AI Open 6 (2025) 183–203

Contents lists available at ScienceDirect

AI Open

journal homepage: www.elsevier.com/locate/aiopen

Full length article

Robust emotion recognition using hybrid Bayesian LSTM based on Laban movement analysis

Shuang Wu [a], Daniela M. Romano [a,b,c,*]

[a] Department of Civil, Environmental and Geomatic Engineering, University College London, London, WC1E 6AP, United Kingdom
[b] Department of Information Science, University College London, London, WC1E 6BT, United Kingdom
[c] Institute of Artificial Intelligence, De Montfort University, Leicester, LE1 9BH, United Kingdom

ARTICLE INFO

Keywords:
Emotion recognition
Machine learning
Body movements
Deep learning
Bayesian LSTM

ABSTRACT

Emotion recognition has become increasingly significant in artificial intelligence; however, the impact of body movements on emotion interpretation remains under-explored. This paper presents a novel Hybrid Bayesian Pre-trained Long Short-Term Memory (HBP-LSTM) framework that combines low-level pose data with high-level kinematic features, utilising Bayesian inference to enhance the accuracy and robustness of emotion recognition. The proposed model is trained on high-quality laboratory data to capture the fundamental patterns of emotional expression through body movements. We introduce noise and employ adversarial attack methods such as the Fast Gradient Sign Method (FGSM) to evaluate the model's robustness during testing. This approach assesses the HBP-LSTM's ability to maintain performance under data degradation and adversarial conditions, common challenges in real-world scenarios. We validated the HBP-LSTM on two public datasets, EGBM and KDAEE, demonstrating that the model exhibits high robustness against noise and adversarial perturbations, outperforming traditional models. The HBP-LSTM accurately identifies seven basic emotions (happiness, sadness, surprise, fear, anger, disgust, and neutrality) with accuracies of 98% and 88% on the EGBM and KDAEE datasets, respectively. HBP-LSTM is a noise-resistant model with a reliable emotion recognition framework, which lays the foundation for future applications of emotion recognition technology in more challenging real-world environments.

1. Introduction

In the current literature, facial expressions (Canal et al., 2022), body language (Oğuz and Ertuğrul, 2024), voice (Zhang et al., 2023), and physiological changes (Tang et al., 2024) are the primary methods utilised to analyse people's expression of feelings. Also, there are efforts to interpret facial expressions in connection with the voice (Zambeli et al., 2024). According to Ekman (1984), humans are prone to acknowledge facial expressions and disregard body language to empathise with others. Nevertheless, non-verbal communication cannot be underscored enough in conveying emotions and body language and posture contribute to accurately communicating one's intentions and feelings to another person (Ahmed et al., 2019). There is a growing interest in utilising bodily movement, posture, and gesture to comprehend emotions. Multiple fundamental factors underpin this trend. Recent developments in motion capture technology and its enhanced accuracy have resulted in a surge in the volume and quality of data valuable for the automatic recognition of expressive movements (Elansary et al., 2024; Khare et al., 2023). At a distance, facial expressions might not be apparent; thus, physical movement presents a viable method for emotion recognition (Oğuz and Ertuğrul, 2024).

Recent research has focused on creating systems that autonomously identify emotions by examining cues from body posture and forecast emotions by reviewing an individual's body language (Geetha et al., 2024). These advancements seek to optimise and enhance communication effectiveness between humans and robots. However, despite the increased interest, the significance of body language in the automated analysis of emotions is still not yet fully acknowledged (Ebdali Takalloo et al., 2022). A lot of the current research work identifies emotions through several modalities, such as facial expressions, head movements, and hand gestures (Ebdali Takalloo et al., 2022). Yet, movements are also crucial in emotional expression and recognition. For example, we open our arms while experiencing positive emotions like joy, anger, or surprise (Shaarani and Romano, 2007). Faster bodily reactions related to fear, joy, anger, or surprise, and slow movements and responses fall under the category of sadness (De Meijer, 1989).

Laban Movement Analysis (LMA) has been employed to date primarily to analyse physical activity (Wang et al., 2024b; Shafir, 2023).

* Corresponding author at: Department of Civil, Environmental and Geomatic Engineering, University College London, London, WC1E 6AP, United Kingdom.
E-mail address: d.romano@ucl.ac.uk (D.M. Romano).

https://doi.org/10.1016/j.aiopen.2025.09.002
Received 20 March 2025; Received in revised form 9 September 2025; Accepted 24 September 2025
Available online 29 September 2025

LMA is a comprehensive system for observing and notating movement which offers nuanced insights into human behaviour through the categorisation of movements based on body parts, movement dynamics, and spatial pathways (Shafir, 2023). Previous research has frequently restricted their examination to only a small range of characteristics amongst the possible extensive set of measurable qualities connected with a physical activity, failing to consider the correlation and hierarchical significance of other measurable elements (Sapiński et al., 2019a). Moreover, occlusion (objects or people blocking parts of the subject from view) presents a significant challenge (Surace et al., 2017).

We advocate a holistic approach to automatic emotion recognition that includes not just individual gestures or facial expressions but also investigating the whole bodily apparatus's contribution and movement to the communication of emotional states. While there are some early attempts to explore the entire body's emotional experience automatically (Wang et al., 2024b; Ahmed et al., 2019; Wang et al., 2015), this is still an under-researched area.

All the above described challenges highlight the need to for a robust approach, and a machine-learning algorithm, that can reliably detecting and predicting human emotions only from bodily movement, particularly in cases where the data are partially occluded or missing.

This publication makes the following contribution to knowledge:

- A novel framework for structuring bodily motion into low-level and high-level features based on Laban Movement Analysis (LMA);
- The Hybrid Bayesian Pre-Learned Long Short-Term Memory (HBP-LSTM) architecture, a pioneering approach for emotion recognition from body movements that integrates Bayesian methods with Long Short-Term Memory neural networks; and
- Extensive experiments and their results to validate the robustness and accuracy of the proposed HBP-LSTM model against established models using two benchmark laboratory datasets, MDESVG and KDAEE, focusing on its performance under various simulated noise conditions and adversarial attacks.

2. Related work

2.1. Automatic emotion recognition from gross body movement

Gross body movements (movements involving the whole body) are the primary vehicle for expressing emotion (De Meijer, 1989). Body language, such as changes in posture, movements, the manner of walking (or gait), and more, can transfer realistic and substantial emotional information that can hardly be read in the musculature of the face or even materialised by words (YuMeng et al., 2024), highlighting the importance of the body in emotion communication (Reed et al., 2020).

Recent studies have explored various approaches to automatic emotion recognition from body movements, mainly focusing on low-level or high-level features. Sapiński et al. (2019a) introduced an algorithm for emotion recognition utilising the low-level characteristics obtained from the spatial arrangement and orientation of the joints in the complete skeletal structure. However, the authors did not fully explore how to integrate dynamic motor features with high-level kinematic features that could enhance the fine-grained understanding of emotional expression. YuMeng et al. (2024) proposed the Affective-Pose Gait Network (APGN), a novel approach for analysing emotions from gait. APGN employs a Spatio-Temporal Graph Convolutional Network (ST-GCN) to draw out a pose's features and a Convolutional Neural Network (CNN) to extract the affective features, emphasising the necessity of integrating both pose and affective data for more accurate emotion recognition from body movements. However, their model did not fully address the challenges posed by noisy or adversarial data, which can affect the robustness of emotion recognition systems.

2.2. Laban movement analysis

Human symbolic representation systems provide an effective method for analysing and interpreting body movements. In psychology, such coding techniques help to recognise emotional states such as boredom or interest (Wang et al., 2024b) by categorising different body postures and gestures, such as proximity, upper body posture, and hand movements. Laban Movement Analysis (LMA) is a widely used system of movement notation created by choreographer and theorist Rudolf Laban. LMA emphasises the relationship between internal states, intention and attention, and human movement forms, and it can provide insight into the expressive characteristics of movement. LMA has been used effectively in the analysis of emotional and behavioural patterns. High-level movement analysis based on LMA (Laban and Ullmann, 1971) has been employed in several studies to gain nuanced insights into human behaviour. LMA offers a comprehensive system for observing and notating movement, categorising it based on factors such as body parts, dynamics, and spatial pathways. Bartenieff and Lewis (2013) and other researchers compiled and expanded LMA to enhance its comprehensiveness and intricacy. Table 1 summarises the main categories of LMA.

Laban Movement Analysis (LMA) provides a rich body of terminology for understanding body movement as an expression of an individual's feelings and emotional reflections. The categories in its framework play a key role in this understanding (Melzer et al., 2019):

Body: By observing "what is in motion", one can infer emotional states. For instance, a slouched posture might indicate sadness or defeat, while an erect posture could signify confidence.

Space: The "where the body moves" can signify an individual's intent or emotional relation to their environment. Retreating actions may suggest fear or evasion, whilst advancing actions could denote hostility or eagerness.

Shape: Changes in the body's shape can provide insight into a person's emotional response to their surroundings. Contraction or shrinkage may indicate fear or introspection, while expansion indicates openness or joy.

Effort: This category is particularly relevant to emotions. Intrinsic attitudes towards exercise can provide direct clues about emotional states:

- *Weight*: Heavy movements may be firm or aggressive, while light movements may be tentative or gentle.
- *Space*: Direct movements can indicate decisiveness or focus, while indirect ones may suggest distraction or uncertainty.
- *Time*: Sudden movements often correlate with impulsiveness or surprise, and sustained motions with deliberation or calmness.
- *Flow*: Bound flow may be associated with control or tension, while free flow may suggest relaxation or spontaneity.

2.3. Automatic gross body emotion recognition from Laban movement analysis

Ahmed et al. (2019) employed a genetic algorithm for feature selection and presented a two-layer framework that executed feature selection computationally considering the human action descriptors, including from LMA, which was relevant in identifying significant emotional aspects. Wang et al. (2024b) proposed an LMA-based emotion recognition method for dance movements. The method accurately captures the emotional expression in dance by analysing the body's spatial distribution, structural features and movement patterns. They achieved recognition accuracy of 79.74% using deep neural networks (DNN). These studies demonstrate the potential application of LMA in gross body emotion recognition. However, the methods in the literature usually rely on high-quality datasets and do not adequately consider the impact of noise or adversarial perturbations on model performance.

Table 1
Laban movement analysis components.

Component	Description
Body	Neuromuscular patterns, Movement initiation, Movement sequence
Space	Kinesphere, Geometry, Spatial intention
Shape	Connection of body parts, Shape forms, Shape change, Shape flow support
Effort	Qualities of motion including Flow, Weight, Time, and Space

Despite the progress described above in automatic body movement emotion recognition, the robustness of the models still faces serious challenges when dealing with noisy data and adversarial attacks. Data in real environments are often interfered with by various factors such as sensor noise, occlusion, and deliberate perturbation that seriously affect the model's actual performance. Some studies have started addressing some of these problems in other domains. For instance, Goodfellow (2016) came up with the Fast Gradient Sign Method (FGSM) as a way of developing adversarial examples that trick neural networks, which highlights the importance of constructing robust models that can withstand such attacks. However, in the context of emotion recognition, few studies have systematically evaluated the robustness of the model under noisy or adversarial conditions, and there is a gap in research focusing on developing and testing models that maintain high performance when subjected to data degradation or adversarial perturbations.

3. Novelty

This publication presents the Hybrid Bayesian Pre-trained LSTM (HBP-LSTM) framework that automatically recognises emotion from gross body movements based on LMA, ameliorating the reliability and precision of the current automatic emotion recognition models trained on high-quality data. The use of Bayesian inference with the LSTM was purposely incorporated to emphasise the uncertainty and develop a more robust model than the one presented in the literature.

Based on Laban Movement Analysis (LMA) (Bernardet et al., 2019), we segment body movements as low and high-level features. Low-level features are angular and linear distances that display the general body configuration and volume data obtained from bounding boxes (Larboulette and Gibet, 2015). High-level features comprise aspects like speed, acceleration, and jerk (the difference between the acceleration of an object and the rate of change of its acceleration), depending on various body parts. Furthermore, high-level characteristics are also unobservable, e.g., timing, weight (effort), spatial orientation, moving in symmetry, and the trunk being perpendicular to the ground plane (Larboulette and Gibet, 2015). These advanced characteristics offer a detailed comprehension of motion, showcasing the intricate and delicate nature of human emotions expressed through body language. In addition, we provide a deep analysis of the model performances under different simulated noise conditions and adversarial attacks such as FGSM, showing that the HBP-LSTM surpasses the other models in terms of robustness.

The proposed framework, named hybrid Bayesian pre-trained long short-term memory (HBP-LSTM) architecture, aims to reinforce the emotion detection model's robustness. The model utilises a Pre-trained Bayesian LSTM to process low-level features, capturing temporal dynamics and modelling uncertainty through Bayesian inference. Also, to stress the temporal features that are the most important for sentiment recognition, an attention mechanism is utilised on the Pre-trained Bayesian LSTM output. The considered high-level features are processed through an adapter architecture, which changes the high-level features to be related via the output of the attention layer. The induction of the combined outputs of both the low-level and high-level feature handling is followed by inputting them into a Post-fusion Bayesian LSTM layer, which still learns temporal relations in the fused space and deals with uncertainties, also using probabilistic modelling. This approach enables a more nuanced and reliable analysis of complex affective cues, enhancing the robustness of affective detection in the presence of data uncertainty.

To check the robustness of the HBP-LSTM model, we added noise and used the Fast Gradient Sign Method (FGSM) as an adversarial attack during the testing phase. This particular evaluation strategy gauges the model's performance in the presence of degraded data conditions, which are, hence, a common challenge in real-world applications. While the current study is conducted using laboratory datasets, the framework is designed with the potential to handle challenges that may arise in more variable settings.

4. Methodology

4.1. Overview

In human movement analysis, discerning the emotional intent behind various movements is a complex challenge.

The automatic gross body emotion recognition presented in this paper is based on LMA. By integrating the four core dimensions of LMA (Shape, Effort, Body, and Space), the model can identify subtle differences in emotional expression more accurately.

The Shape dimension of LMA was quantified by analysing changes in the body structure and spatial relationships between body parts as in Wang et al. (2024b). Specifically, this included extracting high-level features such as height, width and depth measurements of the torso boundary volume, which are the core elements of the shape dimension.

The Effort dimension of LMA focuses on the dynamic characteristics of the movement in terms of time, weight, space and flow. We quantify changes in these dynamic features through metrics such as velocity, acceleration, and body part jerking. These effort-related features reflect the intensity and style of movement and play a key role in emotion recognition.

The Body dimension captures the low-level features in neuromuscular patterns and movement sequences by quantifying the spatial relationships between body parts, while the Space dimension focuses on the spatial intent and geometric properties of movements, revealing how individuals interact with their environment.

Our proposed solution, the Hybrid Bayesian Pre-Learned LSTM (HBP-LSTM) framework, innovatively combines Bayesian neural networks with Pre-trained feature encodings to enhance emotion recognition from body movement data. HBP-LSTM adopts a novel structure: a Pre-trained Bayesian LSTM processes low-level bodily features, and a subsequent Bayesian LSTM handles the fused features alongside high-level LMA features. This approach is geared towards capturing the intricacies inherent in human emotions and ensuring robustness in the face of data uncertainties by modelling uncertainty through Bayesian methods. A comprehensive discussion of the HBP-LSTM framework and its resilience in handling uncertain data is presented in Section 5. Fig. 1 provides a conceptual overview of the entire HBP-LSTM process for emotion recognition, from data preprocessing to emotion type output.

4.2. Feature extraction

Wallbott (1998) underscored the pivotal role of hand and arm movements in conveying emotions, suggesting that such upper body movements are fundamental, rather than merely supplementary, in embodying emotional states through bodily expressions. Based on this, Wang et al. (2015) proposed an approach that combines low-level

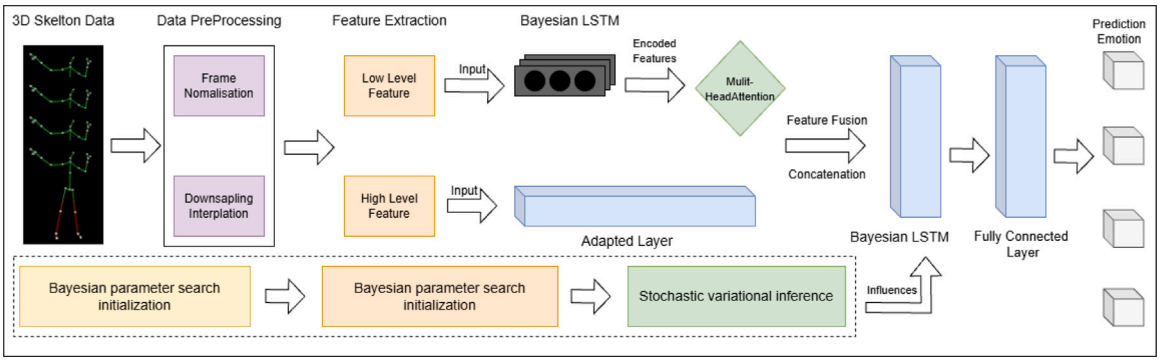


Fig. 1. Conceptual overview of the HBP-LSTM for emotional recognition.

gesture features with high-level movement features, paying special attention to the evolution of these features in the time dimension to capture the dynamic nature of emotional expression. On this basis, Ahmed et al. (2019) further extends the analytical framework by subdividing body movements into ten different categories. These categories aim to reveal the complexity and overlapping features of emotions reflected in body movements. The categorisation lays the foundation for a deeper understanding of the physical expression of emotions and marks an important advance in emotion recognition research.

Drawing inspiration from these groundbreaking studies, the current methodology merges low-level postural features with comprehensive high-level kinematic and geometric features of the body. The features extracted are computationally derived from individual frames or sequences of frames, enabling a detailed depiction of the body's movements. This approach aims to provide a multifaceted analysis of how emotions are manifested through bodily gestures and postures, leveraging both static and dynamic aspects of human movement to achieve a deeper understanding of emotional expression.

4.2.1. Low-level feature

In our study, we represent body postural patterns by computing low-level, context-independent features. Specifically, for the body's three-dimensional skeletal model, we calculate the spatial distances between the hands, elbows, and feet relative to each other and the shoulders. This includes measuring Euclidean distances from each hand to the opposite shoulder, elbow, foot, and between the feet. We also measure the distances from each elbow to the opposite hand and foot and from the head to each hand. All these lead to 41 postural features calculated on a per-frame basis as detailed in Table 2.

4.2.2. High-level feature

In high-level feature extraction, we extracted dynamic and qualitative movement features related to emotional expression based on LMA. These features cover the four main dimensions of LMA: Body, Effort, Shape, and Space.

Velocity, Acceleration, and Jerk: This set of features represents the dynamics of movement in motion sequence, as indicated by Larboulette ~(Larboulette and Gibet, 2015), consider a motion sequence X, which is represented by a series of n sequential postures $\{x(t_1), x(t_2), x(t_3), \ldots, x(t_n)\}$. The velocity of this sequence is given by Eq. (1), and the magnitude of velocity is defined in Eq. (2). The temporal evolution of X through these postures determines the motion's velocity profile. Specifically, the velocity for the kth joint at time t_i is denoted as $v^k(t_i)$, and its x-component is expressed as $v_x^k(t_i)$. The term δt signifies the infinitesimally small time interval between successive frames. For the experiments, we utilised two datasets: one acquired with Kinect V2 at a frame rate of 30 fps, and the other with a frame rate of 125 fps. Therefore, δt is set to the time interval corresponding to a single frame, approximately $\frac{1}{30}$ s and $\frac{1}{125}$ s, respectively.

$$v^k(t_i) = \frac{x^k(t_{i+1}) - x^k(t_{i-1})}{2\delta t} \tag{1}$$

$$\|v^k(t_i)\| = \sqrt{v_x^k(t_i)^2 + v_y^k(t_i)^2 + v_z^k(t_i)^2} \tag{2}$$

Bounding Volume: We extended this set of features to analyse body motion using a 3D bounding volume approach, where the space utilisation of the human body can be estimated from the bounding volumes of each part of the body defined over time (Hachimura et al., 2005). We structure the human body into a hierarchical model that allows for a detailed investigation of movement throughout distinct body regions, as depicted in Fig. 2. Our segmentation subdivides the body into four primary regions, enabling us to evaluate the spatial dynamics of human motion systematically. This segmentation is visually summarised in Figs. 2 and 3. Fig. 2 presents a hierarchical representation of the human body, delineating the main divisions, such as the upper and lower body, as well as the right and left sides. The lower body is further categorised into the right and left legs. In contrast, the upper body is bifurcated into the right and left arms, each with their respective subdivisions, including the shoulder, elbow, wrist, and hand for the arms, and hip, knee, ankle, and foot for the legs.

Fig. 3 illustrates the lateral segmentation in detail, specifying the key joints that make up the left and right sides of the body. The left side joints include the left side of the shoulder, left side of the elbow, left side of the wrist, left side of the hand, left side of the hip, left side of the knee, left side of the ankle, and left side of the foot; the right side joints correspond. This transversal perspective is important for analysing asymmetries in unilateral movements or gestures and different ways of using body space. Based on the equations (Ahmed et al., 2019) in Eqs. (3) and (4), we computed the boundary volumes of the four specified body regions on a frame-by-frame basis in order to quantify the performance characteristics of these regions in motion.

$$\begin{aligned} dx &= \max_{j \in \text{Joints}}(x_j) - \min_{j \in \text{Joints}}(x_j) \\ dy &= \max_{j \in \text{Joints}}(y_j) - \min_{j \in \text{Joints}}(y_j) \\ dz &= \max_{j \in \text{Joints}}(z_j) - \min_{j \in \text{Joints}}(z_j) \end{aligned} \tag{3}$$

$$BoundingVolume(BV) = dx \cdot dy \cdot dz \tag{4}$$

Time, Weight, Space, Flow: This set of features reinforces the Effort component by characterising the motion's dynamics, energy, and expressiveness, with intensity levels that vary continuously across a spectrum of opposing characteristics (Larboulette and Gibet, 2015). As expounded in Section 2.2, the temporal subcategory within the Effort component captures the urgency of the movement, ranging from sudden to sustained. Quantitatively, this aspect is gauged by the accelerations of body parts, with smaller summative values over the sequence indicating greater movement stability.

Reflecting the body's hierarchical segmentation illustrated in Figs. 2 and 3, we calculate the temporal feature for each body segment as

Table 2
The Definition of the 41 low-level features.
Source: Adapted from Weiyi Wang (Wang et al., 2015).

ID	Meaning	ID	Meaning
1	Euclidean Distance of Two Feet	22	Right Hand - Left Feet in Y
2	Euclidean Distance of Two Hands	23	Right Hand - Left Feet in Z
3	Euclidean Distance of Two Elbows	24	Left Hand - Right Shoulder in X
4	Euclidean Distance of Left Hand and Head	25	Left Hand - Right Shoulder in Y
5	Euclidean Distance of Right Hand and Head	26	Left Hand - Right Shoulder in Z
6	Right Hand - Right Shoulder in X	27	Left Hand - Right Elbow in X
7	Right Hand - Right Shoulder in Y	28	Left Hand - Right Elbow in Y
8	Right Hand - Right Shoulder in Z	29	Left Hand - Right Elbow in Z
9	Right Hand - Right Elbow in X	30	Left Hand - Right Feet in X
10	Right Hand - Right Elbow in Y	31	Left Hand - Right Feet in Y
11	Right Hand - Right Elbow in Z	32	Left Hand - Right Feet in Z
12	Right Hand - Right Feet in X	33	Left Hand - Left Shoulder in X
13	Right Hand - Right Feet in Y	34	Left Hand - Left Shoulder in Y
14	Right Hand - Right Feet in Z	35	Left Hand - Left Shoulder in Z
15	Right Hand - Left Shoulder in X	36	Left Hand - Left Elbow in X
16	Right Hand - Left Shoulder in Y	37	Left Hand - Left Elbow in Y
17	Right Hand - Left Shoulder in Z	38	Left Hand - Left Elbow in Z
18	Right Hand - Left Elbow in X	39	Left Hand - Left Feet in X
19	Right Hand - Left Elbow in Y	40	Left Hand - Left Feet in Y
20	Right Hand - Left Elbow in Z	41	Left Hand - Left Feet in Z
21	Right Hand - Left Feet in X		

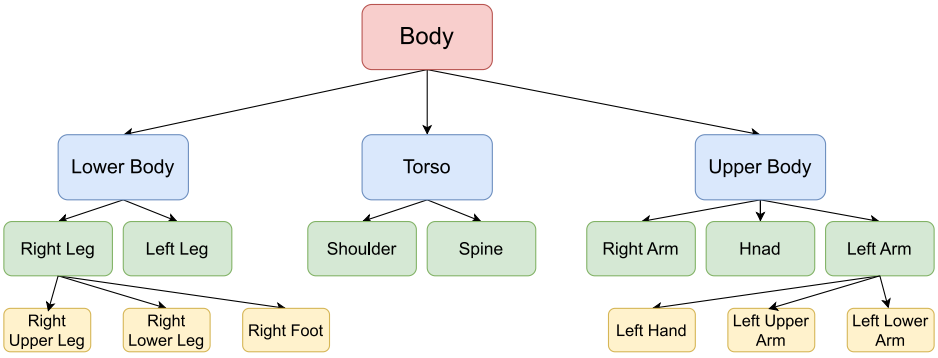


Fig. 2. Hierarchical segmentation of the human body for movement analysis.

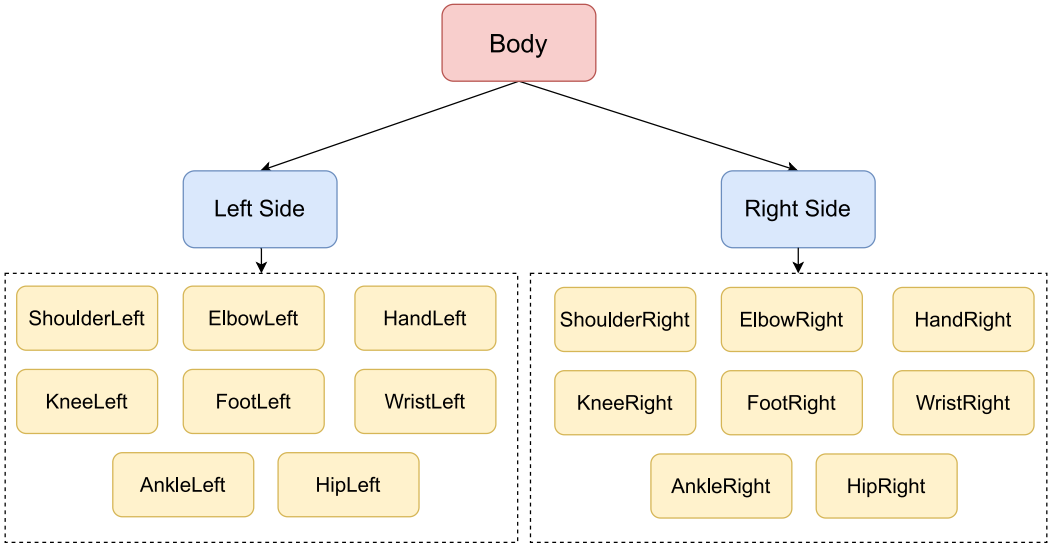


Fig. 3. Body joints distribution for left and right side analysis.

detailed by Eq. (5) (Larboulette and Gibet, 2015). In this equation, $a^k(t_i)$ represents the acceleration of the kth joint at the ith time frame, across a sequence of T frames. The resulting computed value serves as an index of movement stability, with lower scores signifying increased steadiness.

$$Time^k(t_i) = \frac{1}{T}\sum_{i=1}^{T} a^k(t_i) \tag{5}$$

The concept of "Weight" characterises the physical attributes of movement, delineating the action's force, with "light" and "strong" as its polar dimensions. It is quantitatively assessed by aggregating the kinetic energy of body parts over a specified duration. Eqs. (6) and (7) facilitate the computation of kinetic energy, denoted as $E(t_i)$, for designated body segments at the moment t_i (Larboulette and Gibet, 2015). Within the confines of this experiment, which analyses four distinct body regions, α_k signifies the normalised weights assigned to each joint k, uniformly set to 1 to simplify the analysis. As per Eq. (7), with T indicating the time window, the method entails pinpointing the peak kinetic energy within this timeframe. Such calculations are meticulously executed for each frame throughout the experimental procedure.

$$E(t_i) = \sum_{k \in K} E_k(t_i) = \sum_{k \in K} \alpha_k v^k(t_i)^2 \tag{6}$$

$$Weight(t_i) = \max_{i \in [1,T]} E(t_i), \quad i = 1, 2, 3, \ldots, N \tag{7}$$

Space describes whether the motion intake is related to its surroundings, whether it is direct (focused) or indirect (multi-focused). A lower value would suggest a more direct path (less space is used), while a higher value indicates a more indirect path (more space is used) (Larboulette and Gibet, 2015). This feature can be calculated according to Eq. (8).

$$Space^k(t_i) = \frac{\sum_{i=1}^{T-1} \|x^k(t_{i+1}) - x^k(t_i)\|}{\|x^k(t_T) - x^k(t_i)\|} \tag{8}$$

Fluency characterises the degree of continuity in an action, distinguishing between the dimensions of freedom and constraint. Actions that demonstrate greater freedom typically exhibit smoother, more fluid motion, as indicated by lower computed jerk values. Conversely, actions marked by constraint tend to have higher jerk values, reflecting less smooth motion (Laban and Ullmann, 1971). The total jerk for a joint, accumulated over time, allows us to assess the action's continuity. This cumulative measure is formalised in Eq. (9) (Larboulette and Gibet, 2015). In Eq. (9), $j^k(t_i)$ denotes the jerk of the kth joint at time t_i, and T represents the total number of frames in the observed sequence.

$$Flow^k(t_i) = \frac{1}{T} \sum_{i=1}^{T} j^k(t_i) \tag{9}$$

The calculations for the Effort features are based on the segmentation of the body into four parts as depicted in Figs. 2 and 3, with the analysis conducted for each frame in the sequence.

4.3. Statistical feature

Torso Height, Torso Width, Torso Depth: This set of features describes changes in the trunk of the body and can indicate body rotation, and body orientation. Torso width, height, and depth give an idea of space use and can be calculated for each frame in a sequence as per Eqs. (10), (11), and (12).

$$W = \left| x_{\text{ShoulderRight}} - x_{\text{ShoulderLeft}} \right| \tag{10}$$

Here, $x_{\text{ShoulderRight}}$ and $x_{\text{ShoulderLeft}}$ represent the x-coordinates of the right and left shoulder joints, respectively. The absolute difference between these coordinates gives the width of the torso at that frame.

$$D = \left| z_{\text{SpineMid}} - z_{\text{SpineBase}} \right| \tag{11}$$

In this equation, z_{SpineMid} and $z_{\text{SpineBase}}$ denote the z-coordinates of the mid-spine and base-spine joints, respectively. The absolute difference provides the depth of the torso along the z-axis.

$$H = \left| y_{\text{Neck}} - y_{\text{SpineBase}} \right| \tag{12}$$

Here, y_{Neck} and $y_{\text{SpineBase}}$ are the y-coordinates of the neck and base-spine joints, respectively. The absolute difference calculates the height of the torso from the base of the spine to the neck.

We can track dynamic changes in torso dimensions over time by analysing these measurements for each frame. Such changes may reflect movement characteristics such as bending, leaning or twisting. These movements are closely related to specific emotional states, thus supporting more accurate emotion recognition.

5. Model design

5.1. Architectural synthesis of HBP-LSTM

Decoding emotional intent from limb movements requires an advanced analysis method that captures the inherent time dependence and uncertainty in motion data. To this end, we propose a HBP-LSTM framework that combines a Bayesian neural network with a long short-term memory (LSTM) network to enhance emotion recognition performance from limb movement data. Fig. 4 illustrates the specific architecture of the HBP-LSTM framework, focusing on describing the interactions between the Bayesian LSTM modules and the fusion process of low-level and high-level features.

The HBP-LSTM framework uses a dual Bayesian LSTM network architecture for processing and fusing low-level body features with high-level LMA features. Firstly, an initial Pre-trained Bayesian LSTM is used to process the low-level kinematic data sequences to capture the body movements' temporal dynamic features while modelling the parameters' uncertainty through Bayesian inference. Subsequently, a four-headed multi-attention mechanism is applied to this Pre-trained Bayesian LSTM output to highlight key temporal features. This configuration enables the model to focus on different temporal dimensions of the sequence and capture complex temporal relationships while effectively mitigating the impact of noise or missing values in the data on model performance.

Meanwhile, high-level LMA features are processed through an adapter module consisting of a linear layer, ReLU activation and Dropout. The role of the adapter module is to convert the high-level features into a representation compatible with the output of the multi-head attention mechanism, thus enabling seamless fusion with the low-level feature representation.

Subsequently, the output of the multi-attention mechanism is concatenated with the output of the adapter module in order to generate a fused feature vector. The fused feature vector is then fed into the fused Bayesian LSTM layer, which is used to model temporal dependencies further and quantify uncertainty in the combined feature space. The Bayesian LSTM can effectively capture the intricate interactions between low-level and high-level features while quantifying uncertainty in the joint feature space, thereby markedly enhancing the robustness of the model. The classification task is then completed by the standard fully connected layer, which outputs the predicted sentiment labels.

By modelling the parameters of the LSTM layer as probability distributions rather than fixed values, the HBP-LSTM framework can express confidence in predictions and deal with the inherent uncertainty in motion data. This probabilistic approach is particularly effective in the presence of imperfect data quality, enabling robust emotion recognition in the presence of occluded or missing data. The Bayesian LSTM layer is trained by variational inference, employing an optimised Evidence Lower Bound (ELBO) as the loss function, thus efficiently learning from the data while capturing uncertainty. This design significantly improves the model's performance in complex contexts.

5.2. Detailed configuration and training paradigm

Low-level feature processing. The initial kinematic data sequence consisted of 41 low-level features and was processed through a Pre-trained Bayesian LSTM model. The model models the network weights as probability distributions and effectively captures the temporal dynamics inherent in body movements while modelling uncertainty. During training, a Dropout Rate (DR) of 0.5 was used to prevent overfitting.

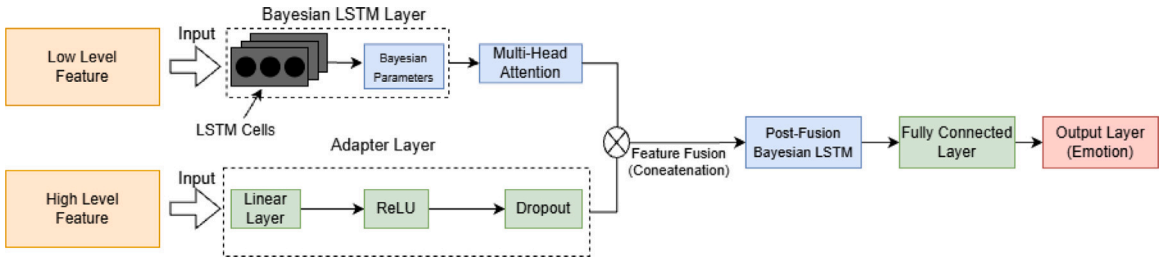


Fig. 4. Detailed architecture of the HBP-LSTM framework.

Upon training completion, the Pre-trained Bayesian LSTM parameters were frozen to preserve their learned representations. To enhance the model's ability to attend to critical temporal features in emotion recognition, we applied a multi-head attention mechanism to the output of the Pre-trained Bayesian LSTM, which contains four attention heads. The choice of four attentional heads is designed to help the model learn the complex relationships between different temporal steps in a human movement sequence while striking a balance between capturing temporal dependencies and maintaining computational efficiency. In addition, the multi-head attention mechanism allows the model to focus on multiple information dimensions in the sequence, thus effectively mitigating the effects of data noise or missing values.

High-level feature processing. The high-level LMA features (13 622 dimensions) are first passed through an adapter — implemented as a fully connected layer — that projects them to a 64-dimensional space, i.e. $13\,622 \rightarrow 64$ parameters plus bias. A ReLU activation and a dropout layer ($p = 0.5$) follows this linear projection. This adapter maps the advanced features into a representation that is dimensionally compatible with the output of the multi-head attention mechanism, enabling seamless fusion with the low-level features.

Feature fusion and post-fusion processing. The outputs of the attention mechanism (4×64 channels) and the adapter module (64 channels) are concatenated, yielding a fused feature vector of length $4 \times 64 + 64 = 320$. This 320-dimensional vector is fed into a post-fusion Bayesian LSTM layer with 256 hidden units, a standard normal prior $\mathcal{N}(0, 1)$, and recurrent dropout of 0.1. The Bayesian LSTM further captures temporal dependencies and propagates uncertainty within the fused feature space, enabling the model to exploit the complementary information carried by low-level and high-level features.

Classification Layer. Following the post-fusion Bayesian LSTM, a standard fully connected layer is used for classification, outputting the predicted emotion labels. Unlike the Bayesian LSTM layers, this fully connected layer uses fixed weights, providing a deterministic mapping from the learned representations to the output classes. The layer takes the output from the post-fusion Bayesian LSTM, which has a size of 256 and maps it to the number of emotion classes.

Training methodology. The HBP-LSTM model employs variational inference to approximate the posterior distributions of the weights in the Bayesian LSTM layers. The network is optimised with the Evidence Lower Bound (ELBO), which combines a likelihood term with a Kullback–Leibler (KL) regularisation term (Papatheodorou, 2024). Following common practice, the weight of the KL term is linearly annealed from 0 to 1 during the first 30% of epochs. Each mini-batch is evaluated with `sample_nbr`=3 forward Monte-Carlo samples, and the KL coefficient is set to $\lambda_{\mathrm{KL}} = 1/|D|$, where $|D|$ denotes the number of training sequences. Gradient norms are clipped to 5.0 to prevent exploding gradients.

Training uses the Adam optimiser with a learning rate of 1×10^{-3}, for 50 epochs and a batch size of 64. All experiments were conducted on an NVIDIA RTX 4090 GPU (PyTorch 2.1 + Blitz-Bayesian-DeepLearning 0.4.0, Python 3.10). Random seeds are fixed to 42 and the best model checkpoint is selected according to validation ELBO, ensuring reproducibility. Dropout and layer normalisation are applied to further improve robustness to distributional shifts.

6. Experiment

6.1. Dataset

This research utilises two databases, each containing diverse emotional data, to evaluate the effectiveness of the HBP-LSTM method for recognising emotion from whole-body movements.

EGBM. The EGBM dataset based on the multimodal database introduced by Sapiński et al. (2019b), comprises emotional speech, video, and gestural data captured using the Kinect v2 sensor. This database contains both video and motion-capture data, recorded using the Kinect v2 sensor, from professional actors simulating seven different emotional states. A total of 16 participants — equal numbers of males and females ranging in age from 25 to 64 — contributed to the dataset. The actors were instructed to enact the emotions sequentially: starting with neutral, followed by sadness, surprise, fear, disgust, anger, and happiness. Each emotion was repeated five times, with the actors bringing their own interpretation to the expression of each emotional state without specific guidelines. The dataset comprises 560 instances, evenly distributed, with 80 samples for each emotional state. The Kinect v2 sensor ensured comprehensive capture of the actors' movements, including the legs, as depicted in Fig. 5 (Sapiński et al., 2019b).

KDAEE. The KDAEE (Zhang et al., 2020) is a kinematic dataset that has a total 1402 recordings from 22 college students performing seven emotion states(happiness, sadness, anger, fear, disgust, surprise, and neutral), gathered using motion capture data 125 Hz and full body kinematic data through 17 sensors placed on the actor's key anatomical points, such as arms, legs, spine, and head. Actors were completed two types of movements: spontaneous (based on actors' understanding of emotional expression) and within-a-scenario movements(using predefined scenarios created by the dataset developers). Each performance lasted six seconds and was repeated as needed to ensure high data quality. This dataset only provides raw kinematic data consisting of 72 anatomical nodes as shown in Fig. 6.

6.2. Data preprocessing

6.2.1. Normalisation

For the EGBM dataset, collected using the Kinect V2 sensor, the raw data comprises 3D positions and orientations of the joints relative to the sensor's coordinate system, specified as $[x, y, z]$. Variability in the distance between the actor and the sensor during the recording sessions may affect the data quality. Consequently, remapping the skeletal coordinates from the sensor-defined space to a body-centric coordinate system denoted as $[u, v, w]$, is crucial. This remapping anchors the local coordinate system at the *SpineBase* joint within the Kinect skeletal model, aligning the u-axis to the left, the v-axis upward, and the w-axis forward about the *SpineBase* joint. Fig. 7 displays the 25 joints that the Kinect v2 tracks. This reorientation process generates a vector that encapsulates the relative positions and orientations of all joints concerning this central joint for each frame. For the 16 individuals represented in the dataset, the reference state for each person's movements is established by the first frame of each emotional state, setting

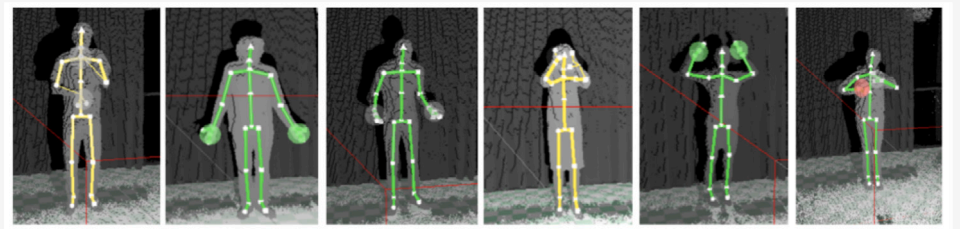


Fig. 5. Illustrative poses of actors expressing the six basic emotions: fear, surprise, anger, sadness, happiness, and disgust. (Sapiński et al., 2019b).

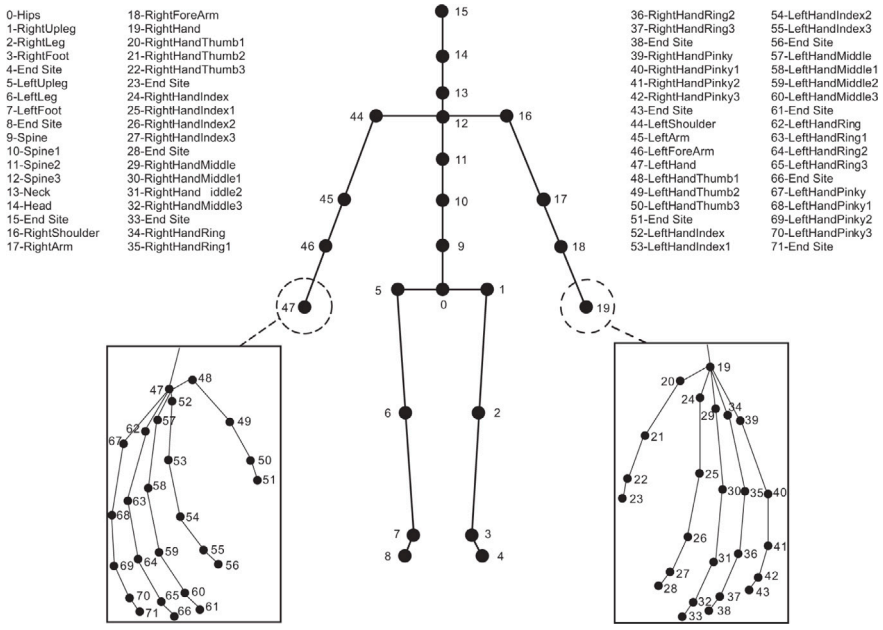


Fig. 6. Illustration of the 72 anatomical nodes used in KDAEE for full-body motion capture (Zhang et al., 2020).

a consistent baseline for subsequent frames. This approach ensures a personalised and uniform reference for movement comparison across the various emotions and subjects.

In contrast, the KDAEE dataset was retained in its original coordinate system due to its use of a different motion capture system, which captures a more extensive set of 72 anatomical nodes compared to the 25 tracked by Kinect V2 in EGBM. For feature extraction, we selected the same set of nodes as used in EGBM, focusing on those involved in low-level and high-level feature calculations. For further details on feature extraction, see Section 4.2.

6.2.2. Downsampling and interpolation

To solve the problem of inconsistent sequence lengths under different emotional states, we normalised the data to obtain uniform sequence lengths for effective analysis. Table 3 shows the average length of processed sequences for each emotion state in both datasets. Sequences exceeding the target length are downsampled to reduce redundancy, thus optimising computational efficiency while preserving core motion features. Conversely, shorter sequences are expanded by linear interpolation to ensure consistency across emotional states.

For EGBM, collected at 30 fps, downsampling is applied minimally to retain smooth motion transitions, while linear interpolation is used sparingly for consistency in sequence length. In the KDAEE dataset, captured at 125 fps with more anatomical nodes, downsampling contributes significantly to noise reduction, helping clarify body movement data without losing crucial details. The interpolation applied here further ensures that subtle body movements are retained, enabling smooth transitions across frames.

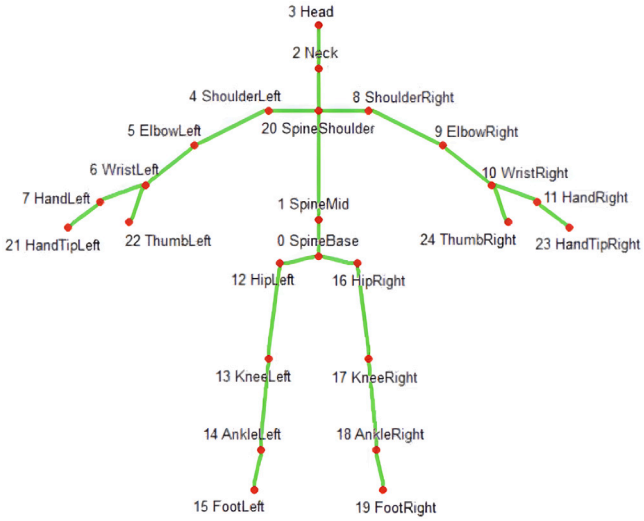


Fig. 7. The 25 skeleton joints tracked by the Kinect v2 Sensor (Cao et al., 2019).

The combination of downsampling and linear interpolation maintains the integrity of the motion data and allows for a consistent representation of the sequence across actors and emotional states, resulting in robust emotion recognition.

Table 3
The average number of frames and their equivalent duration in seconds for each emotional state, based on a 30 fps rate for EGBM and a 125 fps rate for KDAEE.

Emotion	EGBM (30 fps)		KDAEE (125 fps)	
	Frames	Seconds	Frames	Seconds
Anger	107	3.6	891.2	7.31
Disgust	140	4.7	924.9	7.39
Fear	115	3.8	855.7	6.84
Happiness	111	3.7	836.1	6.68
Neutral	93	3.1	90.3.7	7.23
Sadness	110	3.7	1064.5	8.51
Surprise	117	3.9	849.4	6.79

Table 4
Top 10 Features (Mean ± SD) and Levels.

Feature	Mean	SD	Level
`Right_Side_Flow`	0.309	0.108	High-level
`Left_Side_Flow`	0.272	0.134	High-level
`Lower_Body_Flow`	0.198	0.061	High-level
`Upper_Body_Flow`	0.145	0.069	High-level
`Distance_RightHandHead`	0.132	0.097	Low-level
`Distance_LeftHandHead`	0.124	0.105	Low-level
`Right_Side_BV`	0.108	0.071	High-level
`Lower_Body_BV`	0.102	0.029	High-level
`Left_Side_BV`	0.098	0.078	High-level
`HandLeft_FootRight_Distance_X`	0.089	0.015	Low-level

6.3. Feature analysis

In this section, we comprehensively analyse the selected features for emotion recognition. We evaluate their overall importance, compare high-level and low-level features, analyse contributions from different body parts, and examine their significance across various emotional categories.

We employed a Random Forest classifier with 100 estimators to assess feature importance. Prior to training, features were normalised using Scikit-learn's StandardScaler for consistent scaling. We formulated a binary classification task for each emotion by labelling the target emotion as 1 and others as 0. The dataset was split into training and testing sets using an 80/20 ratio. The Random Forest algorithm was chosen for its robustness with high-dimensional data and its ability to provide intrinsic measures of feature importance.

The importance of each feature was computed based on its contribution to the reduction of Gini impurity across all trees in the forest. Specifically, a feature's importance score reflects the total decrease in impurity it provides, aggregated over all trees. These scores were normalised so that the sum of all feature importances equals one.

After training the model for each emotion, we extracted the feature importance scores. To summarise their overall contributions, we calculated the mean and standard deviation of these scores across all emotions. The mean importance indicates a feature's average contribution, while the standard deviation highlights the variability of its importance across different emotional states.

6.3.1. Overall feature importance analysis

We have ranked the selection features for emotion recognition based on their mean importance scores across all emotions. The top 10 features are shown in Fig. 8, with corresponding statistics summarised in Table 4. The black bars represent standard deviations, highlighting the variability in the importance of each feature.

Dynamic features related to overall body movements (such as `Right_Side_Flow`, `Left_Side_Flow`, and `Lower_Body_Flow`) are consistently the most informative. These features reflect how smoothly and extensively different body regions move during emotional expressions, underscoring the importance of capturing dynamic information rather than static positions alone. Conversely, features based on distances between body parts (e.g., `Distance_RightHandHead` and `HandLeft_FootRight_Distance_X`) are still helpful but generally exhibit lower average importance and smaller variances. This indicates that while distance metrics provide reliable and stable spatial cues, their discriminative power for distinguishing emotional states is limited compared to dynamic movement features.

To further explore why certain selection features exhibit a high standard deviation, we conducted a detailed, emotion-specific analysis of feature importance. Fig. 9 presents the top five most influential features separately for each emotional category. The corresponding detailed rankings and statistics per emotion are presented in Table 5, which reveal consistent dominance of dynamic features across most emotions. Dynamic flow features consistently dominate across emotions, especially for *Anger*, *Fear*, *Happy*, and *Sad*, explaining their high global variability noted previously. These dynamic features show pronounced emotional specificity, suggesting that particular emotions manifest uniquely through certain body parts or movement intensities. For example, `Right_Side_Flow` prominently appears across multiple emotions, highlighting the dominant role of right-side movement dynamics in conveying affective states.

In contrast, distance-based features like `HandLeft–FootRight_Distance_X` demonstrate smaller but more consistent contributions across several emotions, notably in *Neutral* and *Surprise*. This indicates that distance features primarily encode general spatial configurations of the body rather than emotion-specific nuances, contributing to baseline or common emotional information. To summarise the relationships among emotions based on feature-importance patterns, we performed Ward hierarchical clustering using the z-scored importance vectors. The dendrogram in Fig. 10 visualises the resulting structure.

As it can be seen in Fig. 10 emotions are divided into two primary clusters:

- First cluster - *Neutral* and *Happy* form one cluster characterised by significant reliance on dynamic flow features and relatively low reliance on distance metrics. This grouping suggests similarities in the movement patterns associated with positive and neutral emotional expressions.
- Second cluster - encompasses predominantly negative emotions (*Anger*, *Disgust*, *Fear*, *Sad*, and *Surprise*). Within this negative cluster, emotions such as *Sad* and *Surprise* care closely aligned, reflecting a common reliance on spatial distance cues, particularly those involving head-hand configurations.

This clustering provides deeper insights into how emotional expressions share or differ in their underlying body language. It highlights how specific dynamic and spatial cues are selectively utilised across affective categories. Such insights can inform targeted improvements in computational models for emotion recognition, enhancing their sensitivity to both universal and emotion-specific movement patterns.

To further validate the observations above, we compared global feature-importance scores across the three categories: Flow, BV, and Distance, using the non-parametric Kruskal–Wallis test (Fig. 11; Table 6).

A Kruskal–Wallis H test indicated a statistically significant difference in feature importance across the three groups, $H(2) = 218.62$, $p < .001$, $\eta^2 = .16$ (medium effect). Post-hoc pairwise comparisons using Mann–Whitney U tests showed that *Flow* features had significantly higher importance than both *BV* features ($U = 2717$, $p < .001$) and *Distance* features ($U = 70\,387$, $p < .001$). Additionally, *BV* features were rated as more important than *Distance* features ($U = 58\,792$, $p < .001$), as illustrated in Fig. 11.

These findings underscore the superior value of movement dynamics over static positional information in affective-computing applications.

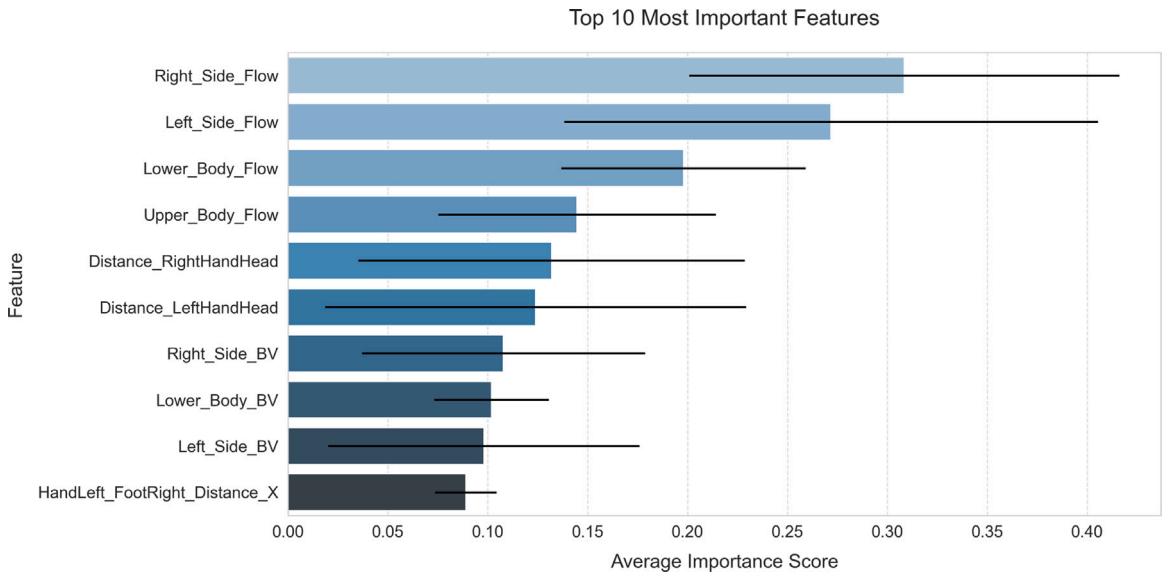


Fig. 8. Top 10 features ranked by average importance in emotion recognition.

Table 5
Top-5 important features for each emotion (mean ± SD).

Emotion	Rank	Feature	Category	Mean	SD
Anger	1	Right_Side_Flow	Flow	0.306	0.172
	2	Left_Side_Flow	Flow	0.246	0.206
	3	Lower_Body_Flow	Flow	0.173	0.060
	4	Upper_Body_Flow	Flow	0.122	0.080
	5	Distance_LeftHandHead	Distance/Other	0.115	0.119
Disgust	1	Right_Side_Flow	Flow	0.290	0.193
	2	Left_Side_Flow	Flow	0.275	0.221
	3	Lower_Body_Flow	Flow	0.204	0.062
	4	Distance_RightHandHead	Distance/Other	0.118	0.105
	5	Lower_Body_BV	BV	0.114	0.042
Fear	1	Right_Side_Flow	Flow	0.305	0.143
	2	Left_Side_Flow	Flow	0.270	0.184
	3	Lower_Body_Flow	Flow	0.229	0.109
	4	Distance_RightHandHead	Distance/Other	0.140	0.144
	5	Distance_LeftHandHead	Distance/Other	0.130	0.151
Happy	1	Right_Side_Flow	Flow	0.353	0.068
	2	Left_Side_Flow	Flow	0.296	0.077
	3	Upper_Body_Flow	Flow	0.221	0.129
	4	Lower_Body_Flow	Flow	0.197	0.064
	5	HandRight_FootLeft_Distance_X	Distance/Other	0.116	0.021
Neutral	1	Right_Side_Flow	Flow	0.300	0.100
	2	Left_Side_Flow	Flow	0.271	0.098
	3	Lower_Body_Flow	Flow	0.206	0.116
	4	Upper_Body_Flow	Flow	0.175	0.105
	5	Distance_RightHandHead	Distance/Other	0.158	0.165
Sad	1	Right_Side_Flow	Flow	0.318	0.126
	2	Left_Side_Flow	Flow	0.282	0.197
	3	Lower_Body_Flow	Flow	0.191	0.078
	4	Distance_RightHandHead	Distance/Other	0.143	0.139
	5	Upper_Body_Flow	Flow	0.140	0.055
Surprise	1	Right_Side_Flow	Flow	0.287	0.163
	2	Left_Side_Flow	Flow	0.264	0.223
	3	Lower_Body_Flow	Flow	0.186	0.043
	4	Distance_RightHandHead	Distance/Other	0.142	0.139
	5	Distance_LeftHandHead	Distance/Other	0.131	0.147

6.3.2. Comparison of high-level and low-level features

Building upon the overall feature-importance analysis, we compared the relative significance of high-level (dynamic movement) and low-level (static spatial) features in predicting emotional states.

A Mann–Whitney U test revealed that high-level features were significantly more important than low-level features, $U = 73\,638$, $p < .001$ (see Fig. 12).

This substantial difference, accounting for approximately 17% of the variance, highlights the superior predictive value of dynamic movement characteristics over static spatial measurements.

Further investigation of emotion-specific patterns (Fig. 13) demonstrated consistent dominance of high-level features across all emotional categories: Anger ($U = 2\,653$, $p < .001$), Disgust ($U = 2\,665$, $p < .001$), Fear ($U = 2\,599$, $p < .001$), Happy ($U = 2\,593$, $p < .001$), Neutral

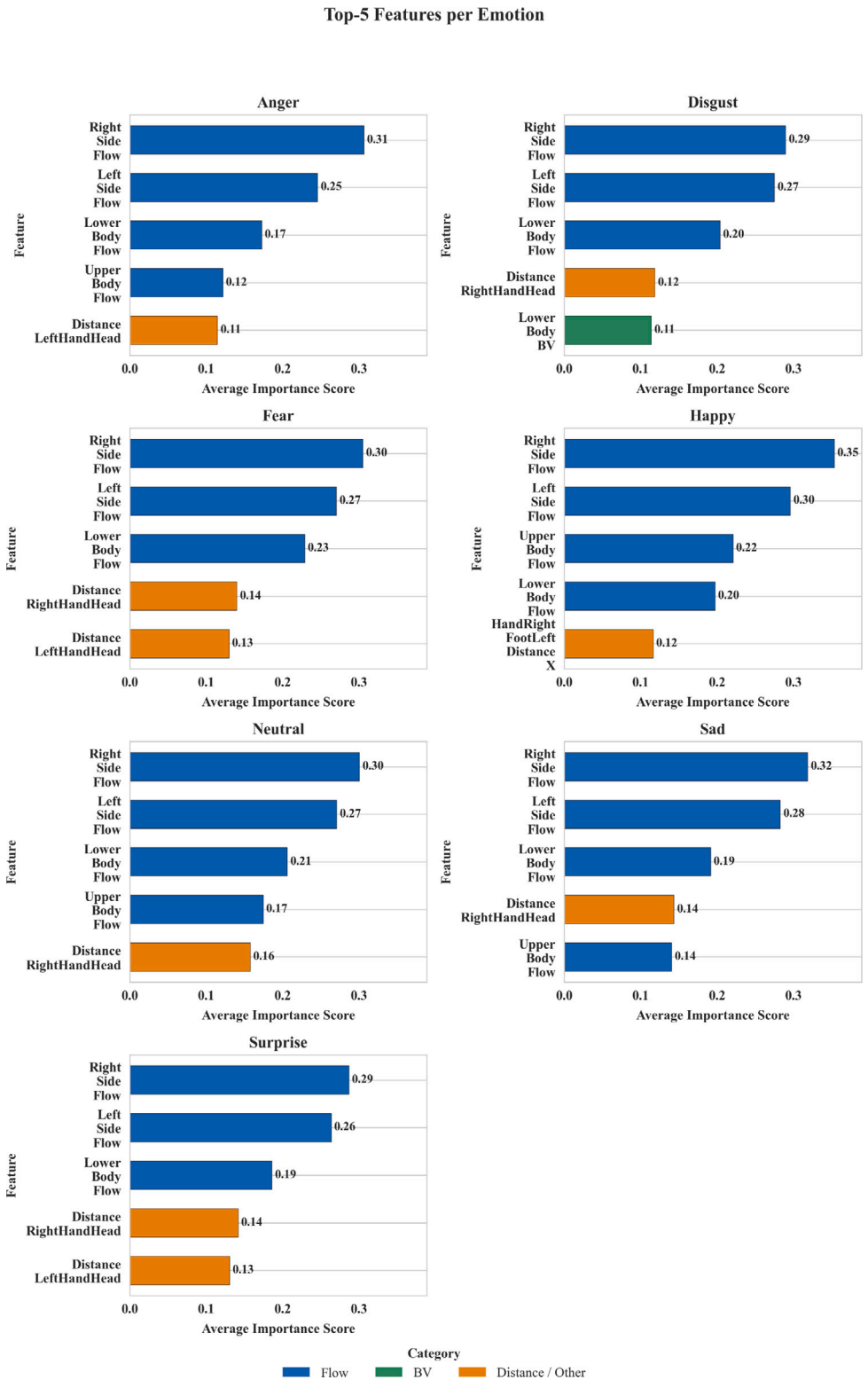


Fig. 9. Top-5 ranked features for each emotion.

($U = 2681$, $p < .001$), Sad ($U = 2644$, $p < .001$), and Surprise ($U = 2647$, $p < .001$). These consistently significant results underscore the robustness of high-level features in capturing dynamic emotional expressions across the full affective spectrum.

These combined results underline a fundamental insight: high-level features effectively encapsulate dynamic and holistic characteristics of emotional body movements, such as fluidity, expansiveness, and movement coherence. They significantly outperform low-level distance metrics, which predominantly capture static spatial relationships between specific body parts. Low-level features, though useful, consistently exhibit lower importance, likely due to their inherent limitations in

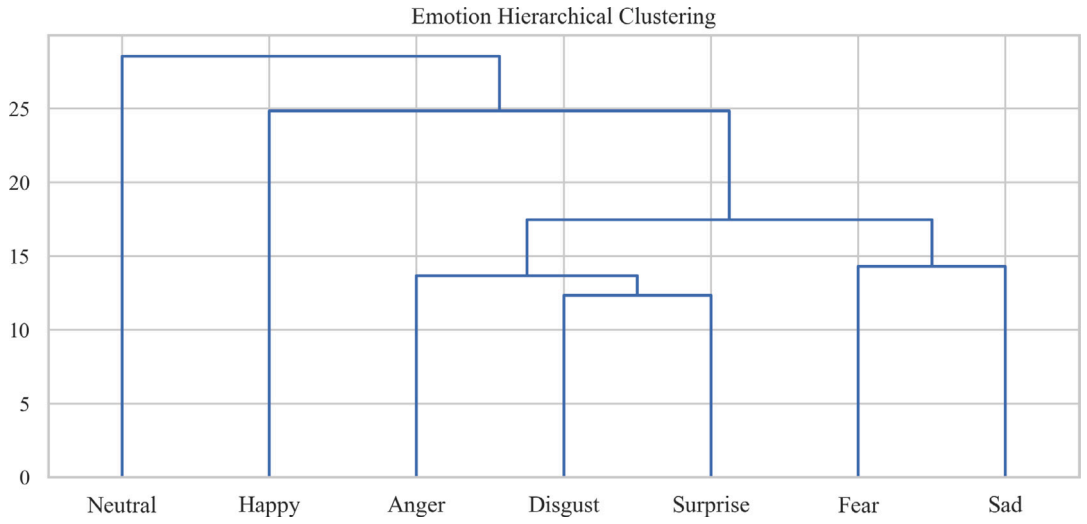


Fig. 10. Hierarchical clustering of emotions based on feature-importance profiles (Ward linkage).

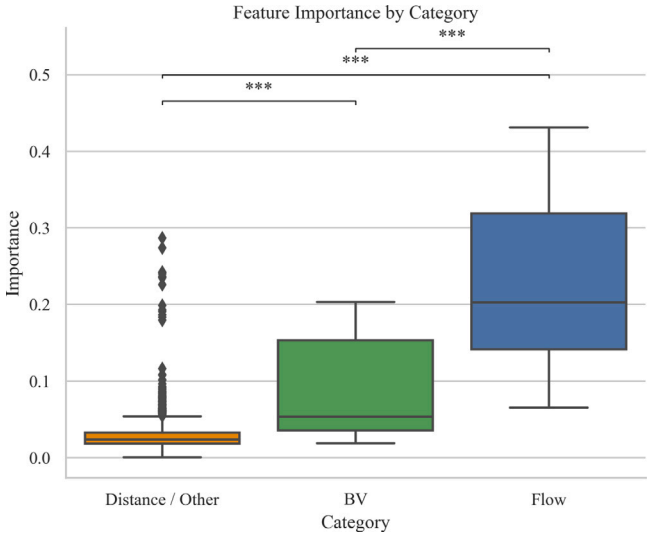


Fig. 11. Feature importance by category. Horizontal bars indicate Mann–Whitney significance tests (*** $p < 0.001$).

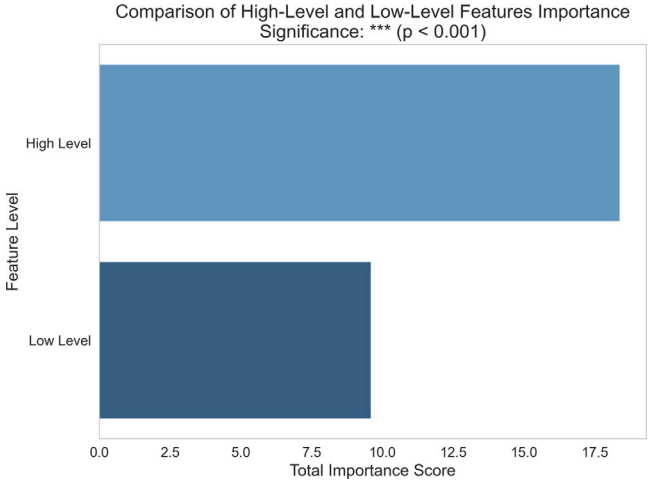


Fig. 12. High-level vs. Low-level feature importance (total importance score). Mann–Whitney $U = 73\,638$, $z = -5.00$, $r = .17$, $p < .001$.

Table 6
Descriptive statistics of feature importance by category (mean ± SD).

Category	Mean	SD	Feature count
Flow	0.231	0.114	56
Distance/Other	0.030	0.027	1267
BV	0.085	0.061	56

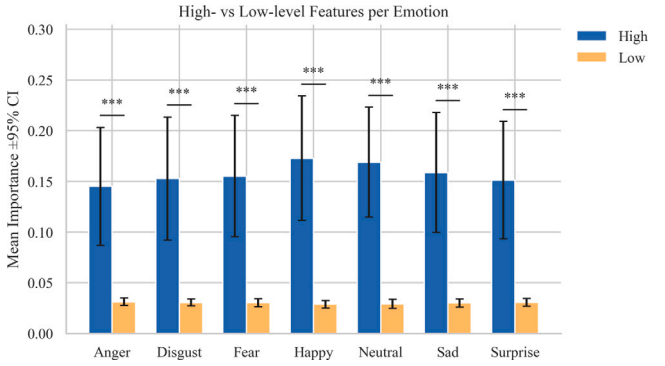


Fig. 13. High-level vs. Low-level feature importance per emotion (mean ± 95% CI). All Mann–Whitney U tests: $p < 0.001$.

accounting for individual variability in physical structure and habitual posture.

These findings advocate the prioritisation of high-level dynamic features in future computational models for emotion recognition, significantly enhancing their sensitivity, accuracy, and robustness. Such models can more effectively accommodate variability in emotional expression, thus improving their reliability and applicability in real-world human–computer interaction scenarios.

6.3.3. Feature analysis by body part

To gain deeper insights into the role of different body regions in emotional expression, we categorised body movements into four distinct regions: left side, right side, lower body, and upper body (Figs. 2 and 3). This analysis builds on previous findings emphasising the superior importance of dynamic and coordinated features over static spatial metrics.

Fig. 14 presents the average feature importance across four body regions. A Kruskal–Wallis test revealed significant differences among regions, $H(3) = 172.26$, $p < .001$. Pairwise Mann–Whitney U tests

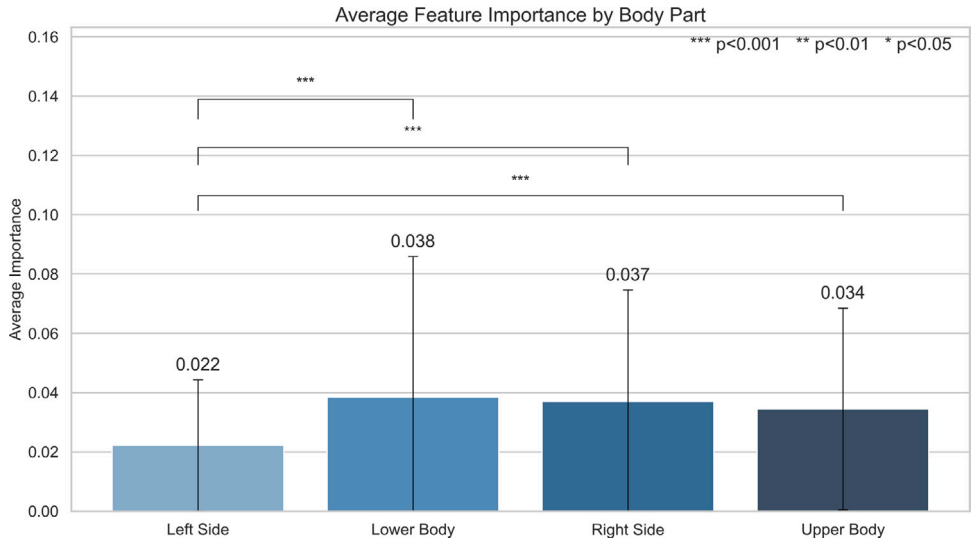


Fig. 14. Average feature importance by body part.

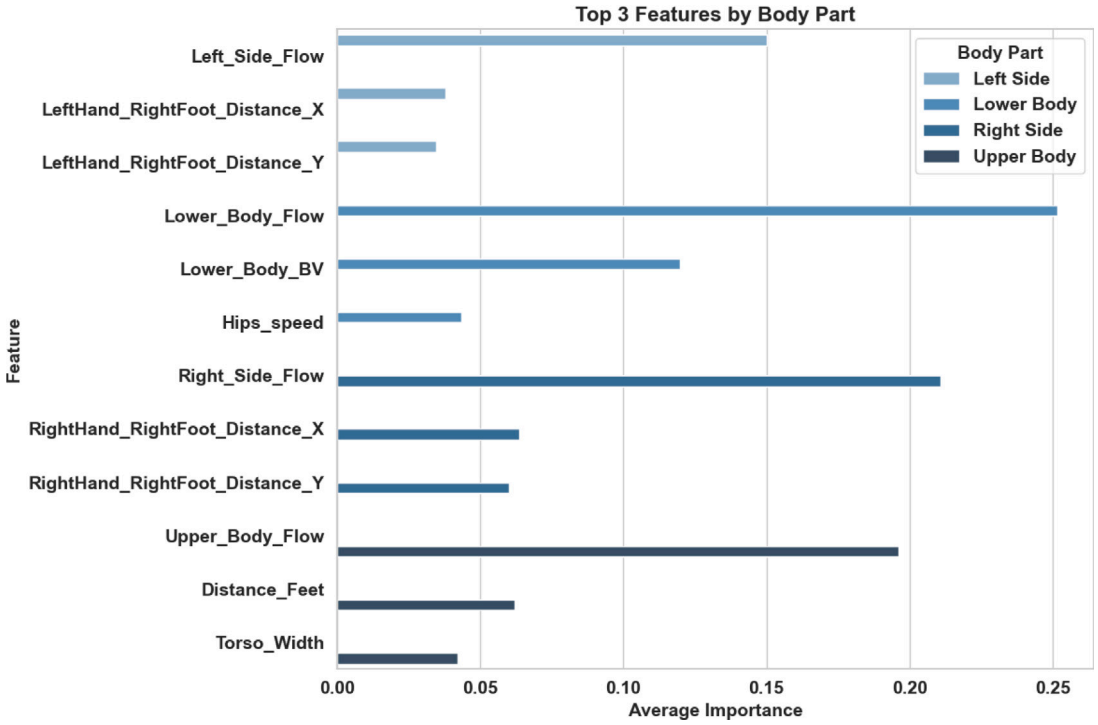


Fig. 15. Top 3 features in each body part.

showed that the left side had significantly lower importance than the lower body ($U = 12\,682$, $p < .001$), the right side ($U = 14\,458$, $p < .001$), and the upper body ($U = 15\,219$, $p < .001$). No significant differences were found between lower body versus right side ($p = 1.00$), lower body versus upper body ($p = 1.00$), and right side versus upper body ($p = 1.00$).

These findings underscore the crucial role of expressive movements associated with the lower body (e.g., stepping, stomping, weight shifting) and dynamic gestures involving the dominant limb (right side), both of which showed significantly higher importance than the left side. The lack of significant differences among lower body, right side, and upper body suggests these regions contribute similarly to emotion recognition when considering their most important features.

Fig. 15 further clarifies this pattern by identifying the top three features per body part. Consistent with earlier results, dynamic flow-related features, notably `Lower_Body_Flow` and `Right_Side_Flow`, dominate feature importance rankings, reinforcing the necessity of capturing motion fluidity and continuity. In contrast, static features such as inter-limb distances (`LeftHand_RightFoot_Distance`) show comparatively lower relevance, highlighting limitations of purely spatial information for robust emotional classification.

Taken together, these analyses suggest a clear prioritisation strategy for future model development. Emphasising dynamic and coordinated movements, particularly those originating from the lower body and dominant side, promises the greatest improvement in emotional recognition accuracy. By targeting these informative body parts and movement types, computational models can achieve finer sensitivity to subtle

Table 7
Model accuracy comparison between two datasets.

Model	EGBM accuracy	KDAEE accuracy
Low-Level LSTM	0.89	0.79
High-Level LSTM	0.93	0.85
Low-Level Bayesian LSTM	0.91	0.82
High-Level Bayesian LSTM	0.93	0.87
All-Level LSTM	0.96	0.87
HBP-LSTM	0.98	0.88

All accuracy values are reported in decimal format (e.g., 0.88 = 88%).

emotional nuances, ultimately enhancing their real-world applicability and effectiveness in human–computer interaction scenarios.

7. Results

Several models are constructed for comparative analysis to validate the robustness of the HBP-LSTM framework. The distinctive aspects of each model are summarised as follows:

- *Low-Level LSTM*: Utilises 41 low-level kinematic features.
- *High-Level LSTM*: Incorporates 80 high-level features derived from LMA (Laban Movement Analysis) to capture the emotional essence in body dynamics.
- *Low-Level Bayesian LSTM*: Extends the Low-Level LSTM by integrating Bayesian inference into the LSTM layers.
- *High-Level Bayesian LSTM*: Enhances the High-Level LSTM with Bayesian LSTM layers for probabilistic modelling.
- *All-Level LSTM*: Combines both low and high-level features in a non-Bayesian LSTM framework, serving as a baseline for comparison.
- *HBP-LSTM*: The proposed framework that synergises LSTM and Bayesian inference across all feature levels for enhanced emotion recognition.

For each model, we used the leave-one-subject-out cross-validation method, in which we systematically removed one participant from the dataset in each round of training and evaluation. This measure assesses the generalisability of the model to participants not included in the training data. All models were trained on a high-performance machine equipped with an NVIDIA RTX 4090 GPU, utilising CUDA acceleration for optimised computation.

As shown in Table 7, after examining the performance on both the EGBM and KDAEE datasets, most models achieve high accuracy levels in emotion recognition, with the exception of the Low-Level LSTM on the KDAEE dataset, which has an accuracy of 79%. This highlights the importance of high-level features for sentiment recognition. The All-Level LSTM, which pools both high and low-level data, achieved 96% accuracy on EGBM and 87% accuracy on KDAEE, demonstrating the impact of integrated data quality on model performance.

The accuracy of the models remained high after adding the Bayesian layer. The accuracy of the low-level LSTM model increased by 2 percentage points on both datasets, indicating the effectiveness of Bayesian inference in enhancing the performance of low-dimensional data processing. In contrast, the HBP-LSTM model achieves an accuracy of 98% on the EGBM dataset and 88% on the KDAEE dataset, which is an improvement of 2 percentage points and 1 percentage point over the All-Level LSTM, respectively. This indicates that the combination of Bayesian inference and LSTM has significant advantages in interpreting complex emotional states and adapting to different datasets.

To assess the robustness of the model, we adopt a multidimensional approach that simulates the challenges associated with data quality and integrity in a real-world environment by introducing multiple perturbations in the data. Specifically, we design and implement three complementary strategies to comprehensively assess the ability of the model to maintain performance in the face of different unpredictable conditions.

Introducing Controlled Noise: We introduced additive Gaussian noise into the data with zero mean and standard deviations varying over the ranges [0, 1], [−5, 5] and [−10, 10], respectively. These ranges correspond to different noise intensities and are intended to simulate signal attenuation and distortion encountered in real-world data acquisition. By evaluating the model's performance under these controlled perturbation conditions, we further analysed its adaptability and robustness to moderate noise environments common in real-world applications.

Unrestricted Noise Introduction: We introduced additive Gaussian noise with zero mean and large variance (e.g., variance > 100) across the feature set and did not set a priori limits on the noise intensity. This high-intensity noise is intended to simulate severe data corruption or extreme environmental changes to fully assess the robustness of the model under highly unpredictable conditions. With this strategy, we can test whether the model is effective in maintaining its performance despite significant degradation in the quality of the input data.

FGSM Adversarial Testing: To assess the vulnerability of the model to adversarial attacks, we use the Fast Gradient Sign Method (FGSM) (Naqvi et al., 2023). FGSM works by calculating the gradient of the loss function for the input features and adding small perturbations in the direction that increases the loss (scaled by the coefficient ϵ) to generate adversarial samples. In our experiments, we selected $\epsilon = 0.01$ and $\epsilon = 0.1$ as two perturbation strengths to simulate how minor but intentional input modifications can significantly affect model performance. Through this test, we can deeply analyse the model's robustness in the face of malicious attacks and small input changes.

Together, these three strategies constitute a rigorous test of the model, designed to simulate a variety of challenges that may be encountered in real-world emotion recognition scenarios. These strategies not only cover the issue of data corruption due to noise, but also test against deliberate adversarial manipulation. In practice, data quality may be affected by a wide range of factors, and therefore, evaluating the performance of the models under these conditions is essential to verify their adaptability, robustness and reliability. The following section describes each robustness testing strategy in detail, and their specific impact on model performance is analysed.

7.1. Comparison with standard machine learning approaches

Both datasets (EGBM and KDAEE) use the same seven emotion categories: anger, disgust, fear, happiness, neutral, sadness, and surprise, framing the task as a 7-way classification problem. We employ Leave-One-Participant-Out (LOPO) cross-validation, where for a dataset with S participants, each fold reserves one participant for testing and uses the remaining $S-1$ for training/validation. This yields S_{EGBM}=16 folds for EGBM and S_{KDAEE}=22 folds for KDAEE. The random-chance accuracy is $\frac{1}{7} \approx 0.143$ under uniform class distribution. All results report macro-averaged accuracy per fold, presented as mean ± standard deviation values across all dataset folds.

We used two datasets (EGBM and KDAEE) to compare the performance of the HBP-LSTM model with traditional machine learning algorithms including Random Forests (RF) (Wu and Chang, 2024), K-Nearest Neighbours (KNN), Support Vector Machines (SVM) (Khan et al., 2024), and Multi-Layer Perceptrons (MLPs). For these traditional algorithms, we extracted six statistical features from the body movement data: minimum, maximum, standard deviation, variance, skewness and kurtosis. These features are based on the descriptions in Figs. 2 and 3, and are computed for different body parts such as upper body, lower body, left side, right side, and torso, using velocity and acceleration data collected during the experiment. The extracted features provide a comprehensive characterisation of the movement

Table 8
Emotion recognition performance on the **EGBM** dataset (LOPO; mean ± SD).

Algorithm	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	Mean ± SD
SVM	0.5250	0.4267	0.3750	0.5823	0.5467	0.3784	0.2875	0.4457 ± 0.041
MLP	0.5000	0.4533	0.4000	0.5570	0.5733	0.3784	0.4125	0.4678 ± 0.038
KNN	0.3250	0.3067	0.2625	0.5696	0.6800	0.3649	0.2000	0.3849 ± 0.052
RF	0.3250	0.4133	0.4625	0.6329	0.7733	0.2703	0.3500	0.4604 ± 0.049
HBP-LSTM	0.9850	0.9800	0.9750	0.9900	0.9950	0.9750	0.9600	0.9817 ± 0.006

SD is the sample standard deviation across all LOPO folds (16 folds for EGBM).

Table 9
Emotion recognition performance on the **KDAEE** dataset (LOPO; mean ± SD).

Algorithm	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	Mean ± SD
SVM	0.4100	0.4350	0.3850	0.4550	0.4250	0.4200	0.4660	0.4230 ± 0.035
MLP	0.4150	0.4300	0.4000	0.4400	0.4250	0.4325	0.4140	0.4220 ± 0.036
KNN	0.3400	0.3600	0.3300	0.3550	0.3700	0.3450	0.3535	0.3505 ± 0.045
RF	0.4600	0.4800	0.4750	0.4500	0.4850	0.4700	0.4700	0.4700 ± 0.039
HBP-LSTM	0.8750	0.8800	0.8650	0.9000	0.8850	0.8700	0.8900	0.8807 ± 0.008

SD is the sample standard deviation across all LOPO folds (22 for KDAEE).

patterns and are used to train machine learning models for emotion recognition on both datasets.

The performance of the models for emotion recognition is shown in Tables 8 and 9. For both datasets, all algorithms perform better than chance level ($\frac{1}{7} \approx 0.143$), although true positive rates are consistently lower for emotions like sadness and surprise. This aligns with previous research (Castellano et al., 2007; Visi et al., 2017; Kleinsmith and Bianchi-Berthouze, 2013) that suggests that sadness and surprise have less distinctive motion patterns, leading to lower classification accuracy.

In both datasets, the KNN algorithm generally shows lower accuracy across most emotion classifications, except for neutral emotion. The SVM model demonstrates reasonable performance for most emotions but struggles with low-energy emotions such as sadness and high-energy emotions like surprise. The MLP model offers slightly better accuracy for happiness and neutral emotions but faces challenges with less distinctive emotions.

Across both datasets, none of the four traditional algorithms compares favourably to the HBP-LSTM in terms of accuracy. The HBP-LSTM achieves significantly higher accuracy levels — over 98% mean accuracy on the EGBM dataset and 88% on the KDAEE dataset — highlighting the importance of temporal information for emotion recognition. By leveraging the sequential nature of LSTM networks, HBP-LSTM effectively models the temporal dynamics of emotions, capturing subtle changes and patterns in body movements over time that traditional algorithms may miss.

Comparing the experimental results on the two datasets, we find that HBP-LSTM consistently outperforms traditional machine learning algorithms. The performance of the KDAEE dataset is degraded, which can be attributed to differences in data characteristics, such as the diversity of participants or variations in recording conditions. In contrast, the traditional algorithms showed greater fluctuations in performance between the two datasets, suggesting that they are more sensitive to dataset-specific features and, thus, less able to generalise across different populations or recording conditions.

Overall, the HBP-LSTM model significantly outperforms both dataset's traditional machine-learning methods for emotion recognition. These results highlight the importance of combining time series and probabilistic modelling techniques in developing emotion recognition systems, providing strong support for improving model robustness and adaptability.

7.2. Comparison with other state-of-the-art methods

In this section, we compare our results with existing state-of-the-art methods on two datasets. To ensure the fairness of the comparison, we strictly follow the cross-validation strategies set out in the literature for each dataset. For both datasets, we used 10-fold cross-validation and a 'Leave-One-Participant-Out' (LOPO) protocol. The following is a specific description of the methods used in the comparison:

(1) EGBM Dataset: Sapiński et al. (2019a) proposed an approach based on body joint analysis using sequential models to capture effective movement for emotion recognition. Zhang et al. (2021) developed an attention-based LSTM network that improves accuracy by focusing on key joint movements. Wang et al. (2024a) introduced a framework using a body expression energy model and a multi-input symmetric positive definite matrix network to extract temporal and spatial features.

(2) KDAEE Dataset: Avola et al. (2022) proposed a pipeline utilising multi-view representation learning (MVRL) for affective action recognition. Ghaleb et al. (2021) represented posture sequences as graphs and employed spatio-temporal graph convolutional networks (ST-GCNs) for emotion recognition.

Our proposed Hybrid Bayesian Pre-trained LSTM (HBP-LSTM) model is compared with the above methods on the respective datasets. Table 10 summarises the comparison results.

As shown in Table 10, our proposed HBP-LSTM model achieves superior performance on both datasets. Specifically, on the EGBM dataset, our method achieves an accuracy of 98.17%, surpassing the previous best result of 97.43% by Wang et al. On the KDAEE dataset, our model attains an accuracy of 88.07%, significantly outperforming the methods by Avola et al. and Ghaleb et al.

In contrast, previous methods, such as those by Sapiński et al. and Zhang et al. primarily relied on low-level joint data without incorporating high-level movement analysis, which may limit their ability to capture the full spectrum of emotional expressions. While Wang et al.'s method achieved high accuracy by using energy models and SPD networks, it may not effectively model uncertainty or handle noisy data.

On the KDAEE dataset, the methods by Avola et al. and Ghaleb et al. employed multi-view learning and graph convolutional networks, respectively, but they may not fully capture the temporal dynamics and uncertainty modelling provided by our HBP-LSTM. Our method's superior performance on KDAEE demonstrates its effectiveness in generalising across different datasets and handling variations in data quality.

Overall, our HBP-LSTM enhances emotion recognition accuracy and robustness by combining multi-level feature integration, Bayesian inference, and attention mechanisms.

7.3. Component contribution via ablation study

To quantify the contribution of each architectural block in HBP–LSTM, we conduct a controlled ablation on the EGBM seven-class

Table 10
Comparison of methods on EGBM and KDAEE Datasets.

Research	Dataset	Methodology	Recognition performance
Sapiński et al. (2019a)	EGBM	RNN-LSTM	0.6900
Zhang et al. (2021)	EGBM	AS-LSTM*	0.7410
Wang et al. (2024a)	EGBM	BEEM* + SPDnet*	0.9743
Avola et al. (2022)	KDAEE	MVRL*	0.6410
Ghaleb et al. (2021)	KDAEE	ST-GCN*	0.6500
Our Method	EGBM, KDAEE	HBP-LSTM*	EGBM: 0.9817; KDAEE: 0.8807

* Abbreviations: AS-LSTM: LSTM network based on attention, GCN: Graph Convolution Network, HMM: Hidden Markov Model, L-GrIN: Learnable Graph Inception Network, LMA: Laban Movement Analysis, LSTM: Long Short-Term Memory, MVRL: Multi-View Representation Learning, Pos: Position, RegNetY-800MF: Regular Network Y-800 MegaFLOPs, RF: Random Forest, RNN: Recurrent Neural Network, Rot: Rotation, ST-GCNs: Spatio-Temporal Graph Convolutional Networks, SVM: Support Vector Machine.

Table 11
Ablation on the EGBM seven-class dataset (LOPO, n=8 folds).

Variant	Accuracy	ΔAcc	Sig.
LSTM–Base	0.2749 ± 0.0423	−0.3875	‡
Bayes–LSTM–1	0.2674 ± 0.0581	−0.3950	‡
Bayes–LSTM–1 + Attn (no Adapter)	0.3069 ± 0.1075	−0.3555	‡
HBP–LSTM (full)	$\mathbf{0.6624 \pm 0.1144}$	*Ref.*	–
HBP–LSTM (– SecondBayes)	0.6171 ± 0.1132	−0.0453	†
HBP–LSTM (– FirstBayes)	0.6270 ± 0.0889	−0.0354	n.s.

Notes. ΔAcc = variant − full; Sig.: Holm–Bonferroni corrected paired t-tests (†$p < 0.05$, ‡$p < 0.01$; n.s.: not significant).
Variant definitions: **LSTM–Base** deterministic 2-layer LSTM (no Bayesian/attention/adapter); **Bayes–LSTM–1** Bayesian weights in the *first* LSTM only; **Bayes–LSTM–1 + Attn (no Adapter)** adds multi-head temporal attention, adapter removed; **HBP–LSTM (full)** first Bayesian LSTM + attention + adapter + second Bayesian LSTM; **(– SecondBayes)/(– FirstBayes)** replace the corresponding Bayesian LSTM with a deterministic one.

task under the same *Leave-One-Participant-Out (LOPO)* protocol (8 folds), preprocessing, and optimiser as in Section 6. All variants use a lightweight configuration of 25 training epochs and hidden size H=64; absolute accuracies are therefore lower than Table 7 but comparable *across* variants. We report mean±std over folds and assess significance via paired two-tailed t-tests with Holm–Bonferroni correction. For clarity, we also report the marginal difference

$$\Delta\text{Acc} = \text{Acc}_{\text{variant}} - \text{Acc}_{\text{full}},$$

where the "full" model is the complete HBP–LSTM.

Table 11 yields four observations. (1) *Post-fusion Bayesian inference is crucial*: removing the *second* Bayesian LSTM reduces accuracy by 4.53 pp (†), indicating that modelling uncertainty *after* low–high feature fusion is key for cross-participant generalisation. (2) *Early Bayesian uncertainty is secondary*: replacing the *first* Bayesian LSTM causes a smaller, non-significant drop (3.54 pp), suggesting that under a lightweight setting (25 epochs, H=64) early Bayesian modelling is less influential than later-stage inference. (3) *Attention helps even without the adapter*: adding multi-head attention to a single-Bayesian model improves accuracy by 3.95 pp; the adapter remains necessary for peak performance. (4) *Synergy is indispensable*: removing all advanced modules (LSTM–Base) leads to a 38.75 pp loss (‡), confirming that robustness stems from the *combination* of Bayesian inference, attention, and adapter rather than any single component.

7.4. Noise tolerance test

7.4.1. Model performance under fixed noise ranges

In this approach, noise was generated by sampling from a uniform distribution within specified ranges, such as U(0,1), U(−5,5), and U(−10,10), and then added to the original feature values. This ensures that the noise has different magnitudes, enabling us to test the model under varying degrees of data corruption.

The generated noise was independently applied to each feature across the entire dataset. In other words, for each data point, all feature values were adjusted by adding a randomly selected noise value from the corresponding distribution. This process was repeated for each noise intensity level, creating multiple noisy versions of the dataset.

The choice of different noise intensity ranges was made to simulate various levels of signal degradation that might occur in real-world scenarios. For example, noise within the [0,1] simulates minor fluctuations, such as those caused by sensor precision limitations or environmental factors, leading to minor data collection errors. Noise in the [−5,5] represents moderate degradation, possibly due to transient hardware issues or temporary environmental disturbances. Noise in the [−10,10] range introduces severe noise, analogous to sensor malfunctions or widespread environmental interference.

The initial trial employed a noise range of 0 to 1, corresponding to a relatively mild level of distortion. As illustrated in Table 12, at a 10% noise level, the Low-Level (LL) LSTM model achieved accuracies of 89.63% on the EGBM dataset and 79.03% on the KDAEE dataset. In comparison, the High-Level (HL) LSTM model attained accuracies of 86.00% on EGBM and 85.00% on KDAEE. These results indicate that the LL LSTM exhibits greater robustness to low-intensity noise than the HL LSTM. However, the performance of the HL LSTM model demonstrated a higher dependency on data quality. As the noise level increased to 70%, both LL and HL LSTM models experienced declines in accuracy across both datasets. Specifically, at 70% noise, the LL LSTM maintained accuracies of 74.71% on EGBM and 69.05% on KDAEE, whereas the HL LSTM's accuracies decreased to 42.73% on EGBM and 45.73% on KDAEE. These findings underscore the superior resilience of the LL LSTM model under severe noise conditions compared to the HL LSTM model.

To further clarify which part of our model contributes most to noise robustness, we performed a targeted comparison in which Gaussian noise was added exclusively to the low-level branch (41-dimensional kinematic inputs) or exclusively to the high-level branch (LMA-derived features). The results, see Table 12, show that when only the low-level inputs are corrupted, the drop in accuracy for the deterministic LL-LSTM (e.g. from 89.6%→83.7% on EGBM at 30% noise) is only slightly larger than for the Bayesian LL-LSTM (from 92.3%→84.7%), indicating a modest gain. By contrast, when only the high-level inputs are corrupted, the deterministic HL-LSTM's accuracy falls much more steeply (e.g. 86.0%→72.6% on EGBM) than the Bayesian HL-LSTM (93.8%→84.8%). In other words, introducing Bayesian inference brings a substantially larger robustness benefit to the high-level branch. This occurs because high-level LMA features are aggregated statistics; thus noise perturbs them, and the network can only recover by leveraging weight uncertainty. Low-level kinematic sequences, however, already contain strong temporal redundancy, so even a deterministic LSTM can partially "average out" the frame-wise noise. Hence, the Bayesian HL-LSTM (together with its Adapter+Attention+second-stage Bayesian LSTM) is the key driver of robustness when noise exceeds 30%.

Incorporating Bayesian inference into the LSTM models (*LL Bayesian LSTM* and *HL Bayesian LSTM*) results in a consistent performance improvement across most noise levels for both the EGBM and KDAEE datasets. At a 30% noise level, the *LL Bayesian LSTM* outperforms the standard *LL LSTM* by approximately 1.0% on EGBM (84.71% vs. 83.74%) and by 1.8% on KDAEE (72.01% vs. 70.23%). Similarly,

Table 12
Testing results for various models with noise range 0–1.

Model*	10% noise		30% noise		50% noise		70% noise	
	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE
LL LSTM	0.8963	0.7903	0.8374	0.7023	0.8073	0.7532	0.7471	0.6905
HL LSTM	0.8600	0.8500	0.7256	0.7250	0.6483	0.6034	0.4273	0.4573
LL Bayesian LSTM	0.9227	0.8223	0.8471	0.7201	0.8398	0.7622	0.6961	0.7010
HL Bayesian LSTM	0.9380	0.8733	0.8485	0.7402	0.6754	0.7011	0.6036	0.6505
All-Level LSTM	0.9663	0.8734	0.8600	0.7701	0.7661	0.7400	0.5783	0.6300
HBP-LSTM	0.9863	0.8304	0.8637	0.7536	0.8619	0.7323	0.8619	0.7136

* LL LSTM = Low-Level LSTM, HL LSTM = High-Level LSTM, etc.

Table 13
Testing results for various models with noise range −5–5.

Model*	10% noise		30% noise		50% noise		70% noise	
	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE
LL LSTM	0.6519	0.6012	0.2947	0.2021	0.2431	0.1941	0.2118	0.1941
HL LSTM	0.8729	0.7301	0.4751	0.3243	0.3370	0.2341	0.2947	0.2240
LL Bayesian LSTM	0.8416	0.7322	0.4052	0.3543	0.2449	0.2351	0.2063	0.1963
HL Bayesian LSTM	0.8802	0.7652	0.5046	0.5123	0.3966	0.3542	0.3016	0.2502
All-Level LSTM	0.9024	0.7833	0.5930	0.5520	0.4254	0.3970	0.4070	0.3540
HBP-LSTM	0.9466	0.8133	0.7974	0.7236	0.7422	0.6543	0.6262	0.6132

* LL LSTM = Low-Level LSTM, HL LSTM = High-Level LSTM, etc.

Table 14
Testing results for various models with noise range −10 to 10.

Model*	10% noise		30% noise		50% noise		70% noise	
	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE	EGBM	KDAEE
LL LSTM	0.3591	0.3233	0.2118	0.2013	0.2173	0.1941	0.1860	0.1531
HL LSTM	0.5064	0.4563	0.2910	0.02510	0.2615	0.1942	0.2357	0.1831
LL Bayesian LSTM	0.5801	0.5501	0.3131	0.2812	0.2560	0.2341	0.2449	0.1832
HL Bayesian LSTM	0.6072	0.5814	0.3324	0.3021	0.2816	0.2513	0.2672	0.1941
All-Level LSTM	0.6538	0.6516	0.4088	0.3875	0.3757	0.2316	0.3407	0.1941
HBP-LSTM	0.8122	0.7532	0.6372	0.7032	0.5304	0.5103	0.4088	0.3980

* LL LSTM = Low-Level LSTM, HL LSTM = High-Level LSTM, etc.

the *HL Bayesian LSTM* achieves a 12.3% accuracy improvement over the standard *HL LSTM* on EGBM (84.85% vs. 72.56%) and a 1.5% improvement on KDAEE (74.02% vs. 72.50%) at the same noise level, indicating the robustness of the Bayesian approach in mitigating the impact of noise.

In the comparative analyses, the HBP-LSTM model demonstrates excellent adaptability with Bayesian inference and multi-level feature integration. Even at a high noise level of 70%, the model still achieves robust accuracies of 86.19% and 76.61% on the KDAEE dataset and the EGBM dataset, respectively, which demonstrate the model's excellent robustness under data distortion conditions and highlight its stable performance in challenging environments.

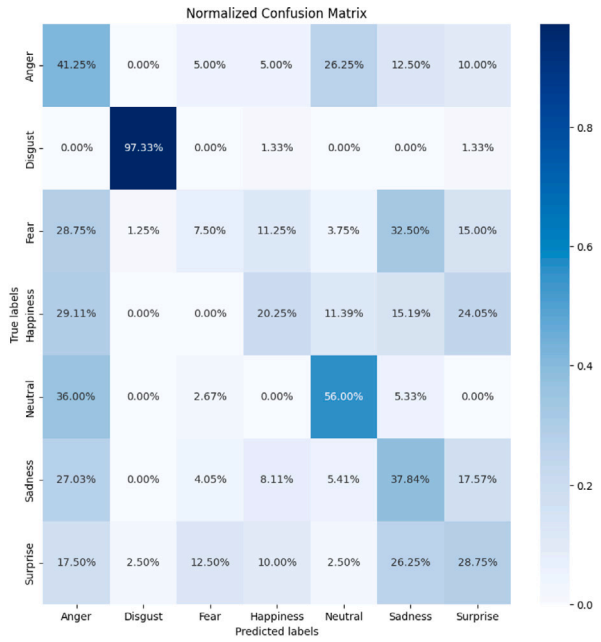
When the noise range is extended to [−5, 5], as presented in Table 13, the performance trajectories of the low-level (LL) and high-level (HL) LSTM models exhibit significant differences. Initially, at lower noise levels, the LL LSTM demonstrates robustness; however, at a 70% noise level, its accuracy sharply declines to 20.63% (EGBM) and 19.63% (KDAEE). In contrast, the HL LSTM exhibits greater noise robustness, with a more gradual decrease in accuracy under the same noise intensity. Notably, in lower-noise environments, the LL model appears to be as fault-tolerant as, or even more fault-tolerant than, the HL model, as shown in Table 12. This robustness may be attributed to the effectiveness of direct low-level features in capturing essential affective information when the data is relatively clean. However, as noise levels increase, the advantages of high-level features become more pronounced. High-level features provide a broader context and intrinsic connections, which are crucial for extracting meaningful patterns from complex and noisy data. This shift underscores the importance of integrating low- and high-level features to enhance model resilience and interpretability under varying noise conditions.

Taking the All-level LSTM as an example, it generally outperforms single-layer models (LL and HL), particularly under moderate noise conditions. Within the noise range of [−5, 5], at a noise level of 30%, the All-level LSTM maintains relatively high accuracies of 79.74% (EGBM) and 72.63% (KDAEE), indicating its superior ability to mitigate noise by leveraging a more comprehensive perspective of the data. However, as shown in Table 14, under the most extreme noise condition of [−10, 10], the full-layer LSTM, while still outperforming single-layer models, experiences a significant performance decline, with accuracy dropping to approximately 40% at a noise level of 70%.

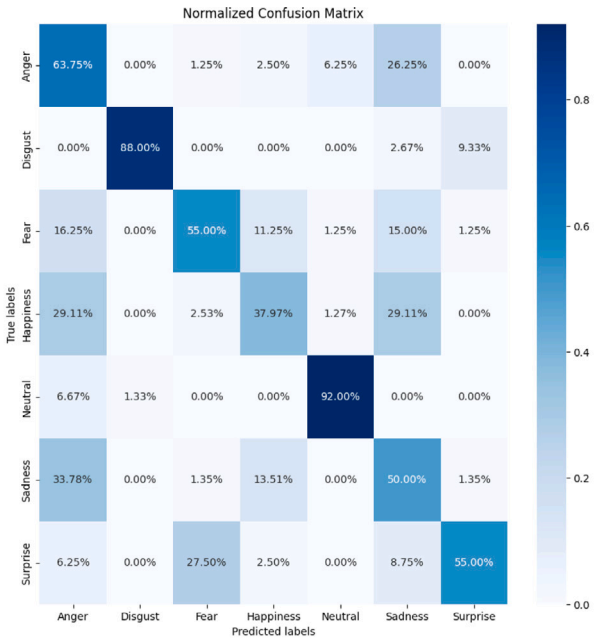
Fig. 16 compares the normalised confusion matrices of the All-Level LSTM and HBP-LSTM models at different noise levels across the EGBM and KDAEE databases. Fig. 16(a) shows the confusion matrix for the All-Level LSTM model on the EGBM database at a 70% noise level. The model achieves a high accuracy of 97.33% in recognising the "Disgust" emotion. However, the model exhibits significant confusion in distinguishing "Happiness" from "Surprise" and "Fear" from "Sadness", leading to higher misclassification rates. This indicates that, despite its overall good performance, the All-Level LSTM struggles to distinguish between emotions with similar expressive patterns in high-noise environments.

In contrast, Fig. 16(b) shows that the HBP-LSTM model exhibits greater robustness at the same database and noise level. The model achieves over 88% accuracy in recognising "Disgust" and significantly reduces the confusion rates between "Happiness" and "Surprise" as well as "Fear" and "Sadness". This suggests that HBP-LSTM has a better discriminatory ability in handling noisy data and is more effective at parsing complex emotional expressions.

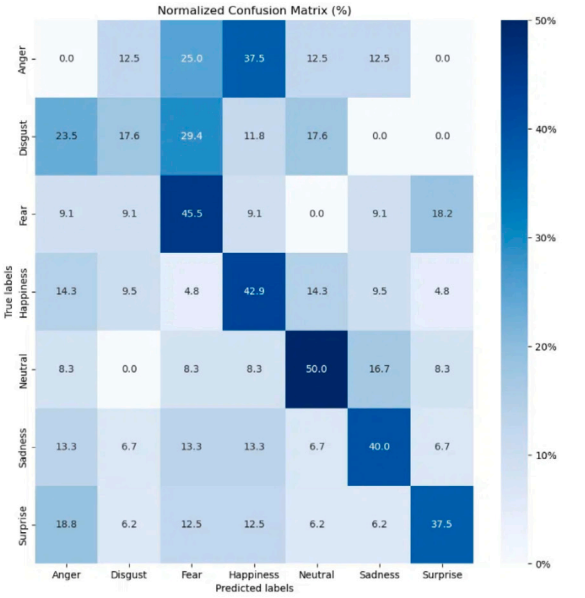
Fig. 16(c) shows the confusion matrix for the All-Level LSTM model on the KDAEE database at a 50% noise level. The model achieves recognition accuracies of 0.0% for "Anger" and 17.6% for "Disgust",



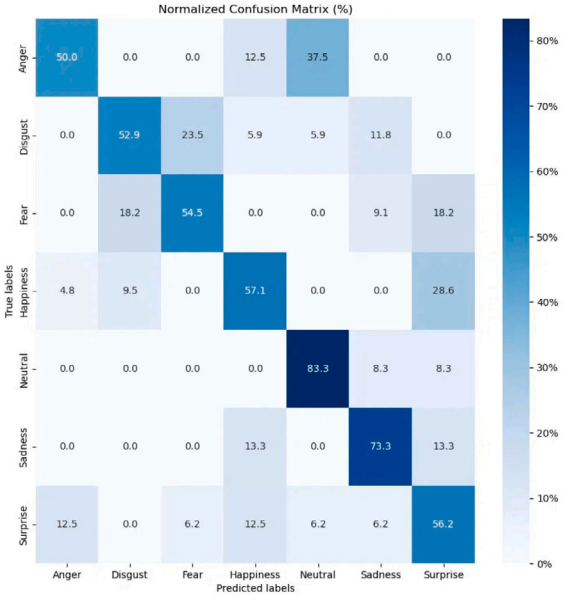
(a) All-Level LSTM Normalized Confusion Matrix for EGBM at 70% Noise Level (-5 to 5)



(b) HBP-LSTM Normalized Confusion Matrix for EGBM at 70% Noise Level (-5 to 5)



(c) All-Level LSTM Normalized Confusion Matrix for KDAEE at 50% Noise Level (-5 to 5)



(d) HBP-LSTM Normalized Confusion Matrix for KDAEE at 50% Noise Level (-5 to 5)

Fig. 16. Comparison of normalised confusion matrices for all-level LSTM and HBP-LSTM models on EGBM and KDAEE datasets at different noise levels.

with 45.5% and 42.9% accuracy for "Fear" and "Happiness", respectively. Notably, "Neutral" and "Sadness" are recognised with moderate accuracies of 50.0% and 40.0%. However, "Surprise" is accurately identified only 37.5% of the time. The All-Level LSTM model exhibits significant challenges in accurately recognising "Anger", which is wholly misclassified, and shows overlaps between "Disgust", "Fear", and other emotions like "Neutral" and "Sadness". Additionally, there is confusion between "Happiness" and "Surprise", indicating that the All-Level LSTM struggles to distinguish nuanced emotional states within the KDAEE dataset under high-noise conditions.

Fig. 16(d) illustrates the confusion matrix for the HBP-LSTM model on the KDAEE database at the same 50% noise level. The HBP-LSTM model demonstrates improved recognition accuracies across most emotion categories, achieving 50.0% for "Anger", 52.9% for "Disgust", 54.5% for "Fear", and 57.1% for "Happiness". "Neutral" and "Sadness" are recognised with high accuracies of 83.3% and 73.3%, respectively, while "Surprise" is accurately identified 56.2% of the time. Although "Anger" remains a challenging category, the HBP-LSTM model significantly reduces misclassification rates for "Disgust" and "Surprise", minimising confusion with other emotions. These improvements suggest that the HBP-LSTM model offers better resilience and discriminatory power in handling noisy data, leading to more accurate emotional recognition in the KDAEE database. However, further optimisation is needed to enhance the recognition of "Anger" and "Surprise".

Table 15
Model performance evaluation under unrestricted random noise conditions for EGBM and KDAEE Datasets.

Model/noise level*	EGBM				KDAEE			
	3%	5%	7%	10%	3%	5%	7%	10%
All-Level LSTM	0.3757	0.3462	0.3278	0.2910	0.3625	0.3401	0.3155	0.2882
HBP-LSTM	0.4880	0.4346	0.4144	0.3554	0.4752	0.4298	0.3981	0.3491

* Noise levels are simulated with completely random values, without fixed range constraints, to evaluate model resilience in unpredictable conditions. Performance is shown separately for EGBM and KDAEE datasets.

7.4.2. Model robustness under unrestricted noise conditions

To assess the model's robustness under extreme and unpredictable conditions, we introduce noise with no predefined limits. This noise is randomly generated and imposed on the entire feature set and designed to simulate real-world scenarios such as sensor failures, system errors, or extreme environmental disturbances.

In contrast to the first strategy, this method's noise intensity is not systematically controlled but randomly sampled from a broad spectrum that may contain outliers and extreme values. The high noise size and direction variability pose a more significant challenge for the model.

Applying noise indiscriminately across the entire feature set means that each feature may be disturbed to varying degrees in magnitude and type. This setup simulates a realistic situation in which some data streams may have severe errors while others remain stable, thus providing a severe test of the model's ability to process under chaotic and unexpected input conditions. We aim to examine the model's ability to generalise and maintain performance under the most challenging conditions by introducing noise without predetermined limits. Such tests provide an essential basis for understanding the robustness and adaptability of the model in real-world applications.

This subsection explores the model's performance under unrestricted random noise conditions, reflecting the unpredictable data corruption encountered in real-world scenarios. Observing the model's performance decline in such a challenging environment provides valuable insights into its robustness.

Table 15 presents the results from this rigorous testing, showcasing how each model variant contends with a spectrum of noise levels from mild (3%) to significant (10%).

The results highlight the challenge that unrestricted noise poses to model robustness. At a noise level of 10%, HBP-LSTM demonstrates a stronger adaptive capability with an accuracy of 35.54% (EGBM dataset) and 34.91% (KDAEE dataset), compared to 29.10% and 28.82% for All-Level LSTM. This result demonstrates the superior performance of HBP-LSTM in dealing with unstructured noise.

Despite the performance degradation observed, the relative stability of the HBP-LSTM in adverse conditions is promising. It accentuates the model's applicability in emotion recognition within complex and unpredictable environments. Future endeavours may concentrate on refining the model's architecture, potentially through optimising its Bayesian elements or adopting advanced regularisation techniques to further attenuate the noise impacts. Such enhancements are pivotal for bolstering the model's resilience and ensuring the reliability of emotion recognition applications.

The above analysis shows that the HBP-LSTM model maintains a significant advantage in different noise levels, consistently delivering high emotion recognition rates. This consistent performance across various noise levels, especially when data integrity is at stake, speaks volumes about the model's potential for application in different environments. The superiority of the HBP-LSTM is most evident when compared to the All-Level LSTM, which is significantly less sensitive to accuracy degradation.

Furthermore, these results reaffirm the importance of robust model design in the field of emotion recognition. As the technique is integrated with dynamic, real-world environments, the ability to maintain high recognition rates in the presence of imperfect data will be critical.

7.5. FGSM adversarial testing

Adversarial attacks were conducted using FGSM with perturbation magnitudes $\epsilon = 0.01$ (minor) and $\epsilon = 0.10$ (significant). This setup aimed to evaluate how each model withstands varying levels of adversarial noise, simulating scenarios where inputs are subtly or heavily manipulated to degrade performance.

Figs. 17(a) and 17(b) illustrate the models' accuracies under FGSM attacks on the EGBM and KDAEE datasets, respectively. The HBP-LSTM (blue solid line) consistently outperforms the Normal LSTM (red dashed line) at lower perturbation levels ($\epsilon = 0.01$ and 0.05), indicating superior resilience to mild adversarial noise. However, at the higher perturbation level ($\epsilon = 0.10$), the Normal LSTM achieves higher accuracy — approximately 22% for EGBM and similar trends for KDAEE — compared to the HBP-LSTM's 15%. This suggests that while HBP-LSTM excels under moderate adversarial conditions, the simpler architecture of the Normal LSTM may better handle extreme, generalised noise. Fig. 17 provides a comparative overview across both datasets. It highlights that the HBP-LSTM maintains higher accuracy at lower ϵ values but is outperformed by the Normal LSTM when perturbations are severe. This outcome underscores a trade-off between model complexity and generalisation capabilities.

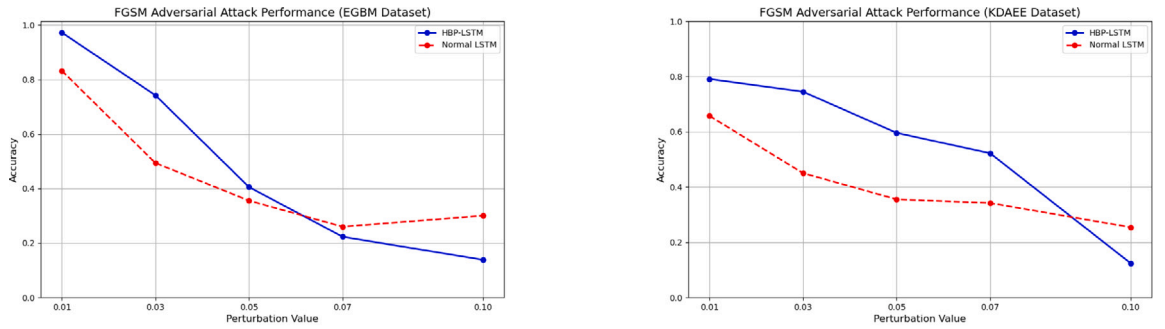
In the previous noise test (see Section 7.4), HBP-LSTM outperformed ordinary LSTM at all noise levels, demonstrating its superior ability to handle random and unstructured noise. However, in the FGSM adversarial test, although the HBP-LSTM still performed well at low to medium perturbation strengths, it was no match for the ordinary LSTM at high perturbation strengths. This phenomenon indicates that the complex architecture of the HBP-LSTM may have limitations when dealing with intentional perturbations in the gradient direction, especially at high perturbation strengths, where the model may be more susceptible to severe interference due to its complexity. The Normal LSTM performs better under FGSM high perturbation strength, possibly due to its simple structure reducing the risk of overfitting, giving it better generalisation ability under extreme noise conditions. This comparison highlights the impact of different noise types on model robustness and emphasises the need to trade off complexity and generalisation ability in model design for various application scenarios.

8. Discussion

The research presented explores the use of body movements for emotion recognition, a rapidly growing area within affective computing. Our proposed Hybrid Bayesian Pre-trained LSTM (HBP-LSTM) framework significantly enhances emotion recognition accuracy and robustness by integrating low-level pose features with high-level kinematic features and incorporating Bayesian inference.

Our experimental results demonstrate that the HBP-LSTM model achieves accuracies of 98.17% on the EGBM dataset and 88.07% on the KDAEE dataset, outperforming existing state-of-the-art methods. These results validate the effectiveness of our approach in handling noisy and uncertain data, and its superior performance in emotion recognition tasks.

The high accuracy achieved by the HBP-LSTM model underscores the importance of integrating both low-level and high-level features for comprehensive emotion recognition. By leveraging Bayesian inference, our model effectively manages data uncertainty and noise, which are



FGSM Adversarial Testing on EGBM Dataset FGSM Adversarial Testing on KDAEE Dataset

Fig. 17. Comparison of model performance under FGSM adversarial attacks.

prevalent in real-world scenarios. This robustness makes our framework highly suitable for practical applications where data quality cannot always be controlled.

Unlike previous approaches focusing solely on low-level joint data or incorporating attention mechanisms without addressing uncertainty, our model's combination of multi-level feature integration and Bayesian inference provides a more nuanced and reliable emotion recognition system. This dual-level feature approach captures both the detailed movements and the overarching kinematic patterns that convey emotional states. Incorporating Bayesian inference within the LSTM framework enhances the model's ability to handle variability and unpredictability in movement data. This adaptability is crucial for deploying emotion recognition systems in dynamic environments where data can be highly variable and subject to noise.

The success of the HBP-LSTM model paves the way for deploying emotion recognition systems in real-world applications, particularly in scenarios where data quality cannot be guaranteed. Applications include real-time monitoring, human–computer interaction, and assistive technologies. Our model can contribute to more reliable and effective emotion-aware systems in these domains by ensuring high accuracy and robustness.

Through comprehensive feature analysis using a Random Forest classifier, we evaluated the importance of various features, compared high-level and low-level features, and analysed contributions from different body parts across emotional categories. The results revealed that certain body parts and features play a more significant role in accurately conveying emotional states. Specifically, high-level kinematic features related to the upper body and hands were particularly influential. This analysis enhances our understanding of emotion recognition mechanisms and provides valuable insights for further refining our model.

8.1. Limitations

Although the HBP-LSTM model exhibits excellent noise immunity and high accuracy in sentiment recognition, it still has some limitations that need to be further explored:

First, the model's performance in the face of severe or uncontrolled random noise suggests that there is still room for improvement in dealing with extreme data corruption. This implies that more advanced noise filtering techniques or more robust model architectures are needed to ensure that relevant features can be effectively extracted despite severe data distortion.

Second, this study relies on a specific dataset for training and validation, which may limit the model's ability to generalise to different real-world scenarios. The existing dataset may not be able to cover all emotional expressions and environmental changes encountered in real-world applications. Therefore, future research should consider incorporating a wider range of datasets that cover different cultural backgrounds, emotional nuances, and diverse environmental conditions to enhance the model's adaptability and robustness.

In addition, although Bayesian methods help to enhance model robustness and uncertainty management, they have high computational complexity and consume a large amount of computational resources. This somewhat limits the feasibility of the model in real-time applications and large-scale deployment. Therefore, exploring more efficient Bayesian inference techniques or developing lightweight model architectures would help alleviate this problem and make the approach more practical in resource-constrained environments.

Finally, although the present model performs well in controlled experimental environments, its performance in real-world applications still needs to be thoroughly evaluated. In real-world applications, data may be affected by various unforeseen perturbations and complexities, so extensive field testing and user studies will provide valuable insights into understanding the model's real-world effectiveness and point the way to further improvements.

8.2. Responsible AI development

This study develops the Hybrid Bayesian Pre-trained LSTM (HBP-LSTM) framework, a unique technology that involves both low-level posture characteristics and high-level kinematic features and includes Bayesian inference. Our way brings about considerably improved body movement-based emotion recognition, especially in the presence of serious noise. The results present the model's resistance and flexibility in uncertain, noisy data ecosystems, demonstrating its potential for real-world applications where data integrity may vary.

Through comprehensive feature analysis using a Random Forest classifier, we evaluated the importance of various features, compared high-level and low-level features, and analysed contributions from different body parts across emotional categories. The outcomes unveiled that certain body parts are the most critical body features accurately reflecting emotions. Specifically, the upper body and hands have been mentioned as higher-level kinematic features that contribute mainly to this end. This analysis enhances our understanding of emotion recognition mechanisms and provides valuable insights for further refining our model.

Regarding the findings from the feature analysis, further research will concentrate on the most influential features and body sections to optimise the HBP-LSTM model. By concentrating on the most influential features, we aim to develop a more efficient model with reduced computational complexity without compromising performance.

Further, we want to explore more advanced deep learning algorithms to deal and analyse data more comprehensively. This might involve employing convolutional neural networks (CNNs) to find and extract local information or transformer to bottle the links between distant processes in motion sequences. These enhancements may provide

deeper insights into the nuances of the data's features and contribute to a more sophisticated understanding of emotion recognition.

Moreover, we recognise the importance of evaluating our model in real-world settings. Future work will involve deploying the HBP-LSTM model in practical applications, such as real-time emotion recognition systems, to validate its generalisation capabilities and robustness in diverse environments. Lastly, considering the computational demands of Bayesian methods, we will investigate efficient inference techniques or develop lightweight model architectures to facilitate real-time applications. This will make the approach more practical for deployment in resource-constrained environments.

CRediT authorship contribution statement

Shuang Wu: Writing – original draft, Software, Conceptualization. **Daniela M. Romano:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to extend our gratitude to Dr. Tomasz Sapiński for providing access to the Multi-modal Dataset of Emotional Speech, Video, and Gestural Data (Sapiński et al., 2019b). Furthermore, we thank Mingming Zhang and colleagues for making the Kinematic Dataset of Actors Expressing Emotions (Zhang et al., 2020) publicly available.

References

Ahmed, F., Bari, A.H., Gavrilova, M.L., 2019. Emotion recognition from body movement. IEEE Access 8, 11761–11781.

Avola, D., Cascio, M., Cinque, L., Fagioli, A., Foresti, G.L., 2022. Affective action and interaction recognition by multi-view representation learning from handcrafted low-level skeleton features. Int. J. Neural Syst. 32 (10), 2250040.

Bartenieff, I., Lewis, D., 2013. Body Movement: Coping with the Environment. Routledge.

Bernardet, U., Fdili Alaoui, S., Studd, K., Bradley, K., Pasquier, P., Schiphorst, T., 2019. Assessing the reliability of the laban movement analysis system. PloS One 14 (6), e0218179.

Canal, F.Z., Müller, T.R., Matias, J.C., Scotton, G.G., de Sa Junior, A.R., Pozzebon, E., Sobieranski, A.C., 2022. A survey on facial emotion recognition techniques: A state-of-the-art literature review. Inform. Sci. 582, 593–617.

Cao, W., Zhong, J., Cao, G., He, Z., 2019. Physiological function assessment based on kinect v2. IEEE Access 7, 105638–105651.

Castellano, G., Villalba, S.D., Camurri, A., 2007. Recognising human emotions from body movement and gesture dynamics. In: International Conference on Affective Computing and Intelligent Interaction. Springer, pp. 71–82.

De Meijer, M., 1989. The contribution of general features of body movement to the attribution of emotions. J. Nonverbal Behav. 13, 247–268.

Ebdali Takalloo, L., Li, K.F., Takano, K., 2022. An overview of emotion recognition from body movement. In: Computational Intelligence in Security for Information Systems Conference. Springer, pp. 105–117.

Ekman, P., 1984. Expression and the nature of emotion. Approaches Emot. 3 (19), 344.

Elansary, L., Taha, Z., Gad, W., 2024. Survey on emotion recognition through posture detection and the possibility of its application in virtual reality. arXiv preprint arXiv:2408.01728.

Geetha, A., Mala, T., Priyanka, D., Uma, E., 2024. Multimodal emotion recognition with deep learning: advancements, challenges, and future directions. Inf. Fusion 105, 102218.

Ghaleb, E., Mertens, A., Asteriadis, S., Weiss, G., 2021. Skeleton-based explainable bodily expressed emotion recognition through graph convolutional networks. In: 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition. FG 2021, IEEE, pp. 1–8.

Goodfellow, I., 2016. Nips 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160.

Hachimura, K., Takashina, K., Yoshimura, M., 2005. Analysis and evaluation of dancing movement based on LMA. In: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. IEEE, pp. 294–299.

Khan, T.A., Sadiq, R., Shahid, Z., Alam, M.M., Su'ud, M.B.M., 2024. Sentiment analysis using support vector machine and random forest. J. Inform. Web Eng. 3 (1), 67–75.

Khare, S.K., Blanes-Vidal, V., Nadimi, E.S., Acharya, U.R., 2023. Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. Inf. Fusion 102019.

Kleinsmith, A., Bianchi-Berthouze, N., 2013. Affective body expression perception and recognition: A survey. IEEE Trans. Affect. Comput. 4 (1), 15–33.

Laban, R., Ullmann, L., 1971. The mastery of movement. ERIC.

Larboulette, C., Gibet, S., 2015. A review of computable expressive descriptors of human motion. In: Proceedings of the 2nd International Workshop on Movement and Computing. pp. 21–28.

Melzer, A., Shafir, T., Tsachor, R.P., 2019. How do we recognize emotion from movement? Specific motor components contribute to the recognition of each emotion. Front. Psychol. 1389.

Naqvi, S.M.A., Shabaz, M., Khan, M.A., Hassan, S.I., 2023. Adversarial attacks on visual objects using the fast gradient sign method. J. Grid Comput. 21 (4), 52.

Oğuz, A., Ertuğrul, Ö.F., 2024. Emotion recognition by skeleton-based spatial and temporal analysis. Expert Syst. Appl. 238, 121981.

Papatheodorou, D., 2024. Functional Bayesian neural networks with non-stationary Gaussian process priors and belief matching.

Reed, C.L., Moody, E.J., Mgrublian, K., Assaad, S., Schey, A., McIntosh, D.N., 2020. Body matters in emotion: restricted body movement and posture affect expression and recognition of status-related emotions. Front. Psychol. 11, 1961.

Sapiński, T., Kamińska, D., Pelikant, A., Anbarjafari, G., 2019a. Emotion recognition from skeletal movements. Entropy 21 (7), 646.

Sapiński, T., Kamińska, D., Pelikant, A., Ozcinar, C., Avots, E., Anbarjafari, G., 2019b. Multimodal database of emotional speech, video and gestures. In: Pattern Recognition and Information Forensics: ICPR 2018 International Workshops, CVAUI, IWCF, and MIPPSNA, Beijing, China, August 20-24, 2018, Revised Selected Papers 24. Springer, pp. 153–163.

Shaarani, A.S., Romano, D.M., 2007. Perception of emotions from static postures. In: Affective Computing and Intelligent Interaction: Second International Conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings 2. Springer, pp. 761–762.

Shafir, T., 2023. Modeling emotion perception from body movements for human-machine interactions using laban movement analysis. In: Modeling Visual Aesthetics, Emotion, and Artistic Style. Springer, pp. 313–330.

Surace, L., Patacchiola, M., Battini Sönmez, E., Spataro, W., Cangelosi, A., 2017. Emotion recognition in the wild using deep neural networks and Bayesian classifiers. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction. pp. 593–597.

Tang, J., Ma, Z., Gan, K., Zhang, J., Yin, Z., 2024. Hierarchical multimodal-fusion of physiological signals for emotion recognition with scenario adaption and contrastive alignment. Inf. Fusion 103, 102129.

Visi, F., Coorevits, E., Schramm, R., Miranda, E., 2017. Musical instruments, body movement, space, and motion data: music as an emergent multimodal choreography.

Wallbott, H.G., 1998. Bodily expression of emotion. Eur. J. Soc. Psychol. 28 (6), 879–896.

Wang, W., Enescu, V., Sahli, H., 2015. Adaptive real-time emotion recognition from body movements. ACM Trans. Interact. Intell. Syst. (TiiS) 5 (4), 1–21.

Wang, T., Liu, S., He, F., Du, M., Dai, W., Ke, Y., Ming, D., 2024a. Affective body expression recognition framework based on temporal and spatial fusion features. Knowl.-Based Syst. 112744.

Wang, H., Zhao, C., Huang, X., Zhu, Y., Qu, C., Guo, W., 2024b. Emotion recognition in dance: A novel approach using laban movement analysis and artificial intelligence. In: International Conference on Human-Computer Interaction. Springer, pp. 189–201.

Wu, Y.c., Chang, Y.l., 2024. Ransomware detection on linux using machine learning with random forest algorithm. Authorea Prepr..

YuMeng, Z., Zhen, L., TingTing, L., YuanYi, W., YanJie, C., 2024. Affective-pose gait: perceiving emotions from gaits with body pose and human affective prior knowledge. Multimedia Tools Appl. 83 (2), 5327–5350.

Zambeli, J.G.A., Lira, A.A.d.M., Cassol, M., 2024. Recognition of emotions by voice and facial expression by medical students. Audiol.-Commun. Res. 29, e2889.

Zhang, S., Yang, Y., Chen, C., Zhang, X., Leng, Q., Zhao, X., 2023. Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects. Expert Syst. Appl. 121692.

Zhang, H., Yi, P., Liu, R., Zhou, D., 2021. Emotion recognition from body movements with as-lstm. In: 2021 IEEE 7th International Conference on Virtual Reality. ICVR, IEEE, pp. 26–32.

Zhang, M., Yu, L., Zhang, K., Du, B., Zhan, B., Chen, S., Jiang, X., Guo, S., Zhao, J., Wang, Y., et al., 2020. Kinematic dataset of actors expressing emotions. Sci. Data 7 (1), 292.