Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 1 (2025)

Comprehensive Survey on Recognition of Emotions from Body Gestures

[1]Ramakrishna Gandi, [2]Dr.A.Geetha, [3]Dr.B.Ramasubba Reddy

[1]Research Scholar,Computer Science & Engineering Department,Annamalai University, Annamalainagar,Tamil Nadu,608002,INDIA
E-mail: gandiramakrishna2@gmail.com
[2]Professor,Computer Science & Engineering Department,Annamalai University, Annamalainagar,Tamil Nadu,608002,INDIA
E-mail:aucsegeetha@yahoo.com
[3]Professor, Department of CSE, MOHAN BABU UNIVERSITY,Sree Sainath Nagar,Tirupati,Andhra Pradesh,517102,INDIA
E-mail:rsreddyphd@gmail.com

ABSTRACT:

Automatic emotion identification has emerged as a prominent area of research during the past decade, with applications in healthcare, human-computer interaction, and behavioral analysis. Although facial expressions and verbal communication have been thoroughly examined, the identification of emotions via body gestures is still inadequately investigated. Body gestures, an essential aspect of "body language" offer significant contextual indicators shaped by gender and culture variations. Recent breakthroughs in deep learning have facilitated the development of robust models capable of accurately capturing complex human movements, hence enhancing emotion recognition precision and adaptability. This study presents a thorough framework for the automatic recognition of emotional body gestures, encompassing essential elements such as individual detection, position estimation, and representation learning. High computational costs and the need for advanced algorithms to fuse multimodal data add to these hurdles. Recent advancements in deep learning, have shown great potential to overcome these issues and improve accuracy. This work highlights the applications, challenges, and future directions in emotion recognition from body gestures, emphasizing the need for scalable, robust, and real-world-ready systems that can enable emotionally intelligent technologies.

KEYWORDS: Body gestures, Emotional recognition, Deep learning, Human-computer interaction, Behavioral analysis

INTRODUCTION

Human-computer interaction (HCI) has advanced dramatically with the introduction of machine learning (ML) and deep learning (DL) technologies, to develop more intuitive techniques for machines to understand and respond to human needs. Emotion identification is critical in this interaction because it allows systems to identify human emotions using a variety of modalities such as facial expressions, voice intonations, and, as recent research has highlighted, body gestures. Historically, emotion detection algorithms have focused primarily on facial expressions, mapping minute facial movements to widely recognized emotions including anger, disgust, fear, happiness, sorrow, surprise, and neutrality [1]. However, this technique ignores the huge array of nonverbal messages given by body language. Body language, which includes gestures, posture, and eye movements, indicates an individual's emotional and cognitive states. This study examines the literature on the automatic identification of emotional body expressions, with a focus on gestures and postures. Figure 1 depicts a handful of the most prevalent body signals, with images from reference [2]. The multidimensional approach to emotions acknowledges their complexity, influenced by elements such as personal experiences, culture, and individual characteristics.

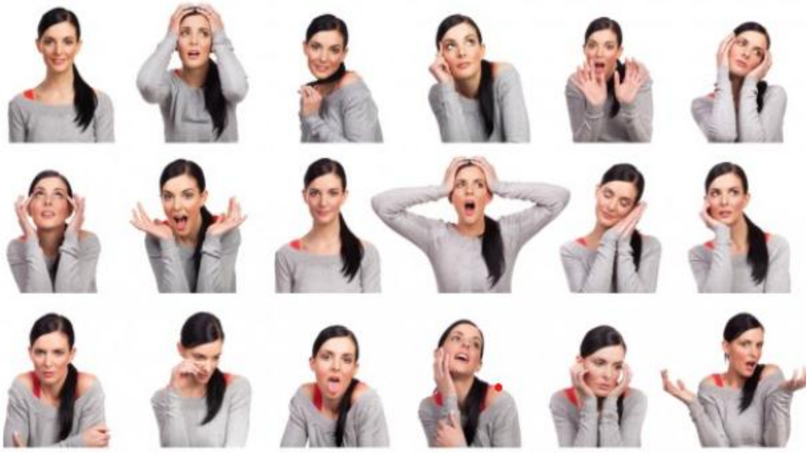

Figure 1: Nonverbal indications and expressions from the human.

It provides a framework for comprehending emotional states by employing models such as the 2D and 3D emotional spaces. The 2D model classifies emotions based on their valence (positive or negative) and arousal (high or low activation). Russell's 2D emotional space model, seen in Figure 2, shows emotions mapped along these two dimensions [3, 4].

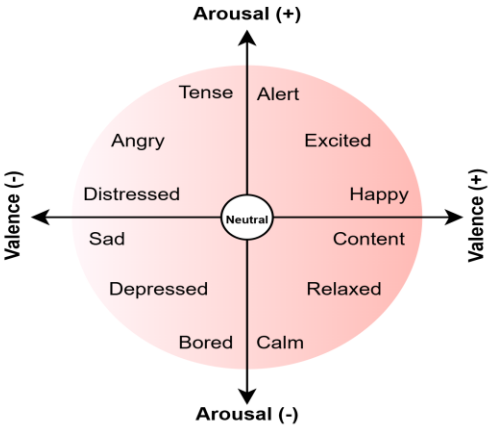

Figure 2: 2D VA Emotional model

Studies in affective computing suggest that body gestures may offer equal or even greater insights into an individual's emotional state, particularly in situations where facial expressions are unclear or suppressed due to social contexts. These non-facial cues enhance the accuracy of emotion detection algorithms and provide valuable context, making emotion recognition critical in domains such as healthcare, marketing, human-robot interaction, mental health monitoring, and security. Non-verbal communication, including gestures, facial expressions, and bodily movements, constitutes a fundamental aspect of human interaction. Research indicates that spoken communication accounts for only 35% of the total message, while non-verbal signals such as gestures and postures comprise 65%. This highlights the pivotal role of body language in bridging verbal communication and understanding its underlying emotional intent. Gestures, in particular, are rich in emotional significance and play an essential role in social relationships.

Despite the importance of body gestures in human communication, much of the research on emotion recognition for intelligent systems has focused on facial expressions and speech. The existing research highlights that body gestures provide substantial emotional information often overlooked in favor of facial and vocal cues. This gap is illustrated in Figure 3, which categorizes emotional cues into verbal and non-verbal methods. Verbal cues include vocal tone and tactile interaction, while non-verbal cues encompass facial expressions and bodily gestures. Body gestures can be further classified into illustrative gestures and micro-gestures, both of which convey subtle emotional nuances. Incorporating body gestures into emotion recognition algorithms is essential for enabling intelligent systems to replicate human emotional understanding more accurately. Expanding the scope of emotional signals assessed, beyond facial expressions and speech, can significantly enhance the emotional intelligence and adaptability of robots and intelligent agents. This broader approach can improve their effectiveness in interpreting and responding to human emotions, ultimately advancing applications in healthcare, human-robot collaboration, and beyond.

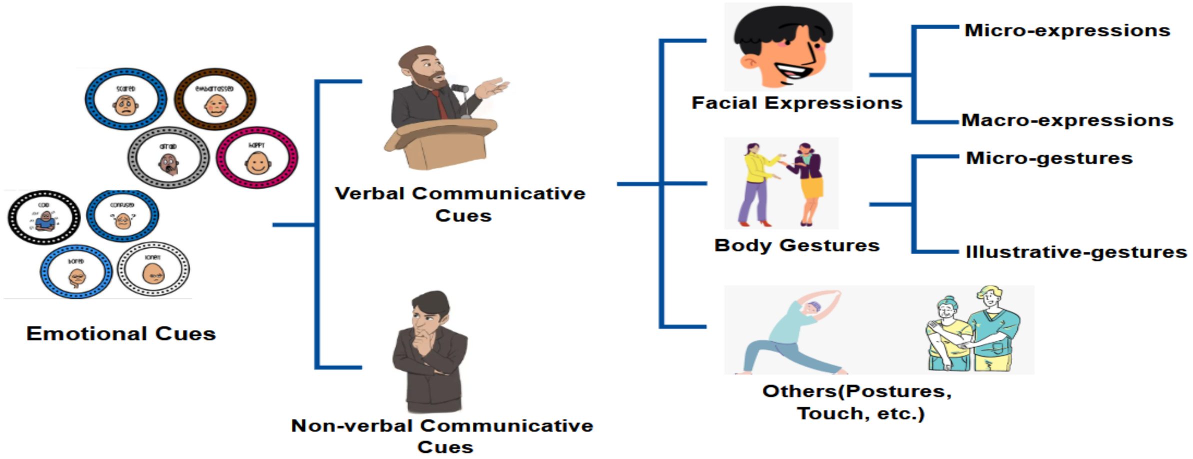

Figure 3: Micro-gestures as one of the emotional cues.

The systematic examination of body language originated in the 19th century, with Charles Darwin's influential publication, The Expression of the Emotions in Man and Animals, establishing the groundwork for contemporary research in this field. Darwin noted that facial expressions and bodily motions exhibit universal similarities throughout cultures. In 1978, Paul Ekman established the Facial Action Coding System (FACS), which continues to be fundamental for comprehending emotional expressions. Although facial expressions and vocal communication have been thoroughly examined for emotion identification, bodily gestures a crucial component of nonverbal communication have traditionally garnered less focus [4].

Gestures, including movements of the hands, head, and other body parts, are fundamental to communication, allowing individuals to express a diverse array of emotions, thoughts, and attitudes. Specific movements, such as nodding to denote agreement or turning away to show rejection, are either instinctual or acquired in early life. Furthermore, certain movements, such as dilated nostrils in moments of enthusiasm or stress, may represent evolutionary vestiges of natural selection. Gestures such as smiling in happiness or frowning in distress are universally recognized, enhancing their significance in emotion recognition [5]. Recent advancements in motion capture technology and the dependability of body gesture analysis have intensified interest in automated emotion identification systems. Although prior research mostly concentrated on facial expressions and verbal communication, breakthroughs in deep learning have facilitated the creation of sophisticated models capable of accurately interpreting complex body movements. These models have substantially advanced domains such as healthcare, human-computer interaction, and behavioral analysis by providing improved adaptability and precision [6].

1. Automatic Emotion Recognition Systems

Figure 4 presents a detailed outline of the procedures involved in an automated emotion detection system, illustrating the progression from data collecting to the ultimate emotion classification and assessment. Each block signifies a critical phase in the process, methodically converting raw data into useful information regarding emotional states [7].

(i) Source and stimuli

The procedure begins with the collection of data, which includes physical and physiological signals including EEG (brain activity), ECG (heart activity), GSR (skin conductance), facial expressions, eye tracking, and speech. These sources document different aspects of human reactions to emotional stimuli. External stimuli that elicit emotions in individuals include virtual reality, visuals, video games, music, and audio/video clips. These stimuli are carefully selected to trigger specific emotions, which are then quantified with instruments such as questionnaires or rating systems.

(ii) Input Signals and Preprocessing

The input signals from the sources are pre-processed to improve data quality and remove background noise. This stage entails removing artifacts, combining data from numerous sources, and scaling or modifying the data before analysis. Pre-processing guarantees that the signals are consistent and ready for efficient feature extraction and modeling, which is essential for successful emotion recognition.

(iii) Feature extraction and selection

After the data has been preprocessed, the next step is to extract useful features. This entails discovering patterns or qualities in the data that are most useful for understanding emotions, such as frequency-domain features, entropy features, or deep learning-based features. Following extraction, feature selection narrows down the most informative characteristics, minimizing redundancy while boosting model efficiency and interpretability. Dimensionality reduction and statistical analysis are useful techniques for identifying optimal features.

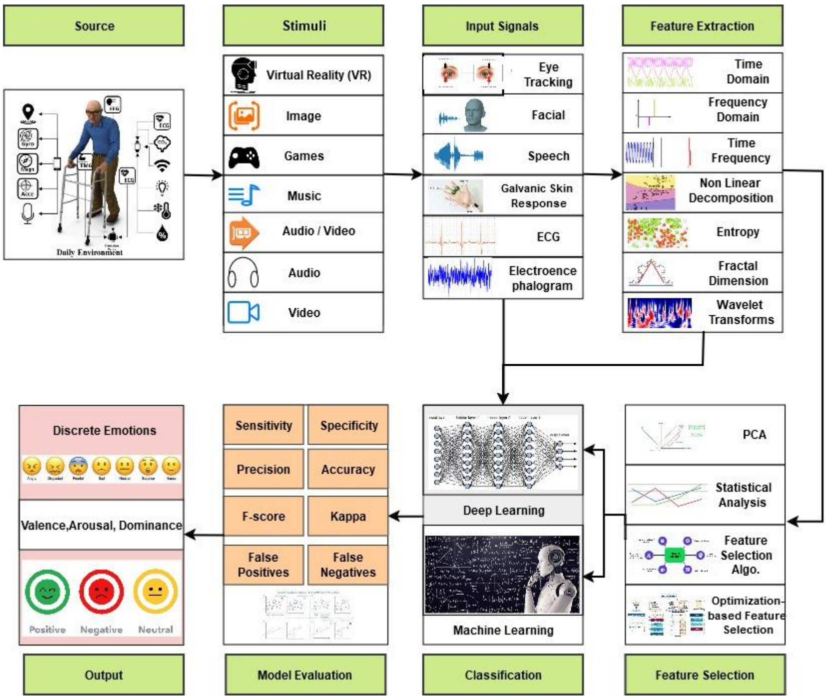

Figure 4: The steps involved in a conventional automated emotion recognition system

(iv) Classification and Model Evaluation

The classification stage divides the processed data into emotion classes using machine learning (SVM, decision trees) or deep learning models (CNNS, RNNs, transformers). These models employ the chosen features to forecast emotions like happiness, sadness, and rage. Finally, the system is assessed using measures such as accuracy, precision, recall, and F1-score to confirm its dependability and efficacy. This stage examines how successfully the system generalizes to previously unseen data, indicating its suitability for real-world applications.

Figure 5 depicts the process of emotion recognition using body motions, from input data collecting to final emotion detection. The procedure begins with Human Detection, in which individuals are identified with visual input utilizing RGB and depth cameras. This stage positions the person within the frame, ensuring precise focus on the subject of interest. Body Pose Estimation then uses part-based or kinematic methodologies to describe the human body structure, identifying crucial locations such as joints and limbs. These pose estimations are then processed in the Feature Extraction stage, which generates either pre-designed features (e.g., geometric or motion-based properties) or learned features (e.g., deep learning embeddings) to accurately depict body movements.

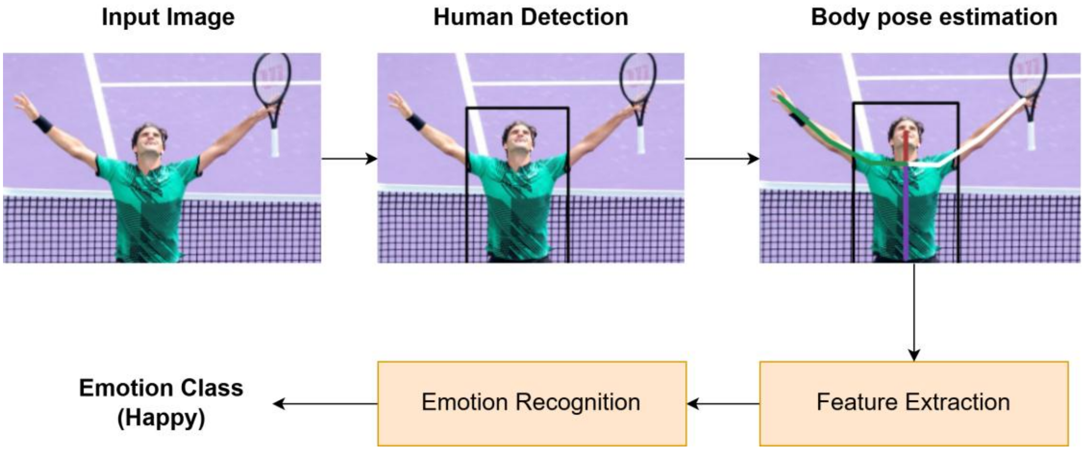

Figure 5: Emotion recognition using input image with body motion

In the following stage, the system uses Emotion Recognition techniques to identify the emotional state using models of emotion such as categorical (e.g., joyful, sad), dimensional (e.g., valence-arousal), or componential. The result displays the recognized emotion, such as "Happy," based on the retrieved features and pose analysis. This systematic workflow shows how body motions are processed to accurately understand emotional states.

2. Benchmark Datasets

The availability of diverse and well-structured datasets is crucial for furthering research on body gesture-based emotion identification. These datasets are critical resources for training and assessing models, allowing the exploration of neutral and emotive gestures in a variety of scenarios. They serve several research aims, including multi-modal integration, dynamic gesture analysis, and culturally diverse emotion recognition, by providing comprehensive annotations and samples.

Table 1 lists body gesture-based emotion recognition datasets. It highlights each dataset's facial, torso, or skeleton body component, emotional categories, individuals, and sample size. FABO, GEMEP, and HUMAINE focus on face and body movements, whereas LIRIS-ACCEDE and POSEC3D focus on upper-body motions and full-body skeletal data. GEMEP and BP4D+ provide extensive face expression, bodily motion, and physiological signal data for multi-modal emotion identification. Due to their concentration on body gestures, theatre, and EMELYA datasets are appropriate for investigations on body movement as the major emotional trigger. Human3.6M and EMOTIC, with their bigger samples and improved context coverage, are appropriate for big data-driven deep learning models. This range lets researchers choose datasets for multimodal integration, dynamic gesture analysis, or culturally varied emotional recognition.

Based on related body language, Table 2 lists the overall movement patterns for six main emotions—fear, anger, sadness, surprise, happiness, and disgust. It emphasizes how emotions show themselves through certain nonverbal signals including posture, gestures, and movement patterns, so offering a clear guide for comprehending emotional manifestations through body behavior. Studies in emotional recognition and nonverbal communication analysis would find great value in this reference.

Figure 6 shows examples of RGB images depicting diverse body motions divided into two categories: neutral gestures and emotional gestures. Neutral gestures are motions like jumping, squatting, throwing, receding, and walking away that are performed without any emotional intention. These motions are frequently employed to record overall body movements in a neutral state. In contrast, emotional gestures portray distinct emotional states. For example, happiness is shown with outstretched hands in a celebratory gesture, but melancholy is conveyed with a hunched posture and a hand covering the face. The photographs illustrate the variety of human body motions and their relationship to emotional states, emphasizing body gestures' potential as a medium for automated emotion recognition. This dataset is a significant resource for exploring the relationship between gestures and emotions, which will help to design robust recognition systems.

Table 1: List of Body Gesture Datasets for Emotion Recognition

Name	**Body Parts**	**Emotions**	**Subjects**	**Samples**
EMOTIC	Body and context	Context-dependent	~23,000 images	~23,000

LIRIS-ACCEDE	Face and upper body	6	64	NA
HUMAINE	Face and Body	8	10	240
THEATER	Body	8	NA	NA
GEMEP	Face and Body	18	10	>7,000
EMILYA	Body	8	11	NA
FABO	Face and Body	10	23	206
POSEC3D	3D skeletal	Context-dependent	~1,500 sequences	~1,500
AFEW	Face and Body	7	>1,800 clips	>1,800
SIG	Body	Context-dependent	1,000 videos	~1,000
CMU Multimodal	Skeletal and body	Context-dependent	Several hours	Several hours
BP4D+	Face, body, and physiology	Context-dependent	>40 participants	NA
ChAirGest	Upper body	Context-dependent	~200 videos	~200
Human3.6M	Skeletal and body	Context-dependent	~3.6M poses	~3.6M
ED-GE	Body	8	~500 videos	~500
SBU Interaction	Body	Context-dependent	~300 sequences	~300

Table 2: Six basic emotions and associated body language [8]

Emotion	Body Language
Fear	Increased heart rate, tense body, arms crossed or wrapped, bouncy or restrained movements, shallow breathing.
Anger	Widened stance, hands on hips or fists, clenched jaw, pointing or shaking finger, aggressive posture.
Sadness	Drooped shoulders, head down, leaning forward, covering face with hands, slow or minimal movement.
Surprise	Sudden backward movement, hands to face or head, head shaking, abrupt posture change.
Happiness	Arms open, relaxed posture, legs apart, smiling, eye contact, pointing with interest.
Disgust	Backing away, hands covering neck or mouth, turning head, body shifting, avoiding contact.

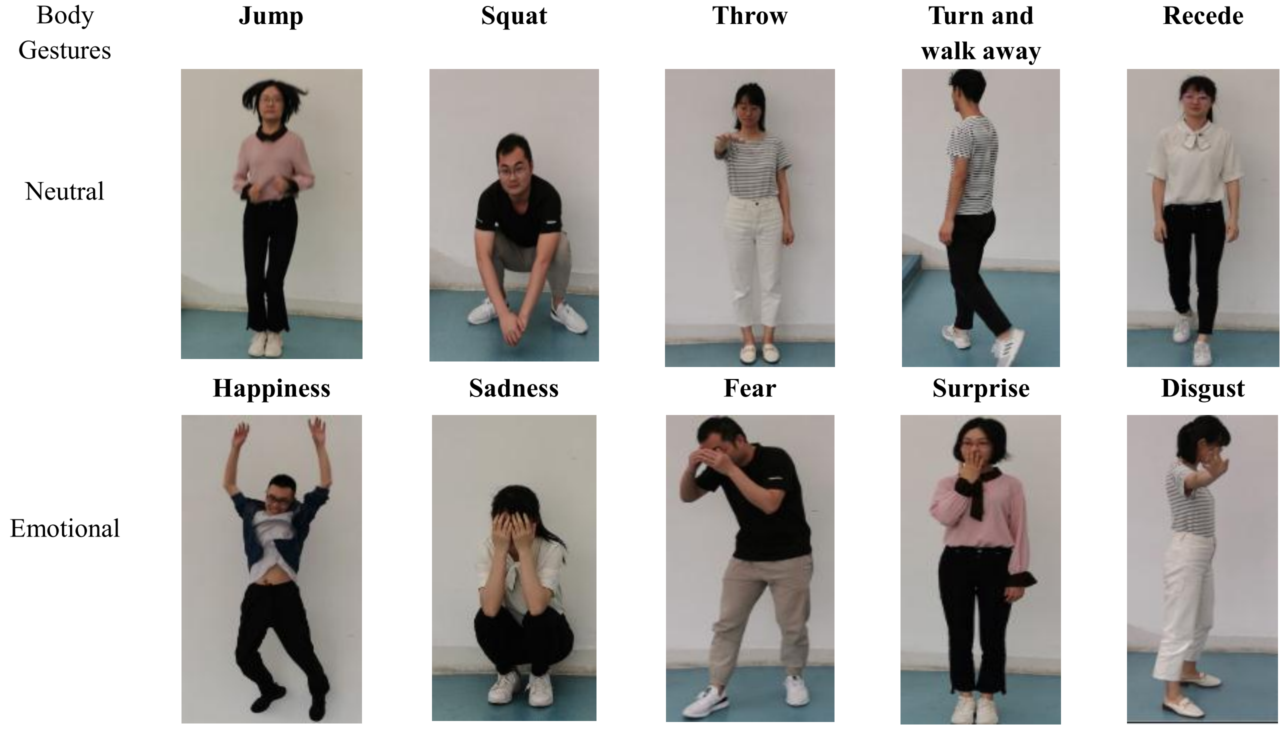

Figure 6: Sample images of body gestures

3. Applications and Challenges

4.1 Applications

Emotion recognition via body motions is a rapidly expanding topic with numerous applications in a variety of domains. One important application is in healthcare, where emotion detection systems can monitor patients' emotional well-being and stress levels, particularly during mental health therapy and rehabilitation. These systems are also used to help people with autism understand and interact with social cues.

Body gesture recognition helps security and surveillance personnel recognize potentially aggressive or violent behavior, boosting public safety. Similarly, emotion detection can improve communication in video conferencing and telephony by providing insights into participants' emotional states, resulting in more effective interactions in remote contexts. Furthermore, the entertainment and virtual reality businesses use this technology to generate emotionally responding virtual agents and avatars, which enhance user experiences. Moreover, emotion recognition is crucial for effective human-machine interaction. Systems that integrate bodily movements with additional modalities, such as facial expressions and vocalizations, provide perceptual user interfaces for e-commerce and virtual assistants. These systems may respond to emotional cues, providing personalized replies and increasing user engagement. In education, emotion-aware technologies are being created to detect learners' emotional states and provide timely feedback to improve motivation and learning outcomes.

4.2 Challenges of Recognising Emotions via Body Gestures

Despite its enormous promise, emotion identification via body motions confronts numerous hurdles. One key challenge is the scarcity of large-scale, diversified datasets that include different cultures, genders, and age groups. Body motions expressing the same emotion can differ greatly between cultures, limiting the applicability of existing models. Furthermore, these datasets frequently fail to replicate real-world circumstances, making it challenging to train models that function effectively in changing environments. Another problem is the dynamic and nuanced nature of body motions, such as micro-gestures or fleeting movements, which require high-resolution sensors and advanced algorithms to accurately recognize. Real-time recognition adds another layer of difficulty by requiring systems to analyze huge amounts of input efficiently while retaining high accuracy. High computing costs and scalability concerns are significant restrictions, especially in resource-constrained settings. Furthermore, multi-modal integration is underexplored in many systems. Body motions give important emotional information; however, integrating them with facial expressions, voice intonations, and physiological signals can increase accuracy. However, synchronizing and interpreting data from many modalities poses major technical problems. Furthermore, real-world variables like as occlusions, lighting conditions, and ambient noise can impair system performance.

To address these issues, the researchers need to create standardized datasets, more robust and scalable models, and improved multi-modal fusion methods. By addressing these constraints, emotion recognition from body movements can reach its full potential in healthcare, security, education, and other fields, resulting in more intuitive and emotionally intelligent systems.

4. Literature Review

ML and DL have transformed emotion recognition by allowing systems to analyze and assess complex data with enhanced accuracy. These tactics have proven crucial in understanding human emotions through several modalities, including facial expressions, vocal tone, body language, and physiological signals. Physical movements might disclose nonverbal emotional cues that facial expressions and verbal interactions could overlook. Computer vision and machine learning have empowered researchers to identify emotions through physical movements, enhancing emotion detection systems.

Deep learning, especially models like Convolutional Neural Networks (CNNs), has demonstrated remarkable efficacy in the automatic extraction and classification of features from large datasets. Multi-view RGB video datasets enable CNNs to interpret body motions in dynamic circumstances, capturing nuanced emotional transitions. Moreover, hybrid architectures that integrate CNNs with RNNs or Transformers are progressively utilized to articulate temporal connections in gestures, hence enhancing the precision of emotion identification. Techniques such as 2D and 3D pose estimation, combined with multi-modal data integration (e.g., mixing movements with facial and voice cues), have also gained popularity, illustrating the effectiveness of deep learning in analyzing complicated emotional patterns.

Support Vector Machines (SVMs) remain an excellent alternative for traditional machine learning-based emotion recognition, especially when datasets are small and simple. However, DL models outperform classic ML methods in terms of unstructured data processing and deep hierarchical feature capture. Recent research emphasizes the importance of large, labeled datasets for effectively training deep learning models, as well as the integration of multi-modal signals to achieve robust performance across a wide range of applications, including psychological analysis, interactive gaming, assistive technologies, and security systems. Researchers are taking substantial steps towards more lifelike and complete emotion identification systems by focusing on the possibilities of body motions and advances in deep learning.

Emotion recognition from body motions is primarily reliant on good feature extraction and classification approaches that map human motion to match emotional states. Most approaches use geometrical representations to extract features from bodily parts such as joints, heads, hands, and torso. Common characteristics include displacement, rotation, velocity, acceleration, and silhouette-based measurements like Quantity of Motion, and Contraction Index[9]. Dynamic characteristics, such as velocity and acceleration, are especially useful because they capture the temporal intricacies of body movement, resulting in greater identification rates when paired with static features [10].

Hand gestures have garnered significant attention in emotion recognition due to their rich expressive content. Palm orientation, hand shape, motion trajectory, and joint configurations (elbow, wrist, shoulders) are often examined on the body for reference. Contemporary techniques, including 3D kinematics of monitored joints and Fourier processing using PCA, enhance the dependability of retrieved characteristics. The features are further examined using advanced machine learning and deep learning models, such as Random Forest, Neural Networks, and DL architectures, which have shown superior performance in classifying emotions from gestures [11].

The efficacy of emotion recognition systems is contingent upon the number of emotion categories and the classifier employed. Reducing emotion classes simplifies the output space and increases classification accuracy. Ensemble-based classifiers, such as RF, outperform Naïve Bayes and SVM models on specific datasets [12]. Similarly, Deep Learning models like multichannel CNNs and Spiking Neural Networks have shown strong performance, with recognition rates over 90% in some circumstances. These findings highlight the growing demand for deep learning architectures because of their capacity to handle complicated spatiotemporal information and learn representations effectively.

Strong testing procedures are also necessary to ensure the dependability of emotion recognition systems. While several systems achieve excellent accuracy (e.g., 93% for five emotion classes using Neural Networks), testing across a wide range of data settings, including lighting, backdrop, and motion dynamics, is critical for generalization. Future research should look into using unsupervised learning to pre-train broad representations of human motion and address difficulties such as dataset fragmentation and low-quality data. By overcoming these issues, emotion detection algorithms can have broader applicability and greater accuracy in real-world circumstances. Table 3 presents an overview of recent studies on emotion recognition using body gestures, highlighting the methodologies, datasets, and key findings, showcasing advancements in the field, and identifying current trends and challenges.

Table 3: Recent works on emotion recognition using body gestures

Author (Year)	**Description**	**Methodology used**	**Metrics considered**	**Limitations**
Haoyu Chen (2023) [13]	Hidden emotional state analysis using subtle body gestures (micro-gestures, MGs) as a novel approach to emotion recognition**.**	Introduced Spontaneous Micro-Gesture (SMG) dataset; proposed AED-BiLSTM framework for robust online MG recognition and graph-based network for MG pattern representation.	Accuracy in MG classification, online recognition performance, and emotional stress state recognition.	Limited scope of SMG dataset; requires integration with other non-verbal cues (e.g., facial expressions) to enhance performance.
David C Jeong et.al. [14]	Introduced MoEmo, a vision transformer (ViT) for emotion detection in robotics systems, leveraging	Combines cross-attention fusion of movement vectors and environmental context using feature maps	Accuracy in emotion detection compared to state-of-the-art methods.	Limited dataset size and lack of diversity in contexts; potential inadequacy for real-time human-robot

	3D human pose estimation and context-aware cross-attention.	from CLIP encoder; trained on the Naturalistic Motion Database.		interaction.
Yu Du et.al. (2023)[15]	Proposed Heuristic Multimodal Real-Time Emotion Recognition (HMR-TER) to enhance e-learning by analyzing emotions through facial expressions, hand gestures, and vocal intonations.	Utilized a multimodal approach combining face, hand gestures, and speech analysis; employed Bayesian classifier for emotion recognition.	Face detection (84.25%), hand gestures (92.70%), voice recognition (82.26%), emotion problem reduction (84.5%), and e-learning efficiency (93.85%).	Limited sample size (20 samples); potential generalization issues for diverse learning environments.
Dimitrios Kollias (2023) [16]	Proposed C-EXPR-DB, a large in-the-wild database for compound expression recognition (CER), and C-EXPR-NET, a multi-task learning model for CER and AU detection (AU-D)	Used multi-label formulation, KL-divergence loss for CER, and distribution matching loss to enhance task coupling; AU-D incorporates semantic descriptions and visual information.	Outperformed state-of-the-art methods for CER and generalization in zero-shot contexts.	Database imbalance and limited availability of other in-the-wild datasets for compound expression validation.
Haoyang Liu et.al. (2024) [17]	Emotion recognition using non-facial cues like gestures and posture, leveraging the Aff-Wild2 and DFEW databases.	OpenPose for pose estimation, ResNet, ANN for classification, and fully connected layers for regression.	Accuracy (emotion classification), valence-arousal regression performance.	**H**igh computational cost, limited real-time processing, and scalability issues with large datasets.
U Bhattacharya et.al. (2024) [18]	Transformer-based Text2Gestures framework for generating emotive full-body gestures aligned with natural language inputs.	Utilized biomechanical features, gender, and handedness; trained on MPI Emotional Body Expressions Database; evaluated using Pearson correlation (min. 0.77 in valence dimension).	Plausibility rating (91% positive user feedback on Likert Scale); Pearson correlation for intended emotions.	Limited to sentence-level gesture mapping; lacks continuity between gestures; does not integrate facial expressions or vocal tones.
Dan Guo et.al. (2024) [19]	A comprehensive micro-action dataset (MA-52) and a benchmark model (MANet) for micro-action recognition (MAR) and emotion analysis.	MANet integrates SE and TSM into ResNet for spatio-temporal modeling; and introduced joint-embedding loss for semantic matching between video and action labels.	Accuracy for multi-label micro-action and emotion recognition tasks	Challenges with noise in real-world scenarios (e.g., occlusion, lighting variations); limited multi-modal integration.
Deng Li et.al. (2024) [20]	Proposed a visual-text contrastive learning solution for Micro Gesture Recognition	Utilized Adaptive Prompting for generating context-aware prompts;	Achieved state-of-the-art performance with a 6%+ improvement in	Limited exploration of multi-modal combinations; relies on results from

	(MGR) to enhance emotional understanding without relying on identity-based data.	combined textual and visual modalities for enhanced learning.	emotion understanding using textual predictions.	specific datasets for generalization claims.
Rong Gao et.al. (2024) [21]	Explored micro-gestures (MG) as unintentional behaviors driven by inner feelings, focusing on their role in emotional understanding and recognition.	Proposed a spatio-temporal balanced dual-stream contrastive learning network with novel augmentation strategies for MG attributes; validated emotional reasoning using large language models.	Achieved state-of-the-art performance in MG recognition and demonstrated their significance in enhancing emotional understanding.	Limited exploration of multimodal cues and potential mappings between emotions and MGs; challenges in applying MG strategies to other tasks.
Busra Karatey et.al. (2024) [22]	Proposed a novel deep neural network (DNN) framework for emotion detection from videos and images using facial and body features extracted by OpenPose.	Combines Gaussian mixture model with CNN, LSTM, and Transformer to create CNN-LSTM and CNN-Transformer models; evaluated on FABO and CK+ datasets.	Achieved close to 100% accuracy for the FABO dataset and over 90% accuracy for the CK+ dataset.	Limited generalizability due to high performance primarily on controlled datasets; potential challenges in real-world applications.
Kang, D. et.al. (2024) [23]	Proposed a multi-modal emotion recognition system (MAER) combining internal (bio-signals) and external (video, voice) cues for real-time emotion prediction.	Utilized Asyncio-based asynchronous parallel processing and a non-parametric modality-adaptive fusion module for integrating signals.	Achieved up to 33% higher accuracy compared to systems using only external signals.	Lack of open datasets for comprehensive evaluation; performance verified only through pilot tests.
P. Sriram Kumar et.al. (2024) [24]	Proposed a multimodal automated emotion recognition (AER) system using physiological signals (EDA and ECG) combined with deep learning techniques.	Utilized decomposition of EDA signals, time-frequency representations (TFRs), and deep learning architectures (AlexNet, cCNN, VGG16) in unimodal and multimodal settings.	Achieved 86.66% accuracy for four-class and 83.96% accuracy for three-class emotion classification using multimodal signals.	Limited generalizability due to reliance on two datasets (CASE and WESAD); real-world applicability not fully validated.
Md. Milon Islam et.al. (2024) [25]	Proposed a deep learning-based multimodal emotion recognition system for healthcare analytics to monitor patients' behavior using physiological and video data.	Combined visual features from videos (extracted using DSCNN and ResNet-34) with physiological data (processed using Bi-LSTM) through a soft attention mechanism for	Achieved accuracies of 97.34% (HOG) and 97.92% (ResNet-34) on the Bio Vid Emo DB dataset.	Limited to offline evaluations; not tested in real-time environments; applicability to diverse healthcare scenarios remains unvalidated.

		significant feature extraction.		

5.1 Research Gaps in the Literature

In our evaluation of current research on emotion recognition by body movements, we discovered many constraints that restrict the effectiveness and generalisability of existing systems. These limitations include dataset concerns, cultural and individual differences, dynamic gesture complexities, a lack of multimodal integration, and computational challenges. Below, we explain these restrictions in depth.

Lack of Standardized and Comprehensive Datasets

One major limitation in current research on emotion recognition through body gestures is the absence of standardized and diverse datasets. Most existing datasets are either small or collected in controlled environments, limiting their ability to generalize across different contexts and populations. Additionally, imbalanced datasets, where certain emotional classes are overrepresented while others are underrepresented, further hinder the performance of machine learning and deep learning models. This scarcity of robust and comprehensive datasets restricts the development of models capable of accurately recognizing emotions in real-world scenarios.

Cultural and Individual Variability

Emotional expressions through body gestures often vary significantly across cultures and individuals, creating challenges in designing universally applicable recognition systems. Gestures conveying emotions such as happiness or anger may differ in meaning depending on cultural or personal contexts. For example, a gesture signifying joy in one culture might have a completely different connotation in another. Current models, trained on culturally specific datasets, often struggle to perform well in diverse populations, leading to potential misinterpretation of emotions.

Dynamic and Subtle Nature of Body Gestures

Body gestures expressing emotions are often subtle, dynamic, and transient, making them difficult to capture and analyze accurately. Emotions are frequently conveyed through small, rapid movements or micro-gestures that require advanced modeling techniques and high-quality data to decode effectively. The inability of current systems to detect and interpret these nuances reduces their effectiveness in real-world applications where emotional expression is rarely exaggerated or obvious.

Lack of Multimodal Integration

While body gestures are a critical component of emotional expression, they are often analyzed in isolation. This siloed approach overlooks the complementary information provided by other modalities, such as facial expressions, vocal intonations, and physiological signals. Integrating multiple modalities can provide a holistic understanding of emotions. However, the synchronization and fusion of data from multiple sources remain challenging due to differences in timing, data formats, and processing requirements.

High Computational Complexity

Emotion recognition systems often require extensive computational resources due to the complexity of analyzing spatio-temporal data from body gestures. Models that achieve high accuracy typically involve advanced architectures, such as deep learning frameworks, which are resource-intensive and may not be feasible for real-time applications. This limitation is especially problematic for scenarios like human-robot interaction, where systems need to process emotions quickly and efficiently.

Addressing these constraints is crucial for the progression of emotion identification via bodily gestures. Future research must prioritize the creation of varied and representative datasets that consider cultural and individual diversity to improve model generalisability. Efforts must prioritize the capture of subtleties and dynamics in body motions to accurately represent real-world emotional expressions. Integrating multimodal techniques, such as the combination of bodily gestures, facial expressions, and voice clues, can yield a more comprehensive knowledge of emotions. Furthermore, developing lightweight and scalable models will facilitate effective real-time processing, rendering emotion identification systems more applicable to many real-world scenarios. These developments are essential for realizing the complete potential of emotion recognition systems.

5. Conclusion

Emotion recognition using body gestures is a potential but underexplored subject within the larger field of affective computing. The literature demonstrates tremendous progress in establishing models and approaches for recognizing emotions using nonverbal clues, which address the complexity and subtlety of human emotional expressions. However, significant hurdles remain, including a lack of standardized and diverse datasets, cultural and individual variation in gestures, difficulties recording dynamic and delicate body motions, and inadequate integration of multimodal information. Despite these challenges, advances in machine learning and deep learning, particularly the use of multimodal fusion, sophisticated feature extraction techniques, and novel architectures such as transformers, have shown great promise for improving emotion recognition accuracy and real-world applicability. To solve these problems, we prioritize integrating multimodal data, such as combining bodily gestures with facial expressions and vocal tones, to gain a more complete understanding of emotions. Furthermore, we hope to create robust and scalable models capable of real-time processing, ensuring adaptation across varied contexts while maintaining computing efficiency. These initiatives are intended to pave the way for more accurate and dependable emotion identification systems.

REFERENCES

1. Khare, Smith K., et al. "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations." Information Fusion 102 (2024): 102019.
2. N. Smrti´c, “Asertivna komunikacija i komunikacija u timu,” Ph.D. dissertation, Polytechnic of Me ¯dimurje in ˇCakovec. Management of tourism and sport., 2015.
3. Kalateh, Sepideh, et al. "A systematic review on multimodal emotion recognition: building blocks, current state, applications, and challenges." IEEE Access (2024).
4. G.F. Wilson, C.A. Russell, Real-time assessment of mental workload using psychophysiological measures and artificial neural networks, Hum. Factors 45(4) (2003) 635–644.
5. R. E. Dahl and A. G. Harvey, “Sleep in children and adolescents with behavioral and emotional disorders,” Sleep medicine clinics, vol. 2, no. 3, pp. 501–511, 2007.
6. Kopalidis, Thomas, et al. "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets." Information 15.3 (2024): 135.
7. Atymtayeva, L., et al. "Fast facial expression recognition system: selection of models." *Appl. Math* 17.2 (2023): 375-383.
8. H. Gunes, C. Shan, S. Chen, and Y. Tian, “Bodily expression for automatic affect recognition,” Emotion recognition: A pattern analysis approach, pp. 343–377, 2015
9. H. A. Vu, Y. Yamazaki, F. Dong, and K. Hirota, “Emotion recognition based on human gesture and speech information using middleware,” in Fuzzy Systems (FUZZ), 2011 IEEE International Conference on. IEEE, 2011, pp. 787–791.
10. D. Glowinski, M. Mortillaro, K. Scherer, N. Dael, and G. V. A. Camurri, “Towards a minimal representation of affective gestures,” in Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on. IEEE, 2015, pp. 498–504.
11. S. Senecal, L. Cuel, A. Aristidou, and N. Magnenat-Thalmann, “Continuous body emotion recognition system during theater performances,” Computer Animation and Virtual Worlds, vol. 27, no. 3-4, pp. 311–320, 2016.
12. J. Arunnehru and M. K. Geetha, “Automatic human emotion recognition in surveillance video,” in Intelligent Techniques in Signal Processing for Multimedia Security. Springer, 2017, pp. 321– 342.
13. Chen, Haoyu, et al. "Smg: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis." *International Journal of Computer Vision* 131.6 (2023): 1346-1366.
14. Jeong, David C., et al. "MoEmo Vision Transformer: Integrating Cross-Attention and Movement Vectors in 3D Pose Estimation for HRI Emotion Detection." 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023.
15. Du, Yu, Rubén González Crespo, and Oscar Sanjuán Martínez. "Human emotion recognition for enhanced performance evaluation in e-learning." *Progress in Artificial Intelligence* 12.2 (2023): 199-211.
16. Kollias, Dimitrios. "Multi-label compound expression recognition: C-expr database & network." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023.
17. Liu, Haoyang. "Emotion Detection through Body Gesture and Face." *arXiv preprint arXiv:2407.09913* (2024).
18. Bhattacharya, Uttaran, et al. "Text2gestures: A transformer-based network for generating emotive body gestures for virtual agents." *2021 IEEE virtual reality and 3D user interfaces (VR)*. IEEE, 2021.

19. Guo, Dan, et al. "Benchmarking Micro-action Recognition: Dataset, Method, and Application." IEEE Transactions on Circuits and Systems for Video Technology (2024).
20. Li, Deng, Bohao Xing, and Xin Liu. "Enhancing micro gesture recognition for emotion understanding via context-aware visual-text contrastive learning." IEEE Signal Processing Letters (2024).
21. Gao, Rong, et al. "Identity-free artificial emotional intelligence via micro-gesture understanding." arXiv preprint arXiv:2405.13206 (2024).
22. Karatay, Buşra, et al. "CNN-Transformer based emotion classification from facial expressions and body gestures." Multimedia Tools and Applications 83.8 (2024): 23129-23171.
23. Kang, Dohee, et al. "Beyond superficial emotion recognition: Modality-adaptive emotion recognition system." Expert Systems with Applications 235 (2024): 121097.
24. Kumar, P. Sriram, et al. "Deep learning-based automated emotion recognition using multi-modal physiological signals and time-frequency methods." IEEE Transactions on Instrumentation and Measurement (2024).
25. Islam, Md Milon, et al. "Enhanced multimodal emotion recognition in healthcare analytics: A deep learning based model-level fusion approach." Biomedical Signal Processing and Control 94 (2024): 106241.