IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REVIEW ON HUMAN MOVEMENT ANALYSIS, GESTURE RECOGNITION, AND EMOTION DETECTION USING COMPUTATIONAL AND SENSOR-BASED APPROACHES

[1]Suchithra N P, [2]K. Viswanath

[1]Associate Professor, [2]Professor

[1]Electronics and Communication Engineering, [2]Electronics and Telecommunication Engineering,
[1]Akshaya Institute of Technology, Tumakuru, India, [2]Siddaganga Institute of Technology, Tumakuru, India,

Abstract: Human activity understanding has become a critical area of research due to its wide-ranging applications in healthcare, human-computer interaction, surveillance, and affective computing. This paper presents a comprehensive review of recent advancements in human movement analysis, gesture recognition, and emotion detection, focusing on both vision-based and sensor-based methodologies. Various data acquisition techniques—such as RGB-D cameras, inertial sensors, electromyography (EMG), and electroencephalography (EEG)—are examined in relation to their role in capturing motion and emotional states. Additionally, we explore classical machine learning and deep learning frameworks employed for feature extraction and classification in each domain. The paper provides a comparative analysis of current approaches, identifies key challenges such as inter-subject variability, real-time processing, and multi-modal data fusion, and outlines future research directions aimed at developing robust and adaptive systems.

***Index Terms* -** Gesture Recognition, Emotion Detection, Machine Learning, Deep Learning

I. Introduction

The ability to automatically interpret human motion and emotional states has gained significant attention across multiple disciplines, including computer vision, biomedical engineering, and human-computer interaction (HCI). The authors in [1] carried out the literature survey to analyze the various applications of artificial intelligence in healthcare emphasizing the emotional and cognitive analysis technique. Accurate movement analysis, gesture recognition, and emotion detection are essential components in the development of intelligent systems across diverse application domains—ranging from physical rehabilitation and virtual reality to smart environments and mental health monitoring. Recent advancements in both sensor technologies and computational algorithms have significantly enhanced our ability to interpret human behavior in real time. These technologies facilitate a deeper understanding of user actions and emotional states, enabling systems to interact more naturally, adaptively, and effectively with people. Vision-based methods utilizing RGB and depth cameras have shown promise in capturing complex movements and gestures, while wearable sensors such as inertial measurement units (IMUs), EMG, and EEG offer detailed physiological insights relevant to both gesture execution and emotional expression. The review on state-of-the-art computer vision algorithms and emotion recognition using facial expression and body pose is carried out by Rafael Pereira et al. [2].

In this context, this survey provides a comprehensive overview of the current state-of-the-art in human movement analysis, gesture recognition, and emotion detection. The existing literature is systematically categorized based on sensing modalities, algorithmic approaches, and application domains. Furthermore, the paper discusses key challenges in this field, such as variability in human-generated data, environmental limitations, and the complexities associated with real-time system implementation. Finally, the survey outlines important research directions and opportunities that can guide future developments in this rapidly evolving and interdisciplinary area.

II. Sensing Modalities and Data Acquisition

Sensing modalities and data acquisition play a crucial role in the development of systems designed for human movement analysis, gesture recognition, and emotion detection. The general process involved in the sensing and analysis of human recognition system is shown in the Figure 1.

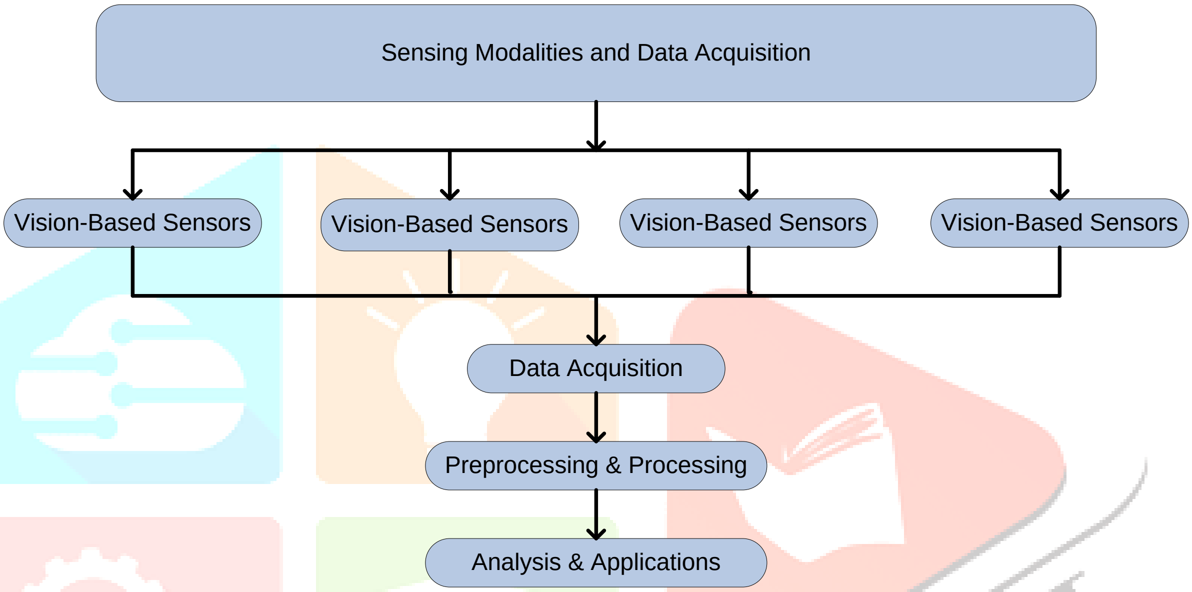

Figure 1: Sensing modalities and data processing pipeline for human activity recognition.

The accuracy, reliability, and real-time performance of these systems are influenced by the selection of the sensing technology. The researchers explored wide range of sensors to capture human activities, physiological signals, and environmental interactions. Different approaches of sensing include vision-based systems, wearable sensors, physiological sensors, multimodal sensors. Each modality offers unique advantages in terms of data richness, cost, portability, and ease of deployment. In addition to sensor selection, effective data acquisition strategies are essential for ensuring high-quality datasets. The performance of subsequent analysis and recognition algorithms are affected by sampling rate, sensor placement, synchronization, noise handling, and environmental conditions.

2.1. Vision-based Sensors

Vision-based sensing is widely used approaches for analyzing human movement and gestures. These systems rely on the cameras to capture visual information such as body posture, facial expressions, and hand movements. A vision based tactile sensor is proposed by [3-4] for continuous sensing. With the advancement of computer vision and deep learning techniques, vision-based methods can effectively detect complex gestures and emotion. The performance of the vision based sensors is affected by lighting conditions, occlusion, and background variations.

2.2. Wearable Sensors

Wearable sensors provide an effective way to monitor human motion and physiological signals continuously. Different devices such as inertial measurement units (IMUs), accelerometers, gyroscopes, and smart wearables are used to capture body movements and activity patterns. These sensors are used in rehabilitation monitoring, sports analysis, and health tracking, since they are portable and collect data in real time. An outline of wearable sensors is given by Aparna K. et al., in the paper [5]. Wearable sensors are used for stress detection method [6], human activity recognition (HAR) in [7]. There are few issues such as sensor placement, user comfort, and battery consumption while choosing the wearable sensors.

2.3. Physiological Sensors

Physiological sensing focuses on capturing signals that reflect a person's internal physical or emotional state. Sensors used for measuring signals such as heart rate, electroencephalography (EEG), electromyography (EMG). These sensors are used in mental health monitoring and stress detection. A physiological sensor is proposed [8-10] to monitor stress level of humans and analysis is carried out. The data must be preprocessed before feature extraction. Filtering and smoothing are the two steps in preprocessing the noise removal signals. The noise removal technique is performed as shown in Figure 2.

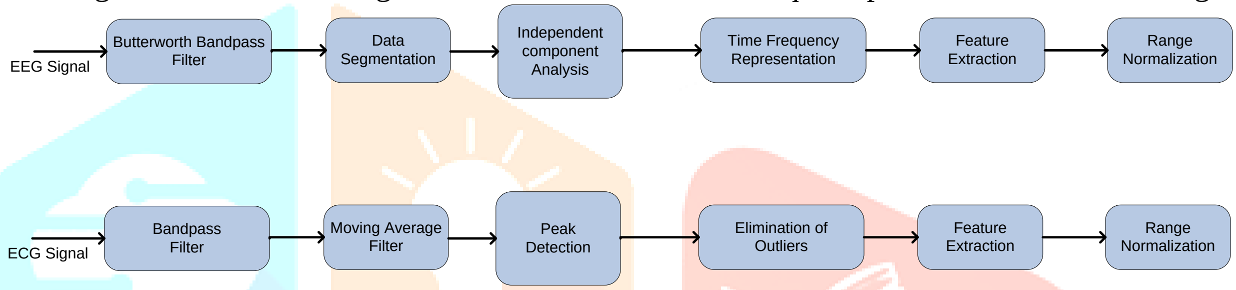

Figure 2: Block diagram of deep learning algorithm workflow.

2.4. Multimodal Systems

Multimodal systems integrate data from multiple sensing modalities, such as vision-based, wearable, and physiological sensors, to improve robustness and accuracy. By combining complementary information from different sources, these systems can overcome the limitations of individual modalities and provide more reliable recognition of human activities, gestures, and emotional states. The integration of data from multiple sources offers complementary information, which enhances the accuracy and robustness of the results. Different deep-based HAR methods using multiple visual data modalities are reviewed in [11].

III. Human Movement and Gesture Recognition Techniques

Human movement and gesture recognition have gained significant attention due to their applications in human computer interaction, healthcare, surveillance, and virtual reality. These techniques can be broadly categorized into traditional machine learning approaches, deep learning approaches, and evaluation based on benchmark datasets and performance metrics.

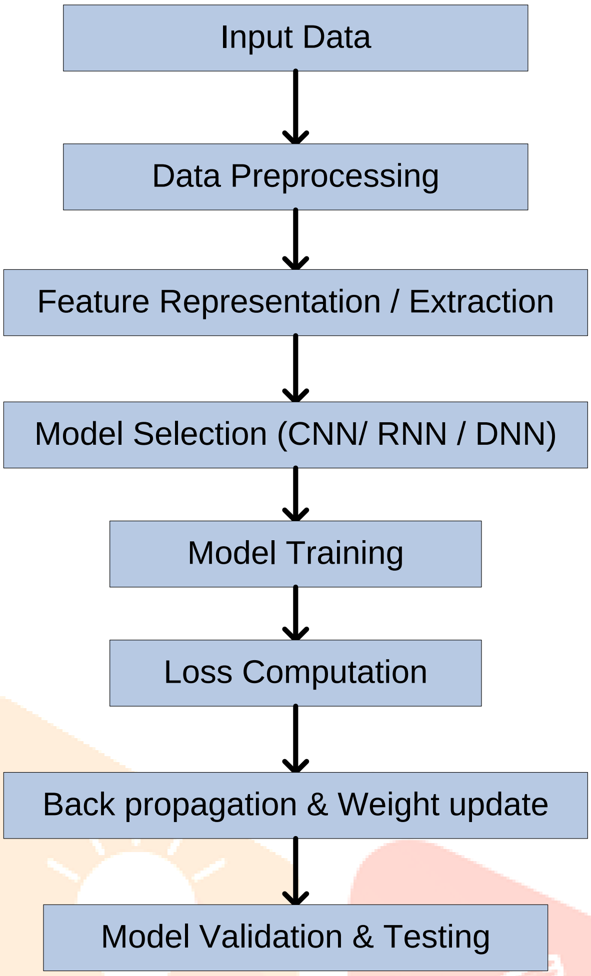

Figure 3: Block diagram of deep learning algorithm workflow.

3.1.Traditional Machine Learning Approaches

Traditional approaches rely on handcrafted feature extraction followed by classification using algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), n d Hidden Markov Models (HMM). In these traditional machine learning approaches, features such as shape, motion, trajectory, and temporal variations are extracted from sensor data or captured images. These methods are computationally efficient but often struggle with complex and dynamic gesture patterns due to limited feature representation capability [12-14].

3.2.Deep Learning Approaches

Deep learning models learn hierarchical features from raw data, significantly improving recognition accuracy, eliminate the need for manual feature extraction. The steps involved in deep learning algorithm is shown in the Figure 3. Deep learning model such as Convolutional Neural Networks (CNNs) for spatial feature extraction, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for temporal sequence modeling, and hybrid CNN-LSTM models are widely used for capturing spatial and temporal features [15], [16]. Deep learning methods perform well in complex scenarios involving variations in lighting, background, and motion. However, they require large datasets, high computational resources, and longer training times. Recent studies demonstrate that attention-based models and 3D CNN architectures further enhance performance in complex environments [17], [18]. These models are highly robust to variations in lighting, background, and motion, though they require large datasets and computational resources.

3.3.Benchmark Datasets and Evaluation Metrics

The performance of human movement and gesture recognition systems is evaluated using standard benchmark datasets and metrics. Standard datasets such as UCF101, HMDB51, and Kinetics are commonly used for evaluation. The evaluation metrics such as accuracy, precision, confusion matrix, recognition rate, processing time /latency are used. These datasets and metrics provide a standardized framework to compare different algorithms and ensure fair evaluation. The features of traditional and deep learning approaches are compared and listed in the Table I.

Table I. Comparison of Traditional Machine Learning and Deep Learning Approaches for Gesture Recognition

Aspect	Traditional Machine Learning	Deep Learning
Feature Extraction	Relies on handcrafted features such as shape, motion, texture, and skeletal descriptors	Automatically learns spatial and temporal features from raw data
Computational Complexity	Low to moderate	High due to deep network architectures
Training Time	Relatively short	Longer training time
Recognition Accuracy	Moderate; depends on feature quality	Generally higher, especially for complex gestures
Scalability	Performance may degrade with increasing gesture complexity	Scales effectively to large and complex gesture datasets
Interpretability	Easier to interpret and analyze	Limited interpretability
Typical Algorithms	SVM, KNN, Random Forest, Hidden Markov Model (HMM)	CNN, RNN, LSTM, GRU, Transformer, CNN-LSTM
Suitable Gesture Types	Static gestures and small-scale datasets	Both static and dynamic gestures in real-world environments

Traditional machine learning methods for gesture recognition depend heavily on handcrafted feature extraction and are suitable for smaller datasets and less complex tasks. Deep learning approaches automatically learn discriminative features from raw gesture data, achieving superior recognition accuracy and robustness for both static and dynamic gestures, particularly in complex real-world scenarios.

IV. EMOTION DETECTION METHODS

Emotion detection has become an important area of research in affective computing and human computer interaction, as it enables systems to better understand and respond to human emotions. Over the years, researchers have developed a variety of techniques to identify emotional states by analyzing both behavioral patterns and physiological responses. Among these, visual emotion recognition, physiological signal-based emotion detection, and multimodal emotion recognition are the most widely explored and adopted approaches.

4.1.Visual Emotion Recognition

Visual Emotion Recognition (VER) is one of the most widely studied approaches due to the availability of cameras and advances in computer vision techniques. This method analyzes facial expressions, eye movements, head poses, and body gestures to infer emotional states. Despite achieving high recognition accuracy under controlled conditions, these systems often face challenges in real-world environments due to variations in illumination, occlusions, pose changes, and cultural differences in emotional expression. The objective of Visual Emotion Recognition (VER) is to enable computers to automatically perceive, understand, and interpret human emotional states from visual cues such as facial expressions, body gestures, and behavioral patterns. By incorporating emotional intelligence into computational systems, VER enhances human–computer interaction and has found applications [19] in healthcare monitoring, surveillance systems, driver assistance systems, intelligent education platforms, social robotics, and affective computing.

4.2.Physiological Signal-Based Emotion Detection

Physiological signal-based emotion detection focuses on analyzing biological responses that reflect an individual's emotional state. Commonly used signals include Electroencephalography (EEG), Electrocardiography (ECG), Galvanic Skin Response (GSR), Electromyography (EMG), and Heart Rate Variability (HRV). Physiological responses are generally involuntary, making them less susceptible to deliberate manipulation. EEG-based emotion recognition, in particular, has gained significant attention due to its direct measurement of brain activity. Physiological signals provide a more dependable approach to emotion detection because they capture the body's automatic and unconscious responses to different emotional states. Recent advancements in wearable sensing technologies, biomedical signal processing techniques, and artificial intelligence have greatly enhanced the ability to monitor, analyze, and interpret these signals, leading to significant progress [20] in physiological signal-based emotion recognition research.

4.3.Multimodal Emotion Recognition

Multimodal Emotion Recognition (MER) integrates information from multiple modalities, such as facial expressions, speech, physiological signals, and textual content, to achieve more robust and accurate emotion recognition. Recent advances in deep learning have enabled sophisticated fusion strategies, including feature-level fusion, decision-level fusion, and attention-based multimodal fusion. Studies consistently report that multimodal approaches outperform unimodal systems in terms of recognition accuracy and reliability. Human emotions are inherently multimodal and are expressed simultaneously through facial movements, vocal characteristics, physiological responses, and behavioral patterns. MER has emerged as a key research area [21] in affective computing, human–computer interaction, healthcare, intelligent transportation, education, surveillance, and social robotics.

V. CHALLENGES AND LIMITATIONS

One of the key challenges in emotion recognition is the wide variation in how people express emotions.

5.1.Data Variability and Generalization

Emotional expressions can differ significantly based on factors such as individual personality, cultural background, age, and gender. For example, two people experiencing the same emotion may display it in very different ways, making it difficult for machine learning models to accurately recognize and generalize emotions across diverse populations. In addition, physiological signals commonly used for emotion detection, such as EEG, ECG, and GSR, vary considerably both between different individuals (inter-subject variability) and within the same individual over time (intra-subject variability). These variations can affect the consistency of extracted features and ultimately reduce the accuracy of emotion classification systems.
Another important limitation is the lack of large-scale, diverse, and well-annotated emotion datasets. Many existing datasets are collected under controlled laboratory conditions and may not adequately represent real-world emotional responses. Consequently, models trained on a specific dataset often struggle to maintain their performance when applied to new users, different environments, or unseen populations. Addressing these challenges is essential for developing more reliable, adaptable, and universally applicable emotion recognition systems.

5.2.Real-Time Processing Constraints

A major challenge in real-time emotion recognition is the need to quickly collect, process, and analyze large amounts of data from multiple sources, such as facial expressions, speech, and physiological signals. To accurately identify emotions, the system must perform tasks such as signal preprocessing, feature extraction, and classification within a very short time frame. However, these operations—especially when using deep learning models—often require substantial computational power and memory, which can lead to processing delays and increased energy consumption.

The challenge becomes even greater in wearable devices, smartphones, and embedded systems, where hardware resources and battery life are limited. Running complex emotion recognition algorithms on such platforms can affect system responsiveness and reduce operational efficiency. Therefore, developing lightweight and efficient models that can deliver accurate emotion recognition with minimal computational overhead is an important area of research. Balancing high recognition accuracy with low latency and energy consumption remains a critical requirement for the successful deployment of real-time emotion recognition systems in practical applications.

5.3.Environmental and Contextual Factors

The effectiveness of emotion recognition systems is often influenced by environmental and contextual factors. In real-world situations, conditions such as poor lighting, facial occlusions (e.g., masks or glasses), variations in head pose, background noise, and differences in camera quality can negatively impact the accuracy of visual and speech-based emotion recognition methods. These factors make it more difficult for systems to correctly interpret emotional cues.

Another important challenge is that emotions are highly dependent on context. The same facial expression, voice tone, or physiological response may represent different emotions depending on the surrounding situation, social setting, or individual circumstances. For example, an increased heart rate may indicate excitement, fear, or physical exertion. Without sufficient contextual information, emotion recognition systems may misinterpret these signals. Therefore, incorporating contextual awareness and improving robustness to environmental variations are essential for building reliable and accurate emotion recognition systems for real-world applications.

VI. FUTURE RESEARCH DIRECTIONS

As emotion recognition technologies continue to evolve, several research challenges and opportunities remain. Future work should focus on improving the scalability, robustness, interpretability, and ethical deployment of these systems in real-world environments.

6.1.Development of lightweight, real-time models for edge deployment

Future research in emotion recognition should focus on developing lightweight and energy-efficient models that can operate effectively on edge devices such as smartphones, wearable sensors, smart glasses, and other embedded platforms. These devices often have limited processing power, memory, and battery capacity, making it challenging to deploy complex deep learning models directly on them.

To address this issue, researchers are exploring techniques such as model compression, network pruning, quantization, and knowledge distillation. These approaches help reduce the size and computational requirements of machine learning models while preserving their performance and recognition accuracy. By optimizing models for resource-constrained environments, emotion recognition systems can process data more quickly and consume less power.

Advances in lightweight model design will play a crucial role in enabling real-time emotion recognition for practical applications, including healthcare monitoring, driver assistance systems, smart environments, and human–computer interaction. Such developments will support the widespread adoption of emotion-aware technologies in everyday life while ensuring efficient and reliable operation on portable and wearable devices.

6.2.Improved cross-subject and cross-environment generalization

A major limitation of many current emotion recognition systems is their difficulty in adapting to new users and unfamiliar environments. Models are often trained on specific datasets collected under controlled conditions, and their performance may decline when used with individuals who differ in age, gender, cultural background, or physiological characteristics. Environmental factors such as lighting conditions, background noise, sensor quality, and recording settings can further impact recognition accuracy.

To overcome these challenges, future research should focus on developing models that can generalize effectively across diverse populations and real-world situations. Techniques such as domain adaptation and transfer learning can help models transfer knowledge from one dataset or environment to another, while self-supervised learning can reduce dependence on large amounts of labeled data. Federated learning also offers a promising approach by enabling models to learn from data distributed across multiple devices while preserving user privacy.

In addition, the creation of large-scale, diverse, and representative datasets is essential for improving the robustness and reliability of emotion recognition systems. By incorporating data from people with different backgrounds and capturing a wide range of real-world conditions, future models can become more adaptable, accurate, and suitable for practical deployment across various applications.

6.3.Emotion-aware gesture recognition

Emotion-aware gesture recognition is emerging as a promising research direction in human-centered sensing and intelligent interaction systems. While conventional gesture recognition focuses on identifying physical movements and actions, human gestures often convey valuable emotional information that reflects an individual's feelings, intentions, and psychological state. Integrating emotional understanding into gesture recognition enables machines to interpret human behavior more comprehensively and respond in a manner that is both natural and contextually appropriate.

Future research should explore the integration of gesture dynamics, body posture, facial expressions, speech cues, and physiological signals such as EEG, ECG, and GSR to create comprehensive multimodal emotion recognition frameworks. Advanced deep learning architectures, including Transformers, Graph Neural Networks (GNNs), and multimodal fusion networks, can be utilized to capture the complex relationships between gestures and emotions. By enabling machines to recognize not only what actions individuals perform but also the emotions associated with those actions, emotion-aware gesture recognition has the potential to significantly enhance the effectiveness, personalization, and empathy of next-generation intelligent systems.

6.4.Integration with AR/VR and digital twin environments

The integration of emotion recognition technologies with Augmented Reality (AR), Virtual Reality (VR), and digital twin environments is emerging as a promising area of research. As immersive technologies become more widely used in education, healthcare, industrial training, entertainment, and remote collaboration, there is an increasing demand for systems that can understand users' emotional states and respond to them in real time.

By incorporating physiological signal-based emotion recognition into AR and VR platforms, virtual environments can become more adaptive and personalized. Information derived from signals such as EEG, ECG, and GSR can provide insights into a user's emotional state, stress level, engagement, or cognitive workload. This enables virtual systems to adjust their behavior according to the user's needs. For example, a training simulation may reduce task complexity when signs of frustration are detected, while a therapeutic VR application may modify its content to promote relaxation or emotional well-being. Such adaptive interactions can improve user engagement, learning effectiveness, comfort, and overall experience.

Future research should focus on developing seamless multimodal sensing frameworks that combine physiological signals, facial expressions, speech, gestures, and contextual information for more accurate emotion recognition. In addition, low-latency processing techniques and efficient data synchronization methods will be essential to enable smooth real-time interaction between users and virtual environments. Researchers must also address important challenges related to data privacy, security, interoperability, scalability, and the ethical use of emotional information.

Overall, the convergence of emotion recognition, AR/VR technologies, and digital twins has the potential to create highly immersive, adaptive, and human-centered systems. By enabling virtual environments to understand and respond to human emotions, these technologies can foster more natural interactions and help bridge the gap between the physical and digital worlds.

6.5.Ethics and privacy in human-centered sensing systems

As human-centered sensing technologies become increasingly integrated into everyday life, addressing ethical and privacy concerns has become a critical research priority. These systems often collect and analyze highly sensitive data, including physiological signals, emotional states, behavioral patterns, and personal interactions. While such information can enable personalized and intelligent services, it also raises important questions regarding data ownership, consent, security, and responsible use.
Another key challenge is addressing potential biases in emotion recognition and sensing algorithms. Variations in age, gender, cultural background, and individual physiological characteristics can affect system performance, potentially leading to unfair or inaccurate outcomes. Researchers must therefore prioritize fairness, inclusivity, and transparency when designing and evaluating human-centered sensing systems.
Explainable Artificial Intelligence (XAI) is expected to play an important role in improving the interpretability of sensing and decision-making processes. By providing understandable explanations for system outputs, XAI can enhance user confidence and support ethical deployment in sensitive domains such as healthcare, education, and workplace monitoring.
Ultimately, future human-centered sensing systems should be designed with a strong emphasis on ethical principles, user autonomy, data protection, and regulatory compliance. Balancing technological innovation with privacy and ethical responsibility will be essential for creating trustworthy, socially acceptable, and sustainable intelligent systems.

VII. CONCLUSIONS

This paper presented a comprehensive review of computational and sensor-based approaches for human movement analysis, gesture recognition, and emotion detection. The study examined a wide range of sensing modalities, including vision-based systems, wearable sensors, inertial measurement units, and physiological signal acquisition technologies, highlighting their roles in capturing and interpreting human behavior. Furthermore, recent advances in machine learning, deep learning, and multimodal data fusion techniques were discussed, demonstrating their significant contributions to improving recognition accuracy, robustness, and real-time performance.

The review revealed that integrating multiple sensing modalities enables a more comprehensive understanding of human actions, intentions, and emotional states than single-modality approaches. Such multimodal frameworks have shown considerable potential in applications including healthcare monitoring, rehabilitation, assistive technologies, human-computer interaction, smart education, intelligent transportation, social robotics, and immersive AR/VR environments.

Despite substantial progress, several challenges remain, including inter-subject variability, environmental sensitivity, data scarcity, computational complexity, privacy concerns, and the need for explainable and trustworthy AI models. Addressing these challenges will be essential for the widespread adoption of human-centered intelligent systems in real-world settings.
Future research is expected to focus on emotion-aware gesture recognition, multimodal learning frameworks, edge and real-time deployment, integration with AR/VR and digital twin environments, and privacy-preserving artificial intelligence techniques. As sensing technologies and AI methodologies continue to evolve, human movement analysis, gesture recognition, and emotion detection systems will become increasingly adaptive, personalized, and context-aware, enabling more natural and effective interactions between humans and intelligent machines. Ultimately, these advancements will contribute to the development of next-generation human-centered systems that enhance quality of life, improve decision-making, and support a wide range of societal and industrial applications.

REFERENCES

[1] P. K. Nag, A. Bhagat and R. Vishnu Priya, "Expanding AI's Role in Healthcare Applications: A Systematic Review of Emotional and Cognitive Analysis Techniques," in IEEE Access, vol. 13, pp. 69129-69160, 2025, doi: 10.1109/ACCESS.2025.3562131.

[2] Pereira, R.; Mendes, C.; Ribeiro, J.; Ribeiro, R.; Miragaia, R.; Rodrigues, N.; Costa, N.; Pereira, A. Systematic Review of Emotion Detection with Computer Vision and Deep Learning. Sensors 2024, 24, 3484. https://doi.org/10.3390/s24113484.

[3] M. A. Mirzaee and A. Sadighi, "Design and Fabrication of a Vision-based Tactile Sensor," 2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, Islamic Republic of, 2023, pp. 562-567, doi: 10.1109/ICRoM60803.2023.10412595.

[4] M. A. Mirzaee, H. -J. Huang and W. Yuan, "GelBelt: A Vision-Based Tactile Sensor for Continuous Sensing of Large Surfaces," in IEEE Robotics and Automation Letters, vol. 10, no. 2, pp. 2016-2023, Feb. 2025, doi: 10.1109/LRA.2025.3527306.

[5] Aparna K, Dayajanaki H. D, Devika Rani P, Devu D, Sreenidhi Prabha Rajeev and Baby Sreeja S D, "Wearable Sensors in Daily Life: A Review," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 863-868, doi: 10.1109/ICACCS57279.2023.10112956.

[6] G. Taskasaplidis, D. A. Fotiadis and P. D. Bamidis, "Review of Stress Detection Methods Using Wearable Sensors," in IEEE Access, vol. 12, pp. 38219-38246, 2024, doi: 10.1109/ACCESS.2024.3373010.

[7] B. P. Singh, P. Baxi, M. Parhi, G. Ahluwalia, S. B. Sollapur and S. DIVAKARAN, "Human Activity Recognition Using Wearable Sensors in IoT Environments: Challenges and Opportunities," 2025 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2025, pp. 447-452, doi: 10.1109/AUTOCOM64127.2025.10957385.

[8] Ö. GÜnaydin and R. B. Arslan, "Stress Level Detection Using Physiological Sensors," 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 2020, pp. 509-512, doi: 10.1109/BIBE50027.2020.00088.

[9] V. H. Ashwin, R. Jegan and P. Rajalakshmy, "Stress Detection using Wearable Physiological Sensors and Machine Learning Algorithm," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 972-977, doi: 10.1109/ICECA55336.2022.10009326.

[10] A. Prabakaran and E. Rufus, "Design of a Multi-Sensor Wearable Physiological Parameters Monitoring System Using Machine Learning Algorithm," 2024 International Conference on Smart Electronics and Communication Systems (ISENSE), Kottayam, India, 2024, pp. 1-6, doi: 10.1109/ISENSE63713.2024.10872213.

[11] Shafizadegan, F., Naghsh-Nilchi, A.R. & Shabaninia, E. Multimodal vision-based human action recognition using deep learning: a review. Artif Intell Rev 57, 178 (2024). https://doi.org/10.1007/s10462-024-10730-5.

[12] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311-324, May 2007, doi: 10.1109/TSMCC.2007.893280.

[13] R. Poppe, “A survey on vision-based human action recognition,” Image and Vision Computing, vol. 28, no. 6, pp. 976–990, 2010.

[14] K. Deshpande, V. Mashalkar, K. Mhaisekar, A. Naikwadi and A. Ghotkar, "Study and Survey on Gesture Recognition Systems," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCUBEA58933.2023.10392214.

[15] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, “Action recognition in video sequences using deep bi-directional LSTM,” IEEE Access, vol. 6, pp. 1155–1166, 2018.

[16] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” Advances in Neural Information Processing Systems, 2014.

[17] Yong Du, W. Wang and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1110-1118, doi: 10.1109/CVPR.2015.7298714.

[18] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[19] Sze Chit Leong, Yuk Ming Tang, Chung Hin Lai, C.K.M. Lee, “Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing”, Computer Science Review, Volume 48, 2023, 100545, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2023.100545.

[20] Ahmad Z, Khan N. A Survey on Physiological Signal-Based Emotion Recognition. Bioengineering (Basel). 2022 Nov 14;9(11):688. doi: 10.3390/bioengineering9110688. PMID: 36421089; PMCID: PMC9687364.

[21] Udahemuka, Gustave, Karim Djouani, and Anish M. Kurien. 2024. "Multimodal Emotion Recognition Using Visual, Vocal and Physiological Signals: A Review" Applied Sciences 14, no. 17: 8071. https://doi.org/10.3390/app14178071