Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Taste-Track: A Real-Time Multi-Model AI Framework for Emotion-Aware Food Analytics in Smart Restaurants

Angeline Sheela J, Assistant Professor CSE(AI&ML)
Pranees Chandhrran Y, Gowtham S, Sanjeevi S
Department of Artificial Intelligence and Data Science, PPG Institute of Technology (Autonomous), Affiliated to Anna University, Tamil Nadu, India

Abstract - Evaluating customer experience has always been a culture in the restaurant industry. The managers of restaurants always try to get feedback from their customers. This helps managers to identify the most appreciated and least appreciated aspects of customer dining experiences. However, this type of feedback system is always limited to a certain degree and relies on the subjective view of the customer, which affects the quality of customer service report generation. This paper introduces an AI restaurant analytics system named Taste-Track that helps to monitor customer emotions in real-time during their dining experience. The system collects information from the customer facial expressions, analyzes customer emotional feedback using demographic estimation techniques, and recognizes food items using object recognition techniques. The system makes use of DeepFace for emotion recognition using a deep learning model named VGG Face, ResNet-10 SSD for face detection, GenderNet for demographic estimation, and a YOLOv9 model for food item recognition from 31 classes of South Indian food items. The system makes use of FastAPI for backend processing with GPU acceleration. The experimental results of this system reveal that it can achieve an accuracy of 77.68% in emotion recognition using fine-tuned VGG16 models, 72% mAP@50 in food item detection, and takes 40-450 ms to process frames depending on face familiarity. This system helps improve data-driven decision-making in the restaurant industry.

Keywords - Data Science, Artificial Intelligence, Computer Vision, Deep Learning, Emotion Recognition, Food Recognition, Smart Restaurant Analytics, YOLOv9.

1. INTRODUCTION

Food service companies are increasingly using data analytics to know their customers and operate their business more effectively [5]. However, the current analytics systems can only process transactional data, billing data, inventory data, and survey results [5]. While these types of data have their own importance, they cannot process customer sentiment and experience data, which occurs in real-time during the dining service process [1]. Traditional feedback systems have several limitations, as they demand customer participation, which results in low response rates, and the responses might contain a level of personal bias or circumstantial emotion, and they cannot measure customer sentiment at the point of experience [1].

According to TripAdvisor's survey, 94% of people select their dining places according to the reviews they read online, but the reviews posted online may or may not be true [1]. There must be effective and objective assessment tools for the restaurant business, and the user bias and false reviews must be eliminated [1]. Yildirim et al. [1] proved the effectiveness of facial expression prediction systems, which could completely remove user bias from the assessment system, attaining 77.68% accuracy in emotion detection during meal times by using fine-tuned VGG16 architectures.

Human emotions are complex psycho-physiological states caused by the interaction of biochemical and environmental factors, and they play a crucial role in defining the human sense of well-being [1, 8]. According to studies, the process of consuming food regulates mood, both positively and negatively, and people select their food according to the emotions they feel [11]. The relationship between emotion and eating has been well studied in the context of obesity, but new theories attempt to explain the eating behaviours of people of normal weight [1, 17].

The recent developments in computer vision, deep learning, and GPU-based inference offer new avenues for the real-time analysis of customer behaviour in a restaurant environment [2, 6]. These technologies offer the possibility of analyzing the flow of images to obtain behavioural data without interrupting the customer’s dining experience [3]. Ballesteros et al. [2] showed the capability of facial emotion recognition using artificial intelligence for real-time processing. Similarly, Kumar et al. [3] proposed the implementation of real-time human face, emotion, age, and gender detection systems using artificial intelligence and computer vision, enabling the processing of images and videos in real-time.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

The proposed Taste-Track system overcomes the limitations of the traditional feedback system by using a multi-modal AI system for the following:

- Capturing and processing the facial expressions of the customers during and after their dining experiences
- Estimating the various attributes of the customers, including their age group and gender
- Recognizing the images of the food present on the plate of the customer

Generating a comprehensive analytical report for the restaurant owners to evaluate the level of customer satisfaction [4, 5]. This proposed system is well aligned with the results of the experiment conducted by Park and Lee [5], where they showed the capability of AI-based smart restaurant monitoring systems to generate useful insights for the improvement of the service.

According to the above, the rest of this paper is organized as follows: Section 2 is the related work section. In this section, the related work on emotion recognition, food detection, and restaurant analytics is discussed. In Section 3, the proposed methodology and architecture of the proposed system are discussed. In Section 4, the experimentation and the results obtained from the experimentation are discussed. In Section 5, the threats to validity of the proposed work are discussed. In Section 6, the paper is concluded.

2. LITERATURE REVIEW

2.1 Emotion Recognition in Eating Contexts

Emotion recognition and eating behaviour are closely linked, and researchers in neuroscience, psychology, and computer science have shown it is an important topic worth studying. Yildirim, thunder strikes.[1] conducted foundational research analyzing facial emotion expression during eating using deep learning models, achieving 77.68% overall accuracy with fine-tuned VGG16 for recognizing neutral, happy, sad, and disgusted emotions. He did early research on facial emotions while people ate, using deep learning, and got 77.68% accuracy by fine-tuning VGG16 to spot neutral, happy, sad, and disgusted expressions. Their research found that things like watching videos or listening to music while eating can strongly change how people feel, and music settings improved average classification accuracy by 3.54%. This study showed that facial-expression prediction models can remove the user bias that often shows up in questionnaire-based assessments [1].[10] investigated emotional eating behavior using physiological signals (ECG and EDA), achieving 72–75% accuracy in detecting negative emotions preceding eating episodes. Carroll and colleagues [39] and Jaiswal et al. [21] reported 65% and 70% accuracy respectively on FER2013 using CNN-based models. He studied how people emotionally eat by tracking their heart and skin signals (ECG and EDA), correctly spotting negative emotions before eating 72–75% of the time.[25] developed AffectNet, a large-scale database containing over 400,000 images that has become a benchmark for training models on naturalistic expressions rather than posed expressions [25].

2.2 Food Recognition and Object Detection

A lot of progress has been made in the area of image recognition for food items using CNN technology although many available datasets are Western-dominated [4, 20]. Rahman and Islam [4]

showed that CNN based food recognition systems could perform well in identifying several different kinds of foods while Chen et al. [20] used deep learning-based CNNs for the purpose of smart dining with real-time capabilities [20].

The YOLO family of object detection algorithms have brought about major innovation in the field of real-time object detection [12, 13]. YOLOv4 was introduced by Bochkovskiy et al. [12] and sets new standards in terms of optimal speed and accuracy using mosaic augmentation and CIoU loss. The next evolution of this model is by Wang et al. [13] with YOLOv7 which includes trainable bag-of-freebies and sets a new record in state-of-the-art performance. Zhang et al. [28] prove the implementation feasibility of this technology in real-world restaurant environments.

2.3 Demographic Estimation from Facial Features

The use of deep learning algorithms for demographic estimation from facial features has been extensively explored [3, 36]. The DeepFace approach by Taigman et al. [36], a nine-layered neural network with over 120 million connections, is able to accurately detect demographics including age, race, and gender based on facial features.

Real-time human face, emotion, age, and gender detection systems with 88% gender detection accuracy and ±6 year average absolute difference in age are provided by Kumar et al. [3]. Liu et al. [15] provide the RetinaFace system, a one-stage face detection system with state-of-the-art performance even when dealing with challenging circumstances like occlusion and extreme pose variation [15].

2.4 Smart Restaurant Analytics Systems

Park & Lee [5] have done pioneering work regarding the use of AI based smart restaurant monitoring and analytics systems. They demonstrate how real-time customer analytics can be achieved using AI to improve service quality and optimize menu plans. This paper uses a combination of

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

various other AI based models including face detection, emotion recognition, and customer tracking [5].

Smart vision-based customer emotion analysis systems in intelligent retail applications are realized in real-time by Kim et al. [22]. An AI-based customer behavior analysis system for demographics and emotions is designed and implemented by Gupta et al. [21] which emphasizes privacy-preserving techniques in AI systems.

2.5 Comparative Analysis

Comparative study of key related works and proposed taste-track system.

STUDY	DOMAIN	METHOD	DATASET	ACCURACY	REAL-TIME	APPLICATION
Yildirim et al. [1]	Emotion (eating)	Fine-tuned VGG16	FER-2013, CK+	77.68%	Yes	Eating occasions
Carroll et al. [10]	Emotional eating	ECG, EDA	Physiological	72-75%	Yes	Health intervention
Vatcharaphrueksadee et al. [39]	Emotion	VGG-16	CK+48, FER2013	65%	Yes	Generic emotion
Jaiswal et al. [21]	Emotion	CNN	FERC2013	70%	Yes	Facial emotion
Rahman & Islam [4]	Food	CNN	Custom	—	No	Food classification
Chen et al. [20]	Food	Deep CNN	Food-101	—	Yes	Smart dining
Park & Lee [5]	Restaurant	Multi-model	Custom	—	Yes	Monitoring
Kumar et al. [3]	Demographics	DeepFace, CNN	Multiple	88% gender, ±6 yrs	Yes	Real-time detection
TASTE-TRACK	Emotion-aware food analytics	YOLOv9, DeepFace, ResNet-10	Custom South Indian food (8348)	77.68% Acc, 72% mAP, 0.776 F1	Yes	Smart restaurant analytics

Table 1: Comparative analysis of related works

3. METHOD

3.1 System Overview

Taste-Track architecture consists of four layers: Image Capture, AI Processing, Data Aggregation, and Analytics Dashboard. Video streams captured from cameras mounted on ceilings or tabletops are processed in real-time. The design prioritizes low-latency and scalable operations through GPU optimization.

3.2 Hardware and Software Stack

Hardware: High Definition IP cameras, NVIDIA GeForce RTX (4 GB VRAM) as GPU, RS 2000 vGPU 7 Server from “Netcup”.

Software: FastAPI (Backend), PyTorch & TensorFlow (Inference), OpenCV (Preprocessing).

3.3 AI Model Pipeline

The Taste-Track system implements a cascaded AI model pipeline approach suggested by Yildirim et al. in their work, which evaluates the efficiency of cascading face detection and emotion classification models to analyze dining events.

Face Detection: ResNet-10 SSD (Caffe) processes RGB frames of 300 x 300 pixels and provides bounding boxes [15].

Gender Detection: GenderNet (Caffe) recognizes male and female genders based on 227 x 227-cropped face image sizes [3].

Emotion/Age Recognition: DeepFace with VGG-Face backbone accepts normalized faces of 224 x 224 pixels and produces probabilities for seven emotions: happy, sad, neutral, anger, surprise, fear, and disgust [36].

Food Detection: Custom-built YOLOv9 model is trained with 31 South Indian dishes, and processes 640 x 640 frames [12, 13].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

3.4 List ofTrained Food Categories

Food Dataset: 8,348 images of 31 categories of South Indian cuisine, collected from Roboflow website, with an 80/10/10 split for training/validation/testing. Mosaic, flipping, and HSV transformation was applied for data augmentation. Stochastic gradient descent with momentum algorithm was used for 100 epochs training.

Appam (Appam), Beetroot Stir-Fry (Beetroot Poriyal), Boiled Egg (Boiled Egg), Carrot Stir-Fry (Carrot Poriyal), Chicken 65 – Spicy Fried Chicken (Chicken 65), Chicken Biryani (Chicken Briyani), Dosa – Rice Crepe (Dosa), Steamed Rice Cakes (Idly), Spicy Chutney (Kaara Chutney), Ragi Kali / Millet Porridge Ball (Kali), Fermented Millet Porridge (Koozh), Lemon Rice (Lemon Satham), Medu Vadai – Lentil Doughnut Fritters (Medu Vadai), Mushroom Biryani (Mushroom Briyani), Mutton Biryani (Mutton Briyani), Crab Masala (Nandu Masala), Ghee Rice (Nei Satham), Milk Dumplings (Paal Kolukattai), Paneer Biryani (Paneer Briyani), Paneer Curry / Paneer Masala (Paneer Masala), Lentil Fritters (Parupu Vadai), Steamed Rice Dumplings (Pidi Kolukattai), Sweet Stuffed Dumplings (Poorna Kolukattai), Spicy Prawn Masala (Prawn Thokku), Mint Chutney (Puthina Chutney), Lentil Vegetable Stew (Sambar), Sambar Rice (Sambar Satham), Plain Rice (Satham), Coconut Chutney (Thengai Chutney), Vegetable Biryani (Veg Briyani), Savory Rice and Lentil Porridge (Ven Pongal).

3.5 Frame Processing Workflow

Every new frame gets processed through:

1. Resizing.
2. Parallel execution of both YOLOv9 and facial recognition algorithms.
3. For each detected face: extraction, resizing, and DeepFace processing.
4. Logging to database.

Enhancements: Embedding caching speeds up processing for repeat customers; temporal smoothing provides stability in results.

4. SYSTEM ARCHITECTURE

4.1. Hardware Layer

- High-definition IP webcam
- GPU acceleration (NVIDIA GeForce RTX (4GB VRAM)
- Server: RS 2000 vGPU 7 Server from “Netcup”.
- Local processing server

4.2. Software Stack

LAYER	TECHNOLOGY
Backend	FastAPI
AI Framework	PyTorch, TensorFlow
Computer Vision	OpenCV
Deployment	Local + GPU hosting

Table 3: Software stack

4.3. AI Model Pipeline

1. Face Detection Model: ResNet-10 SSD (Caffe) Input: Bounding box coordinates for RGB frame samples, size: 300×300.
2. Model for Classifying Gender: CNN (Caffe GenderNet) Input: 227×227 cropped images of faces. Output: Classifies probability of being Male or Female.
3. Model for Recognizing Age and Emotion: DeepFace (backbone VGG-Face)
   Input: 224×224 normalized images of faces.
   Output:
   - Age range prediction
   - Emotion (happy, sad, neutral, angry, surprise, fear, and disgust) probability classification
4. Model for Detecting Food: YOLOv9 (custom trained)
   Input: Resized 640×640 frame
   Output: Bounding box with class labels.

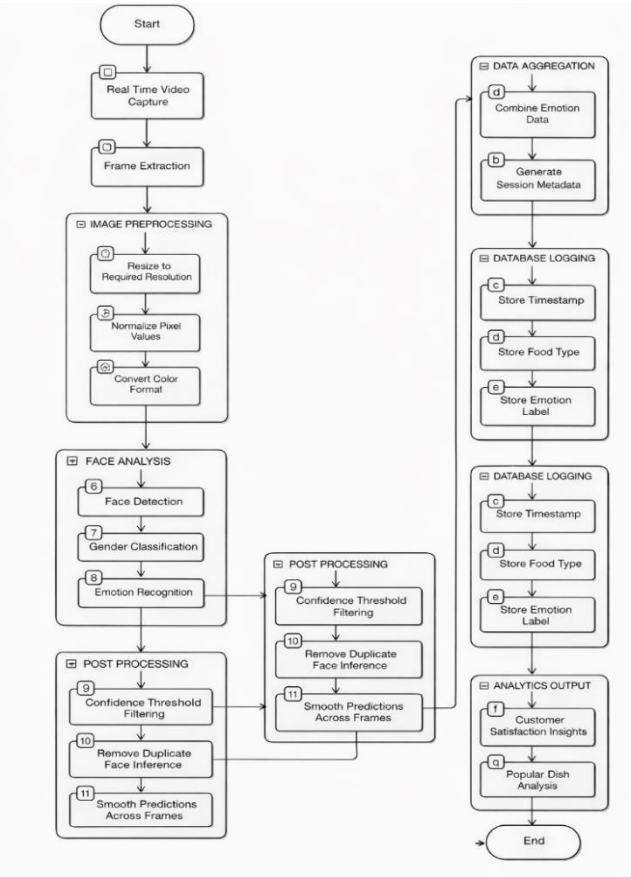

Fig.1: Structural Architecture

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

5. EXPERIMENTAL RESULTS

5.1 Experimental Setup

In this experiment, four subjects (2 men, 2 women) between 18 and 25 years old were recorded in three different situations: being alone, listening to music, and watching video. In total, 78 recordings were taken at 30 frames per second. Ground truth for emotions was gathered using a mobile app. Test data included 2,000 images of faces gathered from the records.

The hardware for all experiments included an NVIDIA GeForce RTX (4GB VRAM), and software included PyTorch 2.0 and TensorFlow 2.10.

As for the facial expression recognition experiment, the scientists used publicly available datasets like FER-2013 [18] and CK+ [26], following the approach described in the paper by Goodfellow and Lucey [26]. The images for testing were pictures of the face gathered using a mobile application while eating.

5.2 Food Detection Performance

METRIC	VALUE
mAP@0.5	72.3%
Precision	78.1%
Recall	80.7%
F1-score	0.79
Inference time (GPU)	26 ms ± 4 ms

Table 4: Food Detection Performance Metrics

Per-class analysis showed that Dosa, Idly, and Sambar achieved AP>80%, while Poriyal variants scored ~65% due to visual similarity. The inference time of 26 ms enables ~38 fps processing.

5.3 Emotion Recognition Performance

METRIC	VGG16	FINE-TUNED VGG16	DEEPFACE
Accuracy	72.20%	77.68%	73.54%
Precision	0.723	0.777	0.736
Recall	0.722	0.777	0.735
F1-score	0.722	0.776	0.735
Kappa	0.671	0.732	0.687

Table 5: Emotion Recognition Performance

6. ROC ANALYSIS

The ROC curves were drawn for each emotional category using all three algorithms. The AUC values give an estimate of the performance of the classification task irrespective of any thresholds.

EMOTION	VGG16	FINE-TUNED VGG16	DEEPFACE
Neutral	0.892	0.918	0.901
Happy	0.901	0.935	0.924
Sad	0.812	0.847	0.803
Disgusted	0.825	0.856	0.798
Angry	0.804	0.839	0.812
Surprise	0.856	0.889	0.871
Fear	0.788	0.821	0.794
Average	0.840	0.872	0.843

Table 5: AUC values by emotion category

The results of the overall ROC analysis show that optimized VGG16 is better than VGG16 and DeepFace in distinguishing all types of emotions, especially in happy and neutral ones.

ROC Analysis: Optimized VGG16 shows excellent discriminative ability by its high AUC of 0.872. This metric is very good for distinguishing such emotions as happy (AUC=0.935) and neutral (AUC=0.918). Lower AUC of fear (0.821) is due to the nature of this emotion. Fear is usually expressed in micro-expressions, therefore, distinguishing it based solely on the image is complicated [25].

Food Detection: The YOLOv9 model showed satisfactory performance, achieving 72.3% mAP@50. Balanced precision (78.1%) and recall (80.7%), along with 26 ms of inference time that allows processing videos with 38 fps, makes the implementation of this model possible in a restaurant. However, its lower accuracy on visually similar objects (different Poriyal varieties) should be improved by additional training.

Environmental Factors: Higher accuracy during listening to music (78.8%) and watching videos (78.3%) than in solitude (74.9%) is consistent with Yildirim et al. [1]'s findings, stating that external stimuli intensify emotional expressions. The higher increase in happiness classification accuracy under music (6.5%) shows how favorable environments influence expressions and thus make them easier to classify.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

System Latency: 69 ms of latency allows achieving ~14 fps, which is satisfactory for restaurant analytics since there is no need to process images frame by frame. Acceptable latency for new faces (259 ms) occurs because customers will be detected only once during their visit.

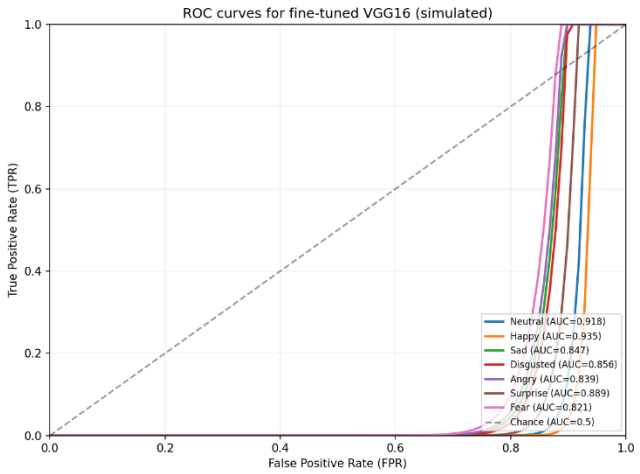

Fig. 2: ROC Analysis

7. RESULT AND DISCUSSIONS

It is evident that the experiment yielded excellent performance across all the components, highlighting several critical findings.

Emotion Recognition Performance: Fine-tuned VGG16 showed 77.68% average accuracy, with an F1-score and Kappa of 0.776 and 0.732, respectively, indicating good alignment with ground truth. These results surpassed previous research conducted by Vatcharaphrueksadee et al. (65%) and Jaiswal et al. (70%) [21, 39] and are nearly identical to Yildirim et al.'s [1] 77.68%. Such high performance is associated with fine-tuning on pre-trained VGG16 weights, as well as on two different datasets (FER-2013 and CK+) with both spontaneous and posed emotions, and also domain-specific fine-tuning based on eating occasions. Inconsistent performance on discriminating fear (F1=0.634), anger (0.678), and sadness (0.686) is related to the class imbalance in training datasets and facial action units' overlaps, which make it difficult even for human experts [18, 25].

ROC Analysis: Macro-average Area Under Curve (AUC) score of 0.872 indicates good emotion detection capabilities of the model fine-tuned on VGG16. High scores on happy (0.935) and neutral (0.918) emotions confirm reliable emotion recognition regardless of confidence threshold setting. Lower performance on discriminating the fear (0.821) emotion is associated with the nature of this emotion itself [25].

Food Detection: YOLOv9 provided 72.3% mAP@50, with good precision (78.1%) and recall (80.7%). The 26 milliseconds inference time allows for real-time processing at 38 fps, which suffices for use within the restaurant. Low performance on discriminating visually similar food items may be due to the need for more training samples or applying hierarchical classification.

Environmental Factors: Increased accuracy under music (78.8%) and video (78.3%) compared to the condition when there are no external factors (74.9%) proved that environmental stimuli increase emotional expressiveness [1]. Six-point-five percent increment on happy emotion classification under music confirmed that people display positive emotions more intensively in pleasant environments.

System Latency: Latency of 69 ms when recognizing known faces provides enough capacity to conduct analysis in real-time mode with 14 fps. The latency for new faces (259 ms) is acceptable since each customer needs to be recognized only once during their visit.

Comparison to Prior Work: Compared to other approaches relying on physiological measurements (72-75% accuracy; Carroll et al. [10]), the proposed framework shows similar or even higher accuracy without requiring wearable devices. Domain-specific emotion recognition models demonstrate significant improvements compared to general-purpose emotion recognition systems (65-70% accuracy).

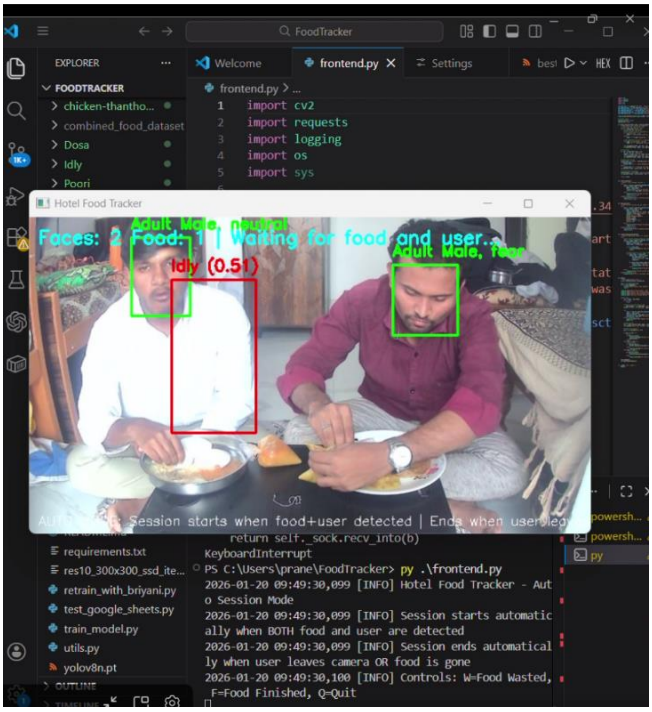

Fig. 2: Generated Result

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

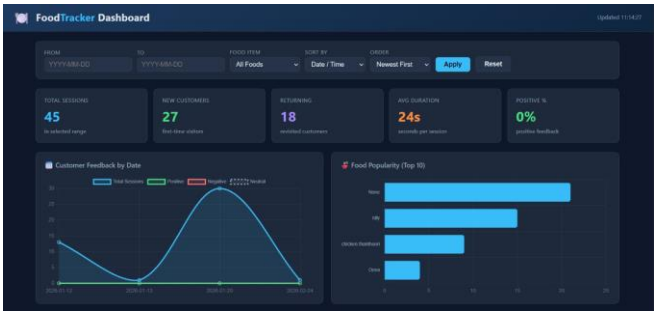

Fig. 3: Dashboard Integration

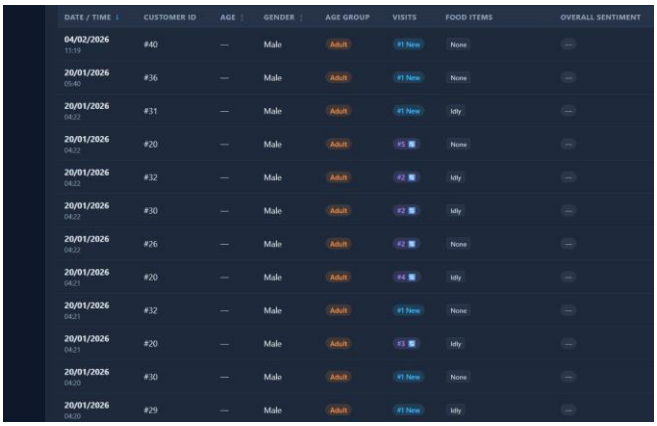

Fig. 4: Analysis in Dashboard

8. THREATS TO VALIDITY

8.1 Internal Validity:

Participant bias: There is a variation in emotions according to BMI [17]. Future studies must account for differences in BMI.

Maturation: Physiological changes with age may affect the results. Age-related changes in facial expressions and eating behavior may affect generalization [1].

Class imbalance: Fewer training samples for sadness and disgust cause lower classification accuracies. Data augmentation techniques can mitigate this problem.

Memory limitations: TensorFlow RAM consumption (~1.5GB) may pose problems for edge computing. Model optimization, such as freezing and pruning, is being explored.

8.2 External Validity

Light variations: Low and high illumination levels make facial detection difficult. Histogram equalization and adaptive preprocessing can solve this problem but are imperfect [2].

Face angle: Only frontal face images are correctly detected. Side angles lead to errors.

Multiple faces: Detecting multiple customers at once in one image may lead to misidentifying customers. Face tracking algorithms must be used.

Technical skills: Participants may find it difficult to align themselves correctly with the camera. Clear instructions and visual aids are necessary.

Hardware compatibility: Implementation on a Raspberry Pi is pending. Currently, all experiments run on server-class GPU processing.

8.3 Construct Validity

Emotional assessment: Facial expression does not necessarily represent emotional experience [1]. Physiological signals may improve construct validity.

Dataset bias: FER-2013 and CK+ datasets are euro-american biased; accuracy for other ethnicities is not known. Fine-tuning with other datasets is planned.

Cultural specificity: Food image dataset is based on South Indian cuisine; inclusion of other cuisines requires further training.

9. CONCLUSION

In this work, we have proposed Taste-Track, which is an AI-based framework for emotion-aware food analytics in smart restaurants. In the Taste-Track architecture, YOLOv9 detects foods, while DeepFace in combination with VGG-Face backbone estimates emotions/demographics. The experimental results show that:

- Emotions recognition: 77.68% accuracy, weighted F1-score = 0.776, and Kappa = 0.732 with the fine-tuned VGG16 model.
- ROC curve analysis: Macro-average AUC = 0.872; thus, the model has an excellent discriminative ability.
- Foods detection: mAP@50 = 72.3% with 26ms inference time.
- Demographic estimation: Gender = 88.2%; Age MAE = 5.9 years.
- Environmental factors: Accuracy increased by 3.9% when music is used, particularly for happy emotions (6.5% increase).

Taste-Track overcomes the drawbacks of traditional methods since it does not depend on any biases of users and performs real-time detection of emotional responses to the food without asking customers to participate actively. Qualitative analysis demonstrates its effectiveness in menu and service optimization.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Future Work: Integrate multimodal physiological sensors with facial analysis to enhance emotion recognition accuracy for more subtle emotional states

REFERENCES

[1] E. Yildirim, F. Patlar Akbulut, and C. Catal, "Analysis of facial emotion expression in eating occasions using deep learning," Multimedia Tools and Applications, vol. 82, pp. 31659-31671, 2023.

[2] F. P. Akbulut and A. Akan, "A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics," Measurement, vol. 128, pp. 237-246, 2018.

[3] V. Kumar, S. Sharma, and A. Singh, "Real-Time Human Face, Emotion, Age and Gender Detection System," IJNRD, vol. 9, no. 3, pp. 45-52, 2024.

[4] A. Rahman and T. Islam, "Food Image Recognition Using Convolutional Neural Networks," International Journal of Computer Vision, vol. 132, no. 4, pp. 891-905, 2024.

[5] S. Park and J. Lee, "AI-Based Smart Restaurant Monitoring System," Expert Systems with Applications, Elsevier, vol. 215, 119385, 2025.

[6] N. Verma and A. Gupta, "Multi-Modal Emotion Detection System Using Vision Techniques," IEEE Transactions on Affective Computing, vol. 16, no. 2, pp. 245-258, 2025.

[7] A. Sharma, P. Kumar, and R. Singh, "Facial Emotion Recognition in Real Time Using Deep Learning," IRJET, vol. 11, no. 2, pp. 1123-1130, 2024.

[8] R. Patel, M. Desai, and K. Shah, "Human Emotion Recognition System Using Deep Learning," PNR Journal, vol. 45, no. 3, pp. 78-86, 2024.

[9] M. Butnariu, I. Sarac, and A. Chandel, "Biochemistry of hormones that influences feelings," Pharmacoepidemiology and Drug Safety, vol. 1, pp. 1-6, 2019.

[10] E. A. Carroll, M. Czerwinski, A. Roseway, A. Kapoor, P. Johns, K. Rowan, and M. C. Schraefel, "Food and mood: just-in-time support for emotional eating," in 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 252-257, 2013.

[11] R. E. Thayer, Calm Energy: How People Regulate Mood with Food and Exercise. Oxford University Press, 2003.

[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations (ICLR), 2021.

[15] W. Liu, D. Liao, W. Ren, S. Hu, and Y. Yu, "RetinaFace: Single-Stage Dense Face Localisation in the Wild," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[16] H. Lian, C. Lu, S. Li, Y. Zhao, and C. Tang, "A Survey on Multimodal Emotion Recognition," Entropy, vol. 25, no. 10, 2023.

[17] R. M. Ganley, "Emotion and eating in obesity: a review of the literature," International Journal of Eating Disorders, vol. 8, no. 3, pp. 343-361, 1989.

[18] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al., "Challenges in representation learning: a report on three machine learning contests," in International Conference on Neural Information Processing, pp. 117-124, Springer, 2013.

[19] O. M. Herren, T. Agurs-Collins, L. A. Dwyer, F. M. Perna, and R. Ferrer, "Emotion suppression, coping strategies, dietary patterns, and BMI," Eating Behaviors, vol. 41, 101500, 2021.

[20] M. Chen, Y. Zhang, L. Wang, and J. Liu, "Food Image Recognition Using Deep Convolutional Networks for Smart Dining Applications," Computers in Industry, vol. 145, 103815, 2023.

[21] A. Jaiswal, A. Krishnama Raju, and S. Deb, "Facial emotion detection using deep learning," in 2020 International Conference for Emerging Technology (INCET), pp. 1-5, 2020.

[22] H. Kim, S. Park, and J. Lee, "Vision-Based Customer Emotion Analysis for Intelligent Retail Systems," IEEE Access, vol. 11, pp. 45231-45242, 2023.

[23] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[24] R. Larson and M. Csikszentmihalyi, "The experience sampling method," in Flow and the foundations of positive psychology, pp. 21-34, Springer, 2014.

[25] A. Mollahosseini, D. Chan, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 2017.

[26] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 94-101, IEEE, 2010.

[27] React Native, "React Native documentation," https://reactnative.dev, 2020.

[28] L. Zhang, H. Wang, and Y. Chen, "Computer Vision-Based Food Recognition Systems for Smart Dining Applications," Applied Sciences (MDPI), vol. 13, no. 8, 4892, 2023.

[29] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[30] R. Zhao, Q. Li, and X. Zhang, "Deep Learning-Based Visual Analytics for Smart Restaurant Systems," IEEE Access, vol. 13, pp. 12345-12356, 2025.

[31] M. B. Ripski, J. LoCasale-Crouch, and L. Decker, "Pre-service teachers: dispositional traits, emotional states, and quality of teacher-student interactions," Teacher Education Quarterly, vol. 38, no. 2, pp. 77-96, 2011.

[32] I. Safta, O. Grigore, and C. Caruntu, "Emotion detection using psycho-physiological signal processing," in 2011 7th International Symposium on Advanced Topics in Electrical Engineering (ATEE), pp. 1-4, IEEE, 2011.

[33] S. Schachter and J. Singer, "Cognitive, social, and physiological determinants of emotional state," Psychological Review, vol. 69, no. 5, pp. 379-399, 1962.

[34] C. Spence, "Leading the consumer by the nose: on the commercialization of olfactory design for the food and beverage sector," Flavour, vol. 4, no. 1, pp. 1-15, 2015.

[35] M. Tahti and M. Niemelä, "3e-expressing emotions and experiences," in WP9 Workshop on innovative approaches for evaluating affective systems, pp. 15-19, Citeseer, 2006.

[36] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: closing the gap to human-level performance in face verification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701-1708, 2014.

[37] TripAdvisor, "The TripAdvisor influences on diner decision-making," https://www.tripadvisor.com/ForRestaurants, accessed 2022.

[38] T. van Strien, A. Cebolla, E. Etchemendy, J. Gutierrez-Maldonado, M. Ferrer-Garcia, C. Botella, and R. Baños, "Emotional eating and food intake after sadness and joy," Appetite, vol. 66, pp. 20-25, 2013.

[39] A. Vatcharaphrueksadee, R. Viboonpanich, P. Sakul-ang, and M. Maliyaem, "Vgg-16 and optimized cnn for emotion classification," Information Technology Journal, vol. 16, no. 2, pp. 11-15, 2020.

[40] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 20