

NHẬN DẠNG DẤU HIỆU NGUY CƠ BẤT THƯỜNG TRÊN KHUÔN MẶT TRONG QUÁ TRÌNH TẬP LUYỆN PHỤC HỒI CHỨC NĂNG

Bùi Quang Vinh*, Lê Quốc Cường#, Hứa Nguyễn Đăng Thy*, Vũ Ngọc Long*,
Nguyễn Thanh Hải*

* Trung tâm Nghiên cứu triển khai Khu Công nghệ cao TP. HCM

Ban quản lý Khu Công nghệ cao TPHCM

Tóm tắt— Trong quá trình tập luyện phục hồi chức năng, việc phát hiện sớm các dấu hiệu nguy cơ bất thường trên khuôn mặt đóng vai trò quan trọng nhằm đảm bảo an toàn trong điều trị. Bài báo này tập trung vào việc ứng dụng các kỹ thuật học máy và xử lý ảnh để nhận diện và phân loại các dấu hiệu bất thường trên khuôn mặt, bao gồm biến dạng 68 điểm trên khuôn mặt, dấu hiệu căng thẳng hoặc bất đối xứng không tự nhiên. Dữ liệu hình ảnh thu thập từ các buổi tập luyện được phân tích và xử lý để xây dựng mô hình nhận dạng, giúp cung cấp cảnh báo sớm và hỗ trợ điều chỉnh phương pháp tập luyện. Kết quả nghiên cứu góp phần nâng cao hiệu quả của hệ thống hỗ trợ phục hồi chức năng, cải thiện chăm sóc người bệnh và giảm thiểu rủi ro trong quá trình điều trị.

Từ khóa—SVM, CNN, Facial Landmark Detection

I. GIỚI THIỆU

Nhận dạng khuôn mặt là một lĩnh vực phổ biến trong công nghệ nhận dạng và bảo mật, với nhiều ứng dụng từ xác thực người dùng cho đến nhận diện hành vi bất thường. Các phương pháp truyền thống như nhận dạng khuôn mặt dựa trên đặc trưng hình học đã được sử dụng rộng rãi, nhưng với các kỹ thuật học máy và học sâu đã mở rộng khả năng nhận dạng và phân loại khuôn mặt, mang lại hiệu quả có tính thực tiễn cao. Một phương pháp chuẩn hóa, khách quan để ghi lại các chuyển động biểu cảm khuôn mặt sử dụng các đơn vị hành động (AU) để mô tả các chuyển động cơ mặt nhỏ nhất có thể quan sát được, dựa trên nguyên tắc giải phẫu. Mỗi AU có thể đại diện cho một cơ hoặc nhóm cơ, tùy theo mức độ khó phân biệt chuyển động. Nhờ khả năng mô tả đầy đủ mọi biểu cảm khuôn mặt của con người, phương pháp này đã trở thành công cụ hữu ích trong nhiều lĩnh vực như tâm lý học, nhận diện khuôn mặt và nghiên cứu hành vi [1].

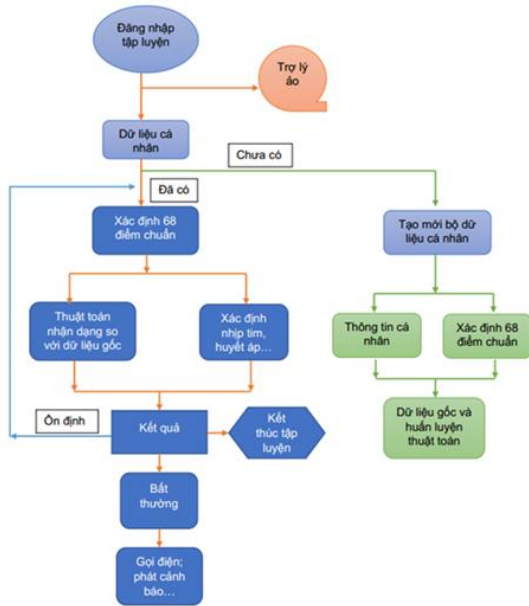
Một nghiên cứu giới thiệu một cơ sở dữ liệu chuẩn hóa về hình ảnh và video của bệnh nhân bị liệt mặt, bao gồm cả liệt mềm hoàn toàn và liệt không mềm sử dụng dữ liệu thu thập gồm ảnh và video được đánh giá bằng các thang đo chuyên biệt như eFACE, House-Brackmann và Sunnybrook. Đồng thời, biểu cảm cảm xúc của bệnh nhân khi thực hiện video tiêu chuẩn được phân tích bằng thuật toán trí tuệ nhân tạo nhận diện cảm xúc tự động [2].

Các phương pháp học máy, học sâu có thể học và tự động trích xuất các đặc điểm cụ thể của chứng liệt, bất đối xứng của khuôn mặt bằng cách áp dụng mạng nơ-ron tích chập (CNN) lên hình ảnh liệt mặt để tự động trích xuất các đặc trưng riêng biệt của bệnh nhân liệt mặt. Các đặc trưng này sau đó được sử dụng để phân loại 5 mức độ liệt mặt [3]. Bên cạnh đó, mô hình Histogram of Oriented Gradients (HOG) đã được chứng minh là hiệu quả trong việc nhận diện các đặc trưng hình học và cấu trúc của khuôn mặt. HOG là phương pháp hiệu quả trong việc trích xuất đặc trưng hình ảnh, đặc biệt phân tích các đối tượng có hình dạng rõ như khuôn mặt người. Mặc dù HOG có thể trích xuất các đặc trưng hình ảnh khá tốt nhưng hiệu quả của mô hình này phụ thuộc vào thuật toán phân loại được sử dụng. Các thuật toán như CNN, K-Nearest Neighbors (KNN) và Support Vector Machine (SVM) đều được áp dụng rộng rãi trong nhận dạng khuôn mặt, mỗi thuật toán có những ưu điểm và hạn chế riêng [4]. Cách tiếp cận này cho phép hệ thống tập trung vào việc phát hiện và đánh giá các yếu tố cụ thể như mức độ bất đối xứng giữa hai bên khuôn mặt hoặc các biểu hiện bất thường trong cấu trúc cơ mặt. Thông qua việc phân tích chính xác các điểm đặc trưng, hệ thống không chỉ cải thiện khả năng nhận diện mà còn hỗ trợ các đội ngũ y tế trong việc điều chỉnh phương pháp tập luyện phù hợp.

Mục tiêu của nghiên cứu là tối ưu hóa hiệu quả nhận dạng dấu hiệu bất thường trên khuôn mặt trong bối cảnh phục hồi chức năng, đồng thời đánh giá hiệu năng của mô hình HOG kết hợp với các thuật toán CNN, KNN và SVM. Bài báo này trình bày phương pháp nghiên cứu, kết quả thực nghiệm và phân tích, cùng các đề xuất cải tiến trong tương lai.

Tác giả liên hệ: Lê Quốc Cường,
Email: lqcuong.shp@tpcm.gov.vn
Đến tòa soạn: 22/01/2025, chỉnh sửa: 24/3/2025, chấp nhận đăng:
12/4/2025.

Quy trình tập luyện trong phương pháp này được thực hiện như sơ đồ trong Hình 1. Quá trình tập luyện và giám sát được thực hiện trong phạm vi nhất định như một phòng tập chuyên biệt, các đối tượng tập luyện sẽ được cá nhân hóa dữ liệu về tiền sử bệnh lý; chế độ, bài tập được khuyến cáo từ y bác sỹ; thông tin người thân cần liên hệ...



Hình 1. Sơ đồ hoạt động của hệ thống

Quy trình của phương pháp cơ bản gồm 4 bước sau:

- Thu thập thông tin cá nhân hóa dữ liệu.
- Kiểm soát quá trình đăng nhập tập luyện.
- Tập luyện cùng trợ lý ảo.
- Giám sát những bất thường qua khuôn mặt.

II. NỘI DUNG PHƯƠNG PHÁP THỰC HIỆN

Một nghiên cứu đánh giá liệt mặt ngoại biên một bên (UPFP) – một dạng tổn thương dây thần kinh mặt – bằng phương pháp khách quan và tự động thông qua chuỗi hình ảnh khuôn mặt được ghi lại khi người bệnh thực hiện các biểu cảm tiêu chuẩn. Các đặc điểm khuôn mặt được trích xuất tự động từ ảnh bằng phân tích điểm mốc khuôn mặt, sau đó chọn lọc những đặc trưng có tương quan cao với thang điểm House-Brackmann (HB). Cuối cùng, các đặc trưng được đưa vào mô hình máy vector hỗ trợ (SVM) để phân loại mức độ liệt mặt. Kết quả cho thấy phương pháp đạt độ chính xác cao (tới 90,01% sau tăng cường dữ liệu), chứng minh tiềm năng của một hệ thống đánh giá UPFP để triển khai thông qua camera web [5]. Bên cạnh đó một mô hình hỗ trợ phát hiện liệt mặt tự động bằng cách kết hợp hai mô hình học sâu để phát hiện khuôn mặt và điểm mốc, cùng với CNN đã được tối ưu tham số để phân loại tình trạng liệt mặt. Hệ thống được thử nghiệm trên tập dữ liệu video YouTube có người bệnh thực tế, cho thấy khả năng phát hiện liệt mặt hiệu quả [6].

Trong bài báo này, nhóm sử dụng camera RealSense của Intel và kết hợp các thuật toán học máy CNN, SVM, mạng nơ-ron nhiều lớp (MLP- Multi-layer perceptrons) để đạt được độ chính xác cao trong nhận diện và phân tích cảm xúc, dự đoán bất thường đồng thời giám sát, cảnh báo các

dấu hiệu bất thường nếu xảy ra thông qua giọng nói của trợ lý ảo.

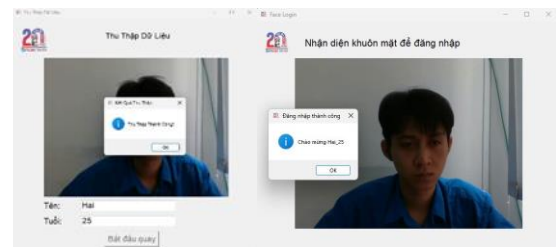
A. Thu thập cá nhân hoá dữ liệu

Ở bước này nhóm sử dụng PCA (Principal Component Analysis) và mạng nơ-ron để trích xuất đặc trưng và phân lớp khuôn mặt để giảm số lượng mẫu [7], người tập thực hiện các tư thế tự nhiên ở nhiều góc độ như khi tập luyện, nhằm đảm bảo dữ liệu đa dạng và đầy đủ. Tiền sử bệnh án cũng được ghi nhận để hệ thống theo dõi và điều chỉnh bài tập phù hợp.

Hệ thống quay video khuôn mặt của người dùng trong 5 giây và lưu trữ các khung hình dưới dạng ảnh xám (Grayscale Conversion) để đơn giản hóa quá trình xử lý hình ảnh.

$$\text{Gray} = 0.299 \times R + 0.587 \times G + 0.114 \times B, \quad (1)$$

R, G, B là các giá trị kênh màu đỏ, xanh lá cây, và xanh dương của từng pixel. Sau khi chuyển đổi, mỗi pixel chỉ còn một giá trị cường độ sáng.



Hình 2. Định vị điểm mốc (Facial Landmark Detection).

Sau khi khuôn mặt được nhận diện trong ảnh, các điểm mốc khuôn mặt (facial landmarks) – là các tọa độ xác định đường viền của hàm, miệng, mũi, mắt và lông mày – thường được trích xuất. Sử dụng thư viện Dlib để phát hiện và định vị 68 điểm mốc trên khuôn mặt, khoảng cách và góc giữa các điểm mốc này có thể được sử dụng để biểu diễn biểu cảm khuôn mặt và suy luận trạng thái cảm xúc, chẳng hạn như cảm giác đau đớn [8].

$$(x_i, y_i) = \text{predictor}(\text{face_rectangle}). \quad (2)$$

Mỗi điểm mốc được biểu diễn dưới dạng tọa độ (x_i, y_i) trên ảnh.

Tích hợp PCA và Mạng nơ-ron trong phân lớp khuôn mặt:

Đầu tiên, PCA được sử dụng để giảm chiều dữ liệu, loại bỏ nhiễu và giữ lại các đặc trưng quan trọng nhất của khuôn mặt. Sau đó, các đặc trưng này được đưa vào mạng nơ-ron để học và phân loại.

Công thức tích hợp:

$$Z = W_{PCA}^T X, \quad (3)$$

$$\text{Output} = f\left(\sum_{i=1}^k w_i Z_i + b\right), \quad (4)$$

với W_{PCA} là ma trận các thành phần chính được chọn từ PCA, Z_i là đầu vào đã giảm chiều qua PCA, f là hàm kích hoạt trong mạng nơ-ron.

Việc tích hợp này đóng vai trò quan trọng trong việc trích xuất đặc trưng và phân loại khuôn mặt trong quá trình thu thập và cá nhân hóa dữ liệu. Sự kết hợp giữa PCA và mạng nơ-ron giúp tối ưu hóa quá trình xử lý, giảm thiểu sai sót và nâng cao hiệu suất của hệ thống.

Chuẩn bị dữ liệu: Dữ liệu được chia thành các tập huấn luyện, kiểm tra và đánh giá với tỷ lệ 70% dành cho huấn luyện, 15% dành cho kiểm tra và 15% dành cho đánh giá mô hình.

Tổng số mẫu huấn luyện: sử dụng bộ dataset mã nguồn mở “CASIA-WebFace” bao gồm 494,414 hình khuôn mặt của 10,575 người và 25 người khác tham gia dự án này, trong đó 7,413 mẫu sẽ được sử dụng cho huấn luyện, tổng số mẫu kiểm tra và đánh giá tương ứng là 1,600 mẫu kiểm tra và 1,600 mẫu đánh giá.

Bảng 1: Kết quả đánh giá hiệu suất mô hình.

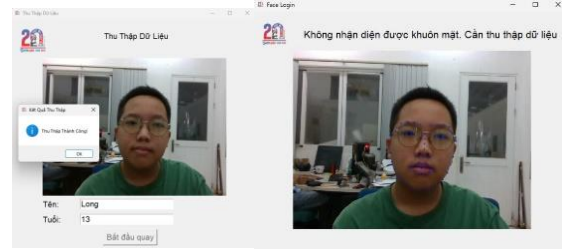
n-Fold	Correct	Errors	Accuracy
1-fold	1204	396	75.26%
2-fold	1446	154	90.38%
3-fold	1432	168	89.50%
4-fold	1423	177	88.94%
5-fold	1439	161	89.93%
Overall	6944	1056	86.80%

Mô hình phân loại thông qua phương pháp K-fold cross-validation cung cấp thông tin về hiệu suất của mô hình trong từng lần khi thực hiện quá trình kiểm tra chéo (cross-validation). Điều này giúp xác định xem mô hình có ổn định hay không khi dữ liệu thay đổi.

Kết quả thực nghiệm cho thấy mô hình nhận diện khuôn mặt hoạt động hiệu quả với độ chính xác tổng thể đạt 86.80%. Qua 5 fold kiểm thử, mô hình thể hiện sự ổn định cao đặc biệt từ fold 2-5 với độ chính xác dao động từ 88.94% đến 90.38%. Mặc dù fold 1 có hiệu suất thấp hơn (75.26%), nhưng các fold còn lại đều duy trì độ chính xác trên 88%, cho thấy khả năng học và thích nghi tốt của mô hình. Với tổng số 6944 mẫu được nhận diện đúng trên 8000 mẫu thử nghiệm, kết quả này cho thấy tính hiệu quả và độ tin cậy của mô hình trong việc nhận diện khuôn mặt, đáp ứng tốt cho các ứng dụng thực tế.

B. Kiểm soát quá trình đăng nhập

Trong quá trình tập luyện, các chỉ số và dữ liệu về thông số thu thập được ghi nhận để phục vụ cho việc đánh giá của đội ngũ y tế, do đó khi cá nhân tham gia vào quá trình tập luyện mà chưa có thông tin, hệ thống sẽ yêu cầu thu thập dữ liệu cá nhân người đó. Điều này đảm bảo rằng quá trình tập luyện luôn được theo dõi và đánh giá chính xác. (Hình 3).



Hình 3. Kiểm soát quá trình đăng nhập.

Để nhận diện khuôn mặt, mô hình CNN được áp dụng, cho phép trích xuất các đặc trưng từ các điểm mốc khuôn mặt. Mô hình này tự động học và phân loại dựa trên các đặc trưng nhận được từ dữ liệu đầu vào. Quá trình nhận diện bao gồm các bước chính như sau:

Biểu diễn dữ liệu điểm mốc khuôn mặt: Các điểm mốc khuôn mặt được biểu diễn dưới dạng tọa độ (x_i, y_i) với $i = 1, 2, \dots, n$, trong đó n là số lượng điểm mốc. Dữ liệu này được chuyển đổi thành heatmap hoặc hình ảnh nhị phân làm đầu vào cho mô hình CNN.

+ Cấu trúc mô hình CNN:

- Lớp tích chập (Convolutional Layer):

Trích xuất các đặc trưng không gian từ dữ liệu điểm mốc.

$$Z_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{i+m,j+n} \cdot W_{m,n,k} + b_k \quad (5)$$

Trong đó, $X_{(i,j)}$: Giá trị pixel tại vị trí (i, j) của đầu vào, $W_{(m,n,k)}$: Trọng số của kernel (bộ lọc) tại kênh k và b_k : Bias của kernel k .

- Hàm mất mát (Loss Function)

Để tối ưu hóa mô hình CNN, hàm mất mát phổ biến được sử dụng là Categorical Cross-Entropy:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \log(\hat{y}_{i,j}) \quad (6)$$

Trong đó N : Số lượng mẫu trong tập luyện, C : số lớp, $y_{i,j}$: nhãn thực tế (ground truth) của mẫu i cho lớp j , $\hat{y}_{i,j}$: xác suất dự đoán của mô hình cho mẫu i thuộc lớp j .

- Khoảng cách Euclidean để so sánh đặc trưng: Sau khi trích xuất đặc trưng từ CNN, khoảng cách Euclidean được sử dụng để so sánh vector đặc trưng của khuôn mặt cần phân loại với các vector trong tập huấn luyện:

$$d(x, x_i) = \sqrt{\sum_{j=1}^m (x_j - x_{ij})^2} \quad (7)$$

Trong đó $x = (x_1, x_2, \dots, x_m)$: Vector đặc trưng của điểm cần phân loại, $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$: Vector đặc trưng của mẫu trong tập huấn luyện, $d(x, x_i)$: Khoảng cách Euclidean.

Thông qua các bước trên, mô hình CNN giúp nhận diện khuôn mặt chính xác, hỗ trợ trong quá trình kiểm soát và quản lý thông tin trong các bài tập luyện.

C. Xây dựng trợ lý ảo

Chức năng này sẽ tích hợp với dữ liệu hồ sơ bệnh án của bệnh nhân để phát triển thông tin bài tập điều trị cũng như các khuyến cáo của bác sĩ dành cho bệnh nhân. Đồng thời hệ thống dựa trên kết quả nhịp tim so với hồ sơ bệnh án sẽ cảnh báo ở 2 mức: Mức cao để điều chỉnh hoạt động và mức rất cao để cảnh báo nguy hiểm và gọi sự hỗ trợ từ người thân hay bác sĩ.



Hình 4. Sơ đồ giải thuật Trợ lý ảo

Ở nội dung này nhóm sử dụng các kỹ thuật nhận dạng giọng nói (Speech Recognition), xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP), xử lý truy vấn (Query Processing), phản hồi (Response Generation). Sau khi xác định được phản hồi phù hợp, Trợ lý ảo tạo ra câu trả lời cho người dùng bằng cách sử dụng mô hình tạo ngôn ngữ để tạo ra câu trả lời tự nhiên và rõ ràng, định dạng lại câu trả lời để phù hợp với ngữ cảnh và yêu cầu của người dùng.

Bảng 2: Độ trễ trung bình của API

STT	Độ trễ Google-STT (s)	Độ trễ Google-TTS (s)	Số từ
0	1.06	0.43	10
1	1.68	0.42	20
2	2.62	0.62	30
3	3.43	0.54	50

Độ trễ của Google-STT tăng dần khi số lượng từ tăng lên: từ 1.06s (10 từ) lên 3.43s (50 từ). Điều này cho thấy rằng hiệu suất nhận dạng giọng nói bị ảnh hưởng đáng kể bởi độ dài của câu nói.

Độ trễ của Google-TTS (Text-to-Speech) tương đối ổn định, dao động từ 0.42s đến 0.62s khi số lượng từ tăng từ 10 lên 50. Sự ổn định này cho thấy khả năng chuyển đổi văn bản thành giọng nói của hệ thống ít bị ảnh hưởng bởi độ dài của câu, đảm bảo thời gian phản hồi nhanh và hiệu quả.

Bảng 3: Kết quả hiệu suất thực nghiệm

STT	Kịch bản	Trước bổ sung (%)	Sau bổ sung (%)
0	Kịch bản 'Nhịp tim vượt ngưỡng'	78%	85%
1	Kịch bản 'Nhịp tim và cảm xúc bất thường'	72%	88%

Ở kịch bản "Nhịp tim vượt ngưỡng", hiệu suất của hệ thống đã tăng từ 78% lên 85% sau khi bổ sung dữ liệu học trước. Ở kịch bản "Nhịp tim và cảm xúc bất thường", hiệu suất cũng tăng đáng kể từ 72% lên 88%. Sự cải thiện này cho thấy việc sử dụng dữ liệu lịch sử từ các buổi tập trước không chỉ giúp hệ thống hiểu rõ hơn về trạng thái sức khỏe của từng bệnh nhân mà còn tăng khả năng nhận diện chính xác các dấu hiệu bất thường.

Hệ thống đạt được mức độ cá nhân hóa tốt hơn, phản ánh khả năng xử lý dữ liệu theo từng trường hợp cụ thể, việc tích hợp dữ liệu từ các buổi tập trước đã cải thiện độ tin cậy đặc biệt trong các tình huống phức tạp kết hợp cả nhịp tim và cảm xúc bất thường.

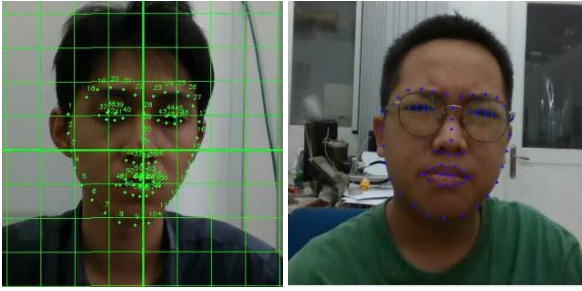
D. Giám sát, đánh giá dấu hiệu bất thường qua khuôn mặt:

Một hệ thống đánh giá và phân loại liệt mặt sử dụng mô hình học tổ hợp dựa trên SVM, kết hợp cả kỹ thuật stacking và bagging, đồng thời xử lý dữ liệu mất cân bằng. Mô hình được đánh giá qua cross-validation 5 lần, cho thấy độ ổn định và khả năng khái quát cao cho việc đánh giá sự bất thường qua khuôn mặt [3]. Đối với nghiên cứu của bài báo này, ta cần trích xuất đặc trưng không gian và thời gian: sử dụng các điểm mốc trên khuôn mặt để tính toán các khoảng cách và góc độ giữa chúng, các đặc trưng này bao gồm vị trí và tỷ lệ của lông mày, mắt, miệng và mũi. Sau khi trích xuất các đặc trưng không gian, thời gian từ video, SVM được sử dụng để phân loại các chuỗi video dựa trên sự hiện diện của biểu cảm đau. Đồng thời sử dụng công cụ mã nguồn mở viết bằng Python nhằm hỗ trợ việc phát hiện, tiền xử lý, phân tích và trực quan hóa dữ liệu biểu cảm khuôn mặt, Py-Feat còn tích hợp các mô hình phát hiện cảm xúc được huấn luyện từ các biểu cảm khuôn mặt cảm xúc do con người tạo ra hoặc được khơi gợi tự nhiên [9].

Nghiên cứu của bài báo này sử dụng 02 bộ dữ liệu [10]: UNBC-McMaster Shoulder Pain Expression Archive: Bộ dữ liệu này bao gồm các video biểu cảm đau vai của các đối tượng, thường được sử dụng trong nghiên cứu nhận diện đau. Các video được ghi lại trong các tình huống gây đau vai tự nhiên. Và BioVid Heat Pain Database: Bộ dữ liệu này chứa các video biểu cảm đau gây ra bởi nhiệt độ cao. Các video bao gồm nhiều mức độ đau khác nhau, từ nhẹ đến rất đau, do nhiệt độ cao gây ra.

Hệ thống sử dụng thư viện Dlib để trích xuất các đặc trưng từ 68 điểm mốc và áp dụng các mô hình học máy để phân loại biểu cảm, sự thay đổi trong phạm vi sai số về đặc trưng này trên khuôn mặt để đưa ra kết quả.

Nhóm cũng sử dụng mạng nơ-ron nhiều lớp (MLP-Multi-layer perceptrons) để phân loại các cử chỉ khuôn mặt. Mạng nơ-ron này được huấn luyện với bộ dữ liệu công khai và có khả năng nhận diện sáu cử chỉ: trung tính, nâng lông mày, nhắm mắt, cười mở rộng, cười khép miệng và chu môi. Cơ sở dữ liệu được sử dụng: Hệ thống được đánh giá trên cơ sở dữ liệu MEEI (Massachusetts Eye and Ear Infirmary), bao gồm 480 hình ảnh có độ phân giải cao từ 60 đối tượng, bao gồm cả người khỏe mạnh và bệnh nhân bị liệt mặt [11].



Hình 5. Trích xuất đặc trưng với 68 điểm mốc.

Trong một mạng MLP, mỗi neuron thực hiện một phép tính tuyến tính theo sau là một hàm kích hoạt phi tuyến:

$$a_j^{(l)} = \sum_{i=1}^n w_{ij}^{(l)} x_i^{(l-1)} + b_j^{(l)}, \quad (8)$$

với

- $a_j^{(l)}$: Đầu ra của neuron thứ j tại lớp l .
- $w_{ij}^{(l)}$: Trọng số kết nối từ neuron thứ i tại lớp $l - 1$ đến neuron thứ j tại lớp l .
- $x_i^{(l-1)}$: Đầu vào của neuron thứ i tại lớp $l - 1$.
- $b_j^{(l)}$: Bias của neuron thứ j tại lớp l .

Hàm kích hoạt (Activation Function)

Hàm kích hoạt quyết định đầu ra của một neuron dựa trên đầu vào tổng hợp. Một số hàm kích hoạt phổ biến trong MLP bao gồm:

Hàm Sigmoid:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (9)$$

Hàm Tanh:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (10)$$

Hàm ReLU (Rectified Linear Unit):

$$\text{ReLU}(x) = \max(0, x). \quad (11)$$

Truyền ngược (Backpropagation)

Truyền ngược là một phương pháp để tính toán gradient của hàm mất mát đối với các trọng số trong mạng nơ-ron. Đây là quá trình chính để huấn luyện MLP:

Hàm mất mát (Loss Function): Đối với bài toán phân loại, một trong những hàm mất mát phổ biến là Cross-Entropy:

$$L(y, \hat{y}) = - \sum_{i=1}^n \left[y_i \log(\hat{y}_i) \right], \quad (12)$$

với y_i : Giá trị nhãn thực tế và $\hat{y}_i$: Xác suất dự đoán.

Gradient Descent (Giảm Dốc) Gradient descent được sử dụng để điều chỉnh các trọng số nhằm giảm hàm mất mát:

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta \frac{\partial L}{\partial w_{ij}^{(l)}}, \quad (13)$$

với η là tốc độ học (learning rate) và $\frac{\partial L}{\partial w_{ij}^{(l)}}$ là đạo hàm của hàm mất mát theo trọng số $w_{ij}^{(l)}$.

Phân loại cử chỉ khuôn mặt:

Để phân loại các cử chỉ khuôn mặt và nhận diện sự bất thường, đầu ra của MLP sẽ là một tập hợp các xác suất tương ứng với các lớp cử chỉ hoặc cảm xúc khác nhau. Cử chỉ có xác suất cao nhất sẽ được chọn làm đầu ra của mô hình.

Hàm Softmax được sử dụng để chuyển đổi kết quả từ mô hình thành các giá trị xác suất, với công thức:

$$\text{softmax}(z_i) = \frac{\exp(z_i - c)}{\sum_{k=1}^C \exp(z_k - c)}, \text{ với } c = \max(z) \quad (14)$$

z_i : Giá trị đầu ra từ mô hình cho trạng thái cảm xúc i .

C : Tổng số trạng thái cảm xúc.

$c = \max(z)$: Giá trị lớn nhất trong vector đầu ra z , được sử dụng để ổn định số học.

Nhận diện sự bất thường :

Nhận diện sự bất thường có thể dựa trên ngưỡng xác suất hoặc sự kết hợp của nhiều yếu tố, chẳng hạn như xác suất dự đoán thấp hoặc mô hình không chắc chắn.

Ngưỡng xác suất được xác định :

If $\max(\text{Softmax}(z)) < \text{threshold}$, thì nhận diện bất thường

Ngưỡng (threshold) có thể được xác định bằng cách tối ưu hóa trên tập huấn luyện hoặc dựa trên kinh nghiệm.

III. KẾT QUẢ NHẬN ĐỊNH

Bài báo đã trình bày phương pháp tích hợp để nhận diện khuôn mặt, thu thập dữ liệu và phân tích cảm xúc, sự bất thường do đau, căng thẳng... gây ra từ sự thay đổi vị trí đặc trưng trên khuôn mặt bằng cách sử dụng camera RealSense và các thuật toán học máy. Hệ thống đã chứng minh được tính khả thi khi ứng dụng vào lĩnh vực hỗ trợ đội ngũ y tế trong việc giám sát, theo dõi trong quá trình tập luyện phục hồi chức năng. Việc tích hợp các công nghệ tiên tiến như Dlib và RealSense đã mang lại độ chính xác ổn định và khả năng hoạt động trong thời gian thực cho hệ thống.

Bảng 4: Kết quả nhận diện phân loại cảm xúc.

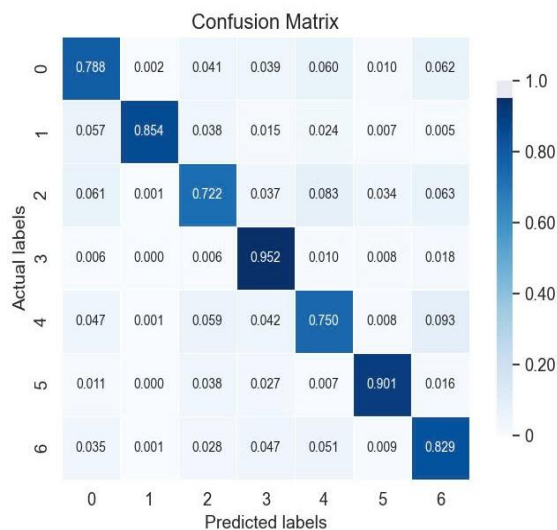
Emotion	Test	Average
Angry	84.20%	85.50%
Disgust	80.40%	85.50%
Fear	77.10%	85.50%

Happy	98.40%	85.50%
Sad	82.00%	85.50%
Surprise	94.50%	85.50%
Neutral	89.00%	85.50%
Overall	86.51%	85.50%

Kết quả thử nghiệm cho thấy mô hình đạt độ chính xác tổng thể là 86.51%, tuy nhiên, hiệu suất giữa các cảm xúc không đồng đều. Các cảm xúc "Happy" (98.40%) và "Surprise" (94.50%) đạt độ chính xác cao, nhờ vào đặc trưng khuôn mặt dễ nhận diện và phân biệt. Trong khi đó, các cảm xúc "Fear" (77.10%) và "Disgust" (80.40%) có độ chính xác thấp hơn, chủ yếu do sự tương đồng trong các đặc trưng khuôn mặt và khả năng bị nhầm lẫn với các cảm xúc khác. Kết quả này phản ánh những thách thức trong việc nhận diện các cảm xúc phức tạp hoặc có đặc điểm tương tự nhau.

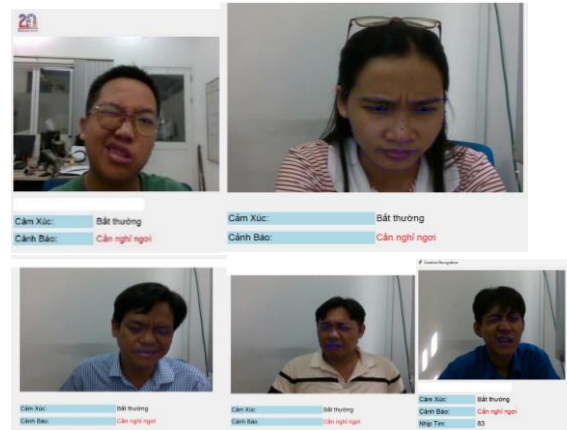
Để cải thiện hiệu suất mô hình nhóm áp dụng thuật toán hợp nhất dựa trên hồ sơ cảm xúc cải tiến (IEPs) và độ tin cậy tối đa được đưa ra để nhận diện cảm xúc [12], trong đó cần tập trung vào việc mở rộng và tăng cường nhiều thêm bộ dữ liệu huấn luyện cho các cảm xúc khó phân biệt, đặc biệt là những cảm xúc dễ bị nhầm lẫn. Việc giảm thiểu sự nhầm lẫn giữa các cảm xúc này sẽ nâng cao độ chính xác và cải thiện hiệu quả ứng dụng của mô hình trong các tình huống thực tế.

Bảng 5 : Ma trận nhầm lẫn.



Biểu đồ ma trận nhầm lẫn cung cấp cái nhìn chi tiết về hiệu suất của mô hình nhận diện cảm xúc. Các giá trị trên đường chéo chính thể hiện tỷ lệ dự đoán đúng, với "Happy" đạt độ chính xác cao nhất (98.4%) nhờ đặc trưng rõ ràng, và "Surprise" đạt 94.5%. Ngược lại, "Fear" có độ chính xác thấp nhất (77.1%) do dễ nhầm lẫn với "Surprise" (3.5%) và "Sad" (5.4%). Các cảm xúc như "Disgust" (80.4%) và "Sad" (82.0%) cũng có hiệu suất trung bình, trong khi "Neutral" đạt 89.0% nhưng bị nhầm với "Sad" (2.5%). Tỷ lệ nhầm lẫn ngoài đường chéo chính cho thấy thách thức trong việc phân biệt các cảm xúc tương tự, như "Angry" (84.2%) thường bị nhầm với "Fear" (3.5%) và "Disgust"

(2.5%). Mô hình đạt độ chính xác trung bình 86.51%, tuy hiệu suất chưa đồng đều giữa các cảm xúc. Để mô hình cải thiện đạt kết quả cao hơn, ta cần phải tăng cường dữ liệu huấn luyện và áp dụng các kỹ thuật học sâu. Ma trận nhầm lẫn giúp xác định điểm mạnh và yếu của mô hình, từ đó cải thiện hiệu suất bằng cách điều chỉnh mô hình hoặc thu thập thêm dữ liệu cho các lớp khó phân biệt.



Hình 6. Kết quả nhận diện bất thường.

Bài báo này hướng đến xây dựng giải pháp hỗ trợ đội ngũ y tế trong giám sát, theo dõi quá trình tập luyện phục hồi chức năng, hạn chế sự cố bất thường xảy ra nhờ vào hệ thống nhận diện khuôn mặt. Các yếu tố như dữ liệu nghiên cứu về các bệnh như liệt mặt, đột quỵ... sẽ ảnh hưởng khá nhiều đến kết quả nghiên cứu, cũng như sự đa dạng đối tượng chưa được thực nghiệm nhiều ở bài báo này.

Hướng phát triển tiếp theo nhóm sẽ phối hợp cùng các đơn vị liên quan để cùng xây dựng cơ sở dữ liệu về sinh trắc học cho đối tượng có tiền sử bệnh có ảnh hưởng đến sự thay đổi dấu hiệu khuôn mặt và tiếp tục phát triển các phương pháp tối ưu, phù hợp trong lĩnh vực chăm sóc sức khỏe, góp phần vào xây dựng thành phố thông minh, công cuộc chuyển đổi số, đổi mới sáng tạo trong y tế.

Nhóm tác giả cũng cảm ơn những đóng góp, những kết quả thực nghiệm mà em Lê Nam Long, học sinh lớp 8A3 Trường phổ thông chuyên Trần Đại Nghĩa đã phối hợp trong suốt quá trình thực hiện công trình này.

REFERENCES

- [1] P Ekman, WV Friesen, "Measuring facial movement, Environmental psychology and nonverbal behavior", 1976 Springer
- [2] JJ Greene, DL Guarin, J Tavares, E Fortier, M Robinson, J Dusseldorp, O Quatela, N Jowett, "The spectrum of facial palsy: The MEEI facial palsy photo and video standard set", The Laryngoscope, 2020
- [3] Amira Gaber, Mona F. Taher, Manal Abdel Wahed, Nevin Mohieldin Shalaby & Sarah Gaber, "Classification of facial paralysis based on machine learning techniques", BioMedical Engineering OnLine volume 21, Article number: 65 (2022)
- [4] Tawsin Uddin Ahmed; Sazzad Hossain; Mohammad Shahadat Hossain; Raihan ul Islam; Karl Andersson, "Facial Expression Recognition using Convolutional Neural

Network with Data Augmentation”, International Conference on Informatics, Electronics & Vision (ICIEV), IEEE, 2019.

- [5] Z Guo, G Dan, J Xiang, J Wang, W Yang, H Ding, O Deussen, Y Zhou, "An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis", IEEE journal of biomedical and health informatics, 2017.
- [6] Anuja Arora, ORCID, Jasir Mohammad Zaeem, Vibhor Garg, Ambikesh Jayal and Zahid Akhtar, "A Deep Learning Approach for Early Detection of Facial Palsy in Video Using Convolutional Neural Networks: A Computational Study", MDPI Computers, 2024
- [7] Đậu Trọng Hiền, Ngô Quốc Cường, Trần Tùng Giang, "Thiết kế hệ thống tách và nhận dạng khuôn mặt từ video", Tạp chí khoa học giáo dục kỹ thuật số 22/2012.
- [8] Werner, P., Al-Hamadi, A., Limbrecht-Ecklundt, K., Walter, S., Gruss S., Traue, H. C. "Automatic pain assessment with facial activity descriptors". IEEE Transactions on Affective Computing, 8, 286–299, 2017.
- [9] Jin Hyun Cheong, Eshin Jolly, Tiankang Xie, Sophie Byrne, Matthew Kenney, Luke Chang, "Py-Feat: Python Facial Expression Analysis Toolbox", Volume 4, pages 781–796, Affective Science, SpringerLink, 2023
- [10] R. Yang, S. Tong, M. Bordallo, E. Boutellaa, J. Peng, X. Feng and A. Hadid, "On Pain Assessment from Facial Videos Using Spatio-Temporal Local Descriptors" The Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA2016).
- [11] Gemma S. Parra-Dominguez, Raul E. Sanchez-Yanez and Carlos H. Garcia-Capulin "Towards Facial Gesture Recognition in Photographs of Patients with Facial Palsy", Artificial Intelligence Developments for Medical Diagnosis and Monitoring, Healthcare, MDPI, 2022
- [12] Qi-rong Mao, Xin-yu Pan, Yong-zhao Zhan & Xiang-jun Shen, "Using Kinect for real-time emotion recognition via facial expressions", Frontiers of Information Technology & Electronic Engineering, Volume 16, pages 272–282, SpringerLink. 2015,

DETECTION OF ABNORMAL RISK SIGNS ON THE FACE DURING THE REHABILITATION TRAINING PROCESS

Abstract: In the rehabilitation training process, early detection of abnormal risk signs on the face plays a crucial role in ensuring safety and optimizing treatment effectiveness. This study focuses on applying machine learning and image processing techniques to identify and classify abnormal facial signs, including facial muscle deformation, signs of stress, or unnatural asymmetry. Image data collected from training sessions are analyzed and processed to build a recognition model, which provides early warnings and supports the adjustment of treatment methods. The research results contribute to enhancing the effectiveness of the rehabilitation support system, improving patient care, and minimizing risks during the treatment process.

Key words— SVM, CNN, KNN, Facial Landmark Detection.



Bùi Quang Vinh – Thạc sỹ kỹ thuật, hiện là trưởng phòng Cơ khí Tự động hóa của Trung tâm R&D Khu Công nghệ cao TPHCM, tham gia nghiên cứu trong lĩnh vực cơ khí, tự động hóa, hệ thống máy thông minh, AioT.



Lê Quốc Cường nhận bằng Tiến sỹ viễn thông năm 1996 tại Đại học Saint Petersburg, được phong hàm phó giáo sư năm 2024. Lĩnh vực nghiên cứu: Viễn thông, MEMs, IOTs, hiện là Phó trưởng ban Ban quản lý Khu Công nghệ cao TPHCM, phụ trách đào tạo, nghiên cứu triển khai, đổi mới sáng tạo và phát triển doanh nghiệp.



Trần Thanh Hải – Kỹ sư kỹ thuật phần mềm Đại học FPT – Nghiên cứu viên phòng cơ khí chính xác và tự động hóa ở Trung tâm Nghiên cứu Triển khai khu công nghệ cao. Lĩnh vực nghiên cứu: Phần mềm hệ thống, AI.



Hứa Nguyễn Đăng Thy - Thạc sỹ kỹ thuật - Đại học Bách Khoa TpHCM. Hiện là Nghiên cứu viên tại Trung tâm R&D Khu Công nghệ cao TpHCM. Lĩnh vực nghiên cứu: Tự động hóa, Robot, IOTs.



Vũ Ngọc Long – Kỹ sư Công nghệ Kỹ thuật Cơ khí. Hiện là Nghiên cứu viên phòng cơ khí chính xác - Tự động hóa. Thuộc Trung tâm Nghiên cứu Triển khai Khu Công nghệ cao. Lĩnh vực nghiên cứu: CAD/CAM/CNC, máy thông minh.