

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering



Project Report on

TasteFeels: The Next Frontier of dining enhancement with facial Recognition

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2023-24

Submitted by

Vinit Patil (D17B , Roll no - 55)
Chirag Panjwani (D17B , Roll no - 52)
Tanishq Harchandani (D17A , Roll no - 21)
Sahil Ramchandani (D17C , Roll no - 47)

Project Mentor

Mrs. Abha Tewari

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering



Certificate

This is to certify that ***TasteFeels: The Next Frontier of dining enhancement with facial Recognition*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on “***TasteFeels: The Next Frontier of dining enhancement with facial Recognition***” as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Mrs. Abha Tewari*** in the year 2023-24.

This project report entitled ***TasteFeels: The Next Frontier of Dining enhancement with facial Recognition*** by ***Vinit Patil, Chirag Panjwani, Sahil Ramchandani, Tanishq Harchandani*** is approved for the degree of **B.E Computer Engineering**.

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Mrs. Abha Tewari

Project Report Approval For B. E (Computer Engineering)

This project report entitled TasteFeels: The Next Frontier of dining enhancement with facial Recognition by Chirag Panjwani, Vinit Patil, Sahil Ramchandani, Tanishq Harchandani is approved for the degree of ***B.E Computer Engineering***.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

(Name of student and Roll No.)

(Signature)

(Name of student and Roll No.)

(Signature)

(Name of student and Roll No.)

(Signature)

(Name of student and Roll No.)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Chapter No.	Title	Page No.
1	Introduction	11
1.1	Introduction of the project	11
1.2	Motivation for the project	11
1.3	Problem Definition	12
1.4	Existing Systems	12
1.5	Lacuna of the existing systems	12
1.6	Relevance of the Project	13
2	Literature Survey	14
A	Brief Overview of Literature Survey	14
B	Related Works	15
2.1	Inference drawn	17
2.2	Comparison with the existing system	18
3	Requirement Gathering for the Proposed System	19
3.1	Introduction to requirement gathering	19
3.2	Functional Requirements	19
3.3	Non-Functional Requirements	20
3.4	Hardware, Software , Technology and tools utilized	20
3.5	Constraints	21

4	Proposed Design	22
4.1	Block diagram of the system	22
4.2	Detailed Design	23
4.3	Project Scheduling & Tracking using Timeline / Gantt Chart	26
5	Implementation of the Proposed System	27
5.1	Methodology employed for development	27
5.2	Algorithms and flowcharts for the respective modules developed	27
5.3	Datasets source and utilization	28
6	Testing of the Proposed System	29
6.1	Introduction to testing	29
6.2	Types of tests Considered	29
6.3	Various test case scenarios considered	30
6.4	Inference drawn from the test cases	30
7	Results and Discussion	31
7.1	Screenshots of User Interface (UI) for the respective module	31
7.2	Input Parameters / Features considered	32
7.3	Comparison of results with existing systems	33
7.4	Inference drawn	33
8	Conclusion	34
8.1	Limitations	34
8.2	Conclusion	35
8.3	Future Scope	35

	References	36
1	Appendix	37
a)	Paper published	37
b)	Plagiarism report	43
c)	Project review sheet	44

List of Figures

Figure No.	Heading	Page No.
4.1	Block Diagram	22
4.2.1	Data flow Diagram 0	23
4.2.2	Data flow Diagram 1	24
4.2.3	Data flow Diagram 2	25
4.3	Gantt Chart	26
7.1	Screenshot of user interface	31
7.2	Screenshot of Emotion Detected using Upload video	31
7.3	Screenshot of Real Time Emotion Detection	32
7.4	Input Parameters for Facial Emotion Detection	33

Abstract

The interface between technology and cooking always holds promise to enhance dining experiences, and the emergence of facial recognition technology has opened a new frontier in this regard. This paper explores the concept of "TasteFeels", a new approach to food enhancement ML- Using its real -time and facial analytics, TasteFeels enables restaurants to create personalized menus, flavors and presentations based on diners' sensory and physiological responses, ultimately elevating the overall dining experience. In this paper we explore aspects of TasteFeels technology in more detail, including the use of facial recognition techniques and their integration into cooking practices.

By tackling these challenges head-on, the proposed TasteFeels system aims to provide a more accurate and user-friendly dining enhancement experience through facial recognition technology, benefiting both restaurant-goers and culinary professionals as they explore the next frontier of dining innovation.

1. Introduction

1.1 Introduction

The existing restaurant reviews that use the responses from textual questionnaires are often those that have a low number of participants and possibly have an opinion bias as well. This project proposes a novel approach: integrating face-detection technology to film the genuine feelings of the diners through their facial expressions. Imagine this: after they just savored a mouth-watering meal, all they do is simply face a camera. The system records the movements of the customers by their facial expressions capturing real-time feedback about the food, service, and ambiance.

Secondly, it reduces the process of reviewing and incorporates a mechanism, which is a shorter questionnaire, in turn improving the attendance rate. Besides, through recording the immediate impression, a greater authenticity is attained through reproducing the same experience. Furthermore, facial recognition is able to detect even small features, providing greater and more varied information than just pure satisfaction indexes. Nevertheless, we retain user privacy above anything. The implementation of the data protection regulations should be done by the book to avoid any information being treated as a secondary factor and to protect the individual's rights.

Therefore, this project will present not only the advantages for restaurant-goers but the establishment owners too. Customers get the easiest way to provide feedback while owners and managers are able to see real-time information and act accordingly, to improve the level of service, food quality, or atmosphere. In sum, it is likely to see a general movement in the industry towards more reliable and useful data to help in restaurant reviews.

However, more research and development is required for the tuning of facial recognition algorithms and trial-usage in users for the purpose of refinement of the current system – Moreover, facial recognition-based approaches can potentially change the way we review our dining experiences.

1.2 Motivation

The modern dining experience has moved from mere keeping hunger in check, as it has become a trip through cultures, societies, and emotions. Consequently, the old systems of getting feedback and sustaining development come to a dead end. They often turn out to be unwieldy, slow and inaccurate regarding the detail of the restaurant's environment

- **Effortless Feedback:** We acknowledge the annoyance of paper censuses or full form responses completed far after the meal. Facial recognition technology smooths out this obstacle. Consumers only gaze into a camera right after satisfaction giving spontaneous and easy to provide feedback without having to reflect too much on the experience.
- **Authenticity in Reactions:** It is important to capture genuine feelings on the spot as they will help to focus on results. Facial recognition technology goes beyond the words, analyzing expressions to help create a subtle and precise appraisal. We feel this tactic is more effective as it contains views of the consumers which are not rehearsed.
- **Continuous Improvement for Restaurants:** Culinary venues are always on the lookout for ways to distinguish their offerings and provide more unified service. Our research project allows restaurants to identify important facts learned from the facial data of many diners at different times. This way they can perceive customer feelings and acceptances at another level. Equipped with this knowledge, restaurants can go into a never-ending journey of improvement taking each visit and bettering it than the previous one.
- **Privacy and Data Security as Top Priorities:** We know the users data protection and the security are important to them. With respect to the use of face recognition, no fewer special rights of the individual can be neglected. Our key is to appropriately manage data, and we follow strict privacy rules so that customers will trust us with their information.

1.3 Problem Definition

The project we propose above is concentrated in bringing about a transition concerning dining with the aid of facial recognition technology and video sentiment analysis. We aim to tackle several key challenges in this endeavor:

- **Reading Emotions, Capturing Feedback:** We plan to use facial recognition for data mining purposes when it comes to customers' feelings on a screen. Such involves interpreting peoples' expression and body language to be certain about their emotions; whether good or bad. The goal is to shed light on the true-to-life experiences, through testimonials, to form a rich image of their stay with the hospital.
- **From Reactions to Culinary Refinement:** Being able to apply the emotional data we envision the recipes will stand at a higher level than ever before. Through the observation of consumers' mood during the dining, the restaurants can take the leading role in providing adaptive measures such as changing the food, service and the ambiance . That is, one can have fun and meaningful dining time with relatives and friends.
- **Scaling Up for Wider Impact:** Scalability is cookie. Our product is designed in a way that flexibly accommodates different restaurants setup, be it an elite eatery or just a casual cafe. Execution of the system should be durable enough that the growing number of users and total data which is visualized will not cause any glitches. Generally speaking, all diners benefit because scalability makes it possible to provide great dining to broad audiences.
- **Privacy and Security: A First Priority:** First of all, privacy has to be the rule instead of the exception when it comes to our customers' facial data. We serve a reminder of the necessity of following the regulations and moral norms when using facial recognition technology. Establishing trust with customers is a critical element of our business success and this can be made possible by making their personal information safe and clean.

1.4 Existing Systems

Among all the papers which we reviewed, essentially almost all of them had depicted the systems or their datasets. The best illustration of this would be “iMIGUE” which is a specially made up dataset including different subjects. This model uses Micro Gestures(MG) to identify the emotions instead of the usual emotional expressions. The most significant advantage for this dataset is that it has data diversity, that is, various types of people from different castes and nationalities were selected.

Another already used system had the feature that only emotion recognition was done on the last part of the video, and it was based on a brief review of the rest of the part. It is as if the emotion of a reviewer is written only with a few letters at the end of the video, which is much more convenient for our use case representing our topic.

1.5 Lacuna of existing systems

Even though aiding to improve dining experience is really up to the level of facial recognition technology, present systems cannot fully reach the optimal point this technology is aimed to go. Here's a deeper dive into these challenges:Here's a deeper dive into these challenges:

- **Over-reliance on Multimodal Inputs:** Currently existing emotion recognition systems often go with a wide range which results in more parameters than face. It may consist of audio recordings of the voice, brainwaves which can be recorded through EEG scans and physiological data measured by sensors. Although these extras have additional input to brighten the topic better, they also bring in complexity and logistical barriers. However, the tendency may happen to the system to focus on data sources such as videos less (so-called facial data) than necessary, which may hinder the precise identification of the feelings by that method.

- **Model Accuracy: A Sharp Sword:** The dialect of emotion detection is a critical parameter to possess. There should be the best-fit models and they should be multimodal, i.e., they should be able to integrate the data from multiple sources such as video, and audio. Nevertheless, the multi-modality approach poses a significant challenge and ensures only acceptable accuracy. My point here is, at times, something from a cleaner model with fewer predict variables may be preferred to even superior models for non-technical reasons. This weakens the informational automation and the truthfulness of the contextual perception.
- **Privacy Concerns and Limited Training Data:** The problem of retaining significance is the training data of electronic emotion detection algorithms. The privacy disclosure is usually at the heart of the debates. Currently, datasets like "iMIUGUE" are very much anonymized, which is not good for our endeavor as it will still pose some problems. Wearing a face mask, in turn, prevents the modeling of facial features, which makes it difficult for professionals to carefully perceive the fine-grained expressions that are useful for reading emotions. Then privacy induces a paradox – more privacy means less capable of operating the data.

1.6 Relevance of project

This project isn't only about technology; it's about integrating technology in life which result in revolutionizing the experience we have in dining:

- **Happier Diners, Higher Satisfaction:** For example, restaurants could gain the capability of getting feedback about emotions. They are able to solve the issue and make amendments live in order to enhance visitors' satisfaction with the event.
- **Personalization Made Easy:** A goodbye to dining in, where there is a variety of options for everyone. The work becomes a beginning for tailored solutions according to an individual's taste. The customers' love of restaurants does come with the knowledge that their tastes and preferences will be met and yearn for always.
- **Culinary Renaissance:** Restaurants will be able to obtain true insights into how their customers feel about food and drinks if they conduct analysis of their diners' responses. This kind of data allows for the refining of menus, design, and even the profiles of the taste to make the culinary experience as captivating as it can be.
- **Data-Driven Decisions, Better Outcomes:** Today, data is the spearhead of the competitive environment. This initiative integrates data into the circle of decision-making about menus, pricing, marketing, and so forth for restaurants.
- **Efficiency Boost:** Conventional feed-back methods are awkward and slow. The task which we have set forth progresses the process and makes it easier for eateries to attend needs timely and effectively.
- **Scalability:** The project to be designed takes into account the factor scalability to benefit from a wide range of restaurants with different sizes, from a small cafe to a Michelin-star restaurant. Widespread integration has the potential to drastically change the culinary industry all over.
- **Tech for Good:** Through this project, an instance of ethical and practical uses of facial recognition and sentiment analysis systems is demonstrated. It has shown how technology can be leveraged in daily lives.

We could say that there was a lot to the project concerning emotions. It is all about the creation of smart diners, the improvement of customer experience, the promotion of a data-driven economy, and the advancement of technology etiquette. As eating out is still considered an integral part of social and cultural experiences, new methodologies like this become very crucial for both guests and the restaurants.

2. Literature Review

2.A Brief Overview of Literature Review

In the field of emotion recognition and sentiment analysis from multimodal data sources, several key trends and techniques have emerged as demonstrated in the literature. Research has increasingly focused on combining the analysis of video, audio, and textual data to gain a more comprehensive understanding of emotional states.

One recurring theme is the application of deep learning, particularly models like LSTM, CNN, and DBN. These deep learning architectures have been extensively employed to extract features and classify emotions from different data modalities, such as images, audio signals, and textual content. But the decision of picking a model depends upon the parameters and type of dataset, there can't be a single model which is best for all datasets available.

Many researches have been done in depth in this field which highlight the importance of expressions and micro-gestures. Doing so helps understand human emotions. Techniques for detecting emotions from facial expressions have been a consistent focus, often employing techniques like Action Units (AU) recognition. These techniques are valuable in recognizing emotions, especially in video data.

The role of context and context-dependent sentiment analysis has been gaining prominence. Analyzing the contextual information surrounding emotional expressions is crucial for a more accurate understanding of the emotional states of individuals, especially in user-generated content. Researchers have proposed various methods, including LSTM networks for capturing contextual information and cross-lingual scenarios to assess the generalizability of emotion recognition models.

Multimodal sentiment analysis, incorporating data from video, audio, and text, has become an exciting research avenue. By leveraging these different modalities, researchers aim to overcome the limitations of single-source emotion recognition and provide a more comprehensive view of emotional states. Nevertheless, challenges in model generalization, privacy concerns, and ethical considerations related to the use of video data remain relevant areas for future exploration.

In summary, the literature in emotion recognition and sentiment analysis has witnessed significant advancements through the integration of deep learning techniques, the focus on facial expressions and micro-gestures, the importance of context, and the increasing interest in multimodal data sources. These trends collectively contribute to a more comprehensive understanding of human emotions and sentiment in various domains, from video content to e-learning and daily-life scenarios.

2.B Related Works

NAME & AUTHOR	DESCRIPTION	TECHNOLOGY	EVALUATION PARAMETERS(ADVANTAGES)	LACUNA
<p>Paper name:Sentiment Analysis on User-generated Video, Audio and Text</p> <p>Authors:Akriti Ahuja,Shyam Kansara ,Vrunda Patel</p> <p>Publication - ICCIS May 2021</p>	<p>It uses multimodal sentiment analysis, which has 4 modules: collection,training,testing and frontend. For video analysis, when the main function which detects the emotion runs, a neutral image is used and the deviation is calculated. The values are converted to integers.</p>	<p>LSTM model, and GME-LSTM model is used. Also AU(Action Unit) having AU name, FACS name and muscular basis.In python, deep learning algorithm is also used. fishface recognizer from the Fisherfaces module integrated in OpenCV.</p>	<ol style="list-style-type: none"> 1. The objective in this paper is to build a system which can identify the sentiment categorized into six types: anger, joy, disgust, sadness, fear, and surprise of a video when the data is fed into it as compared to previous model in this which included only 3. 2. Sometimes, the 6 emotions were not enough, and hence, a 7th emotion:contempt was introduced 3. Since our project will have the video analysis as the main part, many Lacuna found in the system will be nullified easily/automatically. 4. Facial expression is captured, and for that, to detect a face properly,and to also have a clear face, different steps have been applied. 	<p>It derives text from video as a part of the modal, whereas our modal will use facial expressions only rather than both.</p> <p>The ratio of video in multimodal analysis is 32.06 of the total whereas text has about 40%</p> <p>The clips are cut down in smaller clips(each 20 seconds), due to which, a word may be cut.</p> <p>The speed is comparatively slower and accuracy is only about 70%.</p>
<p>Paper name:iMiGUE: An Identity-free Video Dataset for Micro-Gesture Understanding and Emotion Analysis</p> <p>Author name:Xin Liu,Henglin Shi,Haoyu Chen,Zitong Yu ,Xiaobai Li,Guoying Zhao</p> <p>Publication - CVPR 2021</p>	<p>A new and different type of dataset, named, iMiGUE is used. An unsupervised model is trained.</p>	<p>RNN model is used, and the iMiGUE dataset is not fetched from somewhere, it is constructed.</p>	<ol style="list-style-type: none"> 1. Uses micro gestures(MGs) to capture the hidden emotions and not just via normal expressions. 2. The proposed unsupervised model comes in great use to handle MGs. 3. The dataset handles an issue, which is it has data diversity.(different nationalities and castes) 	<p>To preserve the privacy of the person, identity is sometimes hidden and that may or may not be the best idea for our use case.</p> <p>Sometimes, the fusion of other models have higher accuracy as compared to a modal built in this paper, which is built to go along the dataset</p>

Automatic Recognition of Facial Actions in Spontaneous Expressions Marian Stewart Bartlett, Gwen C. Littlewort, Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, Javier R. Movellan Sep 2006	The introduction of the paper sets the stage by highlighting the significance of automated facial expression recognition. It emphasizes the importance of decoding facial expressions for understanding human emotions, social interactions, and psychopathology. The introduction also mentions the challenges posed by manual coding, such as subjectivity and time consumption. It presents the goal of developing a real-time automated system that can accurately recognize facial actions and expressions in both posed and spontaneous contexts.	The technology stack combines computer vision, machine learning, and data preprocessing, using Gabor filters for image feature extraction and SVM and AdaBoost classifiers for training and classification, ensuring efficient real-time implementation.	The paper provides evaluation parameters for posed and spontaneous facial expressions. It includes Percent Agreement with Human FACS Codes, Hit and False Alarm rates, and Area Under the ROC curve. Interval analysis is introduced to measure detections over time, enriching understanding. In spontaneous expression evaluation, correlation analysis between automated system outputs and expert-coded facial action intensities is crucial, providing deeper insights into facial expressions.	The system faces challenges in recognizing spontaneous expressions due to pose variations, speech, and muscle movements, as well as potential bias from imbalanced training data, expert coding variability, external factors like lighting and camera quality, and its applicability on larger datasets.
---	---	--	--	---

Emilya: Emotional body expression in daily actions database. Nesrine Fourati, Catherine Pelachaud May 2014	The paper introduces a database for emotional body expressions during daily actions. Understanding nonverbal communication through body movements is vital, prompting the study's focus on creating this database.	The research utilized motion capture technology for accurate movement data recording. Videos were captured and processed using tools like MediaInfo and VirtualDub. 3D Studio MAX created virtual avatars for the study, and Amazon Mechanical Turk aided data collection.	The study's evaluation centered on a perception study where participants rated emotions portrayed by virtual avatars. Statistical analysis, including ANOVA and Tukey-Kramer tests, assessed emotion recognition differences.	Actor Influence: The paper doesn't deeply delve into how different actors affect emotion recognition. This aspect could enhance the study's insights. Cross-Cultural Factors: The study overlooks cultural variations in emotion recognition. Considering cultural influences could enrich the findings. Mixed Emotions: The research focuses on single emotions in specific actions. Examining recognition in mixed-emotion scenarios could provide a richer perspective. Non-Visual Cues: Non-visual cues like audio aren't addressed. Including these cues could offer a more complete analysis. Emotion Intensity: The impact of varying emotion intensities is unexplored. Analyzing intensity's effect on recognition could offer valuable understanding.
--	--	--	---	---

<p>Emotion Analysis Using Audio/Video , EMG and EEG: A Dataset and Comparison Study</p> <p>Farnaz Abtahi Tony Ro Wei Li Zhigang Zhu Publication - IEEE 2018</p>	<p>The paper aims to explore and compare different modalities for emotion recognition, specifically focusing on audio, video, electromyography (EMG), and electroencephalography (EEG) signals. The motivation behind this study is to understand the effectiveness of different sensor-based methods in recognizing emotions. The authors address the challenges of emotion recognition, such as the dynamic and complex nature of emotions, and highlight the potential of bio-sensing techniques (EMG and EEG) alongside more common modalities (audio and video)</p>	<p>Two prominent classification models took center stage: Deep Belief Networks (DBNs) and Long Short-Term Memory (LSTM) networks. DBNs, recognized for their generative capabilities, demonstrated effectiveness with EMG and EEG data. Conversely, LSTM, a discriminative model, displayed superior performance when dealing with audio and video data</p>	<p>The paper used accuracy (mean and standard deviation) as the primary evaluation metric. The accuracy was reported for each modality and various segment combinations within the data. The authors conducted experiments using different segment pairs to understand how well emotions were classified when transitioning from one segment to another.</p>	<p>Limited Data Explanation: Inadequate detail about dataset collection, subjects, diversity, and elicitation methodology limits study reproducibility. Baseline Comparison: Lack of comparison with established methods or baseline results hinders understanding of proposed techniques' performance improvement. Feature Engineering and Selection: Clarifying rationale behind chosen feature extraction methods could provide deeper insights into modality-specific choices. Inter-Observer Agreement: Discussing inter-observer agreement for emotion labeling would clarify the consistency of ground truth labels. Generalization to Real-World Scenarios: Study's focus on controlled environments neglects discussing its applicability to complex real-world situations.</p>
---	--	---	--	--

Table 2.B(i). Referred Works

2.1 Inference Drawn

Our project tackles the shortcomings of existing emotion detection systems for restaurant reviews by focusing on three key areas:Our project tackles the shortcomings of existing emotion detection systems for restaurant reviews by focusing on three key areas:

- **High-Quality, Focused Dataset:** And nowadays, the existing approach is much prone to rely on the cues without facial features or the dataset is not correct, so it is hard to make a right assessment of dine while consuming. We solve this problem by using FER13 which is known for the dataset's high quality and suitability for facial expression recognition. The data set works with video frames that take records of these tiny movements of facial expressions important for our output.
- **Optimal Model Selection:** Often in current techniques, broad emotion detection is preferred to the discrete dining situation. We cash this by diligent choosing the face analysis model which is tailored to the different scientific landscape of subjective restaurant reviews. It guarantees high accuracy of interpretation of emotions of visitors, so this allows a way for the results to be actionable.
- **User-Friendly Interface:** While the tech itself is very complex, we understand that creating a great experience for the consumers comes first. Our project is particularly concerned with the likeliness of a user-friendly UX, which retains the simplicity and intuitiveness of design, which aids users in carrying out required tasks easily. This is an exemplification of the system's straightforward goal: it observes guest facial expressions, thus allowing a guest and a restaurant to talk with technology without difficulties.

Fundamentally, the objective of our project seeks to exceed past systems by emphasizing an exceptionally chosen and tailored dataset for facial recognition only, by integrating an enhanced model to boost precision of sentiment analysis in the restaurant environment, and by ensuring that the customer experience is smooth while at the same time, it is well-endowed with necessary requirements and a friendly service.

2.2 Comparison with the existing system

Unlike those emotional monitoring systems that are designed in the localized way to understand the user's moods, your system features a comparison technique of the most popular happy emotions. So, these kinds of systems, by nature, are limited in their capability to achieve a more elaborate understanding of real life emotions due to the fact that they are based on the recognition of facial expressions of a predefined facial data set, only. The technique you have come up with can perform this function by employing 2 methods - facial expressions technique and audio analysis method. Through this means, a conducive environment of the app is created and the sentiment of the user can be understood easily. It is indeed outstandingly effective when you add the fact that it offers the services for low vision and blind people with the help of text-to-speech output, thus making your system universally applicable. On this note it should not be neglected that the AI algorithms of the future will be much reinforced by user reviews as well as feedback and this means that they will be much better and more accurate as well. As a proposition here, this specific mood detection mode may demand such models to be used which are developed exactly for this purpose and not by going for the model adaptation like CNN or VGG. Adding to that, advanced techniques will be developed which cater for a large dataset of individuals originally from different regions, diversifying with life experiences and belonging to a variety of ethnic groups to ensure that all people are verified. Concerns related to this always are the way in which your system effectively generates the mood of the user, consequently bringing about a perfect tool to measure the mood of the users in a more effective, integrating, and user and vendor-friendly way.

3. Requirement Gathering for the Proposed System

In this chapter, the focus will be on delineating the resources utilized, analyzing user requirements, and delineating our ability to meet those needs. Furthermore, both functional and non-functional requirements will be explored, followed by an overview of the software and hardware employed in the process.

3.1 Introduction to requirement gathering

In the rapidly evolving culinary landscape, TasteFeels emerges as a pioneering solution, set to redefine dining enhancement through facial recognition technology. This innovative platform aims to revolutionize the dining experience by tailoring it to individual preferences.

The requirement gathering process for TasteFeels is crucial, as it involves identifying, analyzing, and documenting the needs and expectations of stakeholders, including end-users and industry professionals. Through this process, TasteFeels aims to align its technological capabilities with the dynamic demands of the modern gastronomic landscape.

At its core, TasteFeels seeks to integrate facial recognition technology seamlessly into every aspect of the culinary journey. This includes providing personalized menu recommendations based on past dining preferences and enabling real-time feedback mechanisms for chefs. However, to effectively utilize this technology, understanding user requirements, preferences, and concerns is essential.

The requirement gathering process for TasteFeels involves various methodologies, such as interviews, surveys, and focus groups with potential users and industry experts. Additionally, ethical considerations and privacy concerns are paramount, and TasteFeels ensures compliance with regulatory frameworks and user consent mechanisms to protect privacy and data security.

In summary, requirement gathering is fundamental to the development of TasteFeels. By actively engaging with stakeholders and prioritizing their needs, TasteFeels aims to create a symbiotic relationship between technology and gastronomy, ushering in a new era of personalized and ethically responsible dining enhancement.

3.2 Functional Requirements

- **Data Ingestion:** The system should be able to ingest video data for emotion recognition.
- **Feature Extraction:** Extract meaningful features from video frames.
- **Model Training:** Train the deep learning model on labeled emotion data.
- **Emotion Recognition:** The model should be capable of recognizing emotions in real-time or on pre-recorded videos.
- **Evaluation:** Assess the model's performance using appropriate metrics.

- **User Interface:** If applicable, create a user-friendly interface for users to interact with the system.
- **Logging and Reporting:** Implement a system for logging and reporting model performance and predictions.
- **Deployment:** Deploy the model for real-world applications.

3.3 Non-Functional Requirements

- **Accuracy:** The model should achieve a high level of accuracy in emotion recognition.
- **Real-time Processing:** If needed, the system should provide real-time or near-real-time processing of video streams.
- **Scalability:** The system should be scalable to handle a large number of concurrent requests if deployed in a production environment.
- **Privacy and Security:** Ensure that sensitive video data is handled securely and ethically.
- **Usability:** If there's a user interface, it should be user-friendly and intuitive.
- **Performance:** The system should perform efficiently, especially if deployed in resource-constrained environments.
- **Robustness:** The model should be robust against variations in lighting, facial expressions, and other environmental factors.

3.4 Hardware, Software , Technology and tools utilized

Hardware Requirements:

- High-performance GPU(s) for training (e.g., NVIDIA Tesla, GeForce RTX).
- Sufficient RAM for handling large datasets and model weights.
- Storage space for datasets, model checkpoints, and logs.

Software Requirements:

- Deep learning framework (e.g., TensorFlow, PyTorch, Keras) for model development and training.
- Python for scripting and data manipulation.
- Video processing libraries (e.g., OpenCV) for video frame extraction.
- Development environment (e.g., Jupyter Notebook, IDE).
- Web framework or UI development tools for user interface.

Technology and Tools utilized:

- **Deep Learning Framework:** TensorFlow or PyTorch for building and training your neural network.
- **Video Processing:** OpenCV for handling video data and frame extraction.
- **Data Manipulation:** Python libraries like NumPy, pandas for data manipulation.
- **Version Control:** Git for tracking code changes and collaboration.
- **Development Environment:** Jupyter Notebook, Visual Studio Code, or any preferred IDE.
- **Deployment:** Depending on the needs, cloud platforms like AWS, Azure, or Google Cloud for deploying the model in production.
- **Web Development :** Flask, Django, or other web frameworks for creating a user interface.

3.5 Constraints

- **Data Availability:** Availability of labeled video data for training can be a significant constraint.
- **Computational Resources:** Access to GPUs and sufficient computing resources for model training.
- **Ethical Considerations:** Handling emotional data requires ethical considerations, including consent and privacy.
- **Real-time Processing:** Achieving real-time performance can be challenging, depending on the hardware and model complexity.
- **Model Complexity:** Complex models may require longer training times and more computational resources.

4. Proposed Design

4.1. Block Diagram of the proposed system:

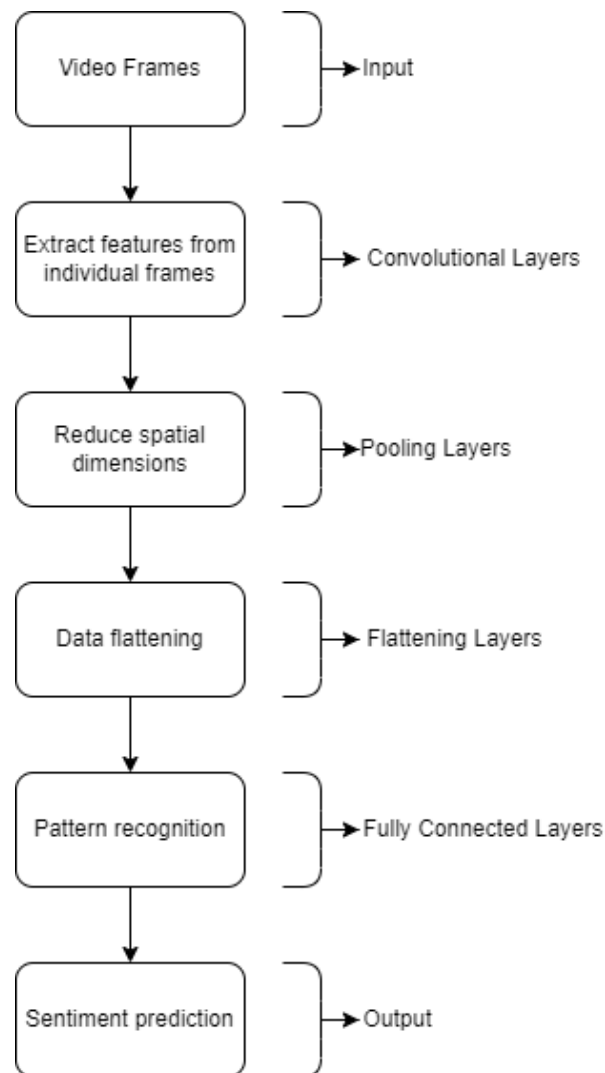


Fig 4.1: Block Diagram

1. Input:

In video emotion analysis, the input typically consists of a sequence of frames extracted from a video clip. Each frame is treated as an individual image. These frames serve as the raw data that your neural network will analyze to predict emotions. The number of frames in your sequence can vary, and you may need to preprocess them to ensure they have consistent dimensions (e.g., resizing or cropping).

2. Convolutional Layers:

Convolutional layers are the backbone of most image and video analysis neural networks. They work by applying a set of learnable filters (kernels) to the input data. Each filter scans

across the input to detect features like edges, corners, textures, and more complex patterns. The output of these layers is a set of feature maps that represent the presence of these features in different parts of the input.

3. Pooling Layers:

After each convolutional layer, it's common to include pooling layers. These layers reduce the spatial dimensions of the feature maps while retaining the most important information. Pooling helps in reducing the computational load and preventing overfitting by discarding less relevant details. MaxPooling and AveragePooling are commonly used pooling techniques.

4. Flattening Layer:

The output of the convolutional and pooling layers is a collection of 2D feature maps. To connect these feature maps to the fully connected layers, you need to flatten them into a one-dimensional vector. This transformation maintains the spatial relationships learned in the earlier layers.

5. Fully Connected Layers:

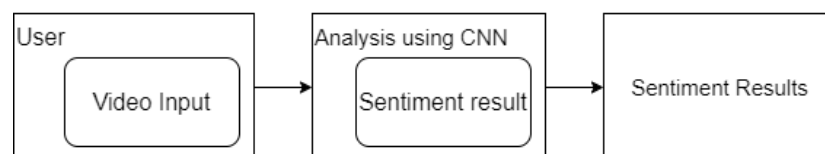
After flattening, the one-dimensional vector is fed into fully connected layers. These layers consist of densely connected neurons, where each neuron is connected to every neuron in the previous layer. Fully connected layers are responsible for learning complex patterns and relationships in the data. You can have multiple fully connected layers, and you often introduce non-linear activation functions (e.g., ReLU) between them to introduce non-linearity into the model.

6. Output:

The final layer of your neural network produces the model's predictions. In the context of video emotion analysis, the number of neurons in this layer depends on the number of emotions or classes you want to predict. For binary emotion classification (e.g., happy or not happy), you would have a single neuron with a sigmoid activation function to output a probability. For multi-class emotion classification (e.g., happy, sad, angry, etc.), you would have multiple neurons, often with a softmax activation function to produce class probabilities that sum to 1.

4.2 Detailed Design:

Data Flow Diagram 0



4.2.1 Data Flow Diagram 0

1. User (Video Input):

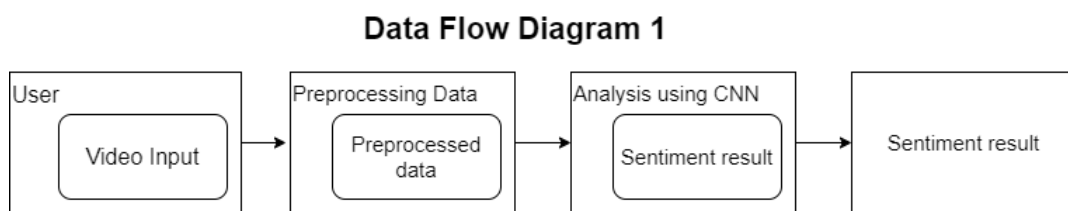
This represents the user as the source of video input data. Users provide the videos that contain the content for sentiment analysis. Video data could be uploaded by users, captured from a camera, or obtained from other sources.

2. Analysis Using CNN (Sentiment Result):

This process represents the core of your video sentiment analysis system. Here, you're using a Convolutional Neural Network (CNN) to analyze the video data and extract sentiment-related information. The CNN processes the video frames and predicts sentiment based on the features it learns from the frames.

3. Sentiment Results:

This is the output of your sentiment analysis process. It represents the sentiment predictions made by the CNN or any other analysis method you're using. Sentiment results could be categorized into different emotions or sentiment labels (e.g., positive, negative, neutral), and the level of granularity depends on the specifics of your sentiment analysis task.



4.2.2 Data Flow Diagram 1

1. User (Video Input):

This represents the user or external entity who provides the video input for analysis. Users might upload video clips or provide access to video data for your system to process.

2. Preprocessing Data (Preprocessed Data):

This process or component is responsible for preprocessing the video data before it's fed into the analysis using Convolutional Neural Networks (CNN). Video preprocessing can include tasks like frame extraction, resizing, normalization, and noise reduction. The output of this step is the preprocessed video data, which is cleaner and more suitable for analysis.

3. Analysis Using CNN (Sentiment Result):

This component represents the core of your video emotion analysis system. It uses Convolutional Neural Networks (CNN) to analyze the preprocessed video data and predict the sentiment or emotion. CNNs are commonly used for image and video analysis tasks because they excel at detecting patterns and features in visual data. The output of this component is the sentiment result, which could be the predicted emotion for each frame in the video.

4. Sentiment Results:

This is where the sentiment or emotion results are collected and stored. The results can be presented in various forms, such as a time series of emotions for the entire video or as a summary (e.g., the predominant emotion detected).

Data Flow Diagram 2

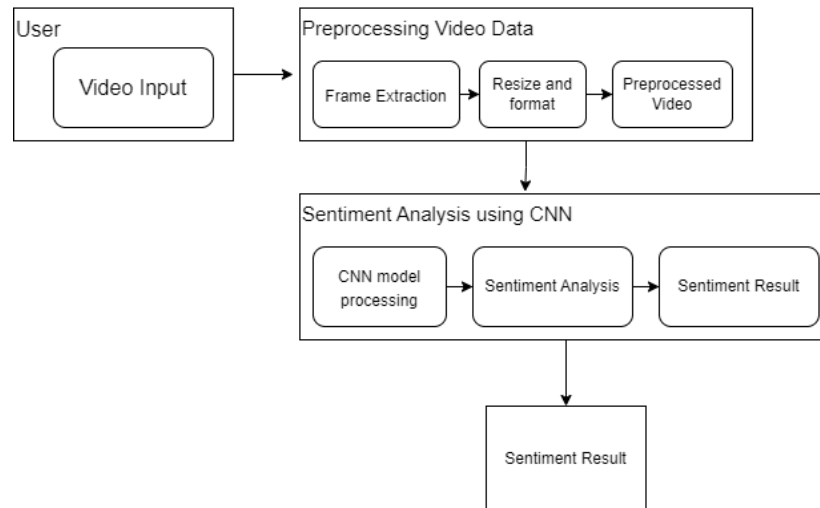


Fig 4.2.3 Data Flow Diagram 2

1. User (Video Input):

This represents the user or system providing the video input for sentiment analysis. Users can upload video clips or provide input in some other way. The video input serves as the raw data for the analysis.

2. Preprocessing Video Data:

- **Frame Extraction:** This process involves splitting the video input into individual frames. In video sentiment analysis, frames are the basic units of analysis, and extracting them allows you to process each frame separately.
- **Resize and Format:** Frames extracted from the video might have varying dimensions. Resizing them to a consistent size is important for input compatibility with your neural network model. Additionally, this step may involve converting the frames into a suitable format for analysis, such as grayscale or RGB.
- **Preprocessed Video:** After frame extraction, resizing, and formatting, you obtain a preprocessed video, which consists of a sequence of preprocessed frames. This preprocessed video is ready for sentiment analysis.

3. Sentiment Analysis using CNN (Convolutional Neural Network):

- **CNN Model Processing:** This step involves feeding the preprocessed video frames into a Convolutional Neural Network (CNN) model. The CNN is responsible for analyzing each frame to detect patterns and features relevant to sentiment analysis.
- **Sentiment Analysis:** The CNN processes each frame and produces sentiment predictions or scores for each frame. It might classify frames into different emotional categories (e.g., happy, sad, angry) or provide a continuous sentiment score.
- **Sentiment Result:** This is the outcome of the sentiment analysis process. It represents the overall sentiment expressed in the video based on the individual frame-level

predictions. Depending on your application, this result could be a sentiment label (e.g., "positive," "negative") or a sentiment score (e.g., on a scale from -1 to 1).

4. Sentiment Results:

This is where the final sentiment results are presented or stored. Users or systems can access this information to understand the sentiment conveyed by the analyzed video. The format of the sentiment results can vary, such as textual labels (e.g., "happy video") or numerical scores (e.g., sentiment score of 0.75).

4.3 Project Scheduling & Tracking using Gantt Chart

The Gantt chart illustrates the timeline of our project, which spanned the entire semester as we endeavored to create this model. It serves as a crucial tool for conceptualizing and organizing the planning process for your topic. Thus, we structured our work according to the Gantt chart displayed, aligning our tasks and milestones with the outlined timeline.



Fig 4.3 Gantt Chart

5.Implementation of the proposed system

5.1 Methodology employed for deployment

The methodology employed for the deployment of emotion recognition using video, utilizing a Convolutional Neural Network (CNN) model, involves several crucial steps to ensure effective implementation. Initially, the trained CNN model undergoes rigorous testing and validation to verify its accuracy and reliability in recognizing emotions from video inputs. Once validated, the model is integrated into a scalable and efficient deployment framework, capable of handling real-time video streams. This integration typically involves leveraging cloud-based services or deploying the model on local hardware with optimized inference capabilities. Furthermore, to enhance accessibility and usability, the deployment process includes developing a user-friendly interface through which users can input video data and receive real-time emotion predictions. Additionally, measures are taken to ensure the security and privacy of the deployed system, particularly when handling sensitive video data. Continuous monitoring and performance evaluation are integral parts of the deployment methodology, allowing for ongoing refinement and optimization of the emotion recognition system to meet evolving user needs and technological advancements.

5.2 Algorithms and flowcharts for the respective modules developed

In the development of an emotion recognition system using video, several modules are crucial for accurate and efficient processing. Each module employs specific algorithms and requires a well-defined flowchart to illustrate its functionality. This section outlines the algorithms and provides corresponding flowcharts for the key modules involved in the emotion recognition system.

Algorithm:

Input: Raw video frames

Output: Preprocessed video frames

Steps:

1. Read video frames.
2. Resize frames to a predefined resolution.
3. Convert frames to grayscale for computational efficiency.
4. Apply noise reduction techniques (optional).
5. Normalize pixel values to enhance consistency across frames.

5.3 Datasets source and utilization

The choice of dataset to begin with was CK+ dataset, which is a widely using image dataset available easily on internet sources. The CK+ (Extended Cohn-Kanade) dataset is a widely used benchmark dataset in the field of facial expression recognition. It consists of both posed and non-posed facial expressions captured in laboratory conditions. The dataset contains grayscale images of subjects displaying a range of six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, as well as neutral expressions. Each expression is annotated with emotion labels and corresponding facial action units. CK+ is notable for its high-quality images and detailed annotations, making it valuable for training and evaluating emotion recognition algorithms. But a more diverse dataset which fits the model well is FER-13 dataset.

The FER 13 (FER-2013) dataset is another commonly used dataset in facial expression recognition research. It comprises over 35,000 images of faces labeled with one of seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. The images are collected from various sources, including the internet and professional databases. FER 13 provides a diverse set of facial expressions captured in real-world conditions, making it suitable for training and testing emotion recognition models. Each image is grayscale and cropped to focus on the face region, with labels assigned based on human judgment. While FER 13 lacks detailed annotations compared to other datasets, its large size and diversity make it a valuable resource for researchers in the field.

The final dataset used for training and testing is a trimmed down version of FER-13 dataset, because for our use case, i.e restaurants, not all the types of images are required. A few thousand images of some of the labeled emotions was found to be the balanced dataset which can do the job and also not be unnecessarily big.

6. Testing of the proposed system

6.1 Introduction to testing

The introduction to testing serves as a primer for understanding the methodologies and objectives behind evaluating the proposed emotion recognition system. It outlines the significance of testing in ensuring the system's reliability, accuracy, and effectiveness in real-world scenarios. This section provides context for the subsequent discussions on different types of tests, test case scenarios, and the inferences drawn from these tests. By setting the stage for testing, the introduction highlights the importance of rigorous evaluation in validating the proposed system's capabilities and addressing potential shortcomings.

6.2 Types of test considered

Each type of test serves a specific purpose and targets distinct aspects of the system's functionality and performance. By categorizing tests based on their scope and objectives, a structured approach is adopted to ensure thorough evaluation.

a. Unit Tests:

Unit tests focus on validating the correctness and functionality of individual components or modules within the emotion recognition system. Each unit, such as algorithms for face detection, facial feature extraction, or emotion classification, is subjected to rigorous testing in isolation. Test cases are designed to verify that each unit performs its intended tasks accurately and reliably. For instance, unit tests for face detection may involve supplying sample images with varying facial orientations and occlusions to ensure robust detection capabilities.

b. End-to-End Tests:

End-to-end tests evaluate the system's overall behavior and performance in a simulated real-world environment. These tests assess the system's ability to process video inputs, detect faces, extract facial features, classify emotions, and provide accurate results. End-to-end tests often involve running the system on a diverse set of video clips containing various facial expressions and environmental conditions. The goal is to validate the system's effectiveness in recognizing emotions under realistic scenarios and to identify any potential issues or limitations.

c. Performance Tests:

Performance tests focus on assessing the system's efficiency, scalability, and responsiveness under different load conditions. These tests measure various performance metrics, such as processing speed, memory usage, and response times, to ensure that the system meets performance requirements.

d. Usability Tests:

Usability tests evaluate the user interface and overall user experience of the emotion recognition system. These tests assess how intuitive and user-friendly the system's interface is and whether it meets the needs and expectations of its intended users. Usability tests may involve conducting surveys, interviews, or observational studies with potential users to gather feedback on the system's design, layout, and functionality.

6.3 Various test case scenarios

Diverse test case scenarios are presented to simulate different usage scenarios and conditions encountered in real-world settings. These scenarios encompass variations in lighting conditions, facial expressions, facial occlusions, and camera perspectives to evaluate the system's robustness and adaptability. Test case scenarios involve single or multiple individuals expressing different emotions at varying intensities, challenging the system's ability to accurately recognize and classify emotions under diverse circumstances. By considering a wide range of test case scenarios, the aim is to comprehensively assess the proposed system's performance across various challenging conditions.

6.4 Inferences drawn from test cases

This focuses on deriving meaningful insights and conclusions from the test results obtained across different test case scenarios. Points to be noted are the observations, trends, and patterns identified during testing, highlighting the system's strengths, weaknesses, and areas for improvement.

Inferences drawn from test cases provide valuable feedback for refining the system's algorithms, enhancing its accuracy, robustness, and generalizability. Additionally, these inferences inform the decision-making process regarding the system's deployment and potential modifications to address any identified shortcomings. Overall, the findings from testing offer valuable insights into the proposed emotion recognition system's performance and its readiness for practical application.

7. Results and Discussion

7.1. Screenshots of User Interface (UI) for the respective module

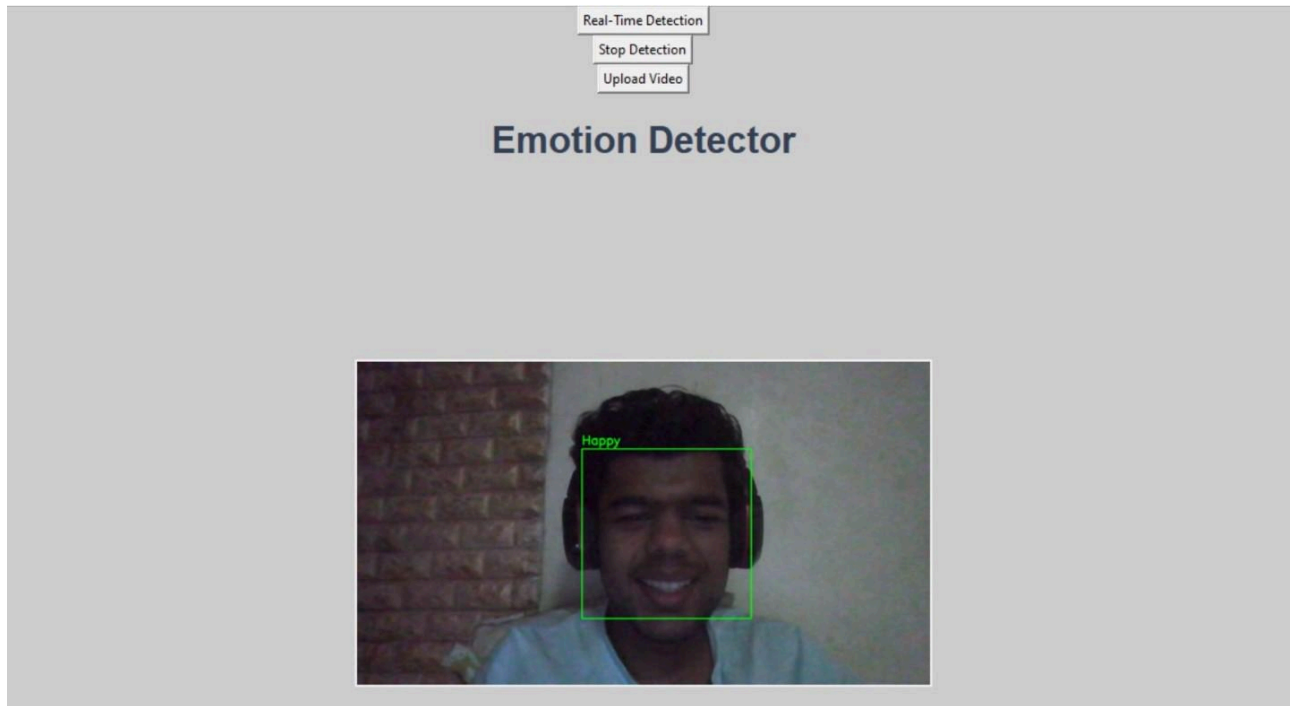


Fig 7.1 - Screenshot of user interface

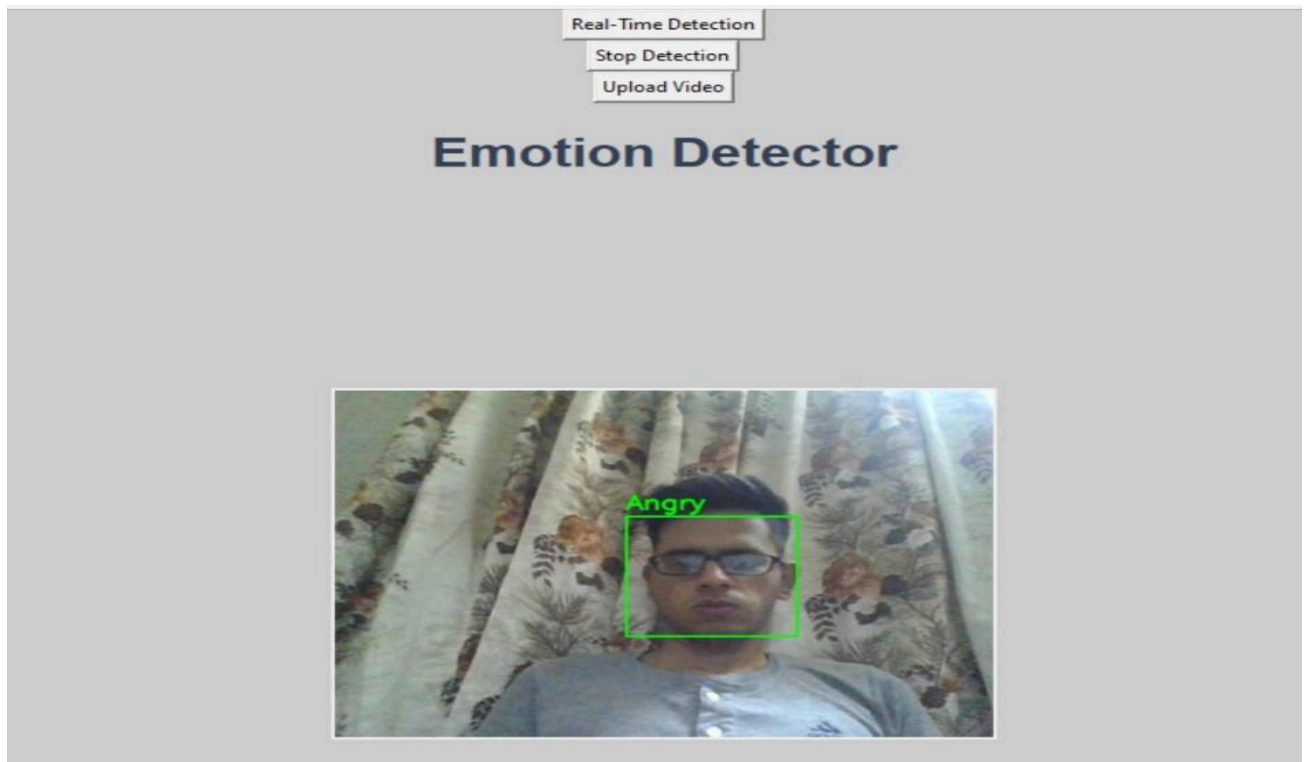


Fig 7.2 - Screenshot of Emotion Detected using Upload Video

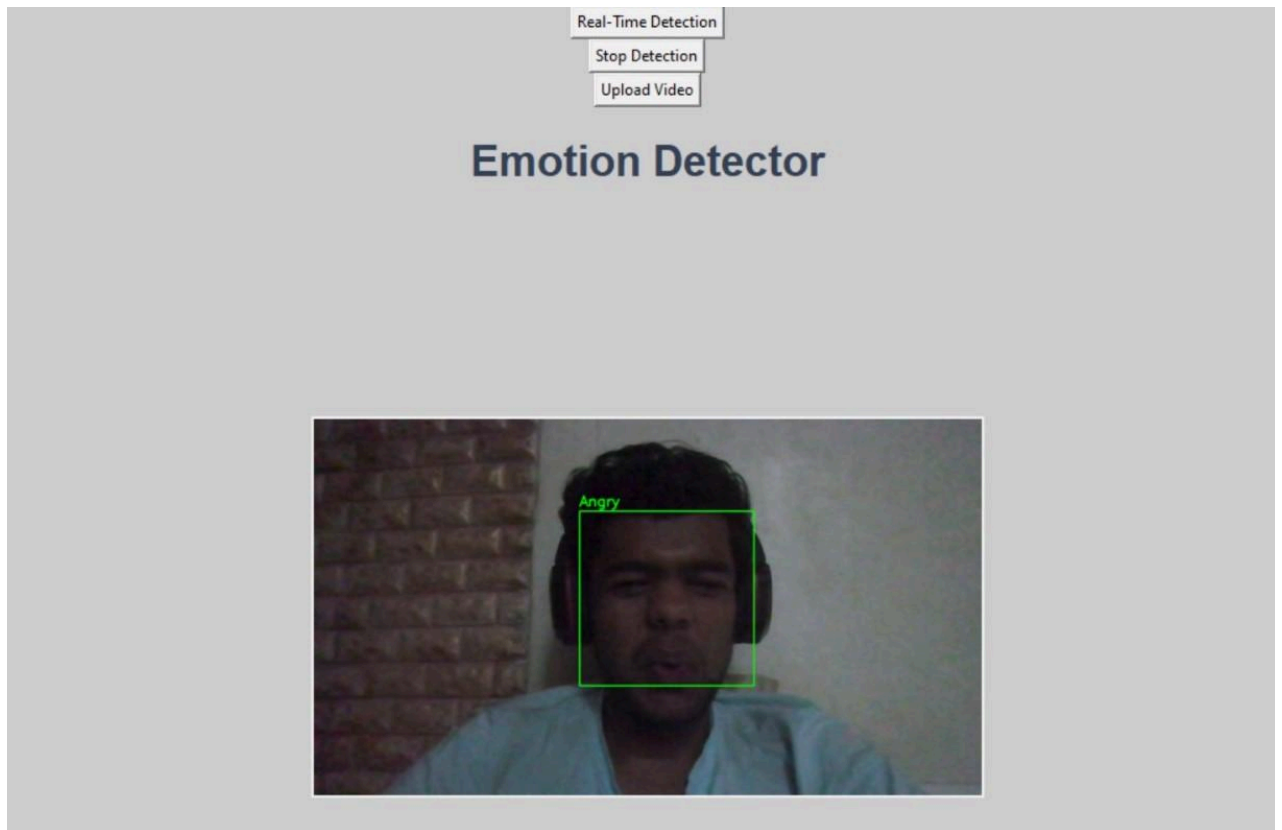


Fig 7.3 - Screenshot of Real time Emotion Detection

7.2. Input Parameters / Features considered

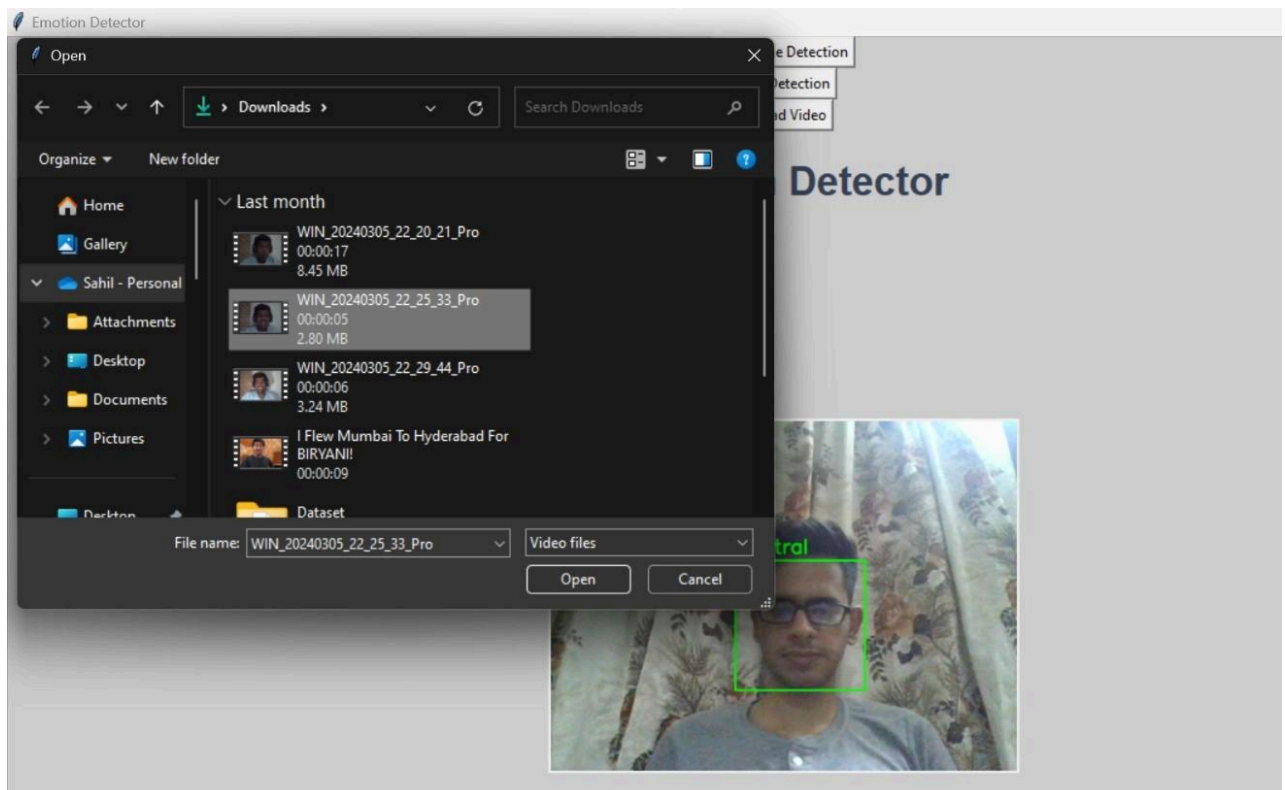


Fig 7.4 Input Parameters for Facial Emotion Recognition System

In fig 7.4 Users can upload the video from which the emotion will be generated by the backend CNN and VGG models and displayed at the frontend.

7.3 Comparison of results with existing systems

Other System	Our System
Only image or video Emotion Detection	Real-Time and Objective Emotion Detection
Only 1 model is used	2 models namely CNN and VGG are used
Dataset is small and not precise	FER2013 Dataset is used which is precise for facial emotion detection

7.4. Inference drawn

Rather than focusing only on image or video facial emotion detection our model can detect emotions in real time as well. TasteFeels provides a user-friendly UI and our model uses CNN and VGG models which give much more precise results than other emotion detection algorithms.

8. Conclusion

8.1 Limitations

Privacy Concerns: One of the primary limitations of facial recognition technology in restaurants is the potential invasion of privacy. Customers may feel uncomfortable or apprehensive about having their facial expressions captured and analyzed without their explicit consent. Ensuring strict adherence to data protection regulations and providing transparent information about data usage is essential to address these concerns.

Accuracy and Reliability: Facial recognition algorithms may not always accurately interpret diners' facial expressions or emotions. Factors such as lighting conditions, facial obstructions (e.g., masks), and cultural differences in expression interpretation can affect the reliability of the technology. Improving the accuracy of facial recognition systems, particularly in diverse dining environments, is a significant challenge.

User Acceptance: Not all diners may be comfortable with the idea of their dining experiences being monitored and analyzed through facial recognition technology. Some individuals may perceive it as intrusive or unnecessary, leading to resistance or reluctance to participate. Ensuring clear communication about the purpose and benefits of the technology is crucial to gaining user acceptance.

Bias and Interpretation: Facial recognition algorithms may exhibit biases in interpreting facial expressions, leading to inaccuracies or misinterpretations. Additionally, emotions are complex and nuanced, making it challenging to accurately capture and analyze them solely based on facial cues. Addressing biases in algorithm training data and improving the ability to interpret subtle emotional cues are ongoing challenges.

Technical Limitations: Facial recognition technology may face technical limitations such as processing speed, scalability, and compatibility with existing restaurant systems.

Implementing real-time facial analysis in busy restaurant environments without causing delays or disruptions requires robust technical infrastructure and optimization.

8.2 Conclusion

- The integration of facial recognition technology into the dining experience presents both exciting opportunities and significant challenges.
- By providing real-time feedback and insights, restaurants can improve service quality, food offerings, and ambiance, ultimately enhancing customer satisfaction and loyalty.
- Moreover, the data generated by facial recognition technology can contribute to more accurate and reliable restaurant reviews, benefiting both patrons and establishments
- In essence, while there are challenges to overcome, the potential of facial recognition technology to revolutionize the dining experience is undeniable.
- By harnessing the power of real-time facial analytics, TasteFeels and similar innovations have the opportunity to shape the future of dining, offering personalized, immersive, and memorable experiences for patrons around the world.

8.3 Future Scope

- To implement Interactive UI for Better User Experience
- Incorporating real-life datasets into facial recognition systems.
- Improving Accuracy of overall system
- Real-Time Feedback Loop

9. References

1. A. Rao, A. Ahuja, S. Kansara and V. Patel, "Sentiment Analysis on User-generated Video, Audio and Text," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 24-28, doi: 10.1109/ICCCIS51004.2021.9397147.
2. Xin Liu, Henglin Shi, Haoyu Chen, Zitong Yu, Xiaobai Li ,”An Identity-free Video Dataset for Micro-Gesture Understanding and Emotion Analysis Finland School of Electrical and Information Engineering”, Tianjin University, China, CVPR 2021
3. Marian Stewart Bartlett, Gwen C. Littlewort Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, Javier R, “Automatic Recognition of Facial Actions in Spontaneous Expressions”, Movellan Journal of Multimedia September 2006
4. Nesrine Fourati, Catherine Pelachaud ,” Emotional body expression in daily actions database”, CNRS LTCI 2014.
5. Farnaz Abtahi, Tony Ro, Wei Li, Zhigang Zhu “Emotion Analysis Using Audio/Video, EMG and EEG: A Dataset and Comparison Study” 2018 IEEE Winter Conference on Applications of Computer Vision
6. Teddy Surya Gunawan, Arselan Ashraf, Bob Subhan Riza, Edy Victor Haryanto, Rika Rosnelly, Mira Kartiwi, Zuriati Janin “Development of video-based emotion recognition using deep learning with Google Colab” TELKOMNIKA Indonesian Journal of Electrical Engineering October 2020
7. A. V. Savchenkoa, I. A. Makarov “Neural Network Model for Video-Based Analysis of Student’s Emotions in E-Learning” ISSN 1060-992X, Optical Memory and Neural Networks, 2022
8. Elif Yildirim, Fatma Patlar Akbulut, Cagatay Cata “Analysis of facial emotion expression in eating occasions using deep learning” Multimedia Tools and Applications (2023)
9. Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Mazumder, Amir Zadeh, Louis-Philippe Morency
“Context-Dependent Sentiment Analysis in User-Generated Videos” 2017 Association for Computational Linguistics

Appendix

A. Paper Published

TasteFeels: The Next Frontier of Dining Enhancement with Facial Recognition

Sahil Ramchandani

*Department of Computer
Engineering(Vivekanand Education
Society's Institute of Technology)
Mumbai, India*
2020.sahil.ramchandani@ves.ac.in

Vinit Patil

*Department of Computer Engineering
(Vivekanand Education Society's
Institute of Technology)
Mumbai, India*
2020.vinit.patil@ves.ac.in

Tanishq Harchandani

*Department of Computer
Engineering(Vivekanand Education
Society's Institute of Technology)
Mumbai, India*
2020.tanishq.harchandani@ves.ac.in

Chirag Panjwani

*Department of Computer Engineering
(Vivekanand Education Society's
Institute of Technology)
Mumbai, India*
2020.chirag.panjwani@ves.ac.in

Mrs. Abha Tewari

*Department of Computer
Engineering(Vivekanand Education
Society's Institute of Technology)
Mumbai, India*
abha_tewari@ves.ac.in

I. ABSTRACT

The interface between technology and cooking always holds promise to enhance dining experiences, and the emergence of facial recognition technology has opened a new frontier in this regard. This paper explores the concept of "TasteFeels", a new approach to food enhancement ML- Using its real -time and facial analytics, TasteFeels enables restaurants to create personalized menus, flavors and presentations based on diners' sensory and physiological responses, ultimately elevating the overall dining experience. In this paper we explore aspects of TasteFeels technology in more detail, including the use of facial recognition techniques and their integration into cooking practices.

By tackling these challenges head-on, the proposed TasteFeels system aims to provide a more accurate and user-friendly dining enhancement experience through facial recognition technology, benefiting both restaurant-goers and culinary professionals as they explore the next frontier of dining innovation.

Keywords: *TasteFeels, Food enhancement, Facial recognition*

II. INTRODUCTION

In today's fast-paced world, food isn't just about food; It's also about making memorable moments and keeping customers happy. Using advanced tools like facial recognition, we're exploring new ways to make dining experiences better and easier to review. Our work focuses on using facial

recognition to review dining experiences. We want to change how we analyze and improve our food. Using this project we are using facial recognition to make the post-meal review process faster, easier and more informative. Imagine enjoying the food and atmosphere at your favorite restaurant. After you eat, instead of composing a long review, you just look at the camera. Facial recognition technology records your emotions and captures your real reactions to food, service and atmosphere. Our work has several important goals. First, we want to simplify the review process, so that you don't have to deal with time-consuming paperwork. Second, we want to capture your emotions and reactions in real time, so your thoughts are authentic. Facial recognition is accurate and can describe your dining experience in detail, capturing details that words alone can't [7]. As we continue to work on this project, we are looking at how facial recognition can improve dining experiences. By simplifying information, we want diners to express their true feelings and help restaurants become even better at serving great food.

III. LITERATURE SURVEY

In the field of emotion recognition and sentiment analysis from multimodal data sources, several key trends and techniques have emerged as demonstrated in the literature [1]-[8]. Research has increasingly focused on combining the analysis of video, audio, and textual data to gain a more comprehensive understanding of emotional states.

One recurring theme is the application of deep learning, particularly models like

LSTM, CNN, and DBN [1]. These deep learning architectures have been extensively employed to extract features and classify emotions from different data modalities, such as images, audio signals, and textual content. But the decision of picking a model depends upon the parameters and type of dataset, there can't be a single model which is best for all datasets available.

Many researches have been done in depth in this field which highlight the importance of expressions and micro-gestures [4]. Doing so helps understand human emotions. Techniques for detecting emotions from facial expressions have been a consistent focus, often employing techniques like Action Units (AU) recognition. These techniques are valuable in recognizing emotions, especially in video data.

The role of context and context-dependent sentiment analysis has been gaining prominence [6]. Analyzing the contextual information surrounding emotional expressions is crucial for a more accurate understanding of the emotional states of individuals, especially in user-generated content. Researchers have proposed various methods, including LSTM networks for capturing contextual information and cross-lingual scenarios to assess the generalizability of emotion recognition models.

Multimodal sentiment analysis, incorporating data from video, audio, and text, has become an exciting research avenue [5]. By leveraging these different modalities, researchers aim to overcome the limitations of single-source emotion recognition and provide a more comprehensive view of emotional states. Nevertheless, challenges in model generalization, privacy concerns, and ethical considerations related to the use of video

data remain relevant areas for future exploration.

In summary, the literature in emotion recognition and sentiment analysis has witnessed significant advancements through the integration of deep learning techniques [1], the focus on facial expressions and micro-gestures [4], the importance of context [6], and the increasing interest in multimodal data sources [5]. These trends collectively contribute to a more comprehensive understanding of human emotions and sentiment in various domains, from video content to e-learning and daily-life scenarios.

IV. PROPOSED SYSTEM

We are proposing the development of two models, both of which will have the same aim, to empower users to submit a video in which we can analyze[4] their facial expressions to discern their emotional state. Our project aims to address the need for a system that can classify various moods based on a user's[3 8 9] facial expressions while they provide feedback. This will involve running different algorithms to determine the user's mood by cross-referencing their expressions with our database. We will then present the user's current mood as feedback.

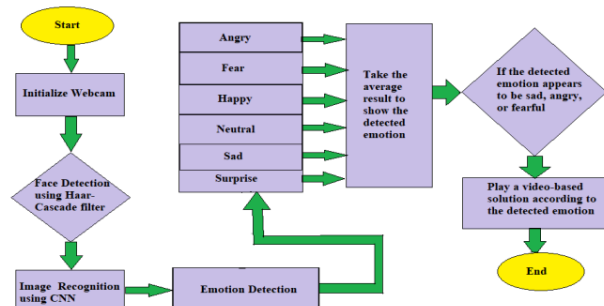


Fig 4.1 - Working of model

Additionally, we plan to offer an alternative method for mood detection and feedback by capturing audio through the system. The

predicted mood will serve as the feedback, which will be conveyed as output, including a text-to-speech conversion for enhanced accessibility. This part of the problem statement will be solved and analyzed differently than the 2 models used right now. The models used are CNN and VGG, a variant of CNN.

V. METHODOLOGY

Analyzing emotions in videos is valuable for various reasons, and it has applications in multiple fields like education, customer service, content recommendation and many more. To build a real-time emotion detection system using Keras, a two-step process was followed: data collection and model training. Data collection involved gathering datasets containing videos of individuals expressing various emotions such as happy, fear, neutral, and surprise. The FER2013 dataset [5] was utilized for training the models. Data preprocessing was performed after data collection, which included converting video data into frames. Videos were transformed into individual frames to create a dataset of images, and pixel values were normalized to bring them within a common range. The dataset was then split into three subsets: training, validation, and test sets, with a common split of 80% for training, 10% for validation, and 10% for testing. The Convolutional Neural Network (CNN) architecture [9] was chosen for model creation due to its suitability for image-based tasks and its ability to capture spatial relationships in data. The CNN model was trained using the training dataset [5], during which the model learned to recognize patterns and features in facial expressions captured in the frames. Subsequently, the model's effectiveness was evaluated on the testing dataset to ensure its performance on unseen data

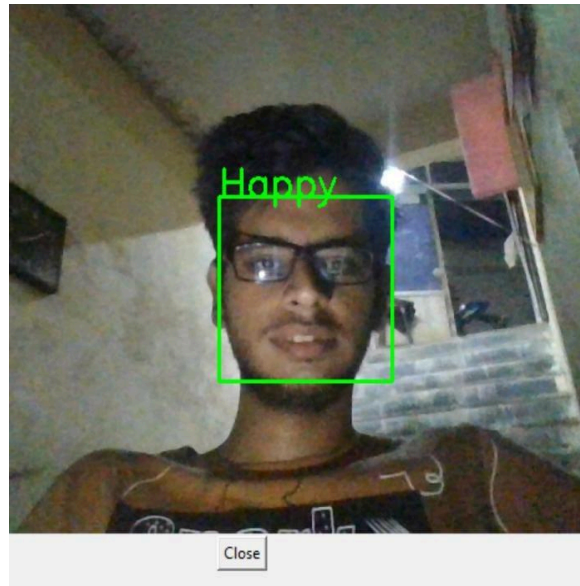


Fig 2.1- Happy emotion

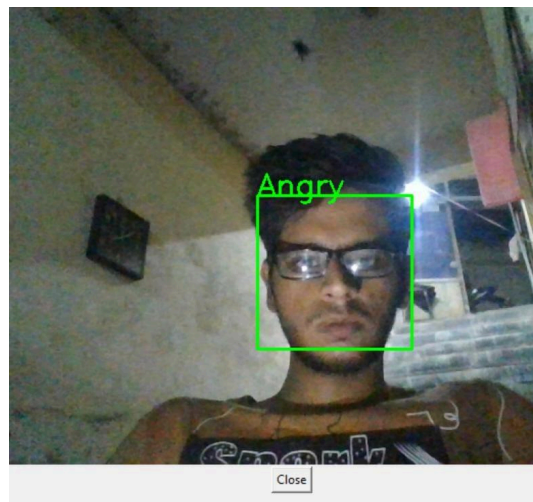


Fig 2.2 - Angry emotion

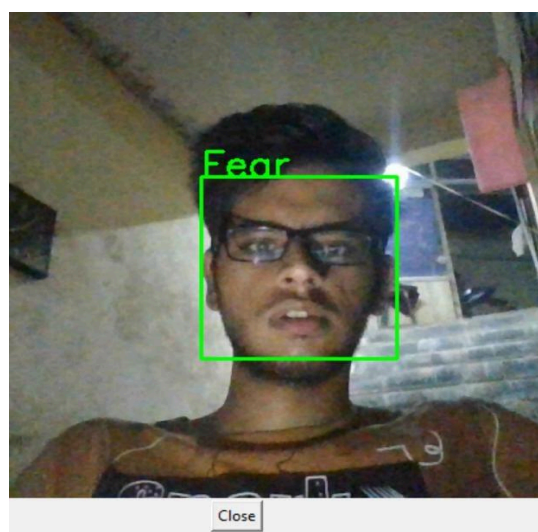


Fig 2.3 - Fear emotion

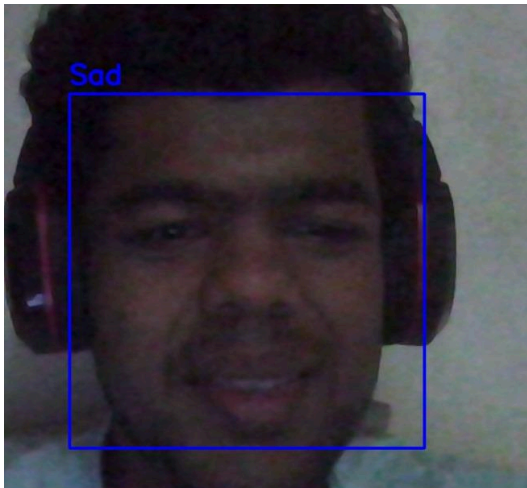


Fig 2.4 - Sad emotion

For VGG, as already mentioned, the dataset used is FER 2013 along with the CK+ dataset, which is the same as CNN[5]. VGG architecture consists of multiple layers, which are convolutional. In the initial layers, low level attributes from the file are extracted while preserving the information. After that, a process[6] called “batch normalization” is applied. The objective of this is to introduce some type of noise in the data so the model[7] reacts better to unseen data and overfitting can be avoided using batch normalization.

The difference can be observed between CNN and VGG,[1] and the difference is that in CNN, the filter size is small. Here in the second layer we increase the filter size, i.e. the network will be able to learn more in terms of spatiality about the input data. After this, an extra layer[8] allows the model to differentiate between more features with a wide scope.

Then comes the final layer, which is the fully connected layer,[3] which is then followed by bath normalization and dropout layers.

And, in the case of VGG,[9] the aggregated extracted features are used to classify the array of emotions.

Usually, the last step [4 6] involves accessing the model and getting to know how correct the model is. [2] This can be done using

evaluation metrics. Examples: Accuracy, precision, recall, F-1 score.

VI. RESULT

This methodology provides a structured approach to developing an emotion recognition model using the FER2013 dataset. Our use case requires real-time video emotion recognition hence, consider factors like model speed and resource requirements.

Our trained emotion recognition model has exhibited strong performance in effectively categorizing emotions based on facial expressions. A notable accomplishment of this project is the seamless integration of real-time emotion recognition using our trained model. This feature allows for immediate emotion analysis in video streams. The model operates efficiently and delivers real-time predictions with minimal delay, rendering it suitable for a wide range of applications that necessitate on-the-spot emotion analysis. These challenges were effectively addressed through the utilization of hardware acceleration and the implementation of model optimization techniques.

In the implementation and evaluation of models, it was evident that emotions depend on the accuracy of the model. That is why for instance, in CNN we have seen a general accuracy of about 60.74%. By any chance, it can be more than this in future. This is demonstrated through an instance whereby “happy” had an accuracy rate that was nearly 80%.

The test results using VGG showed that overall accuracy stood at 75% with primary emotion such as happy, sad and anger being represented better than others by percentage. Therefore, if only important emotions are taken into account, then it means that precision can be easily improved upon.

VII. CONCLUSION:

As we know, the enhancements we aim to achieve in dining experiences, is a use case of the domain of emotion recognition using facial expressions. So, in conclusion, to implement the ML model of emotion recognition precisely, will help increase the accuracy of detection of true and genuine emotions of a customer. For now, we have focused on the domain of facial recognition, which is helping us set the base for testing the model.

In conclusion, it became clear that the dataset selection is important for high accuracy for both CNN and VGG. Also, using the FER dataset and only focusing on the emotions which are primarily meaningful for our use case, the accuracy of VGG model specially can be good enough for it to be used practically. The models chosen work well with the FER-13 dataset as the tuning required is minimum.

The models we have chosen are “Convolutional Neural Network(CNN)” and “Visual Geometry Group(VGG)”. Both have their own set of advantages which are obtained using different sets of layers in depth, And a generic comparison can be made for the moods detected.

VIII. REFERENCES:

1. Rao, A., Ahuja, A., Kansara, S., & Patel, V. (2021). Sentiment Analysis on User-generated Video, Audio and Text. In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 24-28). Greater Noida, India. doi: 10.1109/ICCCIS51004.2021.939714. [Link](#)
2. Liu, X., Shi, H., Chen, H., Yu, Z., & Li, X. (2021). An Identity-free Video Dataset for Micro-Gesture Understanding and Emotion Analysis. CVPR 2021. [Link](#)
3. Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006). Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia*, 1(6). [Link](#)
4. Fourati, N., & Pelachaud, C. (2014). Emotional body expression in daily actions database. CNRS LTCL. [Link](#)
5. Abtahi, F., Ro, T., Li, W., & Zhu, Z. (2018). Emotion Analysis Using Audio/Video, EMG and EEG: A Dataset and Comparison Study. 2018 IEEE. doi:10.1109/WACV.2018.0005 [Link](#)
6. Gunawan, T. S., Ashraf, A., Riza, B. S., Haryanto, E. V., Rosnelly, R., Kartiwi, M., & Janin, Z. (2020). Development of video-based emotion recognition using deep learning with Google Colab. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 18(5), 2587-2594. [Link](#)
7. Savchenkoa, A. V., & Makarov, I. A. (2022). Neural Network Model for Video-Based Analysis of Student's Emotions in E-Learning. *Optical Memory and Neural Networks*, 31(1), 35-42. [Link](#)
8. A. V. Savchenkoa, I. A. Makarov “Neural Network Model for Video-Based Analysis of Student's Emotions in E-Learning” ISSN 1060-992X, *Optical Memory and Neural Networks*, 2022 . [Link](#)
9. Elif Yildirim, Fatma Patlar Akbulut, Cagatay Cata “Analysis of facial emotion expression in eating occasions using deep learning” *Multimedia Tools and Applications* (2023). [Link](#).

B. Plagiarism Report

TasteFeels: The Next Frontier of Dining Enhancement with Facial Recognition

ORIGINALITY REPORT

17%

SIMILARITY INDEX

15%

INTERNET SOURCES

14%

PUBLICATIONS

12%

STUDENT PAPERS

PRIMARY SOURCES

1

www.asianssr.org

Internet Source

4%

2

researchr.org

Internet Source

3%

3

Joshua Dsouza. "Sentiment Analysis on IMDB Movie Reviews using VADER with Lexical Affinity and Semantic Sentiment Expansion", 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), 2023

Publication

1%

4

www.diva-portal.org

Internet Source

1%

5

ibug.doc.ic.ac.uk

Internet Source

1%

6

ebin.pub

Internet Source

1%

7

online-journals.org

Internet Source

1%

C. Review Sheets.

Review 1 :

Inhouse/ Industry Innovation/Research:

Sustainable Goal: Industry Innovation & Infrastructure

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 33

Title of Project: TasteFeels: The next frontier of dining enhancement with facial recognition

Group Members: Chirag Panjwani (D17B/52), Vinit Patel (D17B/55), Sahil Ramchandani (D17C/47), Tanishq Ramchandani (D17A/21)

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	4	4	2	4	1	1	1	1	1	3	3	2	2	3	38

Comments:

Name & Signature Reviewer1

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	3	3	2	4	1	1	1	1	1	3	3	2	2	2	32

Comments:

Date: 10th february, 2024

Name & Signature Reviewer 2

Review 2 :

Sustainable Goal: Industry Innovation & Infrastructure

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 33

Title of Project: TasteFeels: The Next Frontier of Dining Enhancement with Facial Recognition

Group Members: Chirag Panjwani (D17B/52), Vinit Patel (D17B/55), Tanishq Ramchandani (D17A/21), Sahil Ramchandani (D17C/47)

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	4	4	2	4	1	2	2	2	1	2	3	3	2	3	38

Comments:

Name & Signature Reviewer1

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	4	4	2	4	1	2	2	2	1	2	3	3	2	3	38

Comments: Final output/emoion to be displayed, Rating, Probability

Date: 9th March, 2024

Name & Signature Reviewer 2