engineering proceedings

MDPI

Proceeding Paper

Recognizing Human Emotions Through Body Posture Dynamics Using Deep Neural Networks †

Arunnehru Jawaharlalnehru [1,*], Thalapathiraj Sambandham [2,*] and Dhanasekar Ravikumar [3]

[1] Department of Computer Science and Engineering (Emerging Technologies), SRM Institute of Science and Technology, Vadapalani, Chennai 600026, India
[2] Department of Mathematics, SRM Institute of Science and Technology, Vadapalani, Chennai 600026, India
[3] Department of Electrical and Electronics Engineering, Sri Sairam Institute of Technology, Chennai 600044, India; dhanasekar.eee@sairamit.edu.in
* Correspondence: arunnehruj@gmail.com (A.J.); thalapas@srmist.edu.in (T.S.)
† The 5th International Electronic Conference on Applied Sciences, 4–6 December 2024; Available online: https://sciforum.net/event/ASEC2024.

Abstract: Body posture dynamics have garnered significant attention in recent years due to their critical role in understanding the emotional states conveyed through human movements during social interactions. Emotions are typically expressed through facial expressions, voice, gait, posture, and overall body dynamics. Among these, body posture provides subtle yet essential cues about emotional states. However, predicting an individual's gait and posture dynamics poses challenges, given the complexity of human body movement, which involves numerous degrees of freedom compared to facial expressions. Moreover, unlike static facial expressions, body dynamics are inherently fluid and continuously evolving. This paper presents an effective method for recognizing 17 micro-emotions by analyzing kinematic features from the GEMEP dataset using video-based motion capture. We specifically focus on upper body posture dynamics (skeleton points and angle), capturing movement patterns and their dynamic range over time. Our approach addresses the complexity of recognizing emotions from posture and gait by focusing on key elements of kinematic gesture analysis. The experimental results demonstrate the effectiveness of the proposed model, achieving a high accuracy rate of 91.48% for angle metric + DNN and 93.89% for distance + DNN on the GEMEP dataset using a deep neural network (DNN). These findings highlight the potential for our model to advance posture-based emotion recognition, particularly in applications where human body dynamics distance and angle are key indicators of emotional states.

Keywords: gesture; GEMEP; emotions; matrix features

check for updates

Academic Editor: Eugenio Vocaturo

Published: 16 April 2025

Citation: Jawaharlalnehru, A.; Sambandham, T.; Ravikumar, D. Recognizing Human Emotions Through Body Posture Dynamics Using Deep Neural Networks. *Eng. Proc.* **2025**, *87*, 49. https://doi.org/10.3390/engproc2025087049

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Emotion is a mental state linked with thoughts, the intensity of positive or negative experiences, and human behavioral reactions. Defining human emotion precisely remains exclusive at present. Over the past two decades, extensive research has been conducted on human emotion across disciplines such as psychology, neuroscience, and behavior analysis. The literature reveals varying interpretations of human emotion, leading to subjectivity in its definition. The two primary theories, physiological and cognitive neuroscience, seek to elucidate human emotion. The physiological theory posits that bodily responses form the basis of emotions, influenced by external stimuli. For instance, encountering a snake typically triggers fear in most individuals due to physiological changes.

Psychological research has identified a variety of frequently occurring emotions in humans. In a recent study, researchers reported 27 distinct types of human emotions by eliciting subjective responses from participants who viewed 2185 video clips. These emotions include "admiration, adoration, esthetic appreciation, amusement, anxiety, awe, awkwardness, boredom, calmness, confusion, craving, disgust, empathetic pain, entrancement, envy, excitement, fear, horror, interest, joy, nostalgia, romance, sadness, satisfaction, sexual desire, sympathy, and triumph". Robert Plutchik's [1] 'wheel of emotions' model further categorizes human emotions, highlighting eight primary emotions: anger, disgust, fear, joy, sadness, trust, surprise, and anticipation.

W. Gerrod Parrott's [2] classification distinguishes between primary emotions, which are innate and universally present from birth (anger, love, fear, joy, sadness, surprise), and secondary emotions, which individuals learn through interactions with their environment. Additionally, certain emotions are universally recognized across cultures based on facial expressions and behavioral cues, termed universal or basic human emotions. These emotions exhibit similar facial expressions regardless of cultural diversity and are considered hard in the human brain [3,4].

Recent advancements in machine–human interaction and deep learning algorithms offer new opportunities for the efficient and accurate prediction and recognition of micro-emotions. High-resolution frames and frame sequences incorporating machine and deep learning can provide a faster, more precise approach to identifying emotions with better recognition accuracy. Current research on emotion recognition through body posture dynamics has several notable gaps. Firstly, there is limited focus on micro-emotions, which are subtle and fleeting emotional expressions. Secondly, most studies rely on static posture analysis, overlooking the dynamic nature of emotional expressions. Thirdly, existing research often neglects the importance of upper body movements in conveying emotions.

This study addresses these gaps by utilizing kinematic features from the GEMEP dataset to recognize micro-emotions. By focusing on the dynamics of upper body posture, we capture the nuanced and transient nature of emotional expressions. Our approach employs a deep neural network to analyze these dynamic features, enabling more accurate and comprehensive emotion recognition. The unique aspects of our methodology include the emphasis on upper body posture dynamics and the application of advanced machine learning techniques. This approach allows for a more nuanced understanding of emotional expressions, particularly in detecting subtle changes that may be missed by traditional static analysis methods.

Our research has significant potential applications in various fields, including human–computer interaction, mental health monitoring, and social robotics. By advancing posture-based emotion recognition, we contribute to the development of more empathetic and responsive AI systems, enhancing their ability to understand and interact with human emotions.

The primary objectives of this study are as follows:

1. Develop a novel method for recognizing micro-emotions using upper body posture dynamics.
2. Evaluate the effectiveness of deep neural networks in analyzing kinematic features for emotion recognition.
3. Contribute to the broader understanding of how body language conveys emotional states.

These objectives directly address the identified research gaps by focusing on dynamic, subtle emotional expressions and leveraging advanced analytical techniques to improve emotion recognition accuracy.

2. Literature Review

In contrast, body gestures are the most persuasive form of communication, as they convey intense movements of the head, torso, and other body parts. This study focuses on recognizing emotions through body gestures [5]. The video data capture expressive human behavior, including the dynamics of body movements in response to specific emotions.

In their study, S. Piana et al. [6] present a live approach for automatically detecting emotions by analyzing body movements. This approach incorporates a multi-class SVM classifier that makes use of a collection of postural, kinematic, and geometrical data derived from sequences of 3D skeletons. The system's ability to recognize the six basic emotions is evaluated using a leave-one-subject-out cross-validation technique. The study retrieved and evaluated a specific set of emotionally relevant elements from the movement data recorded by RGB-D sensors, such as the Microsoft Kinect, to infer the child's emotional state. The system acquired an accuracy rate of 61.3%, which is similar to the 61.9% estimation rate achieved by human observers.

Jinting Wu et al. [7] present a three-branch Generalized Zero-Shot Learning (GZSL) paradigm to infer novel body gestures and emotional states from semantic descriptions. The prototype-based detector (PBD) in the first branch determines whether a sample belongs to a body gesture category and predicts results. The second branch, a stacked autoencoder (StAE) with manifold regularization, predicts data from new categories using semantic representations. The third branch also has a softmax-layered emotion classifier to improve feature representation learning for emotion classification. A shared feature extraction network, a Bidirectional Long Short-Term Memory Network (BLSTM) with a self-attention module, generates input features for these three branches. Experimental results on the MASR dataset show that the program controls hidden gesture and emotion categories.

Z. Shen et al. [8] utilized a multi-view body gesture dataset (MBGD) specifically created for emotion recognition, which includes 43,200 RGB videos of simplified body gestures depicting six basic emotions along with their neutral control groups. The researchers proposed a novel method for integrating skeleton and RGB modalities using single modalities. In their approach, joint locations are first estimated using the OpenPose toolbox. To extract features from RGB and skeleton data, they employ Temporal Segment Network (TSN) and Spatial–Temporal Graph Convolutional Networks (ST-GCNs). Given that the output vectors of these models differ in length, a residual feature encoder is designed to standardize them. The two different feature vectors obtained are then concatenated into a unified vector, which is subsequently classified using a residual network.

D. Li et al. [9] present a robust visual-text contrastive learning approach that utilizes textual information for Micro Gesture Recognition (MGR). The task of emotion understanding based on micro gestures is divided into two sub-tasks: (1) clip-level Micro Gesture Recognition (MGR), which identifies micro gestures across multiple video clips, and (2) video-level emotion understanding, which assesses the overall emotional state of the video based on MGR results from various clips.

Myeongul Jung et al. [10] compared emotional states using body movement and senses. BMMs are a new way to visualize body part engagement and disengagement connected to emotions using motion capture data. This method compares its emotional aspects to bodily sensation maps (BSMs), which use computerized self-reports to address body activation or deactivation sensations. User research with 29 individuals collected BSMs and BMMs across seven emotional states. Though physiological sensation and body motion differed, BMMs and BSMs shared traits that might distinguish moods. This method is promising for automatic emotion recognition (AER) applications like emotion monitoring, telecommunications, and virtual reality, and it can visualize emotional physical activity to advance emotion-related disorder research.

3. Methodology

Gesture recognition technology has wide-ranging applications in assessing human behavior in public spaces such as airports, train stations, phone booths, parking lots, and in real-time monitoring by security personnel. Figure 1 presents an overview diagram illustrating the process of micro-emotion recognition through gesture analysis, using a deep learning algorithm applied to the GEMEP datasets.

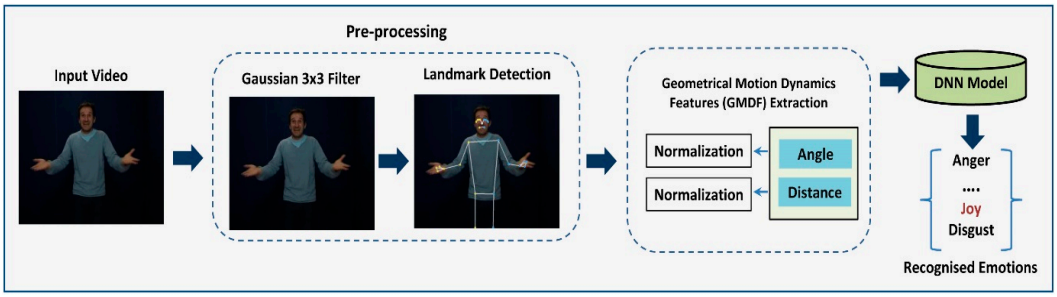


Figure 1. Overview of the emotion recognition using Geometrical Motion Dynamics Features.

Gesture landmark detection is used for identifying emotional patterns to locate Human Pose Estimation (HPE) in a dynamic sequence-based dataset called GEMEP. The technique known as pose estimation detects and categorizes the joints in the human body to recognize various activities in real-time applications. Essentially, it involves acquiring a set of coordinates for each joint (such as the wrist, torso, etc.) referred to as key points or landmarks, which together outline a person's pose [11]. The crucial links formed between these points are known as pairs. With various approaches in estimating the pose, here we use a skeleton-based approach for detecting the pose landmark features in the dataset. The goal of HPE is to depict the human body as a skeleton-like structure and utilize it for various task-specific applications.

The input is the GEMEP video sequence, initially pre-processed using a Gaussian filter (3×3), which smooths the video frame and removes noise, and then landmark features are identified for recognizing the emotional instance. The pre-trained Blaze Pose is incorporated within the Generic Human Model (GHUM) for estimating the precise location of the 33body landmark coordinates (x, y) representing the skeleton structure of the human body.

3.1. Preprocessing

The Gaussian filter in face detection is a specialized tool used in image processing to smooth images by eliminating noise and redundant features. It applies a blur effect, enhancing the identification of facial patterns. The Gaussian filter applies 3×3 kernel matrices in the RGB image, as given in Equation (1), to convolve within the image itself. According to the Gaussian function, the kernel values are determined by assigning a greater weight to the central pixels and a lower weight to the encircling pixels.

$$G_i = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix} \quad (1)$$

A weighted sum of its neighbors, with weights set by the Gaussian kernel, is used to replace each pixel value in the image. The primary benefit of utilizing Gaussian distribution is to eliminate stochastic fluctuations in pixel values (noise) without substantially smudging crucial characteristics.

3.2. Identification of Human Pose Landmark Points

Mediapipe 0.10.10 offers customizable machine learning solutions for live and streaming media across platforms. It provides a machine learning-based solution for accurate body pose detection, using 33 landmarks in RGB video frames and a background segmentation mask. Developed by Google, it uses a lightweight convolutional neural network for real-time human posture estimation, making it ideal for fitness tracking and sign language recognition.

3.3. Geometrical Motion Dynamics Features (GMDF) Extraction

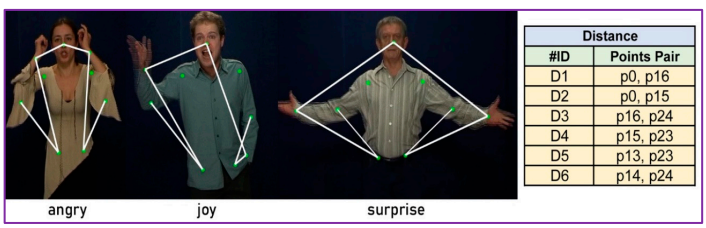
The initial extraction of affective cues from hand, shoulder, head, and waist movements is conducted using the GEMEP corpus dataset. Information about body posture, encompassing kinetics, motion, and contour, is derived from video sequences to discern emotions. Preceding analysis, the input video sequence transforms RGB frames to enable comprehensive examination. Then, 33 landmark points are delineated via the media pipe mesh predictor, imposed across the frame sequence to gauge human pose. Consequently, this study adopts feature vectors derived from the angle, distance, and velocity of skeleton coordinates to effectively differentiate between the seventeen distinct emotions.

Based on prominent research outcomes from diverse spatial contexts, this study presents a formal methodology in gesture geometry features, with the primary objective of isolating three geometric points. This approach leverages distance, angle, and velocity measurements derived from a set of 33 gesture landmark points. The body landmark points' locations are utilized to extract information on the kinetics, motion, and shape of the body stance, which is then used to identify the micro-emotions. Here, the nose is used as a midpoint to determine the expressive features between the hand, elbow, wrist, and hip locations.

3.3.1. GMDF—Distance

The derived geometric feature based on distance metrics encompasses the Euclidean distances between specific sets of landmark points, such as the hand, elbow, wrist, and hip. Visualizing these distances helps in understanding posture dynamics. In Figure 2, the straight line between two points in a Cartesian coordinate system is identified for the estimation of distance points. Calculating the distance metric allows us to quantify the spatial and temporal relationships between the main parts of the human body to recognize emotions accurately. The six Euclidean distances $D_l(p_1, p_2)$ are mathematically computed for each pair of the coordinate gesture landmark feature points $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$ as below in Equation (2).

$$D_{l(p_1,p_2)} = \sqrt{p_1(x_2 - x_1)^2 + p_2(y_2 - y_1)^2} \quad (2)$$



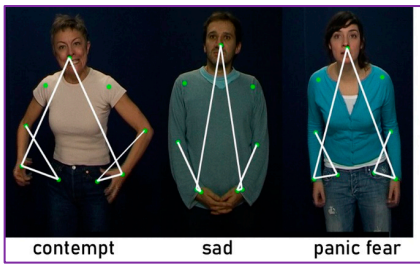
Distance	
#ID	Points Pair
D1	p0, p16
D2	p0, p15
D3	p16, p24
D4	p15, p23
D5	p13, p23
D6	p14, p24

Figure 2. Distance Feature Extraction—Points Pair.

3.3.2. GMDF—Angle

Angle calculation provides valuable insights into the dynamics of emotional expressions and the alignment of body segments, facilitating the recognition of affine body

gestures. Figure 3 below visualizes and analyses these angles, enhances our understanding of body language, and facilitates emotion differentiation in depth.



Angle	
#ID	Points Pair
A1	p16, p14, p12
A2	p11, p13, p15
A3	p14, p12, p24
A4	p13, p11, p23

Figure 3. Angle feature extraction—points pair.

Using two non-zero vector landmarks, the cosine angle measure $A_{l(p_1,p_2,p_3)}$ is calculated with three landmark point coordinates, as given below in Equation (3).

$$A_{l(p_1,p_2,p_3)} = \arccos\left(\frac{(p_3 - p_2)\cdot(p_1 - p_2)}{\|p_3 - p_2\|\|p_1 - p_2\|}\right) \quad (3)$$

3.3.3. GEMEP Dataset

The GEMEP (Geneva Multimodal Emotion Portrayal) dataset is a valuable resource for research in emotion recognition and analysis, developed by the researchers at the Swiss Center for Affective Sciences, University of Geneva by Klaus R. Scherer [12,13]. The main purpose of the dataset is to understand and analyze emotional expressions through multiple modalities like vocal speech, facial expression, and body posture. Figure 4 represents the sample frames of the dataset, which include 145 video sequences of 9272 images (720 × 576 pixels) with 25 frames per second, with a mean duration of 2.5 s of facial expressions and body movements, and records of vocal expressions and prosody (intonation, rhythm, and stress patterns). It comprises 17 classes of portrayed emotions. The subjects are equally partitioned with five male and female candidates aged 25 to 45 years, captured in both frontal and side poses. This dataset provides a rich source of multimodal data for studying emotional expressions, supporting a wide range of research and development in affective computing and related fields.

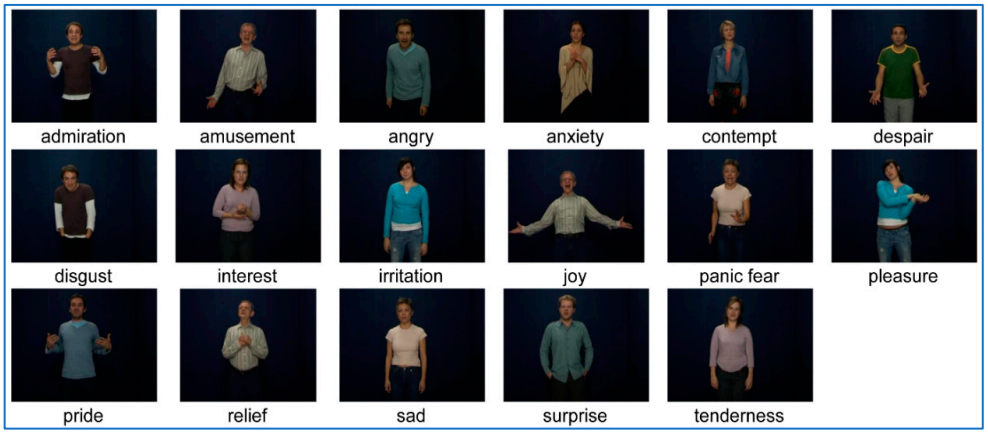


Figure 4. Sample frames from the GEMEP dataset.

The GEMEP dataset, while valuable for emotion recognition, has limitations due to its controlled laboratory setting. The restricted size and diversity of the dataset pose challenges in generalizing findings to real-world scenarios. Participants in such environments may exhibit exaggerated emotional expressions, reducing the dataset's applicability to spontaneous human interactions. To enhance the dataset's effectiveness, it is recommended

to integrate emotional data from natural settings where individuals express emotions authentically. This approach would improve the model's ability to recognize emotions in diverse and realistic contexts, making it more suitable for practical applications in human–computer interaction and affective computing.

4. Experimental Setup and Results

The experiments were performed using MATLAB 2019b (Version number 9.7) on the Windows 10 Professional operating system installed on a computer featuring an Intel i7 Processor and 16 GB of RAM with GPU/CPU resource utilization and memory consumption. For instance, our model was trained on an NVIDIA RTX 3090 GPU, with an average training time of 12 h for 100 epochs and an inference speed of 10 ms per frame.

The obtained feature vectors were fed to the DNN model for it to recognize the micro-emotions in the body movements. The performances of the DL classifier model were evaluated using 70% for training and 30% for testing, and the quantitative evaluation was performed with statistical metrics like accuracy, precision, recall, and f-measure. This session utilizes a deep neural network model to forecast bodily gesture emotions by utilizing a video-based dataset called GEMEP.

The evaluation of the Geometrical Motion Dynamics Features (GMDF) extraction of distance and angle metrics was carried out using the GEMEP dataset. To identify the gesture emotions, the obtained feature vectors were fed to DNN model to evaluate the effectiveness of the method.

Table 1 indicates the performance of distance features in the GEMEP dataset, of which the micro emotion 'tenderness and anxiety' had the highest accuracy with an overall mean precision of 93.85%, a sensitivity of 93.72%, specificity of 99.62%, and an f-measure of 93.76%.

Table 1. Performance of distance features in the DNN.

Emotions	Precision (%)	Sensitivity (%)	Specificity (%)	F-Measure (%)
ADM	95.28	93.80	99.77	94.53
AMU	95.57	93.72	99.65	94.63
ANG	94.92	87.96	99.65	91.30
ANX	96.43	93.75	99.81	95.07
CON	95.41	93.69	99.81	94.55
DES	95.63	93.58	99.69	94.60
DIS	90.91	91.84	99.67	91.37
INT	93.95	96.65	99.50	95.28
IRR	92.14	96.35	99.30	94.20
JOY	90.50	88.04	99.35	89.26
PAN	86.81	92.59	99.28	89.61
PLA	95.53	97.89	99.48	96.70
PRI	91.97	89.36	99.58	90.65
REL	93.16	96.20	99.50	94.65
SAD	96.09	95.56	99.73	95.82
SUR	93.33	96.55	99.78	94.92
TEN	97.78	95.65	99.93	96.70
Mean	93.85	93.72	99.62	93.76

Table 2 indicates the performance of angle features in the GEMEP dataset, of which the micro emotion 'tenderness and pleasure' had the highest accuracy, with an overall mean precision of 91.17%, a sensitivity of 91.32%, specificity of 99.47%, and an f-measure of 91.22%.

Table 2. Performance of DNN model with the angle features.

Emotions	Precision (%)	Sensitivity (%)	Specificity (%)	F-Measure (%)
ADM	92.86	90.70	99.66	91.77
AMU	93.07	90.82	99.46	91.93
ANG	91.67	86.39	99.42	88.95
ANX	93.71	93.06	99.66	93.38
CON	88.79	92.79	99.51	90.75
DES	93.51	92.51	99.54	93.01
DIS	86.54	91.84	99.48	89.11
INT	91.35	90.91	99.30	91.13
IRR	91.52	93.61	99.26	92.55
JOY	88.40	86.96	99.19	87.67
PAN	87.50	88.15	99.36	87.82
PLA	94.31	93.31	99.36	93.81
PRI	87.41	88.65	99.32	88.03
REL	92.23	96.74	99.42	94.43
SAD	92.94	95.00	99.50	93.96
SUR	88.64	89.66	99.63	89.14
TEN	95.46	91.30	99.85	93.33
Mean	91.17	91.32	99.47	91.22

Tables 1 and 2 present the mean performance measure achieved for the distance and angle features, and Table 3 gives the mean recognition accuracy achieved across the DNN for angle and distance features, of which the distance features exhibit superior performance when compared to the angle features. The DNN achieves a mean recognition accuracy of 91.48% and 93.89%, respectively.

Table 3. Average performance measure obtained for GGMF features.

Models	Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
DNN	Angle	91.48	91.17	91.32	91.22
	Distance	93.89	93.85	93.72	93.76

The experimental results highlight the effectiveness of the deep neural network (DNN) model in recognizing bodily gesture emotions using the GEMEP dataset. The study evaluates two different feature extraction approaches—distance-based features and angle-based features—demonstrating the model's performance across various emotions. A direct comparison of Tables 2 and 3 shows that the DNN model achieves better performance with distance features than with angle features. The mean accuracy difference of 2.68% suggests that distance-based representations provide richer information for distinguishing emotions in bodily gestures. This could be due to the fact that distances between key joints reflect broader body movements, which are more indicative of emotion than angular variations. However, the angle-based approach still achieves competitive performance, indicating that it captures complementary information. In future work, a hybrid approach combining both distance and angle features could potentially enhance recognition accuracy by leveraging the strengths of both feature types.

The results demonstrate that the DNN model is highly effective in recognizing emotions based on bodily gestures, with distance-based features performing better than angle-based features. The high mean accuracy values indicate the potential of deep learning techniques for emotion recognition tasks, particularly when leveraging rich spatial features

from human movement. Future research could explore integrating both feature sets and incorporating temporal dynamics to further improve the robustness of emotion classification.

5. Comparative Analysis with State-of-the-Art Methods

The results derived from the proposed Geometrical Motion Dynamics Features (GMDF)-based method were compared with those from previously published methods in the literature. The comparison included methods utilizing the spatio-temporal interest points and the pose net model. Table 4 presents the outcomes from the GEMEP datasets.

Table 4. State-of-the-art recognition accuracy for the GEMEP dataset.

Reference	Method	Examined Emotions	Accuracy
Geetha et al. [14]	Spatio-temporal interest points + RF	[An, Fr, Jy, Re, Sa]	92.10%
Santhosh et al. [15]	HOG feature + KNN		90.30%
Muthana et al. [16]	DCNN	[An, Fr, Jy, Re, Sa]	90.70%
Proposed Approach	Angle metric + DNN	[Ad, Am, An, Ax, Co, De, Di, In, Ir, Jy, PF, Pl, Pr, Re, Sa, Su, Te]	91.48%
	Distance metric + DNN		93.89%

6. Conclusions

Body posture dynamics offer critical insights into understanding emotional states through human movement during social interactions. This paper presents a robust method for recognizing 17 micro-emotions by analyzing the kinematic features of body movements from the GEMEP dataset, focusing on upper body posture dynamics, including skeleton points and angles. Our approach, grounded in kinematic gesture analysis, specifically addresses the challenges of capturing the fluid, evolving nature of body dynamics. By employing a deep neural network (DNN), our model achieves high recognition accuracy, with results showing a mean accuracy of 93.76% for distance features and 91.22% for angle features, particularly excelling in recognizing emotions like 'tenderness' and 'anxiety'. Evaluated on MATLAB 2019b with rigorous training and testing splits, our DNN model demonstrated superior performance metrics across accuracy, precision, recall, and F-measure. The effectiveness of our model signifies the potential of using posture-based analysis in emotion recognition applications, especially where body dynamics play a central role in conveying emotional cues. The comparative analysis with other methods, including those based on spatio-temporal interest points and pose net models, indicates that our Geometrical Motion Dynamics Features (GMDF)-based approach achieves improved accuracy. These findings underscore the utility of DNN-driven kinematic analysis for future advancements in emotion recognition technologies, potentially enhancing various applications in human–computer interaction, social robotics, and psychological research.

Author Contributions: Conceptualization, A.J. and T.S.; methodology, A.J. and D.R.; software, A.J. and T.S.; validation, all of the authors; investigation, all of the authors; writing—original draft preparation, all of the authors; writing—review and editing and supervision, A.J. and T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Plutchik, R. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]
2. Parrott, W.G. The nature of emotion. In *Blackwell Handbook of Social Psychology: Intraindividual Processes*; John Wiley & Sons: Hoboken, NJ, USA, 2001; pp. 375–390.
3. Kuusikko, S.; Haapsamo, H.; Jansson-Verkasalo, E. Emotion recognition in children and adolescents with autism spectrum disorders. *J. Autism Dev. Disord.* **2009**, *39*, 938–945. [CrossRef] [PubMed]
4. Arunnehru, J.; Kumar, A.; Verma, J.P. Early prediction of brain tumor classification using convolution neural networks. In Proceedings of the Advances in Computational Intelligence, Security and Internet of Things: Second International Conference, Agartala, India, 13–14 December 2019; ICCISIoT Proceedings. Springer: Singapore, 2020; pp. 16–25.
5. Pelachaud, C. Modelling multimodal expression of emotion in a virtual agent. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3539–3548. [CrossRef] [PubMed]
6. Piana, S.; Stagliano, A.; Odone, F.; Verri, A.; Camurri, A. Real-time automatic emotion recognition from body gestures. *arXiv* **2014**, arXiv:1402.5047.
7. Wu, J.; Zhang, Y.; Zhao, X.; Gao, W. A generalized zero-shot framework for emotion recognition from body gestures. *arXiv* **2020**, arXiv:2010.06362.
8. Shen, Z.; Cheng, J.; Hu, X.; Dong, Q. Emotion recognition based on multi-view body gestures. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3317–3321.
9. Li, D.; Xing, B.; Liu, X. Enhancing micro gesture recognition for emotion understanding via context-aware visual-text contrastive learning. *IEEE Signal Process. Lett.* **2024**, *31*, 1309–1313. [CrossRef]
10. Jung, M.; Cho, Y.; Kim, J.; Kim, H.; Kim, K. Bodily Sensation Map vs. Bodily Motion Map: Visualizing and Analyzing Emotional Body Motions. *IEEE Trans. Affect. Comput.* **2024**, *15*, 1649–1658. [CrossRef]
11. Bashirov, R.; Ianina, A.; Iskakov, K.; Kononenko, Y.; Strizhkova, V.; Lempitsky, V.; Vakhitov, A. Real-time RGB-D-based extended body pose estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 2807–2816.
12. Banziger, T.; Scherer, K.R. Introducing the Geneva multimodal emotion portrayal (GEMEP) corpus. In *Blueprint for Affective Computing: A Sourcebook*; OUP Oxford: Oxford, UK, 2010; pp. 271–294.
13. Banziger, T.; Mortillaro, M.; Scherer, K.R. Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* **2012**, *12*, 1161. [CrossRef] [PubMed]
14. Arunnehru, J.; Chamundeeswari, G.; Bharathi, S.P. Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos. *Procedia Comput. Sci.* **2018**, *133*, 471–477. [CrossRef]
15. Arunnehru, J.; Kalaiselvi Geetha, M. Difference intensity distance group pattern for recognizing actions in video using support vector machines. *Pattern Recognit. Image Anal.* **2016**, *26*, 688–696. [CrossRef]
16. Mahdi, M. Movements Recognition in the Human Body Based on Deep Learning Strategies. *J. Multiapp* **2023**, *4*, 1–10. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.