Tạp chí Khoa học Đại học Cần Thơ
Phần A: Khoa học tự nhiên, Công nghệ và Môi trường
ISSN 1859-2333 | e-ISSN 2815-5599

DOI:10.22144/ctujos.2026.215

SUY LUẬN TRẠNG THÁI CẢM XÚC TIỀM ẨN DỰA TRÊN BIỂU CẢM KHUÔN MẶT CẦU THỦ BẰNG MÔ HÌNH CHUỖI THỜI GIAN TÍCH HỢP CHẤT LƯỢNG ẢNH

Trần Hùng Cường*, Nguyễn Quang Huy, Trần Hồng Sơn, Trần Thu Phương, Nguyễn Mạnh Quân và Phan Trường Giang
Trường Công nghệ Thông tin và Truyền thông, Đại học Công nghiệp Hà Nội, Việt Nam
**Tác giả liên hệ (Corresponding author): tranhungcuong@haui.edu.vn*

Thông tin chung (Article Information)

Nhận bài (Received): 26/02/2026
Sửa bài (Revised): 21/04/2026
Duyệt đăng (Accepted): 18/09/2026

Title: *Inferring players' psychological states using a time-series model integrated with facial image quality*

Author(s): *Tran Hung Cuong*, Nguyen Quang Huy, Tran Hong Son, Tran Thu Phuong, Nguyen Manh Quan and Phan Truong Giang*

Affiliation(s): *School of Information and Communication Technology, Ha Noi University of Industry, Viet Nam*

TÓM TẮT

Việc phân tích trạng thái cảm xúc tiềm ẩn của cầu thủ trong môi trường thi đấu thực tế là một bài toán thách thức do chất lượng video không ổn định và hạn chế của các phương pháp hiện tại trong việc mô hình hóa động thái cảm xúc theo thời gian. Bài báo được thực hiện nhằm đề xuất một hệ thống tích hợp đánh giá chất lượng ảnh khuôn mặt (FIQA) như một tín hiệu độ tin cậy vào mô hình chuỗi thời gian, cho phép theo dõi liên tục biểu cảm khuôn mặt ngay cả trong các điều kiện hình ảnh bị rung, mờ hoặc che khuất một phần. Trên cơ sở không gian trạng thái tiềm ẩn thu được, ba chỉ số định lượng đã được đề xuất gồm Latent Stability Index (LSI), Temporal Drift Index (TDI) và Latent Variability Index (LVI) nhằm đặc trưng hóa mức độ ổn định và biến động cảm xúc theo thời gian. Kết quả thực nghiệm cho thấy hệ thống có khả năng phản ánh nhất quán diễn biến trạng thái cảm xúc trong các tình huống thi đấu thực tế. Nghiên cứu được định vị như một proof-of-concept, đặt nền tảng cho hướng phân tích trạng thái cảm xúc tiềm ẩn từ video thể thao và mở ra tiềm năng ứng dụng trong hỗ trợ phân tích hiệu suất và ra quyết định chiến thuật.

Từ khóa: *Biểu cảm khuôn mặt, đánh giá chất lượng ảnh, mô hình chuỗi thời gian, phân tích tâm lý cầu thủ, trạng thái cảm xúc tiềm ẩn*

ABSTRACT

Analysing the latent emotional states of players in a real-world competitive environment is a challenging problem due to the inconsistent video quality and limitations of current methods for modelling emotional dynamics over time. This paper proposes an integrated system that assesses facial image quality (FIQA) as a reliability signal within a time-series model, allowing for continuous monitoring of facial expressions even under shaky, blurry, or partially obscured images. Based on the acquired latent state space, the study proposes three quantitative indices: Latent Stability Index (LSI), Temporal Drift Index (TDI), and Latent Variability Index (LVI) to characterise the stability and variability of emotions over time. Experimental results show that the system is capable of consistently reflecting the evolution of emotional states in real-world competition situations. The study is positioned as a proof-of-concept, laying the groundwork for analysing the underlying emotional states from sports videos and opening up potential applications in supporting performance analysis and tactical decision-making.

Keywords: *Facial expression analysis, image quality assessment, latent emotional state, player psychological analysis, temporal sequence modeling*

1. GIỚI THIỆU

Trong bóng đá hiện đại, trạng thái cảm xúc tiềm ẩn của cầu thủ – được biểu hiện qua biểu cảm khuôn mặt trong quá trình thi đấu – có ảnh hưởng trực tiếp đến hiệu suất thi đấu, khả năng ra quyết định và nguy cơ chấn thương. Tuy nhiên, các phương pháp đánh giá hiện nay vẫn chủ yếu dựa trên quan sát chủ quan, do đó thiếu các công cụ định lượng có khả năng theo dõi sự biến đổi trạng thái theo thời gian (Ivarsson et al., 2020). Trong khi đó, các biểu hiện cảm xúc quan sát được qua khuôn mặt có thể thay đổi liên tục trong suốt trận đấu và tác động tức thời đến quyết định chiến thuật cũng như điều chỉnh nhân sự. Do đó, việc xây dựng một phương pháp phân tích định lượng, khách quan nhằm theo dõi trạng thái cảm xúc tiềm ẩn dựa trên biểu cảm khuôn mặt của cầu thủ liên tục, dựa trên dữ liệu video thi đấu thực tế, là một vấn đề có ý nghĩa cả về khoa học lẫn thực tiễn.

Nhiều nghiên cứu gần đây trong lĩnh vực thị giác máy tính đã tập trung vào bài toán nhận diện biểu cảm khuôn mặt và biểu diễn sự thay đổi cảm xúc theo thời gian. Về nhận diện biểu cảm (Facial Expression Recognition – FER), các nghiên cứu tổng quan và phương pháp học sâu (Li & Deng, 2022; Azizan et al., 2024) đã khai thác hiệu quả dữ liệu ảnh và video để suy luận cảm xúc; tuy nhiên, phần lớn vẫn thực hiện phân loại ở từng khung hình độc lập, tính chuyển tiếp và tính liên tục của trạng thái chưa được xem xét đầy đủ. Về mô hình chuỗi thời gian, đã có những nghiên cứu (Kim et al., 2019; Rubanova et al., 2019; Khan et al., 2024) đề xuất các biểu diễn trạng thái tiềm ẩn nhằm mô tả quỹ đạo biến đổi theo thời gian, nhưng các cách tiếp cận này chưa được áp dụng sâu trong bối cảnh phân tích trạng thái cảm xúc biểu hiện qua khuôn mặt cầu thủ từ video thi đấu thực tế. Bên cạnh đó, trong môi trường thi đấu, chất lượng ảnh khuôn mặt thường biến động do chuyển động nhanh, che khuất và thay đổi ánh sáng. Các nghiên cứu về đánh giá chất lượng ảnh khuôn mặt (Face Image Quality Assessment – FIQA) như FaceQnet (Hernandez-Ortega et al., 2019), MagFace (Meng et al., 2021) và CLIB-FIQA (Ou et al., 2024) đã khẳng định vai trò của chất lượng ảnh đối với độ tin cậy của hệ thống nhận diện, song chủ yếu sử dụng cho mục đích sàng lọc hoặc cải thiện hiệu năng nhận dạng. Nhìn chung, các hướng nghiên cứu hiện tại vẫn thiếu một cơ chế tích hợp trực tiếp tín hiệu chất lượng ảnh vào mô hình chuỗi thời gian nhằm hỗ trợ suy luận trạng thái cảm xúc tiềm ẩn liên tục, tạo nên khoảng trống khoa học cần được bổ sung.

Xuất phát từ những hạn chế trên, nghiên cứu này được thực hiện nhằm đề xuất một hệ thống phân tích trạng thái cảm xúc tiềm ẩn dựa trên biểu cảm khuôn mặt (facial expression-based latent affective state) liên tục, dựa trên mô hình chuỗi thời gian, trong đó chất lượng ảnh được tích hợp như một tín hiệu bổ trợ trong mô hình. Hệ thống học một trạng thái tiềm ẩn (được biểu diễn dưới dạng vectơ z_t) và diễn giải thành các chỉ số định lượng LSI, TDI và LVI nhằm mô tả mức độ ổn định và biến đổi của biểu hiện cảm xúc trong suốt trận đấu. Hệ thống này hướng tới việc cung cấp thông tin có giá trị thực tiễn, hỗ trợ công tác huấn luyện và phân tích chiến thuật. Điểm mới của nghiên cứu nằm ở việc tích hợp đánh giá chất lượng ảnh trực tiếp vào mô hình chuỗi thời gian và xây dựng bộ chỉ số định lượng mới cho phân tích trạng thái tiềm ẩn. Việc thực nghiệm trên dữ liệu video thi đấu thực tế cho thấy phương pháp đề xuất cho khả năng mô hình hóa trạng thái liên tục và ổn định hơn so với các phương pháp không tích hợp tín hiệu chất lượng ảnh.

2. DỮ LIỆU VÀ KIẾN TRÚC MÔ HÌNH

2.1. Dữ liệu

Trong nghiên cứu này, ba nguồn dữ liệu độc lập được sử dụng để phục vụ các giai đoạn huấn luyện, tinh chỉnh và đánh giá mô hình. Cách phân chia này giúp đảm bảo mô hình học được đặc trưng biểu cảm tổng quát trước khi thích nghi với ngữ cảnh thể thao đặc thù.

Tập dữ liệu 1 – FER2013 (huấn luyện ban đầu ResNet-50):

Bộ dữ liệu FER2013 (Goodfellow et al., 2013) gồm 35.887 ảnh khuôn mặt (48×48 pixel) với bảy nhãn biểu cảm: vui vẻ, tức giận, ác cảm, sợ hãi, buồn, ngạc nhiên và bình thường. Tập dữ liệu này được sử dụng để huấn luyện ban đầu mô hình ResNet-50 nhằm học các đặc trưng biểu cảm tổng quát trước khi tinh chỉnh theo ngữ cảnh thể thao.

Tập dữ liệu 2 – Tập dữ liệu thể thao tự xây dựng (tinh chỉnh ResNet-50):

Do chưa có bộ dữ liệu công khai về biểu cảm cảm xúc đặc thù trong thi đấu bóng đá nên nhóm nghiên cứu đã xây dựng một tập dữ liệu riêng gồm ba nhãn: phấn khích, căng thẳng và mệt mỏi. Dữ liệu được thu thập từ khoảng 50 video thi đấu bóng đá có sẵn công khai trên YouTube, chủ yếu từ các kênh chính thức của FIFA (https://www.youtube.com/@fifa) và UEFA (https://www.youtube.com/@UEFA).

Chiến lược gán nhãn theo ngữ cảnh tình huống:

Để đảm bảo tính nhất quán và khách quan trong quá trình gán nhãn, phương pháp gán nhãn dựa trên ngữ cảnh tình huống đã được áp dụng thay vì yêu cầu người gán nhãn phán đoán trạng thái cảm xúc của từng cá nhân. Cụ thể, mỗi nhãn được xác định dựa trên bối cảnh sự kiện có thể quan sát khách quan trong video:

– Phấn khích: Trích xuất từ các đoạn video ghi lại phản ứng của cầu thủ ngay sau khi ghi bàn thắng, trong khoảng thời gian ăn mừng bàn thắng.

– Căng thẳng: Trích xuất từ các đoạn video quay cầu thủ trong quá trình chuẩn bị và thực hiện đá luân lưu (penalty shootout) – tình huống áp lực tâm lý cao và có thể xác định khách quan.

– Mệt mỏi: Trích xuất từ các đoạn video ghi lại cầu thủ trong khoảng 5 – 10 phút cuối mỗi hiệp thi đấu – giai đoạn mà mức độ mệt mỏi thể chất và tinh thần được xem là cao nhất trong trận đấu.

Sau khi thực hiện gán nhãn tự động theo ngữ cảnh, 10% số mẫu của mỗi lớp nhãn được lựa chọn ngẫu nhiên để kiểm tra và hiệu chỉnh thủ công, nhằm loại bỏ các trường hợp ngoại lệ rõ ràng như khuôn mặt bị che khuất hoàn toàn hoặc biểu cảm không phù hợp với nhãn đã gán.

Quy trình tiền xử lý và lọc chất lượng ảnh:

Từ các video trên, khuôn mặt cầu thủ được phát hiện bằng RetinaFace (Deng et al., 2020) và trích xuất ở tần suất 25 khung hình/giây. Để đảm bảo chất lượng đầu vào cho giai đoạn tinh chỉnh, mỗi khuôn mặt được đánh giá bằng FaceQnet (Hernandez-Ortega et al., 2019) và chỉ giữ lại những ảnh có điểm chất lượng ≥ 0,6 – ngưỡng này loại bỏ các khung hình bị mờ, che khuất nặng hoặc góc quay quá lớn. Sau bước lọc, tập dữ liệu cuối cùng được thể hiện trong Bảng 1:

Bảng 1. Thống kê tập dữ liệu thể thao tự xây dựng.

Nhãn	Số mẫu
Phấn khích	~2.500 ảnh
Căng thẳng	~2.500 ảnh
Mệt mỏi	~2.500 ảnh

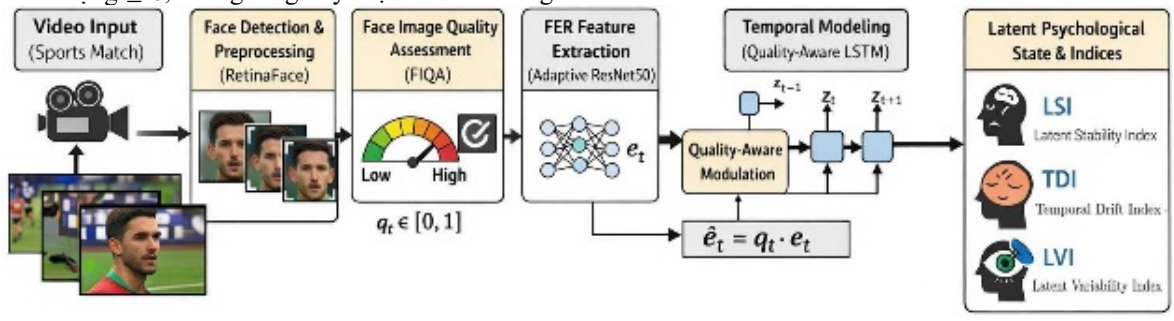
Tập dữ liệu 3 – Video thi đấu thực tế (huấn luyện LSTM):

Bên cạnh tập dữ liệu ảnh tĩnh trên, 50 video thi đấu và tập luyện thực tế được sử dụng để huấn luyện mô hình chuỗi thời gian LSTM (Jailani et al., 2023). Trong đó, khuôn mặt được phát hiện bằng RetinaFace (Deng et al., 2020) và chất lượng ảnh được đánh giá bằng FaceQnet (Hernandez-Ortega et al., 2019), tạo cơ sở cho việc phân tích trạng thái cảm xúc tiềm ẩn theo thời gian của cầu thủ trong quá trình thi đấu.

2.2. Kiến trúc mô hình

2.2.1. Tổng quan hệ thống

Hệ thống đề xuất phân tích sự thay đổi trạng thái cảm xúc biểu hiện qua khuôn mặt của cầu thủ dựa trên chuỗi biểu cảm khuôn mặt trong video thi đấu. Quy trình gồm các bước: phát hiện khuôn mặt, đánh giá chất lượng ảnh, trích xuất đặc trưng biểu cảm thích nghi và mô hình hóa theo thời gian. Khác với các phương pháp truyền thống chỉ phân loại cảm xúc rời rạc, hệ thống học một biểu diễn liên tục để mô tả trạng thái cảm xúc tiềm ẩn và sự biến đổi của nó theo thời gian.



Hình 1. Tổng quan về hệ thống

2.2.2. Phát hiện và tiền xử lý khuôn mặt

Khuôn mặt trong từng khung hình video được phát hiện bằng RetinaFace (Deng et al., 2020), do mô hình này có khả năng nhận diện tốt ngay cả khi khuôn mặt bị nghiêng nên việc quay lệch hoặc bị che khuất một phần là những tình huống thường xuyên xảy ra trong video thi đấu. Các vùng khuôn mặt sau đó được cắt và chuẩn hóa về kích thước cố định trước khi đưa vào khối xử lý tiếp theo.

2.2.3. Tính liên tục của khung hình

Thay vì sử dụng FIQA như bộ lọc cứng (các khung hình có điểm chất lượng thấp hơn một ngưỡng xác định bị loại bỏ hoàn toàn khỏi chuỗi dữ liệu), điểm chất lượng ảnh $q_t \in [0,1]$ được xem như một hệ số tin cậy (confidence score) liên tục cho từng khung hình. Cụ thể, thay vì loại bỏ các quan sát chất lượng thấp, toàn bộ khung hình vẫn được giữ lại và vectơ nhúng $q_t \in [0,1]$ biểu cảm e_t được điều chỉnh theo mức chất lượng theo công thức:

$$\tilde{e}_t = q_t e_t (1)$$

Cách tiếp cận này cho phép mỗi khung hình đóng góp vào mô hình chuỗi thời gian theo mức độ tin cậy tương ứng, trong đó các quan sát nhiễu có ảnh hưởng nhỏ hơn thay vì bị loại bỏ hoàn toàn. Nhờ đó, tính liên tục của chuỗi video được bảo toàn, đồng thời thông tin về chất lượng thị giác được tích hợp trực tiếp vào quá trình học trạng thái theo thời gian.

2.2.4. Học biểu diễn biểu cảm khuôn mặt

Kiến trúc mạng

Để chuyển đổi hình ảnh khuôn mặt thành vectơ đặc trưng biểu cảm giàu ngữ nghĩa, chúng tôi tích hợp hai kỹ thuật, gồm Hybrid Pooling (Shawi et al., 2025) và Adaptive Leaky ReLU (Shawi et al., 2025) vào kiến trúc ResNet50 (He et al., 2016) nhằm khắc phục hạn chế trong việc nhận diện các biểu cảm có sự thay đổi nhỏ.

Thứ nhất, để giải quyết vấn đề mất mát thông tin, Hybrid Pooling (Shawi et al., 2025) đã được sử dụng. Vectơ đặc trưng là tổ hợp lồi của $Pool_{max}$ và $Pool_{avg}$:

$$Pool_h = \alpha \cdot Pool_{max} + (1-\alpha) \cdot Pool_{avg} (2)$$

Trong đó:

$\alpha \in (0,1)$: hệ số học,

$Pool_{max}, Pool_{avg}$: kết quả của phép gộp cực đại và gộp trung bình trên bản đồ đặc trưng.

Thứ hai, để tránh hiện tượng "suy giảm đạo hàm" tại các vùng tối trên khuôn mặt (như hốc mắt, bóng đổ), chúng tôi thay thế hàm kích hoạt ReLU bằng Adaptive Leaky ReLU. Hệ số học $\alpha(t)$ không cố định mà thay đổi theo hàm mũ giảm dần theo thời gian huấn luyện, giúp mô hình học nhanh ở giai đoạn đầu và ổn định dần ở các giai đoạn sau:

$$f(x) = \max(x, \alpha(t) \cdot x) (3)$$

$$\alpha(t) = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot e^{-\lambda t} (4)$$

Trong đó:

$\alpha(t)$: là hệ số điều khiển độ dốc ở phía âm tại lần lặp thứ t.

$\alpha_{max}, \alpha_{min}$: các giá trị biên của hệ số độ dốc.

Hiệu quả của hai cải tiến này đã được kiểm chứng trong nghiên cứu của Shawi et al. (2025) trên hai bộ dữ liệu chuẩn FER2013 và CK+, trong đó mô hình Improved ResNet-50 đạt độ chính xác lần lượt là 96,60% và 96,32%, vượt trội so với ResNet-50 nguyên bản cũng như nhiều phương pháp IRON tiên tiến khác. Trong nghiên cứu này, chúng tôi tích hợp hai kỹ thuật Hybrid Pooling và Adaptive Leaky ReLU (Shawi et al., 2025) nhằm nâng cao hiệu năng mô hình. Kết quả thực nghiệm cho thấy sự kết hợp này giúp cải thiện đáng kể độ chính xác, từ 66,35% lên 92,29% (xem Bảng 3), qua đó cung cấp bằng chứng thực nghiệm vững chắc cho việc sử dụng chúng như một nền tảng hiệu quả trong mô-đun trích xuất đặc trưng.

Chiến lược huấn luyện đa giai đoạn

Mô hình Adaptive ResNet50 được huấn luyện theo chiến lược tinh chỉnh đa giai đoạn (Multi-stage Fine-tuning) (Parthasarathy et al., 2024) trên tập dữ liệu thứ hai, nhằm thích nghi hiệu quả với đặc trưng biểu cảm cảm xúc trong bối cảnh thể thao.

Do sự khác biệt trong phân phối giữa dữ liệu cảm xúc phổ quát và dữ liệu thể thao nên việc tinh chỉnh trực tiếp dễ dẫn đến học không ổn định hoặc quên thông tin và ngữ nghĩa. Vì vậy, trong nghiên cứu này, chiến lược huấn luyện đa giai đoạn đã được áp dụng nhằm chuyển dịch dần không gian biểu diễn.

Cụ thể, mạng ResNet50 được huấn luyện trước trên FER2013 nhằm học các biểu cảm cơ bản, sau đó tiếp tục được tinh chỉnh trên dữ liệu thể thao với ba nhãn đặc thù: mệt mỏi, căng thẳng và phấn khích. Quá trình này giúp mô hình thích nghi tốt hơn với các biểu hiện cảm xúc đặc trưng của cầu thủ trong môi trường thi đấu thực tế. Sau khi hoàn tất tinh chỉnh, lớp phân loại cuối cùng được loại bỏ và ResNet50 được sử dụng như một bộ trích xuất vectơ nhúng biểu cảm, làm đầu vào cho mô hình chuỗi thời gian.

2.2.5. Cơ chế tích hợp điểm chất lượng ảnh vào mô hình chuỗi thời gian

Mô hình chuỗi thời gian nhận đầu vào là chuỗi vectơ biểu cảm đã được điều chỉnh theo chất lượng ảnh. Cụ thể, mỗi vectơ nhúng được nhân với điểm FIQA tương ứng, để những khung hình rõ nét có ảnh hưởng lớn hơn, còn những khung hình kém chất

lượng sẽ đóng góp ít hơn. Nhờ đó, mức độ tin cậy của hình ảnh được gắn trực tiếp với vai trò của từng quan sát trong quá trình phân tích.

Vectơ đầu vào x_t cho LSTM được tính theo công thức:

$$x_t = \tilde{e}_t = q_t e_t \quad (5)$$

Trong nghiên cứu này, phép nhân vô hướng được lựa chọn thay cho các cơ chế phức tạp hơn như Gating (Chen et al., 2024) hay Attention Weighting (Nguyen et al., 2024), xuất phát từ hạn chế về dữ liệu khi không có nhãn trạng thái cảm xúc ở mức từng khung hình (frame-level ground truth) trong video thi đấu. Các cơ chế Gating thường đòi hỏi quá trình huấn luyện có giám sát để tối ưu hóa các trọng số cổng, do đó tiềm ẩn nguy cơ quá khớp khi áp dụng trên dữ liệu thiếu nhãn theo thời gian.

Ngược lại, phép nhân vô hướng khai thác trực tiếp các đánh giá khách quan về chất lượng vật lý của ảnh như một hệ số độ tin cậy, qua đó duy trì kiến trúc mô hình đơn giản, giảm số lượng tham số cần học, đồng thời bảo toàn tính liên tục của chuỗi tín hiệu đầu vào.

Một mạng LSTM được sử dụng để học và mô tả sự thay đổi của biểu cảm theo thời gian (Jailani et al., 2023; Khan et al., 2024). Cụ thể, LSTM nhận vào chuỗi các vectơ biểu cảm $\{x_1, x_2, \dots, x_t\}$ và tạo ra một vectơ trạng thái tiềm ẩn $z_t \in \mathbb{R}^k$. Vectơ này đại diện cho trạng thái cảm xúc tiềm ẩn được suy luận từ quá trình thay đổi biểu cảm khuôn mặt theo thời gian (Kim et al., 2019; Rubanova et al., 2019):

$$z_t = \text{LSTM}(x_{1:t}) \quad (6)$$

z_t không đại diện cho một nhãn cảm xúc rời rạc hay một kết luận tâm lý mang tính chẩn đoán, mà là biểu diễn trạng thái cảm xúc tiềm ẩn dùng để mô hình hóa sự chuyển biến cảm xúc theo thời gian.

Hàm mất mát chính: Temporal Smoothness:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{T-1}\sum_{t=2}^{T} \| z_t - z_{t-1} \|^2 \quad (7)$$

Hàm mất mát phụ: Anti-collapse:

$$\mathcal{L}_{\text{var}} = \beta \frac{1}{T}\sum_{t=1}^{T} \|z_t - x_t\|^2 (8)$$

Hàm mất mát tổng:

$$\mathcal{L}_{\text{LSTM}} = \mathcal{L}_{\text{smooth}} + \mathcal{L}_{\text{var}} (9)$$

Trong đó:

T: độ dài cửa sổ trượt (window size).

$\| z_t - z_{t-1} \|^2$: Đo lường mức độ thay đổi giữa hai trạng thái liên tiếp z_t và z_{t-1} qua đó đánh giá độ mượt của quỹ đạo theo thời gian. Nhờ đó, mô hình có thể hạn chế các dao động tần số cao hoặc hiện tượng "nhấp nháy" không mong muốn trong chuỗi trạng thái.

$\| z_t - x_t \|^2$: Đo lường mức độ sai khác giữa trạng thái tiềm ẩn z_t và vectơ đặc trưng đầu vào x_t, qua đó phản ánh mức độ điều chỉnh hoặc thích nghi của mô hình so với thông tin quan sát ban đầu.

β: Hệ số điều chỉnh (siêu tham số) dùng để kiểm soát sự cân bằng giữa tính mượt của chuỗi trạng thái và mức độ bám sát dữ liệu quan sát, qua đó điều phối sự đánh đổi giữa ổn định theo thời gian và độ chênh lệch so với đầu vào.

2.2.6. Bộ chỉ số định lượng trong không gian trạng thái tiềm ẩn

Để tăng cường khả năng diễn giải của mô hình và hỗ trợ ứng dụng trong bối cảnh huấn luyện thể thao, một hệ thống các chỉ số định lượng được suy dẫn trực tiếp từ không gian trạng thái tiềm ẩn z_t đã được đề xuất. Các chỉ số này nhằm lượng hóa đặc tính ổn định, mức độ biến thiên và xu hướng chuyển dịch của trạng thái cảm xúc tiềm ẩn theo thời gian, qua đó chuyển biểu diễn tiềm ẩn thành các thước đo có ý nghĩa thực tiễn.

Chỉ số ổn định tiềm ẩn (Latent Stability Index – LSI)

Chỉ số ổn định tiềm ẩn được xây dựng trên cơ sở độ lệch chuẩn hóa của các trạng thái trong một cửa sổ thời gian xác định. LSI phản ánh mức độ dao động nội tại của trạng thái cảm xúc tiềm ẩn theo thời gian.

Giả sử trong một cửa sổ quan sát có độ dài W, ta thu được tập các biểu diễn tiềm ẩn:

$$\{z_1, z_2, \dots, z_W\}$$

Trước hết, giá trị trung bình và độ phân tán của các trạng thái trong cửa sổ thời gian được xác định:

$$\mu = \frac{1}{W}\sum_{i=1}^{W} z_i\,, \quad \sigma = \sqrt{\frac{1}{W}\sum_{t=1}^{W}(z_t - \mu)^2} \quad (10)$$

Trong đó, μ đại diện cho trạng thái trung tâm, còn σ phản ánh mức độ biến thiên tổng thể của các trạng thái tiềm ẩn.

Độ lệch của từng trạng thái so với giá trị trung bình được chia cho độ lệch chuẩn của toàn bộ cửa sổ quan sát, nhằm đánh giá mức sai khác trên cơ sở tương đối thay vì dựa vào độ lớn tuyệt đối. Trên cơ sở đó, chỉ số LSI được tính theo công thức:

$$\mathrm{LSI} = \frac{1}{W}\sum_{i=1}^{W}\left\|\frac{\mathbf{z}_i - \mu}{\sigma}\right\|_2 (11)$$

LSI cao cho thấy trạng thái biến động mạnh và kém ổn định, trong khi giá trị thấp biểu thị sự duy trì ổn định tương đối.

Chỉ số thay đổi tạm thời (Temporal Drift Index – TDI)

Chỉ số thay đổi tạm thời được định nghĩa dựa trên chuẩn sai khác giữa các trạng thái liên tiếp và được làm trơn theo cơ chế trung bình lũy thừa.

Tại thời điểm t, độ sai khác Δt giữa hai trạng thái liên tiếp được xác định bằng chuẩn Euclid:

$$\Delta t = \|\mathrm{z}_t - \mathrm{z}_{t-1}\|_2 (12)$$

Để hạn chế nhiễu và làm nổi bật xu hướng chuyển dịch theo thời gian, Δt được làm trơn bằng trung bình lũy thừa với hệ số $\alpha \in (0,1)$. Trên cơ sở đó, chỉ số TDI tại thời điểm t được tính theo công thức:

$$\mathrm{TDI}_t \quad = \alpha\,\mathrm{TDI}_{t-1} + (1-\alpha)\,\Delta_t (13)$$

TDI lớn hàm ý sự thay đổi nhanh hoặc đột ngột, còn TDI nhỏ cho thấy quá trình chuyển trạng thái diễn ra chậm và ổn định.

Chỉ số dao động chung (Latent Variability Index – LVI)

Chỉ số dao động chung được tính bằng phương sai của các trạng thái trong khoảng quan sát. LVI cung cấp thước đo tổng quát về mức độ phân tán của trạng thái cảm xúc tiềm ẩn, qua đó phản ánh xu hướng biến thiên dài hạn.

Vì z_i là vectơ trong không gian d chiều, phương sai tổng quát được định nghĩa thông qua trung bình bình phương độ lệch Euclid so với giá trị trung tâm:

$$\mathrm{LVI} = \mathrm{Var}(z_1, \ldots, z_W) = \frac{1}{W}\sum_{i=1}^{W}\|z_i - \mu\|_2^2 (14)$$

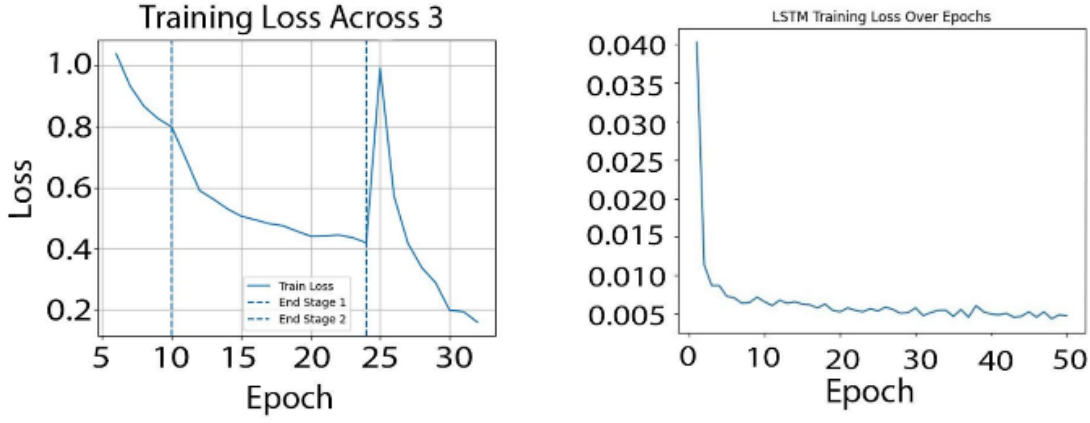
LVI cao cho thấy trạng thái biến động mạnh và kém ổn định, trong khi LVI thấp biểu thị sự duy trì ổn định tương đối.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Kết quả

3.1.1. Phân tích sự hội tụ và tính ổn định của mô hình

Trong quá trình tinh chỉnh ResNet50 trên dữ liệu thể thao, giá trị hàm mất mát giảm dần qua các vòng lặp và tiến tới trạng thái ổn định (Hình 2a). Xu hướng này cho thấy mô hình đã thích nghi hiệu quả với dữ liệu thực tế, đồng thời học được các đặc trưng biểu cảm đặc thù của môi trường thể thao thay vì chỉ dựa vào các mẫu cảm xúc cơ bản từ bộ dữ liệu thông thường.



(a) *(b)*

Hình 2. Giá trị hàm mất mát của mô hình ResNet50 và LSTM trong quá trình huấn luyện

Ghi chú: (a) ResNet50 và (b) LSTM.

Đối với mạng LSTM, quá trình huấn luyện cũng cho thấy sự hội tụ ổn định của hàm mất mát mà không xuất hiện hiện tượng suy biến biểu diễn (Hình 2b). Điều này cho thấy biểu diễn trạng thái cảm xúc tiềm ẩn được duy trì nhất quán theo thời gian và không bị co cụm về một giá trị hằng. Kết quả này

phản ánh khả năng mô hình hóa sự chuyển biến trạng thái theo chuỗi thời gian một cách ổn định và liên tục.

3.1.2. Đánh giá định lượng hiệu năng phân loại biểu cảm

Để đánh giá định lượng hiệu năng của mô-đun trích xuất đặc trưng, mô hình Adaptive ResNet-50 sau khi được tinh chỉnh trên tập dữ liệu thể thao được kiểm chứng trên tập kiểm tra độc lập gồm 1.168 mẫu, tương ứng với tỷ lệ phân chia dữ liệu 80/20. Các kết quả đánh giá chi tiết được trình bày trong Bảng 2 và Bảng 3.

Bảng 2. Kết quả phân loại biểu cảm trên tập dữ liệu thể thao

Nhãn	Precision	Recall	F1-score	Support
Mệt mỏi	0,93	0,97	0,95	368
Căng thẳng	0,94	0,87	0,90	425
Phấn khích	0,90	0,94	0,92	375
Macro avg	0,92	0,93	0,92	1.168
Weighted avg	0,92	0,93	0,92	1.168
Accuracy			**0,92**	**1.168**

Kết quả cho thấy mô hình đạt độ chính xác tổng thể 92,29% với F1-score trung bình (macro) đạt 0,92, thể hiện hiệu năng ổn định và cân bằng trên cả ba lớp nhãn. Nhãn mệt mỏi đạt F1-score cao nhất (0,95), phản ánh tính đặc thù rõ rệt của biểu cảm trong giai đoạn cuối hiệp. Nhãn căng thẳng có recall thấp hơn (0,87), có thể do sự chồng lấn biểu cảm giữa căng thẳng và mệt mỏi trong một số tình huống thi đấu.

Bảng 3. So sánh hiệu năng các biến thể ResNet-50 với các kỹ thuật cải tiến đề xuất

Mô hình	Accuracy (%)	Thay đổi so với baseline
ResNet-50 (baseline)	66,35	–
ResNet-50 + Hybrid Pooling	72,04	+5,69%
ResNet-50 + Adaptive Leaky ReLU	67,60	+1,25%
ResNet-50 + Hybrid Pooling + Adaptive Leaky ReLU	**92,29**	**+25,94%**

Kết quả ablation cho thấy khi kết hợp hai kỹ thuật, hiệu năng tăng đáng kể. Cụ thể, mức cải thiện đạt 25,94 điểm phần trăm, cao hơn nhiều so với tổng mức tăng của từng kỹ thuật riêng lẻ (5,69 và 1,25 điểm phần trăm). Điều này cho thấy tồn tại hiệu ứng tương hỗ giữa Hybrid Pooling và Adaptive Leaky ReLU trong việc học các đặc trưng biểu cảm trong môi trường thể thao.

Cần lưu ý rằng kết quả phân loại này phản ánh hiệu năng của mô-đun trích xuất đặc trưng biểu cảm – một thành phần trong hệ thống tổng thể. LSI, TDI, LVI này.

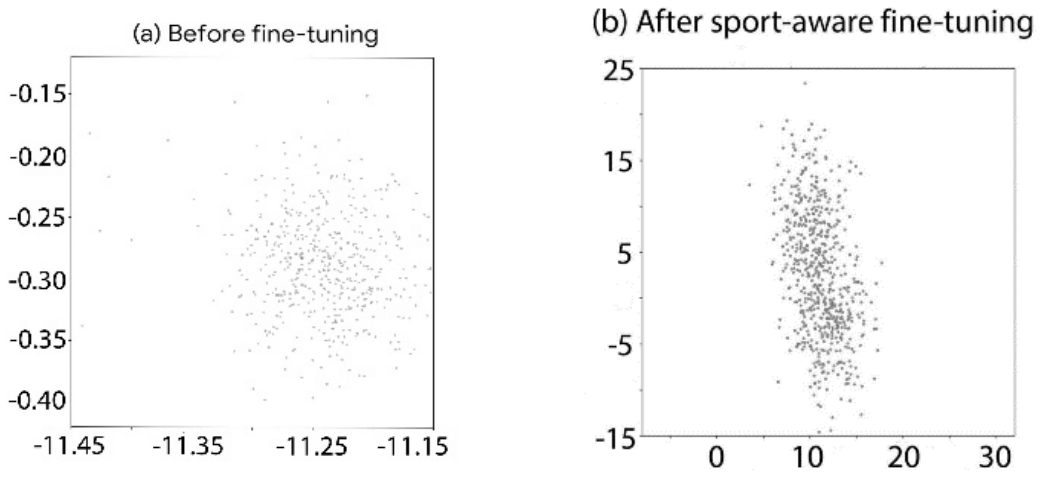
3.1.3. Phân tích sự thay đổi của không gian biểu diễn sau khi tinh chỉnh theo ngữ cảnh thể thao

Để đánh giá tác động của quá trình tinh chỉnh, các vectơ nhúng biểu cảm được trích xuất từ cùng tầng Dense-1024 của mô hình ResNet50 trước và sau khi tinh chỉnh trên dữ liệu thể thao, sau đó được trực quan hóa bằng phương pháp PCA.

Trong Hình 3a, các vectơ nhúng tập trung trong một vùng hẹp và không thể hiện cấu trúc phân tách rõ ràng khi áp dụng dữ liệu thể thao. Điều này cho thấy không gian biểu diễn được học từ các cảm xúc phổ quát chưa phù hợp để mô tả sự đa dạng biểu cảm trong bối cảnh thi đấu thể thao.

Hình 3b với dữ liệu thể thao, không gian biểu diễn được mở rộng đáng kể và hình thành một cấu trúc liên tục theo một hướng biến thiên chính. Các vectơ nhúng không còn tập trung trong một vùng hẹp, mà trải dài liên tục. Điều này cho thấy mô hình đã điều chỉnh không gian biểu diễn theo hướng phản ánh sự thay đổi liên tục của trạng thái, phù hợp hơn với mô hình chuỗi thời gian.

Kết quả này cho thấy quá trình tinh chỉnh không chỉ nhằm cải thiện phân loại mà còn tái tổ chức không gian biểu diễn theo hướng phù hợp với dữ liệu thể thao. Các vectơ nhúng sau tinh chỉnh phản ánh rõ hơn sự thay đổi liên tục của trạng thái cảm xúc tiềm ẩn theo thời gian, thay vì chỉ phục vụ việc tách biệt các cảm xúc rời rạc.



Hình 3. Trực quan hóa không gian vectơ nhúng biểu cảm bằng PCA

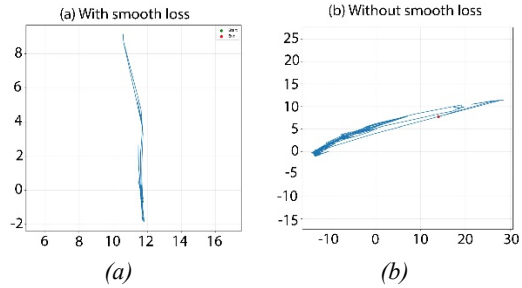
Ghi chú: (a) kết quả trước khi tinh chỉnh, (b) kết quả sau khi tinh chỉnh.

3.1.4. Tính nhất quán theo thời gian của các trạng thái tiềm ẩn

Do không có nhãn trạng thái cảm xúc theo thời gian trong dữ liệu thi đấu thực tế nên việc đánh giá mô hình được thực hiện gián tiếp thông qua phân tích tính liên tục và cấu trúc phụ thuộc thời gian của các biểu diễn trạng thái tiềm ẩn. Cụ thể, chúng tôi xem xét mức độ biến thiên giữa các thời điểm liên tiếp và sự duy trì cấu trúc quỹ đạo theo thời gian như một chỉ báo về tính nhất quán của mô hình.

Để trực quan hóa không gian biểu diễn, các vector trạng thái z_t, vốn nằm trong không gian nhiều chiều, được chiếu xuống không gian hai chiều bằng phương pháp phân tích thành phần chính (PCA). Các điểm sau đó được sắp xếp theo thứ tự thời gian và nối thành quỹ đạo, cho phép quan sát tính liên tục và xu hướng chuyển biến trạng thái trong một đoạn video thi đấu.

Hình 4a minh họa quỹ đạo khi mô hình LSTM được huấn luyện với đầy đủ hàm mất mát $L_{\text{smooth}} + L_{\text{var}}$ cho thấy quỹ đạo trạng thái cảm xúc tiềm ẩn thay đổi liên tục theo một hướng nhất quán, không xuất hiện các dao động đột ngột giữa các thời điểm liên tiếp. Ngược lại, Hình 4b cho thấy kết quả khi loại bỏ thành phần L_{smooth} và chỉ giữ lại L_{var}. Trong trường hợp này, trạng thái cảm xúc tiềm ẩn xuất hiện nhiều dao động đột ngột và thiếu tính liên tục.



Hình 4. Quỹ đạo trạng thái tiềm ẩn trong không gian hai chiều

Ghi chú: (a) Mô hình LSTM được huấn luyện với hàm mất mát tổng $\boldsymbol{L_{smooth} + L_{var}}$, (b) Mô hình chỉ sử dụng hàm mất mát phụ $\boldsymbol{L_{var}}$ (loại bỏ $\boldsymbol{L_{smooth}}$).

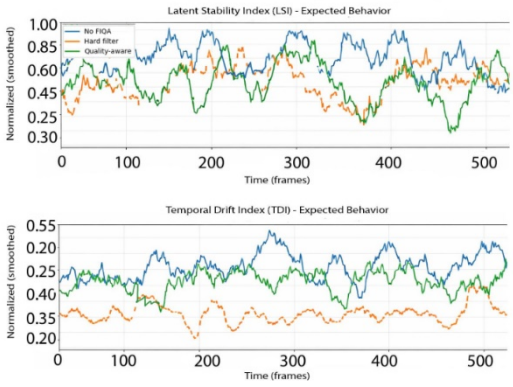
Sự khác biệt này xác nhận vai trò của L_{smooth} trong việc ràng buộc tính liên tục của trạng thái tiềm ẩn, phù hợp với giả định rằng trạng thái tâm lý trong thi đấu có xu hướng thay đổi dần dần thay vì biến động đột ngột giữa các khung hình liên tiếp.

3.1.5. Tác động của việc tích hợp điểm chất lượng ảnh vào mô hình chuỗi thời gian

Nhằm đánh giá ảnh hưởng của chỉ số chất lượng ảnh khuôn mặt (FIQA) đến độ ổn định của chuỗi trạng thái tiềm ẩn, ba phương án được so sánh trên cùng một chuỗi video: (i) không sử dụng FIQA (No FIQA), (ii) áp dụng cơ chế lọc cứng (Hard Filter), và (iii) tích hợp FIQA theo phương pháp đề xuất. Hiệu quả của các phương án được đánh giá thông qua sự biến thiên của hai chỉ số LSI và TDI, với các quỹ đạo tương ứng được minh họa trong Hình 5.

Hình 5 minh họa quỹ đạo biến đổi theo thời gian của hai chỉ số Latent Stability Index (LSI) và Temporal Drift Index (TDI) khi áp dụng ba phương án tích hợp FIQA khác nhau. Để bổ sung bằng chứng định lượng cho phân tích trực quan, bảng 4 tổng hợp các giá trị trung bình của LSI và TDI tính từ quỹ đạo quan sát trên 500 khung hình phân tích.

Với phương pháp No FIQA, chỉ số LSI duy trì ở mức trung bình cao (0,68) và dao động đáng kể trong suốt chuỗi quan sát, trong khi TDI có xu hướng gia tăng theo thời gian (trung bình 0,31), cho thấy trạng thái cảm xúc tiềm ẩn dễ bị chi phối bởi nhiễu phát sinh từ các khung hình có chất lượng thấp.



Hình 5. So sánh sự thay đổi của các chỉ số LSI và TDI theo thời gian giữa ba phương pháp.

Bảng 4. So sánh định lượng giữa phương pháp tích hợp FIQA (đề xuất), No FIQA và lọc cứng

Phương án	LSI trung bình	TDI trung bình	Tính liên tục chuỗi
No FIQA	0,68	0,31	Thấp – dao động lớn liên tục
Lọc cứng (Hard filter)	0,58	0,33	Trung bình – có gián đoạn cục bộ
Tích hợp FIQA (đề xuất)	0,52	0,27	Cao – mượt và liên tục

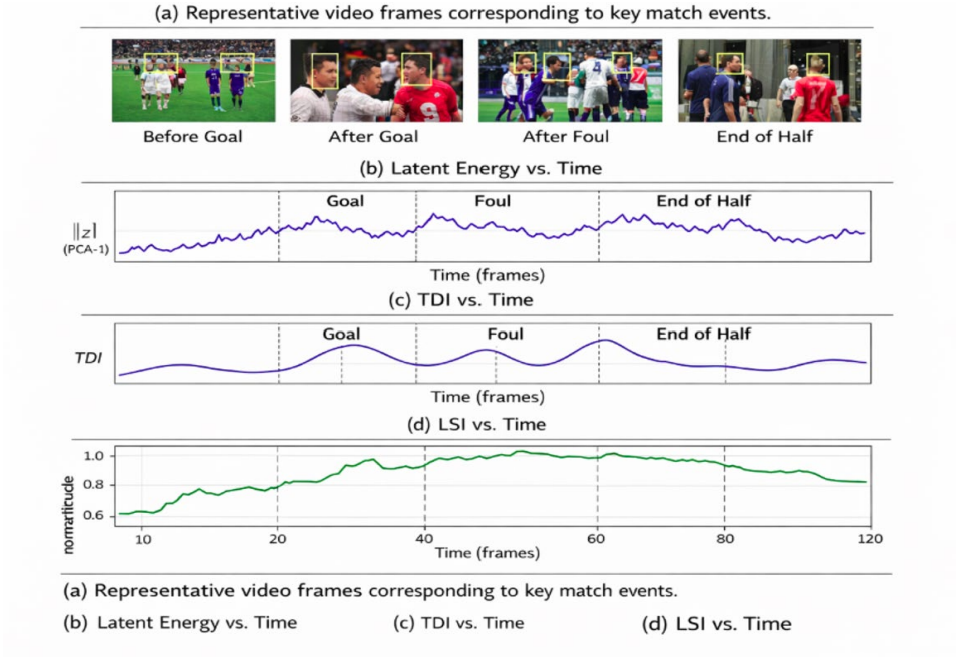
Phương pháp lọc cứng dựa trên FIQA giúp giảm LSI trung bình xuống còn 0,58, tức là cải thiện khoảng 14,7% so với No FIQA. Tuy nhiên, một nghịch lý đáng chú ý là TDI trung bình của phương án này (0,33) lại cao hơn cả No FIQA (0,31). Nguyên nhân nằm ở việc loại bỏ hoàn toàn các khung hình kém chất lượng tạo ra các bước nhảy đột ngột tại điểm gián đoạn trong chuỗi trạng thái, làm tăng mức biến đổi cục bộ dù giảm được nhiễu tổng thể. Điều này cho thấy tính liên tục của chuỗi quan sát đóng vai trò quan trọng không kém việc kiểm soát nhiễu trong bài toán mô hình hóa trạng thái tiềm ẩn.

Ngược lại, khi FIQA được tích hợp dưới dạng hệ số độ tin cậy liên tục (FIQA mềm), mô hình đồng thời đạt LSI thấp nhất (0,52, giảm 23,5% so với No FIQA) và TDI thấp nhất (0,27, giảm 12,9% so với No FIQA), cho thấy phương án này vừa kiểm soát nhiễu hiệu quả vừa bảo toàn cấu trúc thời gian của chuỗi trạng thái. Cả LSI và TDI duy trì ở mức thấp và ổn định hơn, cho thấy trạng thái cảm xúc tiềm ẩn không chỉ ít bị nhiễu mà còn biến đổi mượt mà theo thời gian. Kết quả này nhấn mạnh ưu thế của cách tiếp cận "trọng số mềm" trong các bài toán phân tích chuỗi trạng thái tiềm ẩn, đặc biệt khi dữ liệu đầu vào có chất lượng không đồng nhất.

3.1.6. Đánh giá tính hợp lý và khả năng diễn giải của trạng thái tiềm ẩn

Để đánh giá tính hợp lý và khả năng diễn giải của trạng thái cảm xúc tiềm ẩn trong điều kiện không có nhãn chuẩn, một nghiên cứu tình huống đã được thực hiện trên một cầu thủ trong phân đoạn thi đấu thực tế chứa nhiều sự kiện có mức độ áp lực cao, bao gồm ghi bàn, phạm lỗi và giai đoạn cuối hiệp. Kết quả phân tích định tính được minh họa trong Hình 6, gồm: (a) các khung hình đại diện tại bốn mốc sự kiện; (b) diễn biến theo thời gian của trạng thái cảm xúc tiềm ẩn, được biểu diễn bởi thành phần chính thứ nhất (PCA-1) của $\|z\|$; (c) diễn biến của chỉ số TDI; và (d) diễn biến của chỉ số LSI.

Hình 6 cho thấy mối liên hệ nhất quán giữa các sự kiện thi đấu và sự biến đổi của trạng thái tiềm ẩn. Mức độ biến thiên của trạng thái tiềm ẩn tăng rõ rệt sau các sự kiện áp lực cao và giảm dần về cuối hiệp (Hình 6b), trong khi chỉ số TDI xuất hiện các đỉnh đồng bộ với các mốc sự kiện (Hình 6c). Đồng thời, LSI duy trì ở mức ổn định, cho thấy quỹ đạo trạng thái tiềm ẩn duy trì ổn định và không bị gián đoạn (Hình 6d)



Hình 6. Nghiên cứu trường hợp và phân tích định tính

3.2. Thảo luận

Khác với nhiều nghiên cứu trước đây chỉ dự đoán nhãn cảm xúc rời rạc theo từng khung hình (Li & Deng, 2022; Lee et al., 2023), phương pháp đề xuất mô hình hóa trạng thái cảm xúc tiềm ẩn như một quá trình biến đổi liên tục theo thời gian. Cách tiếp cận này phù hợp hơn với bối cảnh thi đấu thực tế, nơi cảm xúc thường thay đổi dần dưới tác động của áp lực và các sự kiện quan trọng, thay vì biến đổi đột ngột.

So với các phương pháp nhận dạng cảm xúc từ video hiện đại (Kim et al., 2019; Khan et al., 2024), hệ thống đề xuất không nhằm cạnh tranh trực tiếp trên cùng tiêu chí đánh giá do khác biệt căn bản về bài toán. Các phương pháp FER truyền thống chủ yếu tối ưu hóa độ chính xác phân loại nhãn rời rạc trên các tập dữ liệu chuẩn trong môi trường có kiểm soát, trong khi nghiên cứu này tập trung vào việc mô hình hóa quỹ đạo trạng thái liên tục trong điều kiện tự nhiên. Vì vậy, thay vì đóng vai trò thay thế, kết quả nghiên cứu này có thể được xem như một hướng tiếp cận mới cho bài toán phân tích trạng thái theo thời gian mà các phương pháp FER hiện tại chưa giải quyết.

Cụ thể, thay vì chỉ dự đoán cảm xúc tức thời, hệ thống học một không gian trạng thái tiềm ẩn và từ đó suy ra các chỉ số định lượng như LSI, TDI và LVI nhằm đặc trưng hóa mức độ ổn định và biến động của trạng thái cảm xúc theo thời gian. So với các phương pháp dựa trên tín hiệu sinh lý như GPS hoặc nhịp tim, cách tiếp cận này cho phép trích xuất trực tiếp các chỉ số từ dữ liệu video, qua đó hỗ trợ hiệu quả cho phân tích và ra quyết định trong huấn luyện thực tế.

Một số lựa chọn thiết kế trong nghiên cứu này được xác định dựa trên thực nghiệm. Cụ thể, phép nhân vô hướng embedding với điểm FIQA được chọn do tính đơn giản và phi tham số, kích thước cửa sổ thời gian W trong tính toán LSI và hệ số làm trơn α trong TDI được xác định qua quan sát trực quan trên tập dữ liệu thực nghiệm.

4. KẾT LUẬN VÀ KIẾN NGHỊ

Một hệ thống phân tích trạng thái cảm xúc tiềm ẩn dựa trên biểu cảm khuôn mặt cầu thủ theo thời gian được đề xuất trong nghiên cứu; trong đó, trạng thái cảm xúc tiềm ẩn được xem như một quá trình biến đổi liên tục thay vì các nhãn cảm xúc rời rạc. Bằng cách tích hợp điểm đánh giá chất lượng ảnh khuôn mặt (FIQA) như một tín hiệu độ tin cậy liên tục, mô hình có thể giảm ảnh hưởng của nhiễu hình ảnh mà vẫn giữ được tính liên tục của dữ liệu theo thời gian.

Kết quả thực nghiệm cho thấy chiến lược tích hợp điểm chất lượng ảnh vào mô hình chuỗi thời gian giúp học được các quỹ đạo trạng thái ổn định và mạch lạc hơn, đồng thời suy ra các chỉ số biểu hiện cảm xúc tiềm ẩn có khả năng diễn giải trong

bối cảnh thi đấu thực tế. Về hiệu năng phân loại biểu cảm, mô hình Adaptive ResNet-50 đề xuất đạt accuracy 92,29% và F1-score macro 0,92 trên tập kiểm tra, vượt trội so với baseline ResNet-50 gốc (66,35%), cho thấy hiệu quả thực nghiệm rõ ràng của các thành phần kiến trúc được tích hợp. Hệ thống đề xuất không chỉ góp phần nâng cao độ tin cậy của phân tích dựa trên video, mà còn hướng tới ứng dụng hỗ trợ đánh giá hiệu suất và ra quyết định cho huấn luyện viên, mở ra một hướng tiếp cận mới trong phân tích thể thao dựa trên dữ liệu thị giác.

Trong các hướng nghiên cứu tiếp theo, một yêu cầu quan trọng là xây dựng bộ dữ liệu có nhãn trạng thái theo thời gian với mức độ chi tiết cao; trong đó, quá trình gán nhãn cần có sự tham gia của các chuyên gia tâm lý thể thao và được kiểm chứng thông qua các chỉ số đánh giá độ đồng thuận giữa nhiều người gán nhãn, chẳng hạn như chỉ số Cohen's Kappa. Đây sẽ là cơ sở quan trọng để kiểm định thống kê các chỉ số LSI, TDI và LVI trên quy mô dữ liệu lớn. Đồng thời, cần tiến hành việc phân tích độ nhạy của hệ thống đối với các siêu tham số thiết kế — bao gồm kích thước cửa sổ W, hệ số làm trơn α trong TDI và ngưỡng FIQA — nhằm đánh giá tính ổn định và xác định miền giá trị phù hợp cho từng bối cảnh ứng dụng. Bên cạnh đó, việc mở rộng mô hình theo hướng đa phương thức, kết hợp biểu cảm khuôn mặt với các tín hiệu sinh lý và hành vi, cùng với việc xây dựng các giao thức so sánh tương thích với các phương pháp FER video hiện đại, sẽ góp phần đánh giá toàn diện hơn hiệu quả và khả năng tổng quát hóa của hệ thống trên nhiều môn thể thao khác nhau.

LỜI CẢM TẠ

Nhóm tác giả trân trọng cảm ơn Trường Công nghệ Thông tin và Truyền thông, Đại học Công nghiệp Hà Nội đã tạo điều kiện về cơ sở vật chất và môi trường nghiên cứu trong suốt quá trình thực hiện đề tài.

Chúng tôi chân thành cảm ơn các đồng nghiệp tại Khoa Khoa học Máy tính đã hỗ trợ trao đổi chuyên môn, góp ý về phương pháp và kiểm chứng mô hình thực nghiệm.

Cuối cùng, nhóm tác giả cũng xin cảm ơn các tổ chức Fédération Internationale de Football Association (FIFA) và Union of European Football Associations (UEFA) đã cung cấp nguồn video thi đấu công khai phục vụ cho mục đích nghiên cứu học thuật.

TÀI LIỆU THAM KHẢO

Azizan, L. A., Aziz, N. H. A., & Abdullah, S. N. H. S. (2024). Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets. *Information*, *15*(3), 135. https://doi.org/10.3390/info15030135

Chen, Z., Wang, M., Zhang, Y., Deng, W., Cui, X., Shi, H., Wen, D., & Zhao, J. (2024). Confidence-aware RGB-D face recognition via virtual depth synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 1421-1430). IEEE. https://arxiv.org/abs/2403.06529

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5203-5212). IEEE. https://doi.org/10.1109/CVPR42600.2020.00525

Goodfellow, I. J., Boureau, Y.-L., & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing (ICONIP)* (pp. 117-124). Springer. https://doi.org/10.1007/978-3-642-42051-1_16

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Hernandez-Ortega, J., Galbally, J., Fierrez, J., Haraksim, R., & Beslay, L. (2019). FaceQnet: Quality assessment for face recognition based on deep learning. In *Proceedings of the 2019 International Conference on Biometrics (ICB)* (pp. 1-8). IEEE. https://doi.org/10.1109/ICB45273.2019.8987255

Ivarsson, A., Kilhage-Persson, A., Martindale, R., Priestley, D., Huijgen, B., Ardern, C., & McCall, A. (2020). Psychological factors and future performance of football players: A systematic review with meta-analysis. *Journal of Science and Medicine in Sport*, 23(4), 415-420. https://doi.org/10.1016/j.jsams.2019.10.021

Jailani, N. L. M., Dhanasegaran, J. K., Alkawsi, G., Alkahtani, A. A., Phing, C. C., Baashar, Y., Capretz, L. F., Al-Shetwi, A. Q., & Tiong, S. K. (2023). Investigating the power of LSTM-based models in solar energy forecasting. Processes, 11(5), 1382. https://doi.org/10.3390/pr11051382

Khan, S. A., Khan, T. A., & Min-Allah, N. (2024). Learning face expression features from video using spatio-temporal feature extractor and CNN-LSTM. In *2023 International Conference on Computing, Intelligence and Data Analytics (ICCIDA)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCIDA58661.2023.10537794

Kim, D. H., Baddar, W. J., Jang, J., & Ro, Y. M. (2019). Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. *IEEE Transactions on Affective Computing*, *10*(2), 223-236. https://doi.org/10.1109/TAFFC.2017.2695999

Lee, J., Kim, S., & Ro, S. (2023). Facial expression recognition in the wild using face graph and attention. IEEE Access, 11, 60324-60336. https://doi.org/10.1109/ACCESS.2023.3287639

Li, S., & Deng, W. (2022). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, *13*(4), 1195–1215. https://doi.org/10.1109/TAFFC.2020.2981446

Meng, Q., Zhao, S., Huang, Z., & Zhou, F. (2021). MagFace: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 14225-14234). IEEE. https://doi.org/10.1109/CVPR46437.2021.01400

Nguyen, X.-B., Nguyen, H.-T., Nguyen, T.-H., Do, N.-T., & Dinh, Q. V. (2024). Emotic masked autoencoder on dual-views with attention fusion for facial expression recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 4784-4792). IEEE. https://doi.org/10.1109/CVPRW63382.2024.00481

Ou, F., Liu, X., Zhang, Y., & Zhang, G. (2024). CLIB-FIQA: Face image quality assessment with confidence calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2554-2564). IEEE. https://doi.org/10.1109/CVPR52733.2024.00248

Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A. (2024). The ultimate guide to fine-tuning LLMs from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities. *arXiv*. https://doi.org/10.48550/arXiv.2408.13296

Rubanova, Y., Chen, R. T. Q., & Duvenaud, D. K. (2019). Latent ordinary differential equations for irregularly-sampled time series. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (pp. 5320-5330). Curran Associates. https://dl.acm.org/doi/10.5555/3454287.3454765

Shawi, R. T., Abdulkadhim, A. A. A., & Abbas, F. N. (2025). Facial emotion recognition based on improved ResNet50 using hybrid pooling and adaptive Leaky ReLU. *International Journal of Scientific Research in Science, Engineering and Technology*, *12*(3), 728-737. https://doi.org/10.32628/IJSRSET251290