, J., & Zhou, Y. (2011). Sentiment polarity analysis based multi-dictionary. Physics Procedia, 22, 590–596. https://doi.org/10.1016/j.phpro.2011.11.091

(6) Jin, Y., Zhang, H., & Du, D. (2016). Improving deep belief networks via delta rule for sentiment classification. In Proceedings of 2016 IEEE28th international conference on tools with artificial intelligence,ICTAI 2016 (pp. 410–414).

(7) Kaibi, I., & Nfaoui, E. H. (2019). A comparative evaluation of word embeddings techniques for Twitter sentiment analysis.

(8) Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Systems with Applications, 39(5), 6000–6010.

(9) Kasper, W., & Vela, M. (2011). Sentiment analysis for hotel reviews. In Proceedings of the computational linguistics – Applications conference,JACHRANKA

(10) Kauser, N., & Sharma, J. (2017). Automatic facial expression recognition: A survey based on feature extraction and classification techniques. In 2016 international conference on ICT in business industry & government (ICTBIG), Indore, 2016 (pp. 1–4). https://doi.org/10.1109/ICTBIG.2016.7892679.

(11) Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning tech-niques: A review. IEEE Access, 7, 117327–117345. https://doi.org/10.1109/ACCESS.2019.2936124

(12) Khan, A. Z. H., Atique, D. M., & Thakare, D. V. M. (2015). Sentiment analysis using support vector machine. International Journal of Advanced Research in Computer Science and SoftwareEngineering, 5, 105–108.

(13) Koolagudi, S. G., Kumar, N., & Rao, K. S. (2011). Speech emotion recognition using segmental level prosodic analysis. In 2011 international conference on devices and communications (ICDeCom). https://doi.org/10.1109/ICDECOM.2011.5738536

(14) Kranjc, J., Smailović, J., Podpečan, V., Grčar, M., Žnidaršič, M. & Lavracˇ, N. (2014). Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the Cloud Flows platform. In 16th IEEE conference on business informatics, CBI 2014 (Vol. 2, pp. 60–67).

(15) Kunal, S., Saha, A., Varma, A., & Tiwari, V. (2018). Textual dissection of live Twitter reviews using naive Bayes. Procedia ComputerScience, 132, 307–313.

# 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

**SAMARKAND STATE UNIVERSITY, UZBEKISTAN**

IN ASSOCIATION WITH

## PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE
## PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

# Certificate of Presentation

*This is to certify that Mr/Mrs/Dr.* **Adith P Kotian** *from* **Alva's Institute of Engineering and Technology** *has presented a paper titled* **Machine learning Based Sentimental Analysis of Social Media Data using Python** *in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.*

| | | | | |
|---|---|---|---|---|
| **Dr.Akhatov Akmal Rustamovich**<br>Vice Rector of International Affairs,<br>Samarkand State University, Uzbekistan | **Dr. Christo Ananth**<br>Professor,<br>Samarkand State University, Uzbekistan | **Dr. P. Aravinthan,** M.E., Ph.D.,<br>Technical Lead, OSIET<br>Secretary | **Prof. A. Krishnamoorthy,** M.E., (Ph.D)<br>Technical Lead, OSIET<br>Administrator | **K.Janani,** M.Tech.,<br>CEO, OSIET<br>Director |

# 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

**SAMARKAND STATE UNIVERSITY, UZBEKISTAN**

IN ASSOCIATION WITH

**PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE**

**PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY**

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

## Certificate of Presentation

*This is to certify that Mr/Mrs/Dr.* **Afran Sayyed Karim** *from* **Alva's Institute of Engineering and Technology** *has presented a paper titled* **Machine learning Based Sentimental Analysis of Social Media Data using Python** *in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.*

| | | | | |
|---|---|---|---|---|
| **Dr.Akhatov Akmal Rustamovich** | **Dr. Christo Ananth** | **Dr. P. Aravinthan,** M.E., Ph.D., | **Prof. A. Krishnamoorthy,** M.E., (Ph.D) | **K.Janani,** M.Tech., |
| Vice Rector of International Affairs, Samarkand State University, Uzbekistan | Professor, Samarkand State University, Uzbekistan | Technical Lead, OSIET | Technical Lead, OSIET | CEO, OSIET |
| | | Secretary | Administrator | Director |

# 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

**SAMARKAND STATE UNIVERSITY, UZBEKISTAN**

IN ASSOCIATION WITH

**PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE**

**PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY**

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

## Certificate of Presentation

*This is to certify that Mr/Mrs/Dr.* **Ashwin kumar** *from* **Alva's Institute of Engineering and Technology** *has presented a paper titled* **Machine learning Based Sentimental Analysis of Social Media Data using Python** *in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.*

| | | | | |
|---|---|---|---|---|
| **Dr.Akhatov Akmal Rustamovich** | **Dr. Christo Ananth** | **Dr. P. Aravinthan,** M.E., Ph.D., | **Prof. A. Krishnamoorthy,** M.E., (Ph.D) | **K.Janani,** M.Tech., |
| Vice Rector of International Affairs, Samarkand State University, Uzbekistan | Professor, Samarkand State University, Uzbekistan | Technical Lead, OSIET Secretary | Technical Lead, OSIET Administrator | CEO, OSIET Director |

# 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

**SAMARKAND STATE UNIVERSITY, UZBEKISTAN**

IN ASSOCIATION WITH

**PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE**

**PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY**

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

## Certificate of Presentation

*This is to certify that Mr/Mrs/Dr.* Ishwar Pavan *from* Alva's Institute of Engineering and Technology *has presented a paper titled* Machine learning Based Sentimental Analysis of Social Media Data using Python *in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.*

| | | | | |
|---|---|---|---|---|
| **Dr.Akhatov Akmal Rustamovich**<br>Vice Rector of International Affairs,<br>Samarkand State University, Uzbekistan | **Dr. Christo Ananth**<br>Professor,<br>Samarkand State University, Uzbekistan | **Dr. P. Aravinthan,** M.E., Ph.D.,<br>Technical Lead, OSIET<br>Secretary | **Prof. A. Krishnamoorthy,** M.E., (Ph.D)<br>Technical Lead, OSIET<br>Administrator | **K.Janani,** M.Tech.,<br>CEO, OSIET<br>Director |

# 12th INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2024

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN COLLABORATION WITH

**SAMARKAND STATE UNIVERSITY, UZBEKISTAN**

IN ASSOCIATION WITH

**PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE**

**PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY**

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

## Certificate of Presentation

*This is to certify that Mr/Mrs/Dr.* Vasudev S Shahapur *from* Alva's Institute of Engineering and Technology *has presented a paper titled* Machine learning Based Sentimental Analysis of Social Media Data using Python *in the "12th International Conference on Contemporary Engineering and Technology 2024" held on 23rd & 24th March 2024.*

| | | | | |
|---|---|---|---|---|
| **Dr.Akhatov Akmal Rustamovich** | **Dr. Christo Ananth** | **Dr. P. Aravinthan,** M.E., Ph.D., | **Prof. A. Krishnamoorthy,** M.E., (Ph.D) | **K.Janani,** M.Tech., |
| Vice Rector of International Affairs, Samarkand State University, Uzbekistan | Professor, Samarkand State University, Uzbekistan | Technical Lead, OSIET | Technical Lead, OSIET | CEO, OSIET |
| | | Secretary | Administrator | Director |

# APPENDIX-B

## PROJECT ASSOCIATES INFORMATION

| | |
|---|---|
|  | Mr. Adith P Kotian pursuing Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University. He is a student of Alva's Institute of Engineering and Technology, Mijar, Moodbidri.<br>Email-ID: adithkotian2002@gmail.com<br>Contact no: 82177 65199 |
|  | Mr. Afran Sayyed Karim pursuing Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University. He is a student of Alva's Institute of Engineering and Technology, Mijar, Moodbidri.<br>Email-ID: 4al20cs008afran@gmail.com<br>Contact no: 86180 17946 |
|  | Mr. Ashwin Kumar pursuing Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University. He is a student of Alva's Institute of Engineering and Technology, Mijar, Moodbidri.<br>Email-ID: kinnigoliashwin@gmail.com<br>Contact no: 80886 30874 |
|  | Mr. Ishwar Pavan pursuing Bachelor of Engineering in Computer Science and Engineering from Visvesvaraya Technological University. He is a student of Alva's Institute of Engineering and Technology, Mijar, Moodbidri.<br>Email-ID: ishwarpavan4517@gmail.com<br>Contact no: 7899137670 |
|  | Mr. Vasudev S Shahapur, Associate Professor, Department Computer Science and Engineering from Visvesvaraya Technological University. Alva's Institute of Engineering and Technology, Mijar, Moodbidri.<br>Email-ID: vasu.shahapur@aiet.org.in<br>Contact no: 90088 11246 |