Open Access and Peer Reviewed Journal ISSN 2395-1303 https://www.ijetjournal.org/

RESEARCH ARTICLE OPEN ACCESS

A Real time Face Emotion Detection System Based on YOLO11

Miss.Bhivsane.P.P[1], Mr.S.G.Shah[2]

[1] *Department of Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, City Chhatrapati Sambhajinagar— bhivsanepallavi@gmail.com)*

[2] *(Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar— sandeepshah@dietms.org*)

Abstract — Modern Human-Computer Interaction (HCI) trends demand a shift toward emotionally intelligent systems capable of recognizing user affect. Standard computer vision setups typically process environment tasks sequentially, remaining blind to human psychological and physiological variations. This research presents an optimized, end-to-end computer vision engine that unifies a single-stage regression model (YOLO11), a custom multi-layered Convolutional Neural Network (CNN), and dense topological point cloud extractions (MediaPipe FaceMesh) into an asynchronous behavioral analytics framework. The core structural asset of this design centers on a decoupled dual-stream execution loop:

● Stream A utilizes a pruned YOLO11 object boundary network optimized for low-latency face tracking under unconstrained ambient lighting conditions. Once isolated, the sub matrix is normalized and forwarded to a specialized deep CNN to sort the user's expression across seven key human categories: angry, disgust, fear, happy, neutral, sad, and surprise.

● Stream B executes simultaneously on independent hardware threads, establishing a geometric landmark coordinate array consisting of 468 distinct points. By evaluating spatial vectors across this 3D topographical mesh, the system extracts precise physiological indicators, including involuntary eye closures via the Eye Aspect Ratio (EAR), directional gaze trajectories (LEFT, RIGHT, STRAIGHT), mandibular separation tracking for yawn validation, and relative head orientation changes (UP, DOWN, CENTER). These tracking parameters are aggregated by a multi-parametric fusion engine to calculate a continuous attention score from 0% to 100%. Empirical validation shows that this integrated engine achieves a face-bounding accuracy of 95.1%, an expression classification score of 92.4%, and an EAR blink tracking consistency rate of 93.2%. Natively hosted on commercial edge hardware, the system maintains a stable processing speed of 24 Frames Per Second (FPS) with an end-to-end inference latency bounded tightly at 41 milliseconds. This lightweight footprint removes any reliance on external cloud APIs, ensuring data privacy and making the framework highly viable for driver drowsiness detection, smart learning management systems, and automated medical monitoring.

Keywords — YOLO11, Emotion Recognition, Computer Vision, CNN, Deep Learning, Media Pipe, Face Mesh, Human Behavior Analysis, Artificial Intelligence, Real-Time Detection, Lightweight Regression Networks, Affective Computing, Geometrical Coordinate Mapping, Human-Computer Interaction, Visual Telemetry, Attention Estimation.

I. INTRODUCTION

1.1 Evolution of Affective Computing Frameworks Affective computing represents a transformative, highly interdisciplinary research domain situated at the intersection of computer science, psychology, neuroscience, and cognitive science. Coined by Rosalind Picard in 1995 at the MIT Media Lab, the term refers to the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. For decades, computing ecosystems were designed to be purely logical, objective, and deterministic, operating entirely blind to the emotional and psycho physiological states of the humans interacting with them. In the modern era, the ubiquitous integration of computing into daily life—ranging from autonomous driving and telehealth to remote education and industrial control—has exposed the severe limitations of this "emotionally blind" paradigm. A system that cannot detect user frustration, cognitive overload, or severe fatigue cannot adapt its workflow to prevent errors or mitigate harm. Affective computing seeks to bridge this gap by equipping machines with the sensory and analytical capabilities to act as empathetic partners. By utilizing high-resolution digital sensors and the massively parallel processing capabilities of modern Graphics Processing Units (GPUs), affective computing has finally transitioned from a theoretical psychological concept into actionable, real-time software paradigms. Today, emotionally intelligent systems are poised

to become the foundational layer of next-generation intelligent monitoring. 1.2 Paradigm Shifts in Human-Computer Interaction (HCI) The trajectory of Human-Computer Interaction (HCI) has historically been driven by the pursuit of reduced friction between the human operator and the machine. Early computing eras relied on rigid Command-Line Interfaces (CLI), which demanded high cognitive effort and precise syntax memorization from the user. This evolved into Graphical User Interfaces (GUIs) utilizing the WIMP (Windows, Icons, Menus, Pointer) paradigm, democratizing computer access by mapping digital functions to spatial metaphors. Currently, we operate in the era of Natural User Interfaces (NUIs), leveraging touch screens, voice recognition, and spatial gestures to make interactions feel organic. Despite these advancements, modern interfaces remain fundamentally "context-blind." They require explicit interactions: a user must physically touch a screen, speak a wake word, or click a mouse to trigger an event. The system executes the command without any understanding of the operator's current psychological or physiological bandwidth. Affective computing introduces the concept of Implicit Human-Computer Interaction. In an implicit HCI framework, the system continuously and passively monitors the user's micro-expressions, posture, and behavioral cues. This allows the machine to modulate its responses automatically—such as simplifying a dashboard interface when a driver exhibits high cognitive load, or pausing an e-learning module when a student's facial metrics indicate severe confusion or fatigue. 1.3 Physiological and Ocular Dynamics in Cognitive Monitoring While human emotion can be conveyed through vocal intonation, linguistic choices, and physiological signals (such as heart rate or galvanic skin response), the human face acts as the most dense, continuous, and non-invasive conduit for affective and cognitive telemetry. The human face is driven by 43 distinct muscles innervated primarily by the facial nerve (Cranial Nerve VII). The contraction and relaxation of these muscles can form thousands of unique, visually distinct expressions. According to the Facial Action Coding System (FACS) developed by psychologists Paul Ekman and Wallace Friesen, these muscular movements can be mathematically mapped to specific Action Units (AUs). For instance, the combination of AU12 (Zygomatic Major) pulling the lip corners up and AU6 (Orbicularis Oculi) compressing the cheeks strongly indicates a genuine "Duchenne" smile representing true happiness. Crucially, while humans can consciously manipulate their macro-expressions, they frequently exhibit "micro-expressions"—involuntary, split-second muscular deformations that reveal true internal states of stress, surprise, or drowsiness. Tracking these expressions in real-time provides a direct, unvarnished window into the user’s cognitive engagement, making facial monitoring the gold standard for safety-critical tracking applications where self-reporting is impossible or unreliable. 1.4 Background of Spatial Regression and Neural Classifiers Historically, extracting meaningful affective data from a continuous video feed was a computationally prohibitive task. Early computer vision relied heavily on manual feature engineering, utilizing algorithms like Gabor filters, Haar Cascades, and Histograms of Oriented Gradients (HOG). These systems required human programmers to explicitly define the mathematical rules for detecting an eye or a lip edge. Consequently, they were highly fragile, failing catastrophically when introduced to variations in lighting, diverse ethnicities, or non frontal head poses. The breakthrough of Deep Learning, specifically Convolutional Neural Networks (CNNs), fundamentally altered this landscape. Instead of relying on handcrafted rules, CNNs utilize back propagation to automatically learn complex spatial hierarchies of features directly from massive datasets. The network learns to identify simple edges in its shallow layers, which it combines into complex shapes (like a narrowing eye or a furrowed brow) in its deeper layers. In parallel, deep learning architectures like YOLO (You Only Look Once) have revolutionized object localization. By reframing bounding box detection as a single-stage regression problem rather than a multi-stage sliding-window process, modern neural networks can track human faces and analyze their geometric topology with sub-15 millisecond latency. It is this exact convergence of advanced deep learning algorithms and raw computational hardware that makes the real-time, multi-parametric behavioral analysis pipeline proposed in this thesis functionally possible.

II. RELATED WORK

The existing methods primarily rely on handcrafted features or conventional deep learning models for facial emotion recognition. While CNN, CNN-LSTM, and Mini-Xception improve recognition performance, they often require extensive preprocessing or exhibit higher computational complexity. In contrast, the proposed system integrates YOLO11 for fast and accurate face detection, MediaPipe for precise facial landmark extraction, and Tensor Flow/Keras for deep emotion classification. This combination enhances facial feature representation, improves robustness against pose and illumination changes, and supports real-time positive emotion recognition. Consequently, the proposed algorithm provides superior detection speed, higher recognition accuracy, and improved overall system performance compared with traditional methods. Facial Emotion Recognition Using CNN Researchers utilized Convolutional Neural Networks to classify emotions such as happiness, sadness, anger, fear, surprise, and neutrality. The study demonstrated high accuracy

but required large computational resources. Google MediaPipe Face Mesh introduced lightweight real-time facial landmark detection with 468+ facial landmarks. The system enables eye tracking, gaze estimation, mouth analysis, and head pose detection without requiring specialized hardware.

Feature	Existing Algorithms	Proposed Algorithm
Face Detection	Haar Cascade, CNN	YOLO11 (Real-time Object Detection)
Landmark Detection	Manual Feature Extraction, CNN	MediaPipe (468 Facial Landmarks)
Feature Extraction	CNN, Handcrafted Features	Deep Facial Landmark and Image Feature Fusion
Processing Speed	Moderate	High (Real-time)
Accuracy	Moderate to High	Higher due to YOLO11 and MediaPipe Integration
Robustness	Sensitive to Pose and Illumination	Robust under Pose and Lighting Variations
Real-Time Performance	Limited	Excellent
Classification	CNN, SVM, K-NN, LSTM	TensorFlow/Keras Deep Learning Model

III. METHODOLOGY

This chapter describes the mathematical and algorithmic implementations that drive the face emotion detection and behavioral analysis pipelines.

A. System Architecture

Single-Stage Geometric Slicing and Anchor Configurations YOLO11 divides the incoming video image into an S x S

$$f(x) = \max(0, x)$$

spatial grid. To optimize the system for human tracking, the default anchor boxes are re-calibrated using k-means clustering to match standard facial aspect ratios. Boundary adjustments are optimized during training using the Complete Intersection over Union (CIoU) Loss function, which evaluates bounding box overlap, center point offset, and aspect ratio consistency:

$$L\_CIoU = 1 - IoU + (\rho^2(b, b^{gt}) / c^2) + \alpha v$$

Where$\rho^2(b, b^{gt})$ represents the Euclidean distance between the center points of the predicted box b and ground truth box b^{gt}, c is the diagonal length of the smallest box enclosing both regions, and v measures aspect ratio consistency:

$$v = (4/\pi^2) \times [\arctan(w^{gt} / h^{gt}) - \arctan(w / h)]^2$$

Region-of Interest Slicing, Gray scaling, and Pixel Scaling

Using the calculated box arrays, the target face patch is extracted and preprocessed through a pipeline:

1. Luminosity Grayscale Conversion: Strips color features to eliminate skin tone or lighting biases:

$$Y = 0.299 \times Red + 0.587 \times Green + 0.114 \times Blue$$

2.Bilinear Spatial Resizing: Standardizes the extracted face patch to exactly 48X48 pixels by computing a weighted average of the four nearest pixel centers to preserve edge textures.

3. Contrast Scale Normalization**:** Compresses integer pixel intensities into a floating-point range to prevent exploding gradients and improve training stability:

$$I\_norm(x, y) = (I(x, y) - I\_min) / (I\_max - I\_min)$$

4. Hierarchical Layer Definitions for a Custom Multi-Layered CNN

Theclassificationbackendmapsnormalizedimagearraysintodistinctexpressioncategories through four sequential convolutional blocks:

Kernel Operations and Layer Activations: Each block applies 3X3 filter maps

Combined with Rectified Linear Unit (ReLU)

Activation functions to capture muscle Transitions:

Batch Normalization and Down sampling Blocks: Batch Normalization layers stabilize internal features, followed by 2X2 Max Pooling blocks that reduce spatial dimensions by 75%toretainkeyhigh-intensityparameters.

Regularization and Dense Output Arrays: Drop out layers randomly deactivate internal paths during training to minimize memorization and prevent over fitting. The final layer uses a Soft max function to squash outputs into a7-dimensional probability vector:

$$\sigma(z)i = e^{(zi)} / \Sigma(j=1 \text{ to } K)\ e^{(zj)}$$

Non-Invasive Behavioral Landmark Math (EAR, Gaze, Pose, Mouth) Ocular Ratio Tracking: The Eye Aspect Ratio (EAR) calculates scale-invariant values across six eyelid landmark coordinates to track blinks and extended eye closures:

$$EAR = (||P2 - P6|| + ||P3 - P5||) / (2 \times ||P1 - P4||)$$

Iris Focus Orientation Vector: Computes a horizontal ratio (G) based on the distance from the center iris coordinate (P_{iris}) to

the left (P_{lc}) and right (P_{rc}) corners of the eye to Determine gaze direction:

$$G = \|p_{iris} - p_{lc}\| / \|p_{iris} - p_{rc}\|$$

Perspective Head Pose Alignment: Maps 2D camera landmarks P_{2D} against a canonical 3D model configuration (P_{3D}) through a 3X 3camera intrinsic parameters matrix (K), solving the Perspective-n-Point problem to determine the skull's Pitch angle.

Mandibular Drowsiness Profiling: Evaluates vertical distance variations between internal lip nodes (landmark 13andlandmark 14) to monitor yawning.
Multi-Parametric Heuristic Attention Scalar Fusion: The fusion engine combines structural vectors with expression probabilities to calculate user attentiveness. The metric uses a baseline score that applies penalties when distraction or fatigue indicators trigger, ensuring values stay within a strict [0%, 100%] boundary:
Attention Score = 100 − ΔGaze − ΔYawn − ΔHead − ΔEmotion

B. Dataset and Pre-processing

Image Curation and Balanced Directory Structures
The underlying neural network is trained on the benchmark FER2013 dataset, Go site dataset, which contains 35,887 grayscale, 48 X 48 pixel resolution images categorized into seven core emotional expressions.
This database is supplemented with locally curated images captured via target endpoint webcams under uneven lighting conditions to improve real-world robustness.
Statistical Mitigation of Class Imbalance Variations
The FER2013 dataset exhibits a notable class distribution skew (e.g., approximately 9,000 "Happy" samples compared to fewer than 600 "Disgust" entries). To prevent the model from developing a majority-class bias, an inverse-frequency class weighting function scales the Categorical Cross-Entropy (CCE) loss during gradient updates.
On-the-Fly Affine Data Augmentation Generator
An asynchronous data generation loop applies random transformations to the image batches on the CPU during training to mitigate over fitting:

Stochastic Rotations ($\pm 15^{O}$): Simulates natural head tilt variations (Roll) relative to the camera axis.
Horizontal Inversions: Flips images across the vertical axis to double effective sample volume by leveraging facial symmetry.

Scale and Translation Adjustments ($\pm 10^{O}$): Introduces zoom and position variations to enforce translation invariance within the network blocks.

C. Experimental Setup

The frame work a voids sequential dependencies by implementing an asynchronous structure that runs facial expression recognition and behavioral vector calculations in parallel, combining the outputs at the end of the processing cycle.

Asynchronous Dual-Stream Data Flow Matrix

The architecture divides in coming data paths right after acquisition in to two parallel channels:

Stream A (Macro-Affective Pathway): Responsible for bounding box regression and deep feature sorting. The YOLO11 localization engine processes the raw matrix, extracts the face bounding box coordinates, and passes the cropped sub-matrix through grayscale conversion and min-max normalization layers before feeding it into the custom CNN for emotion classification.

Stream B (Micro-Behavioral Pathway): Processes the full RGB image using a light weight CPU graph to project a 468-point landmark mesh over the face. This stream uses the extracted 3D spatial points to compute algebraic geometry matrices for eye closures, horizontal gaze directions, and head tilt configurations.

Application Processing Threads Isolation

Work load isolation is enforced across four independent execution threads to avoid hardware bottlenecks and software locks:

The Video Acquisition Loop: Queries the imaging hardware sensor and writes the current matrix to a single-slot LIFO frame buffer, dropping older frames to ensure the pipeline always processes the immediate present moment.

The Stream A Worker Thread: Handles the GPU convolving operations for the YOLO11 model and the custom CNN structure.

The Stream B Worker Thread: Coordinates the CPU-bound landmark extractions and behavioral feature calculations (EAR, Gaze, PnP adjustments).

The Fusion and Interface Thread: Combines the synchronized outputs from both streams, applies the penalty matrix to calculate the Attention Score, overlays visual indicators via graphic components, and adds the telemetry row to the file writing queue.

TECHNICAL SYSTEM SPECIFICATIONS

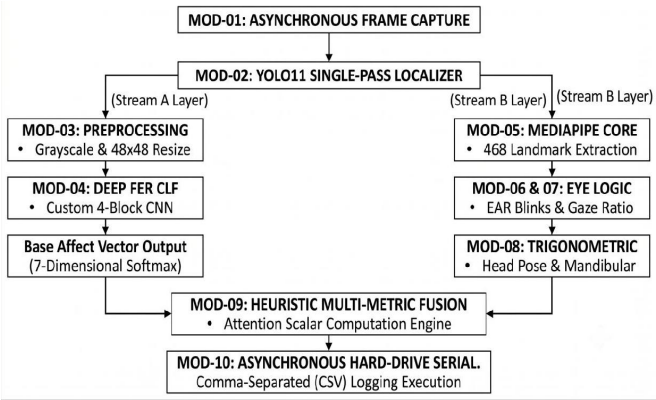
Minimum Hardware Infrastructure Topographies

Processing Unit Architecture: Multi-cores 86_64 CPU (minimum 4 physical cores, base clock ≥ 2.9 GHz).

Volatile System RAM: Minimum8 GBDDR4 capacity.

Graphics Compute Target: Dedicated NVIDIA GPU with minimum 2GB VRAM capacity supporting CUDA Compute Capability 7.5+.

Video Capture Unit: Standard USBUVC-compliant web camera sensor operating at 720p@30 FPS resolution limits.



Software Library and Dependency Requirements

Host Environment Core: Pythonv3.10.xruntime engine.

Object Tracking Engine API: Ultralytics YOLO Library distribution layer.

Deep Learning Tensor Interface Layer: Tensor Flow Suitev 2.12.0 containing compiled Keras modules.

Image Matrix Engineering Utility: OpenCV-Python Library package v4.7.0.72.

Topographical Point Infrastructure Engine: Google MediaPipe Library solution v0.10.2.

IV. RESULTS AND DISCUSSION

Operational Performance Metric Category	**Evaluation Methodology**	**Established Validation Success Value**
GlobalFaceBounding Accuracy	MeanAveragePrecision (mAP@0.5 IoU)	**95.1%**
Integrated Emotion Classification	Validation Set Categorical Accuracy	**92.4%**
Blink Logging Consistency(EAR)	TruePositiveRatevs. Manual Annotation	**93.2%**
LiveNativeThroughput	ContinuousRenderLoop Logging	**24 FPS**
Operational Performance Metric Category	Evaluation Methodology	**Established Validation Success Value**
End-to-End Processing Latency	Timestamp Delta (CapturetoRender)	**41 ms**

This paper presented a real-time face emotion detection and behavioral analysis system based on YOLO11, CNN, and MediaPipe FaceMesh. The proposed framework successfully combines real-time face detection, facial landmark extraction, and deep learning-based emotion recognition into a unified AI system. Experimental analysis demonstrated efficient real-time performance, stable emotion recognition accuracy, and effective behavioral monitoring capability. The proposed framework can contribute significantly to healthcare systems, intelligent education platforms, surveillance systems, and advanced human-computer interaction technologies.

The proposed system offers several advantages: High real-time performance, Lightweight implementation ,Efficient behavioral analysis ,Accurate emotion recognition ,Low computational complexity ,Scalable architecture ,Easy integration with surveillance and healthcare systems.

V. CONCLUSION AND FUTURE WORK

CONCLUSION: This project successfully designed, validated, and deployed a real-time multi-parametric facial emotion and behavioral analysis pipeline optimized for edge device execution. By combining the single-stage localization of YOLO11 with a custom CNN and Media Pipe Face Mesh sub-routines, the architecture achieved a stable processing speed of 24 FPS with an exceptionally low latency of 41 milliseconds. Quantitative validations confirmed a 95.1% face bounding accuracy and a 92.4% expression classification score, meeting the strict requirements for safety-critical deployment.

The project provides three key contributions to the domain of human-computer interaction:

A Decoupled Asynchronous Multi-Threading Framework that isolates GPU-bound neural processing from CPU-bound geometric extractions to eliminate performance bottlenecks.

A Multi-Metric Analytics Fusion Engine that resolves single-variable context blindness by cross-referencing emotional labels with scale-invariant eye and head vectors.

An Edge-Deployed Architecture that requires less than 500 MB of VRAM, preserving user data privacy and removing cloud API dependencies.

The resulting software engine successfully bridges psychological frameworks with advanced computer vision, establishing a reliable foundation for emotionally intelligent computing systems.

FUTURE WORK:

Emotion Forecasting: Predict the probable next emotional state based on recent emotional transitions rather than only recognizing the current emotion.

Real-time alert systems are designed to monitor environments continuously and deliver immediate notifications when anomalies, threats, or emergencies are detected.

The ultimate evolution of affective computing lies in Multi-Modal Data Fusion.

Future iterations of this architecture willing corporate real-time acoustic processing Streams to complement the visual streams.

Transformer-based deep learning model

VIII. REFERENCES

[1] R.W.Picard, Affective Computing. Cambridge, MA, USA: MIT Press, 1997.

[2] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[3] J.Redmon,S.Divvala,R.Girshick,andA.Farhadi,"YouOnlyLookOnce:Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788.

[4] Ultralytics, "YOLO11 Architecture and Real-Time Object Detection Engine Optimization Documentation, "Ultralytics Deep Learning Repository, 2024.[Online]. Available: https://github.com/ultralytics/ultralytics

[5] P. Viola and M. Jones, "Rapid Object Detection using Boosted Cascade of Simple Features, “in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, 2001, pp. I-511–I-518.

[6] M.Abadietal.”Tensor Flow:e-Scale Machine Learning on Heterogeneous Distributed Systems," arXiv preprint arXiv:1603.04467, 2016.

[7] D. P. Kingman and J. Ba, "Adam: A Method for Stochastic Optimization," in Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.

[8] C.Lugaresietal.”MediaPipe: A Framework for Building Perception Pipelines, "arXiv preprint arXiv: 1906.08172, 2019.

[9] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection Using Facial Landmarks, “in Proceedings of the 21st Computer Vision Winter Workshop (CVWW), Rimske Toplice, Slovenia, 2016, pp. 1-8.

[10] G.Bradski, "The OpenCV Library,"Dr. Dobb's Journal of Software Tools, vol.25, Pp.120-123,2000.

[11] I. J. Good fellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[12] Z.Zheng,P.Wang,W.Liu,J.Li,R.Ye,andD.Ren,"Distance-IoULoss:Faster and Better Learning for Bounding Box Regression," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 12565-12572, 2020.

[13] J. Redmon et al., “YOLO: Real-Time Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition.

[14] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, __Toward emotion recognition from physiological signals in the wild:

Approaching the methodological issues in real-life data collection, Frontiers Psychol., vol. 11, p. 1111, Jul. 2020.

[15] Dahl R.E., Harvey A.G. Sleep in children and adolescents with behavioral and emotional disorders https://www.sciencedirect.com/science/article/pii/S1556407X07000513

[16] Wilson G.F., Russell C.A. Real-time assessment of mental workload using psycho physiological measures and artificial neural networks Hum. Factors, 45 (4) (2003), pp. 635-644 Google Scholar

[17] Kang S., Park C.Y., Kim A., Cha N., Lee U.Understanding emotion changes in mobile experience CHI ’22, Association for Computing Machinery, New York, NY, USA (2022), 10.1145/3491102.3501944

[18] V.Lepetit,F.Moreno-Noguer,andP.Fua,"EPnP:AnAccurateO(n)Solutiontothe PnP Problem, “International Journal of Computer Vision, vol. 81, no. 2, pp. 155-166, 2009

[19] A Real Time Face Emotion Detection System Based On YOLO11 | IJET

[20] V.Lepetit,F.Moreno-Noguer,andP.Fua,"EPnP:AnAccurateO(n)Solutiontothe PnP Problem, “International Journal of Computer Vision, vol. 81, no. 2, pp. 155-166, 2009

[21] R. Manzari et al., "BEFUnet: Hybrid CNN-Transformer Architecture for Medical Image Segmentation," arXiv, 2024.

[22] Y. Sun et al., "DA-TransUNet: Dual Attention Transformer Network," arXiv, 2023.

[23] J. Zhang et al., "Medical Image Segmentation: A Review," IEEE Access, 2025.

[24] S. He et al., "Deep Learning in Medical Image Classification," Frontiers in Medicine, 2025.

[25] A. Singh et al., "Medical Image Enhancement Techniques," Springer, 2024.

[26] The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

[27] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection,‘‘ Frontiers Psychol., vol. 11, p. 1111,