1a. Project title

Tennis Emotion Analysis for Mates

1b. Project acronym

TEAM: Tennis **E**motion **A**nalysis for **M**ates

1c. Applicants

Name: Lorenzo Gandini, Muhammad Feroz Din

E-mail: lorenzo.gandini@studenti.unitn.it , muhammadferoz.din@studenti.unitn.it

Student#: 248430, 246722

Institute: University of Trento

2a. Summary of research proposal (max 250 words)

Emotions play a critical role in sports performance, influencing decision-making, focus, and resilience. While individual emotion recognition in sports has been explored [1][2][5], limited research addresses doubles tennis, where the interplay between teammates' emotions and dynamics is pivotal. Studies on emotional intelligence and synchrony in team sports [3][4][11] highlight the importance of chemistry and coordination between players, but an analysis of such factors in doubles tennis remains underexplored.

This project aims to develop a system that analyzes emotions and team chemistry in doubles tennis using advanced computer vision techniques. Facial expressions and body gestures will be extracted from high-quality match footage using ResNet50 for facial emotion recognition [5][6][7], OpenPose for pose estimation [8][9], and YOLO for player detection [10]. Multimodal approaches [11][12] will integrate these features to:

- Detect individual players' emotional states during critical match events (e.g., breakpoints, rallies, unforced errors).
- Assess team chemistry by evaluating synchronization, emotional congruence, and mutual responsiveness to events.

Building upon existing work in emotion recognition [5][6][7][12], gesture analysis [9], and tennis action classification [13][14], the system will utilize broadcast-quality video to ensure real-world applicability. The output will include annotated match timelines, event detection [13], and emotion-chemistry metrics synchronized with match events.

This research will provide actionable insights for athletes and coaches, enabling data-driven post-match evaluations and tailored training programs. By bridging emotion recognition, event detection, and team dynamics, this project advances the understanding of interpersonal performance factors in doubles tennis and sets a foundation for future applications in team sports analytics.

2b. Abstract for laymen (max 250 words)

In every tennis match, emotions play a crucial role. Players deal with pressure, celebrate successes, and navigate setbacks—all of which shape their performance. In doubles tennis, this dynamic becomes even more complex, as success depends not just on individual skill but also on the chemistry between teammates.

Our project seeks to illuminate these emotional and interpersonal dynamics by analyzing facial expressions, body gestures, and synchronized movements using advanced artificial intelligence. Through AI technologies like ResNet50 for facial analysis, OpenPose for body gesture recognition, and YOLO for player detection, we will uncover how emotions influence a match, from the confidence gained after a perfect serve to the frustration of a tough rally. Additionally, we will study the "chemistry" between teammates, evaluating how synchronization, emotional alignment, and mutual responses contribute to their collaboration and results.

Using real-world video footage from professional matches, our system will provide a timeline of emotional states, team chemistry, and critical match events. By integrating this data, we aim to deliver practical insights into the emotional and strategic aspects of doubles tennis.

The ultimate goal is to empower players and coaches with tools to understand the emotional underpinnings of performance and teamwork. These insights can refine strategies, improve training, and foster stronger partnerships on the court. This research represents a step forward in sports analytics, offering a deeper understanding of the interplay between emotions, teamwork, and success in competitive tennis.

3. Classification

Emotion Analysis: facial expression recognition, body gesture recognition, emotion classification

Sport Analytics: Match event correlation, post-match analysis tool

Computer vision: Pose estimation, Facial Landmark analysis, ResNet

4. Description of the proposed research

a. Research topic and envisaged results

In doubles tennis, individual emotions and team chemistry are critical factors influencing performance. While previous research has explored the role of emotional intelligence and synchronization in team sports [1][3], the interplay between emotions and match events remains underexplored. Existing systems for emotion recognition focus on individual sports or controlled settings, leaving a gap in tools for analyzing real-world dynamics in doubles tennis.

This project aims to develop an AI-driven system capable of:

- Recognizing individual emotions through facial expressions and body gestures using advanced deep learning techniques [5][6][8].
- Quantifying team chemistry by evaluating synchronization and congruence between players [3][4][9].
- Detecting key match events (e.g., winning a point, unforced errors) using video analysis.
- Correlating emotional states and team dynamics with match events to provide actionable insights for performance improvement.

The envisaged results include:

- A novel emotion and chemistry analysis framework tailored for doubles tennis, combining facial and body cue detection with team dynamics evaluation.
- A comprehensive dataset derived from professional match footage to train and validate the system, ensuring real-world applicability.
- An interactive dashboard visualizing match footage alongside detailed emotional and event analyses, empowering coaches and players to refine strategies and improve team coordination.

This project leverages state-of-the-art technologies like ResNet50 for facial emotion recognition [7], OpenPose for pose estimation [8], and insights from sports psychology [1][3]. By bridging AI and sports analytics, it aims to revolutionize performance analysis in doubles tennis.

b. Approach

The proposed research will use mainly computer vision techniques to analyze emotions and interactions between team players in doubles tennis. The system will process match footage, frame by frame, to extract body movements and facial expressions, classify emotional states, evaluate team chemistry, detect key match events, and correlate these factors to provide actionable insights to the match analyst.

This system works on 5 different keysteps:

1) Body and Facial Expression Recognition
2) Emotion Mapping
3) Chemistry Evaluation Between Players
4) Match Event Detection
5) Correlation of Events with Emotions

<u>1. Body and Facial Expression Recognition</u>

The first step in our approach involves extracting detailed information about players' body movements and facial expressions from match footage frame-by-frame. This granularity is needed to ensure precise tracking of rapid emotional changes during pivotal moments, such as a tiebreak or breakpoint.
We begin by implementing a YOLO (You Only Look Once) detector that will provide bounding boxes for both the player body coordinates and the player face coordinates [10]. These bounding boxes allow us to isolate the areas of interest in each frame.

To recognize the body posture, we will use OpenPose, a state-of-the-art tool for human pose estimation, to track and analyze the body movements of the players [8]. OpenPose identifies keypoints on the human body, including joints such as shoulders, elbows, hips, knees, and ankles. These physical features provide insights into player engagement (e.g., active vs. passive postures) and emotional states (e.g., upright posture could indicate confidence, whereas slumped shoulders may reflect frustration). For facial expression analysis, starting from the points detected with YOLO, we will use ResNet50, a deep convolutional neural network, to extract hierarchical features from the cropped facial regions [5][6][7]. ResNet50 focuses on minute details such as eyebrow movement, mouth curvature, and eye squinting, enabling precise classification of expressions.

This work of recognition is crucial as it forms the foundation for all subsequent phases of the system, providing as output:

- Keypoints: Skeletal points for each player, describing body posture and movement.
- Facial Features: High-dimensional vectors representing expressions, mapped to emotions like confidence, frustration, or focus.

<u>2. Emotion Mapping</u>

Once the skeletal keypoints and facial features have been extracted, the next phase involves mapping these physical characteristics to emotional states. We aim to classify the emotional states of tennis players into distinct categories that are both relevant and observable in a competitive context. We based this definition on psychological and physiological studies of emotional expression and tailored them to the context of professional tennis [1][3][6]:

Positive Emotions:
- **Confidence**: Relaxed facial muscles, upright posture, steady gaze, and fluid movements.
- **Determination**: Tightened facial features (e.g., focused eyes, slight clenching of the jaw), leaning-forward posture, and purposeful movements.
- **Satisfaction**: Relaxed facial expressions, small smiles, and open gestures, often following a successful rally or point.

Negative Emotions:
- **Frustration**: Tensed facial muscles (e.g., furrowed brows), abrupt gestures, and slumped shoulders.
- **Disappointment**: Drooping facial features (e.g., downturned mouth), slow movements, and a passive stance after errors or missed opportunities.
- **Anger**: Tightened lips, flared nostrils, rapid, forceful gestures, and aggressive postures.

Neutral States:
- **Focus**: Neutral facial expression, steady body posture, and consistent eye direction, often observed during preparation for a serve or rally.
- **Calmness**: Relaxed posture, smooth and controlled movements, with minimal facial tension, usually seen during lulls in the match or before executing a familiar action.
- **Fatigue**: Slouched shoulders, slower movements, drooping eyelids, and lack of engagement, typically emerging late in a match.

To classify emotional states accurately, the system will integrate data from multiple modalities—facial expressions and body posture—using a multimodal framework. This approach is inspired by existing works that demonstrate the effectiveness of combining visual cues for emotion recognition [11][12].
From the facial expression, we will extract using ResNet50, fine-grained expressions such as eyebrow positioning, mouth curvature, and eye squinting since this data captures micro-expressions that are critical for distinguishing between nuanced emotions like satisfaction, frustration, or determination [5][6].
Instead, from the skeletal keypoints extracted with OpenPose, we take posture characteristics (e.g., upright, slouched), movement patterns, and gestures since they are the insights able to describe engagement levels, energy, and overall body alignment [8]. The system will employ a neural network architecture designed to integrate and analyze these distinct data streams, in two branches (one for the face, one for the body) and then concatenate all together. The fused features are passed through fully connected layers, which assign probabilities to each emotion category (e.g., confidence, frustration, fatigue) [11]. The model is trained on a combined dataset of facial and posture annotations to optimize the interaction between modalities. This enables the system to analyze the relationship between modalities, such as how an upright posture with a slight smile indicates confidence, while the same posture with a furrowed brow may signal determination.
In this way, with this multi-modal approach, we will ensure that the system leverages

complementary information from both modalities. Finally each detected emotional state is assigned a precise timestamp, ensuring the future integration with the other system components (like event match and chemistry values).

3. Chemistry evaluation between players

This is a crucial point in our work since it allows us to differentiate between us and other existing works. In doubles tennis, success is not only determined by individual skill but also by the level of synchronization and chemistry between teammates.

Team chemistry can be defined as the degree of synchronization, emotional congruence, and mutual responsiveness between players.
- Synchronization of Movements: How well players coordinate their movements on the court, such as during defensive or offensive plays.
- Emotional Congruence: The degree to which teammates align emotionally in response to match events, such as celebrating a win or recovering from an error.
- Interaction Quality: Non-verbal gestures, signals, and overall responsiveness to each other's actions.

To evaluate team chemistry, we propose a novel chemistry score that combines the previous mentioned metrics. This score is continuously time-stamped, creating a detailed timeline of team dynamics that can will be aligned with the emotional states of each single player and with the match events in the final step.

The chemistry score is obtained with the following formula:

$$Chemistry = \alpha \cdot Sync + \beta \cdot Emotion_{congruence} + \gamma \cdot Interaction_{quality}$$

1) Synchronization (Sync): Since we need to evaluate how well the movements of the two players are coordinated over time, we will give in input the players' trajectories, represented as sequences of keypoints (extracted with OpenPose) and then we will use DTW (Dynamic Time Warping). This algorithm measures the similarity between two temporal sequences by aligning them to minimize time-based differences [8][11]. So for each frame, the coordinates of corresponding keypoints for both players are compared and the normalized minimum temporal distance represents the synchronization value:

$$Sync = 1 - \frac{DTW(player_1, player_2)}{max(DTW)}$$

where $DTW(player_1, player_2)$ is the temporal distance between the players' movement, and *max(DTW)* is a reference value derived from non-synchronized sessions.

2) Emotional Congruence: Measures the alignment of emotional states between teammates during match events. In order to do that, we start from the emotional states vector for both players E_1 and E_2, extracted from the emotion recognition model. These

vectors contain probabilistic values for each emotion (e.g., [confidence: 0.8, frustration: 0.2]). Since they are vectors we can compare them with a cosine similarity:

$$Emotion_{congruenc} = \frac{E_1 \cdot E_2}{||E_1|| \, ||E_2||}$$

Where E_1 and E_2 are the emotional state vectors, $||E_1||$ and $||E_2||$ are their magnitude and $E_1 \cdot E_2$ is the dot product. A value close to 1 indicates highly aligned emotions (e.g., both players showing "confidence"), while a value near -1 indicates opposing emotions (e.g., one player showing "confidence" and the other "frustration") [12]. Take this case as an example: after winning a crucial point, if both players exhibit "satisfaction" with similar emotional vectors, the $Emotion_{congruence}$ will be high. If one player shows "confidence" and the other "frustration," the score will be low.

3) Interaction quality: This score assesses the frequency and effectiveness of non-verbal interactions, such as supportive gestures or strategic signals between the players. This can be achieved as the ratio obtained by the number of interactions between the players during the match (e.g. high-fives, hugs, and finger-pointing) and the number of events (that will be explained in the next point).

$$Interaction_{quality} = \frac{N_{interactions}}{N_{events}}$$

4) Weights ($\alpha, \ \beta, \ \gamma$): This three elements determine the relative importance of each component in the overall chemistry score. Their values depend on the specific priorities of the analysis and the coach/match analyst wants to highlight:
   a) if precise coordination, like in defensive strategies, α will have a higher value
   b) if maintaining aligned emotional states is a key factor for the performance, β will have more relevance.
   c) if the analyst wants to emphasize communication and responsiveness, γ will be the relevant weight.

<u>4. Match Event detection</u>

The fourth component of the system focuses on detecting key match events. This is a critical step in understanding how players' emotions and chemistry evolve throughout the game. In order to identify each event, we will build a deep learning system, following the methodology proposed by Hovad et al. [13], that will produce a list of events and their relative timestamp. The system will utilize a recognition software inspired by the SlowFast architecture [14] which is the best suite for classifying fine-grained temporal actions in sports, allowing us to capture both temporal dynamics (like player movements) and fine details (like racket swings or ball bounces). Training this model with annotated tennis data, such as the THETIS dataset, the model will classify events into predefined categories as reported below [14] :

a. *Winning a point* occurs when a player successfully executes a rally or serve that results in a score. This event is recognized by analyzing the player's body language, such as celebratory gestures (e.g., fist pumps or raised arms), along with posture that shifts into a relaxed state. A scoreboard update validating the point gain further confirms this detection.

b. *Losing a point* is identified by gestures and reactions indicative of frustration or disappointment, such as slumped shoulders, shaking the head, or slamming the racket. The opponent's positive gestures, like fist pumps, provide additional evidence. Scoreboard updates showing a point for the opponent validate the loss.
c. *Unforced errors*, such as hitting the ball out of bounds, are detected by identifying abrupt stops in player movement and inefficient postures during a swing. Frustration gestures, such as racket tosses, and ball trajectory data (if available) showing the ball landing out of play reinforce this classification.
d. *Forced errors* result from an opponent's challenging shot. These are recognized by observing defensive postures, rushed movements such as lunges or leaning backward, and ball trajectories indicating a high-pressure situation. Synchronization between the opponent's aggressive movements and the player's defensive actions is key to this classification
e. *Aces and double faults* are specific to serves. Aces are detected when the ball lands in the service box and the opponent fails to move in response, while double faults are recognized when the server commits two consecutive errors. Scoreboard updates confirm these serve-related events.
f. *Breakpoints* are critical moments where a player has the opportunity to break the opponent's serve. These moments are identified by specific score contexts, such as 40-30 or Advantage-40, and pre-serve behaviors like focused stances or adjusting grips. Emotional tension observed through body gestures provides additional context.
g. *Rallies*, characterized by extended exchanges of shots lasting more than 10 seconds, are detected by tracking the continuous lateral movements of both players and measuring the duration of the sequence.
h. *Celebrations or emotional reactions*, such as fist pumps, jumping, or racket slams, are captured by analyzing exaggerated player movements and corresponding facial expressions. These reactions are linked to preceding events to provide a comprehensive understanding of their context, whether celebratory or frustration-driven.

5. <u>Correlation of events with emotions</u>

The final component of the system focuses on combining all outputs from the previous steps to establish meaningful connections between detected match events, emotional states, and team chemistry. By aligning these elements temporally, the system reveals how events influence player emotions and interactions, and vice versa.

Since we have a precise timestamp for everything that happened in the match (emotional states, chemistry scores, and events) we can align and link all these elements.
By examining emotional states and chemistry before and after the event, the system captures transitions and impacts, revealing the emotional buildup and aftermath. For instance, the emotional shift from focus to frustration during a missed breakpoint or the transition from shared determination to joy after winning a rally is mapped directly to the timeline. This temporal matching ensures that patterns are contextualized and linked to both individual and team dynamics, allowing us to a comprehensive understanding of performance triggers.

The output of this module is a cohesive timeline inserted into a dashboard that integrates match events, emotional states, and chemistry dynamics. Visualizations present these elements side by side, providing a clear and intuitive representation of their interplay. For example, the timeline might highlight a series of rallies with corresponding spikes in chemistry, followed by a breakpoint where frustration emerged due to miscommunication.

c. Scientific or economic relevance

This research bridges a critical gap in the scientific analysis of team sports, with doubles tennis serving as a pivotal case study. Traditional methodologies often do not analyze the emotional and psychological dimensions of performance, focusing instead on physical metrics like shot accuracy or movement patterns. Our research introduces an innovative framework for quantifying team chemistry and correlating it with match events through state-of-the-art computer vision and deep learning techniques.
This scientific contribution of this project lies in expanding the understanding of how emotions and interpersonal dynamics influence performance, providing a foundation for future studies in sports psychology, and artificial intelligence.
The integration of this multimodal analysis (body movements, facial expressions, emotion recognition, match events, and chemistry) offers a new framework for understanding such complex dynamics, setting a new benchmark for studying complex human behavior dynamics. (e.g. studies on how particular events affect a group of people).

From an economic perspective, this system has a direct application for professional sports teams and organizations:
- Coaches and analysts gain access to specific insights of the match and can think and develop ad hoc training sessions, in order to improve or fix specific behavior of the players during their matches.
- Players gain awareness of their emotional and psychological states during matches, helping them identify and address patterns more effectively.

Beyond tennis, this technology has immense potential in other team sports, including basketball, volleyball, soccer, and rugby. Its scalable nature also makes it adaptable for amateur-level applications, widening its market appeal.

d. International developments

Several projects and studies have explored components of this field, though none have fully integrated emotion recognition, chemistry evaluation, and event detection into a single system designed to doubles tennis.

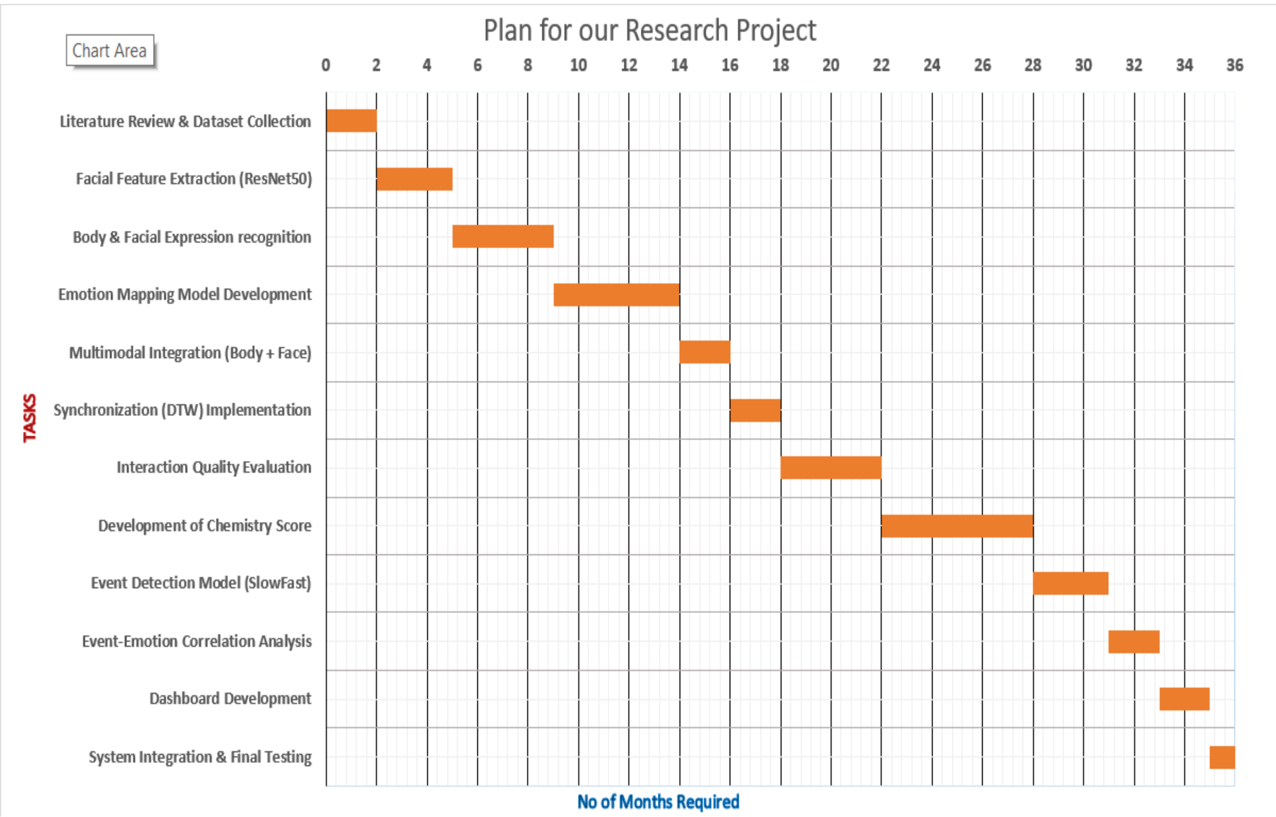
In the domain of emotion recognition, works like those by Bodapati et al. [5] and Singh et al. [6] focus on facial emotion analysis using convolutional neural networks. These studies demonstrate the effectiveness of ResNet-based architectures in capturing micro-expressions but do not extend to full-body analysis or team dynamics. Similarly, Liu et al. [9] have explored multi-view body gesture recognition, showing its relevance in capturing engagement and emotional states, although their methods have not been applied specifically to sports.

In sports analytics, projects like the work by Hovad et al. [13] have demonstrated the potential of deep learning with their use of the SlowFast architecture, providing a strong foundation for event detection. While effective for video analysis, their approach is limited to action classification and does not integrate emotional analysis or investigate team chemistry, leaving a gap in understanding the interplay between physical and psychological factors

KIT University has been the first group in applying AI to single-player emotion detection, developing systems that can track and classify emotions during matches [15]. Their research showcases the feasibility of automated emotion recognition in a sports context. However, their work focuses on individual subjects and does not address the interaction between players, which is critical in doubles tennis.

To our knowledge, no existing system combines these elements into a comprehensive framework designed specifically for doubles tennis. This project can bridge that gap by building on these advancements, integrating state-of-the-art methodologies, and providing a practical, end-to-end solution that addresses both scientific and practical needs.

5. Work Programme



Equipment

6. Expected use of instrumentation

The research requires high-quality video analysis tools and computational resources. We can divide this into 3 main categories:

1. **Hardware**:
   - High-Performance GPUs: Essential for training and inference of deep learning models, such as YOLO, ResNet50, OpenPose, and SlowFast, and for processing large video datasets efficiently. Examples include NVIDIA RTX 3090 or A100.
   - Storage Solutions: Large-scale storage systems, such as NAS or SSD arrays, are required to store raw match footage provided by broadcasters as well as intermediate outputs and processed data.
2. **Software**:
   - Deep Learning Frameworks: Tools like TensorFlow or PyTorch will be used to implement and train the models, ensuring adaptability and scalability.
   - Video Processing Libraries: OpenCV will be utilized for preprocessing match footage, including cropping, resizing, and extracting frames for analysis.
   - Dashboard Development Tools: Platforms like Dash by Plotly or Flask will be employed to develop the final user interface for presenting match insights.
3. **Datasets resources**:
   - Professional Match Footage: Access to high-quality videos of professional doubles tennis matches provided by broadcasters (e.g., Sky, DAZN, CNN) is crucial for developing and validating the system.
   - Training Datasets: The following datasets will be used to train and fine-tune the models:
     i. FER2013 and AffectNet: For emotion recognition training.
     ii. COCO: For pose estimation in body recognition.
     iii. THETIS: For tennis-specific event detection.

Cost estimates

7. Requested Budget:	Amount (euro)
Salary and overhead (3 researcher for 3 years)	270 000
Taxes	4 538
sub-total	**274 538**
Commercial license for video	4 000
GPUs (3 units * + accessories)	20 000
Storage solutions : NAS system	5 000
Miscellaneous equipment: monitors, network and backup system	3 000
Total requested funding	**306 538**

* NVIDIA RTX 3090 costs € 6500

References

References Utilizzate

1. [1] Laborde, A., Dosseville, F., Guillén, F., & Chávez, E. (2016). The influence of emotional intelligence on performance in competitive sports: A meta-analytical investigation. *International Review of Sport and Exercise Psychology*, 9(3), 264–290.

2. [2] Heidari, F., Saeedi, H., & Shahhosseini, M. (2021). The association of emotional intelligence with sport injuries and receiving penalty cards among Iranian professional soccer players. *Asian Journal of Sports Medicine*, 12(2), e97321.
3. [3] Levenson, R. W., Amini, F., & Teymoorian, S. (2016). Emotional synchrony and performance in team sports. *Psychological Bulletin*, 142(1), 55–87.
4. [4] Halevy, A., Shamir, A., & Oron, S. (2019). Chemistry of team dynamics in sports: Analyzing interpersonal relationships and emotional intelligence. *Journal of Sports Analytics*, 5(3), 145–162.
5. [5] Bodapati, J. D., & Garg, S. (2021). Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, 33(4), 1273–1290.
6. [6] Singh, S. K., Roy, S., & Gupta, R. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, 2(5), 779–792.
7. [7] Kaur, H., Sharma, S., & Arora, R. (2023). An optimized facial emotion recognition architecture based on a deep convolutional neural network. *Signal, Image and Video Processing*, 17(1), 45–56.
8. [8] Costa, W., Rocha, M., & Lima, C. (2021). Multi-cue adaptive emotion recognition network for dynamic environments. *arXiv preprint arXiv:2111.02273*.
9. [9] Liu, Y., Zhang, J., & Wang, H. (2020). Emotion recognition based on multi-view body gestures. *IEEE Transactions on Affective Computing*, 11(3), 567–578.
10. [10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.
11. [11] Caridakis, G., et al. (2007). Multimodal emotion recognition from expressive faces, body gestures and speech. Artificial Intelligence Communications, 20(2), 87–96.
12. [12] Soleymani, M., et al. (2017). Multimodal emotion recognition in response to videos. IEEE Transactions on Affective Computing, 3(2), 211–223.
13. [13] Hovad, E., Hougaard-Jensen, T., & Clemmensen, L. K. H. (2024). Classification of Tennis Actions Using Deep Learning. arXiv preprint arXiv:2402.02545.
14. [14] Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). SlowFast Networks for Video Recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 6202–6211
15. [15] Jakub Jakauc, KIT University, "AI-Based Emotion Detection in Single-Player Sports," 2022.