A Neuro-Inspired Computational Framework for Modeling Human Emotional States Using Multimodal Physiological and Thermal Signals

Tanishka Pansare[1], Vivek Bokefode[2]
[1,2]*Department of Computer Science, Dr. DY Patil College of Arts, Commerce and Science*
doi.org/10.64643/ijirt.208599-459

***Abstract*—Human emotion recognition remains a complex challenge due to the subjective, dynamic, and multi-layered nature of affective states. Existing systems that primarily rely on Facial Expression Recognition (FER), facial landmark detection, and iris tracking can be affected by environmental conditions and may not fully capture the involuntary physiological changes associated with emotional responses.**
To address these limitations, this paper presents a neuro-inspired multimodal framework that integrates facial analysis with infrared thermal imaging for human emotional-state monitoring. Thermal imaging provides non-contact information related to facial temperature distribution and physiological changes associated with autonomic nervous system activity. A Convolutional Neural Network (CNN) is proposed for extracting spatial characteristics from thermal patterns, while a Long Short-Term Memory (LSTM) network is employed to model temporal variations in emotional responses.
The framework further proposes a neural transmission pathway model that conceptually connects detected emotional states with underlying neurobiological processes as a computational abstraction. By integrating affective computing with concepts from neuroscience, the framework establishes a foundation for biologically inspired emotional-state modeling, with potential applications in human-computer interaction, adaptive systems, emotional well-being monitoring, and future neuro-rehabilitative technologies.

***Index Terms*—Facial Expression Recognition (FER), Thermal Imaging, Neural Transmission Modeling, Multimodal Emotion Detection, CNN-LSTM Fusion, Affective Computing, Neurotransmitter Mapping, Physiological Signal Processing, Human-Computer Interaction (HCI), Neuro-Inspired Computing.**

I. INTRODUCTION

Emotion recognition is a core enabling capability for affective computing, with applications ranging from human-computer interaction (HCI) to mental well-being monitoring and adaptive systems. Human emotions are subjective, dynamic, and multi-layered, which makes automatic recognition a persistent research challenge [1]. Most existing systems rely on visible-spectrum cues, including facial expression recognition (FER), facial landmark detection, and iris tracking. While effective under controlled conditions, these approaches are sensitive to environmental factors such as illumination, occlusion, and head pose, and they capture only the voluntary, expressive component of emotion. They may not fully capture the involuntary physiological changes associated with emotional responses [1].
Infrared thermal imaging offers a complementary signal. Because it measures facial temperature distribution and the physiological changes driven by autonomic nervous system (ANS) activity, thermal imaging is non-contact, illumination-robust, and difficult to consciously suppress or fake [2]. This makes it a strong candidate for capturing the physiological substrate of emotion that visible imaging misses.

In this paper, we propose a neuro-inspired multimodal framework that fuses facial analysis with infrared thermal imaging. The framework uses a CNN to extract spatial features from thermal patterns and an LSTM to model temporal variations in emotional responses [2]. In addition, we introduce a neural transmission pathway model that conceptually connects computationally detected emotional states with their underlying neurobiological processes, including autonomic responses, neurotransmitter-mediated signalling, synaptic transmission, and interactions between limbic and prefrontal regions.

The main contributions of this work are:
1. A multimodal framework combining visible facial analysis with infrared thermal imaging for emotion recognition.
2. A CNN-LSTM architecture for joint spatial and temporal modeling of thermal and facial signals.
3. A neuro-inspired neural transmission pathway model that maps detected emotional states to neurobiological processes as a computational abstraction.
4. An explicit analysis of the feasibility and limitations of contactless neural detection.
5. An evaluation of the proposed framework against unimodal baselines [complete with your results].

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes thermal physiological sensing. Section IV discusses neural transmission and contactless detection. Section V covers brain-wave characterization. Section VI presents the proposed neuro-inspired computational model. Section VII discusses feasibility and limitations. Section VIII describes thermal and neural signal fusion. Section XI concludes the paper.

II. RELATED WORK

A. Thermal Imaging for Emotion Recognition

Infrared thermography is an established non-contact method for measuring emotional arousal through facial temperature changes. Merla and Romani showed that facial cutaneous temperature and its topographic distribution correlate with emotional arousal and sympathetic activity [5]. Kosonogov et al. demonstrated that facial thermal variations at the nose tip index the arousal dimension of emotional stimuli, comparable to electrodermal responses [6]. Early classification work by Khan et al. used facial thermal feature points with discriminant analysis, reporting person-independent success rates around 56% [7], and later showed thermal features can distinguish pretended from evoked expressions [8]. Yoshitomi et al. combined thermal image processing with a neural network for expression recognition [9]. More recent work applies deep learning directly to thermal images: Gupta and Sengupta proposed a CNN for thermal facial emotion recognition [10], and Goulart et al. used infrared emissivity variations to classify emotions in children [11]. Most of these systems rely on single-frame or statistical temperature features and do not explicitly model temporal dynamics, which motivates the CNN-LSTM design in this work [Figure1].

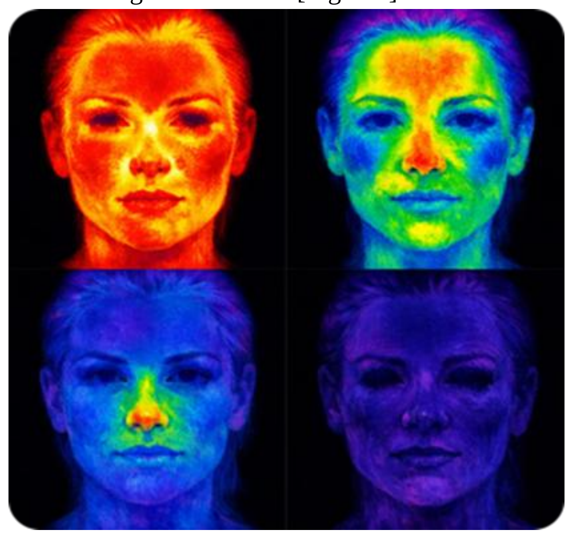


Figure 1: Facial Thermal Differentiation

B. CNN-LSTM for Spatiotemporal Emotion Modeling

CNN-LSTM architectures are a standard approach for capturing spatial and temporal information in facial expression sequences. Rajan et al. combined dual CNNs with an LSTM layer to extract spatial features and temporal relations for real-time FER [12]. Pan et al. proposed a deep spatial and temporal aggregation framework fusing CNN and LSTM streams for video-based FER [13]. Mohana et al. employed a VGG-19 backbone with a bidirectional LSTM for spatiotemporal feature extraction [14]. Kabir et al. applied a CNN-LSTM architecture to posed and non-posed facial expression datasets [15]. These works operate on visible-spectrum imagery; applying the same spatiotemporal paradigm to thermal patterns, as proposed here, remains comparatively underexplored [Figure 2].

C. Multimodal Fusion of Physiological Signals

Multimodal fusion of physiological signals consistently improves emotion recognition over unimodal approaches. Zhang et al. proposed a regularized deep fusion framework for multimodal physiological emotion recognition [16]. Nakisa et al. developed a temporal multimodal deep learning model fusing EEG and BVP signals under early and late fusion strategies [17]. Wang and Wang combined EEG and ECG with an attention-based CNN-GRU model for multimodal classification [18]. Yu et al. introduced GBV-Net, which hierarchically fuses facial expressions with physiological signals [19]. These results support fusing visible facial analysis with thermal imaging in our framework.

D. Affective Computing and Neuroscience

Several efforts integrate neuroscience concepts into affective computing. Vallverdú et al. proposed NEUCOGAR, a neurobiologically inspired architecture that maps the influence of serotonin, dopamine, and noradrenaline onto computational processes [20]. Lim et al. reviewed EEG-based affective recognition from a neuroscience perspective, discussing biologically inspired modeling as a path forward [21]. Gong et al. proposed an amygdala-inspired affective computing framework for personalized emotion recognition [22]. Feng et al. built a brain-inspired affective empathy model grounded in mirror neuron mechanisms [23]. Our neural transmission pathway model follows this line by conceptually linking detected emotional states to neurobiological processes [23], while remaining an explicit computational abstraction.

III. THERMAL PHYSIOLOGICAL SENSING

Long-wave infrared (LWIR) thermal imaging operates in the 8-14 µm band, where the human body emits thermal radiation. Unlike visible-spectrum cameras, which rely on reflected light, an LWIR camera measures radiation emitted directly by the skin, so it is insensitive to illumination and skin color.

The resulting image encodes facial temperature distribution, and because emotional arousal modulates autonomic nervous system (ANS) activity, which alters blood flow in superficial facial vessels, this distribution carries information about affective state.

Functional infrared imaging studies have established that facial cutaneous temperature and its topographic distribution correlate with emotional arousal and sympathetic activity [5]. Specific facial regions are differentially responsive: the nose tip and periorbital areas show robust thermal responses to emotional stimuli [6], and emissivity variations in the nose, cheeks, and periorbital regions have been associated with different emotions [11]. We therefore define thermal regions of interest (ROIs) on the forehead, nose, cheeks, and periorbital region, and extract temperature statistics from each ROI as input features.

The uncooled microbolometer keeps the sensor compact and low-power, and the radiometric mode provides calibrated temperature values suitable for ROI analysis.

IV. NEURAL TRANSMISSION AND CONTACTLESS DETECTION

A. Biological Basis of Neuronal Transmission

Neural signalling follows a well-characterized sequence: a stimulus activates a neuron, which generates an action potential, propagates it along the axon, and transmits the signal across a synapse to downstream neurons.

At rest, a neuron maintains a resting potential: a negative membrane potential established by the unequal distribution of ions across the cell membrane and maintained by ion pumps and channels. When a stimulus depolarizes the membrane past a threshold, voltage-gated channels open and the membrane potential rises rapidly, producing an action potential. This depolarization is followed by repolarization, which restores the resting potential and prepares the neuron for subsequent firing. When the action potential reaches the axon terminal, the neuron releases neurotransmitters into the synapse, the junction between the presynaptic and postsynaptic neurons. The neurotransmitters diffuse across the synaptic cleft and bind to receptors on the postsynaptic membrane, eliciting a postsynaptic response that may excite or inhibit the downstream neuron. This chain, stimulus → neuron activation → action potential → synaptic transmission → downstream neural activity, is the fundamental unit of neural communication.

B. Electromagnetic Signature of Neural Activity

Neuronal signalling is fundamentally electrical: action potentials and postsynaptic currents are moving charges, and moving charges produce magnetic fields.

Although these fields are extremely weak, they are measurable, which yields the chain:

neural current → magnetic field → magnetic sensing → reconstructed neural activity

This is the physical basis of magnetoencephalography (MEG), which reconstructs neural activity from the magnetic fields generated by neural currents. MEG therefore offers a direct, non-contact window onto neural transmission itself, in contrast to techniques that measure downstream consequences of neural activity.

C. Contactless Neural Sensing

Table I compares the main sensing techniques relevant to this work.

It is essential to be precise about what our thermal camera does and does not measure. The thermal camera does not detect neuronal transmission directly. It measures a downstream physiological manifestation: the surface temperature changes driven by autonomic nervous system activity. MEG and OPM-MEG, by contrast, sense the magnetic fields generated directly by neural currents.

Table I: Comparison of neural and physiological sensing techniques

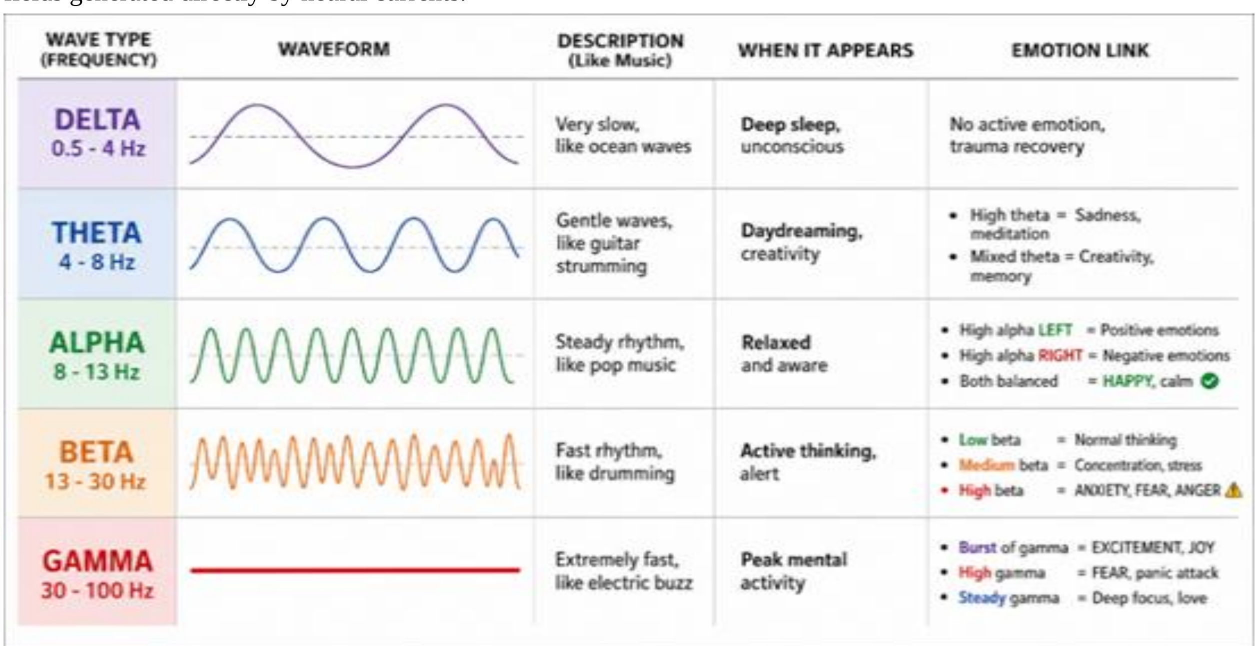
Technique	Contact?	What it measures
EEG	Yes	Electrical potential
MEG	No scalp contact	Magnetic fields from neural currents
Thermal imaging	No	Surface temperature
fMRI	No	Hemodynamic changes
OPM-MEG	No scalp electrodes	Magnetic fields

This distinction is central to the honest framing of this work: our framework is a computational abstraction that connects observable thermal and behavioral signals to a modeled neural pathway, not a direct neural measurement.

V. BRAIN-WAVE CHARACTERIZATION

Population-level neural activity produces oscillatory signals conventionally decomposed into frequency bands as shown in [Figure 3].

The standard bands are delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (above 30 Hz). Exact cutoffs vary across systems and guidelines, so band boundaries are conventions rather than fixed physiological thresholds.



WAVE TYPE (FREQUENCY)	WAVEFORM	DESCRIPTION (Like Music)	WHEN IT APPEARS	EMOTION LINK
DELTA 0.5 - 4 Hz		Very slow, like ocean waves	Deep sleep, unconscious	No active emotion, trauma recovery
THETA 4 - 8 Hz		Gentle waves, like guitar strumming	Daydreaming, creativity	• High theta = Sadness, meditation • Mixed theta = Creativity, memory
ALPHA 8 - 13 Hz		Steady rhythm, like pop music	Relaxed and aware	• High alpha LEFT = Positive emotions • High alpha RIGHT = Negative emotions • Both balanced = HAPPY, calm
BETA 13 - 30 Hz		Fast rhythm, like drumming	Active thinking, alert	• Low beta = Normal thinking • Medium beta = Concentration, stress • High beta = ANXIETY, FEAR, ANGER
GAMMA 30 - 100 Hz		Extremely fast, like electric buzz	Peak mental activity	• Burst of gamma = EXCITEMENT, JOY • High gamma = FEAR, panic attack • Steady gamma = Deep focus, love

Figure 3: The Brain waves frequency

These bands are frequency ranges of population-level neural oscillatory activity, and their interpretation depends on context. A single band does not map deterministically to a single emotion. Elevated beta activity is not "the stress band": beta-range oscillations accompany active motor planning and sensorimotor processing in some contexts and sustained attention in others. The same caution applies to every band. This context dependence is why we treat band power as a neural-state feature to be learned by the model, rather than as a direct emotion label.

Given a raw neural signal, we propose the following processing pipeline: raw neural signal → preprocessing → FFT/time-frequency analysis → band-power extraction → neural-state features

The band power over an interval [f1, f2] [f1, f2] [f1, f2] is computed from the power spectral density of the signal. For a frequency-domain representation X(f)X(f)X(f) of the neural signal:

$$P_{\text{band}} = \int_{f_1}^{f_2} |X(f)|^2 \, df$$

where X(f)X(f)X(f) is the frequency-domain neural signal. This provides a quantitative, mathematically grounded feature set, in contrast to purely qualitative descriptions of brain states. Table II summarizes the standard bands and their approximate ranges.

Table II: Standard EEG frequency bands

Band	Approximate range (Hz)
Delta	0.5-4
Theta	4-8
Alpha	8-13
Beta	13-30
Gamma	>30

VI. PROPOSED NEURO-INSPIRED COMPUTATIONAL MODEL

The proposed model connects the sensing modalities to emotional-state estimation through a single pipeline:

FER features + Thermal features + Neural/brain-wave features (where available) → Multimodal feature fusion → Neural activation representation → Neuro-inspired transmission model → Emotional-state estimation

Facial expression features capture the observable behavioral component of emotion. Thermal features capture the autonomic physiological component via facial temperature distribution. Neural/brain-wave features, where available, capture the brain-level component. These feature sets are combined in a multimodal fusion stage, which produces a neural activation representation: a compact encoding of the modelled state of the relevant neural circuits. This representation is passed through a neuro-inspired transmission model, which maps it onto the emotional-state estimate.

We emphasize that this is a computational abstraction, consistent with the framing of our abstract. The model does not claim to directly measure neuronal activity or neurotransmitter release; it connects computationally detected emotional states with their underlying neurobiological processes as a modeling device. This abstraction follows earlier neurobiologically inspired architectures that map neurotransmitter influences onto computational processes [20].

VII. CONTACTLESS DETECTION: FEASIBILITY AND LIMITATIONS

Contactless neural detection is attractive, but its feasibility degrades under a number of conditions:

- Extremely weak magnetic signals: neural magnetic fields are orders of magnitude weaker than environmental fields, requiring extremely sensitive sensors.
- Electromagnetic interference: ambient fields from power lines, electronics, and moving objects can swamp the neural signal.
- Distance from sensor: signal strength falls off rapidly with distance.
- Head and body movement: motion introduces artifacts and misalignment.
- Environmental noise: thermal and magnetic environments are rarely controlled outside the laboratory.
- Sensor sensitivity: not all sensors achieve the sensitivity needed for neural-scale signals.

- Spatial resolution: resolving which brain region generated a signal requires dense sensor arrays.
- Individual differences in brain-wave patterns: band structure and reactivity vary across individuals.
- Inability to uniquely map a frequency band to an emotion: the same band can appear across different states, and the same emotion can involve multiple bands.
- Inability to directly measure neurotransmitter concentration remotely: no contactless technique currently quantifies neurotransmitter levels at a distance.

Equally important is when this approach should not be used. A contactless neural-sensing approach should not be treated as a clinical diagnostic tool without validated neural measurements and appropriate medical instrumentation. Our framework is intended for affective-state modeling in HCI and well-being monitoring contexts, not for diagnosis.

VIII. THERMAL + NEURAL SIGNAL FUSION

The distinctive contribution of this work is that the different modalities capture different stages of the emotional response:

- Neural activity → brain-level signal
- Autonomic response → thermal changes
- Facial expression → observable behavioral response

This yields a three-level chain, neural → physiological → behavioral, in which each modality observes a different stage of the same underlying process. Facial expression is the observable behavioral endpoint; thermal imaging captures the autonomic physiological substrate; and neural signals, where available, capture the initiating brain-level activity.

Our framework attempts to computationally connect these levels, which is a substantially richer story than detecting emotion with a thermal camera alone. By integrating affective computing with concepts from neuroscience, the framework establishes a foundation for biologically inspired emotional-state modeling.

IX. CONCLUSION AND FUTURE WORK

This paper presented a neuro-inspired multimodal framework that integrates facial analysis with infrared thermal imaging for emotional-state monitoring, combining CNN-based spatial feature extraction from thermal patterns with LSTM-based temporal modeling, and linking detected states to a neural transmission pathway abstraction.

The framework is positioned as a computational connection across the neural → physiological → behavioral chain, with explicit acknowledgment of the feasibility limits of contactless neural detection. By integrating affective computing with concepts from neuroscience, the framework establishes a foundation for biologically inspired emotional-state modeling. Future work will extend the framework toward real-time deployment in human-computer interaction and adaptive systems, and explore emotional well-being monitoring and neuro-rehabilitative technologies.

REFERENCES

[1] Merla and G. L. Romani, "Thermal signatures of emotional arousal: A functional infrared imaging study," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Lyon, France, 2007, doi: 10.1109/IEMBS.2007.4352270.

[2] V. Kosonogov et al., "Facial thermal variations: A new marker of emotional arousal," PLoS ONE, vol. 12, no. 9, 2017, doi: 10.1371/journal.pone.0183592.

[3] M. M. Khan, M. Ingleby, and R. Ward, "Automated facial expression classification and affect interpretation using infrared measurement of facial skin temperature variations," ACM Trans. Auton. Adapt. Syst., vol. 1, no. 1, 2006, doi: 10.1145/1152934.1152939.

[4] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," ACM Trans. Appl. Percept., vol. 6, no. 1, 2009, doi: 10.1145/1462055.1462061.

[5] Y. Yoshitomi, N. Miyawaki, S. Tomita, and S. Kimura, "Facial expression recognition using

thermal image processing and neural network," in Proc. IEEE Int. Workshop Robot. Hum. Commun. (RO-MAN), 1997, doi: 10.1109/ROMAN.1997.647016.

[6] S. Gupta and A. Sengupta, "Unlocking emotions through heat: Facial emotion recognition via thermal imaging," in Proc. Int. Conf. Electr. Electron. Eng. (ICEEFET), 2023, doi: 10.1109/ICEEFET59656.2023.10452206.

[7] C. Goulart, C. Valadão, D. Delisle-Rodríguez, E. Caldeira, and T. Bastos, "Emotion analysis in children through facial emissivity of infrared thermal imaging," PLoS ONE, vol. 14, no. 3, 2019, doi: 10.60692/106ee-6cw15.

[8] S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian, "Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM," IET Image Process., vol. 14, no. 12, 2020, doi: 10.1049/iet-ipr.2019.1188.

[9] X. Pan, G. Ying, G. Chen, H. Li, and W. Li, "A deep spatial and temporal aggregation framework for video-based facial expression recognition," IEEE Access, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2907271.

[10] M. Mohana, P. Subashini, and M. Krishnaveni, "Emotion recognition from facial expression using hybrid CNN-LSTM network," Int. J. Pattern Recognit. Artif. Intell., vol. 37, no. 8, 2023, doi: 10.1142/S0218001423560086.

[11] M. M. Kabir, T. A. Anik, M. S. Abid, M. F. Mridha, and A. Hamid, "Facial expression recognition using CNN-LSTM approach," in Proc. Int. Conf. Smart Comput. Technol. (ICSCT), 2021, doi: 10.1109/ICSCT53883.2021.9642571.

[12] X. Zhang et al., "Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine," IEEE Trans. Cybern., vol. 51, no. 9, 2021, doi: 10.1109/TCYB.2020.2987575.

[13] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Automatic emotion recognition using temporal multimodal deep learning," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3027026.

[14] Z. Wang and Y. Wang, "Emotion recognition based on multimodal physiological electrical signals," Front. Neurosci., vol. 19, 2025, doi: 10.3389/fnins.2025.1512799.

[15] J. Yu, Y. Ru, B. Lei, and H. Chen, "GBV-Net: Hierarchical fusion of facial expressions and physiological signals for multimodal emotion recognition," Sensors, vol. 25, no. 20, 2025, doi: 10.3390/s25206397.

[16] J. Vallverdú, M. Talanov, S. Distefano, M. Mazzara, A. Tchitchigin, and I. Nurgaliev, "A cognitive architecture for the implementation of emotions in computing systems," Biol. Inspired Cogn. Archit., vol. 15, 2016, doi: 10.1016/j.bica.2015.11.002.

[17] G. N. Taur, M. M. Vidhate, and S. S. Shevkari, "Deep learning meets biometrics: A comparative study of modern identification techniques," International Journal of Innovative Research in Technology, vol. 12, no. 2, pp. 545–552, 2025.

[18] H. Chinni, Y. K. Sharma, S. Shevkari, J. L. Vydehi, Saurabh, and M. Kavitha, "Comparative evaluation of custom CNN architectures and transfer learning models for image classification in PyTorch," in 2026 13th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2026, pp. 1–6, doi: 10.23919/INDIACom70271.2026.11525435.

[19] H. Feng, Y. Zeng, and E. Lu, "Brain-inspired affective empathy computational model and its application on altruistic rescue task," Front. Comput. Neurosci., vol. 16, 2022, doi: 10.3389/fncom.2022.784967.

[20] "Electroencephalography," IEEE Technology Navigator. [Online]. Available: IEEE Technology Navigator

[21] C. Babiloni et al., "EEG standards and best practice," International Federation of Clinical Neurophysiology (IFCN) guidelines.