

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - CƠ - TIN HỌC



MÔN HỌC: PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

**ĐỀ TÀI: HỆ THỐNG PHÁT HIỆN TÌNH TRẠNG BUỒN NGỦ CỦA
LÁI XE**

Sinh viên:

Nguyễn Hữu Quang - 21001574

Đặng Đình Thắng - 21001587

Nguyễn Văn Thắng - 21002175

Giảng viên hướng dẫn:

TS. Nguyễn Thị Bích Thủy

Hà Nội - 2024

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - CƠ - TIN HỌC



MÔN HỌC: PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

**ĐỀ TÀI: HỆ THỐNG PHÁT HIỆN TÌNH TRẠNG BUỒN NGỦ CỦA
LÁI XE**

Sinh viên:

Nguyễn Hữu Quang - 21001574

Đặng Đình Thắng - 21001587

Nguyễn Văn Thắng - 21002175

Giảng viên hướng dẫn:

TS. Nguyễn Thị Bích Thủy

Hà Nội - 2024

Lời cam đoan

Nhóm sinh viên xin cam đoan đây là công trình nghiên cứu của nhóm, với sự hỗ trợ tận tình từ giảng viên hướng dẫn là TS. Nguyễn Thị Bích Thủy. Các tài liệu tham khảo, các số liệu phục vụ cho việc xây dựng và dẫn dắt đề tài này được thu thập từ các nguồn tài liệu khác nhau và được ghi chú trong mục tài liệu tham khảo theo đúng quy định. Nội dung và kết quả nghiên cứu trong đề tài này là trung thực và chưa từng được công bố trong bất cứ công trình nghiên cứu nào trước đây.

Lời cảm ơn

Chúng em muốn bày tỏ lòng biết ơn sâu sắc đến TS. Nguyễn Thị Bích Thủy, người đã định hướng, hỗ trợ, trực tiếp hướng dẫn và tận tình chỉ bảo chúng em trong suốt quá trình nghiên cứu, xây dựng và hoàn thiện đề tài.

Một lời cuối, chúng em xin chân thành cảm ơn và kính chúc cô mạnh khỏe và hạnh phúc trong cuộc sống, cùng tiếp tục chia sẻ kiến thức và đạt được nhiều thành công trong công việc giáo dục và nghiên cứu.

Tuy có nhiều cố gắng nhưng do những hạn chế về thời gian, kiến thức và kinh nghiệm nên không thể tránh khỏi những sai sót. Nhóm nghiên cứu rất mong được tiếp thu các ý kiến đóng góp của các thầy, cô, bạn bè, những người quan tâm đến đề tài nghiên cứu này để bài nghiên cứu được hoàn thiện hơn.

Hà Nội, ngày 30 tháng 5 năm 2024

Mục lục

Danh sách hình vẽ	v
1 ĐẶT VẤN ĐỀ	3
1.1 Thực trạng	3
1.2 Phát triển bài toán	3
1.3 Thách thức của bài toán và cách giải quyết của các nhóm nghiên cứu	4
2 PHƯƠNG PHÁP TRIỂN KHAI	6
2.1 Bộ dữ liệu	6
2.1.1 The Real-Life Drowsiness Dataset (RLDD)	6
2.1.2 The Closed Eyes in Wild (CEW)	8
2.2 Phương pháp đề xuất	8
2.2.1 Trích xuất đặc trưng	9
2.2.1.1 Phát hiện nháy mắt	9
2.2.1.2 Trích xuất các đặc trưng nháy mắt	9
2.2.2 Xây dựng mạng LSTM phát hiện buồn ngủ	13
3 KẾT QUẢ - ĐÁNH GIÁ	16
3.1 Triển khai	16
3.2 Thực nghiệm	17
3.3 Đánh giá	19
4 KẾT LUẬN - HƯỚNG PHÁT TRIỂN	20

TÀI LIỆU THAM KHẢO

20

Danh sách hình vẽ

2.1	Video trong trạng thái tỉnh táo.	7
2.2	Video trong trạng thái buồn ngủ.	7
2.3	Ảnh nhắm mắt.	8
2.4	Ảnh mở mắt.	8
2.5	Quy trình hoạt động.	9
2.6	Dự đoán 68 điểm trên khuôn mặt theo góc thẳng và góc nghiêng.	10
2.7	68 điểm trên khuôn mặt	11
2.8	12
2.9	12
3.1	Kết quả thực nghiệm trên fold 1.	17
3.2	Kết quả thực nghiệm trên fold 2.	17
3.3	Kết quả thực nghiệm trên fold 3.	18
3.4	Kết quả thực nghiệm trên fold 4.	18
3.5	Kết quả thực nghiệm trên fold 5.	18
3.6	Kết quả dự báo đúng.	19
3.7	Kết quả dự báo sai.	19

Lời nói đầu

Cuộc sống phát triển nên nhu cầu đi lại, vận chuyển hàng hóa ngày càng nhiều. Các phương tiện giao thông ngày càng được cải tiến và có rất nhiều thiết bị hỗ trợ việc tham gia giao thông tiện lợi và an toàn hơn. Thế nhưng tình trạng an toàn giao thông hiện nay lại trở thành vấn đề đáng báo động. Số vụ tai nạn giao thông ngày càng gia tăng, gây thiệt hại lớn về người và tài sản. Một trong những nguyên nhân dẫn đến tai nạn giao thông nghiêm trọng là do lái xe buồn ngủ.

Phát hiện tình trạng buồn ngủ của lái xe là một lĩnh vực nghiên cứu quan trọng trong an toàn giao thông. Việc phát triển các hệ thống phát hiện tình trạng buồn ngủ của lái xe hiệu quả có thể góp phần giúp người lái xe được cảnh báo sớm về tình trạng buồn ngủ làm giảm nguy cơ xảy ra tai nạn. Đề tài nghiên cứu này tập trung vào việc phát triển một hệ thống sử dụng mô hình LSTM có thể phát hiện sớm tình trạng buồn ngủ trong điều kiện môi trường thực tế.

Báo cáo trình bày chi tiết quá trình nghiên cứu và phát triển phát hiện sớm tình trạng buồn ngủ sử dụng mô hình LSTM, bao gồm các nội dung sau:

- **Chương 1:** Đặt vấn đề
- **Chương 2:** Phương pháp triển khai
- **Chương 3:** Kết quả và đánh giá
- **Chương 4:** Kết luận và hướng phát triển

Chương 1

ĐẶT VẤN ĐỀ

1.1 Thực trạng

Phát hiện sự buồn ngủ là một vấn đề quan trọng. Các giải pháp tốt có ứng dụng quan trọng trong các lĩnh vực như lái xe và nơi làm việc. Ví dụ, trong việc lái xe, Cơ quan An toàn Giao thông Quốc gia ở Hoa Kỳ ước tính rằng mỗi năm có 100.000 vụ tai nạn được báo cáo của cảnh sát là kết quả trực tiếp của tình trạng ngủ gật khi lái xe. Điều này dẫn đến khoảng 1.550 người chết, 71.000 người bị thương và mất 12,5 tỷ đô la Mỹ. Để so sánh, ước tính trong 25 người thì có 1 người báo cáo rằng họ đã ngủ gục trong khi lái xe trong 30 ngày trước đó. Ở Việt Nam cũng xảy ra nhiều vụ tai nạn nghiêm trọng mỗi năm vì tài xế buồn ngủ. Ngoài ra, các nghiên cứu cho thấy rằng, khi lái xe trong thời gian dài, lái xe mất khả năng tự đánh giá mức độ buồn ngủ của họ, và điều này là một trong những nguyên nhân chính khiến nhiều vụ tai nạn xảy ra.

1.2 Phát triển bài toán

Tất cả những sự thật đáng lo ngại này thúc đẩy nhu cầu về một giải pháp có thể phát hiện sự buồn ngủ ở các giai đoạn sớm, hỗ trợ người tài xế ý thức và chủ động hơn trong

việc xử lý các tình huống nguy hiểm tiềm ẩn khi điều khiển phương tiện lưu thông.

Đề tài xây dựng một hệ thống phát hiện các dấu hiệu buồn ngủ của lái xe và đưa ra cảnh báo cho tài xế để có những biện pháp xử lý phù hợp.

1.3 Thách thức của bài toán và cách giải quyết của các nhóm nghiên cứu

Thách thức của bài toán là làm sao để dự đoán chính xác và sớm nhất về tình trạng buồn ngủ của lái xe từ đó có thể ứng dụng rộng rãi trong thực tế.

Thông thường có ba loại dấu hiệu trong việc phát hiện sự buồn ngủ: hiệu suất lái xe, đặc trưng về sinh lý, và đặc trưng hành vi. Trong lĩnh vực lái xe, việc tính toán về hiệu suất lái xe tập trung vào các chuyển động của vô lăng, tốc độ lái xe, mẫu phanh và việc lệch làn đường. Một ví dụ là hệ thống Hỗ trợ tập trung của Mercedes Benz. Mặc dù các phương pháp này có thể ứng dụng thực tế nhưng thường được dành cho các mẫu cao cấp, vì chúng quá đắt đỏ với người tiêu dùng thông thường. Phương pháp này cần có các thiết bị hỗ trợ đặc biệt trên xe nên không khả thi khi nghiên cứu. Đặc trưng về sinh lý như nhịp tim, điện tâm đồ (ECG), điện cơ (EMG), điện não đồ (EEG) có thể được sử dụng để theo dõi sự buồn ngủ. Tuy nhiên, việc thu thập tín hiệu của các phương pháp như vậy là quá phiền phức, dễ bị nhiễu và không thực tế để ứng dụng trong xe hơi hoặc không gian làm việc mặc dù độ chính xác của chúng rất cao.

Các đặc trưng về hành vi được lấy từ các chuyển động và biểu hiện khuôn mặt bằng cách sử dụng các cảm biến như máy ảnh điện thoại là một phương thức thay thế có thể tiếp cận và rẻ tiền so với các phương pháp trên mà độ chính xác cũng rất cao.

Trong các phương pháp đã có sử dụng đặc trưng hành vi thì chủ yếu tập trung vào việc phát hiện sự buồn ngủ với các dấu hiệu rõ ràng như ngáp, ngủ gục và đóng mắt kéo dài. Tuy nhiên, đối với các lái xe, các dấu hiệu rõ ràng như vậy có thể không xuất hiện cho

đến khi chỉ còn vài khoảnh khắc trước một vụ tai nạn. Do đó, việc phát hiện sự buồn ngủ ở giai đoạn đầu tiên có giá trị đáng kể, để cung cấp thêm thời gian cho các phản ứng phù hợp.

Mục tiêu của báo cáo này là tìm hiểu và triển khai một hệ thống sử dụng đầu vào từ camera để phát hiện cảnh báo sớm và rõ ràng hơn về dấu hiệu của sự buồn ngủ trong thời gian thực.(sử dụng dấu hiệu gì

Chương 2

PHƯƠNG PHÁP TRIỂN KHAI

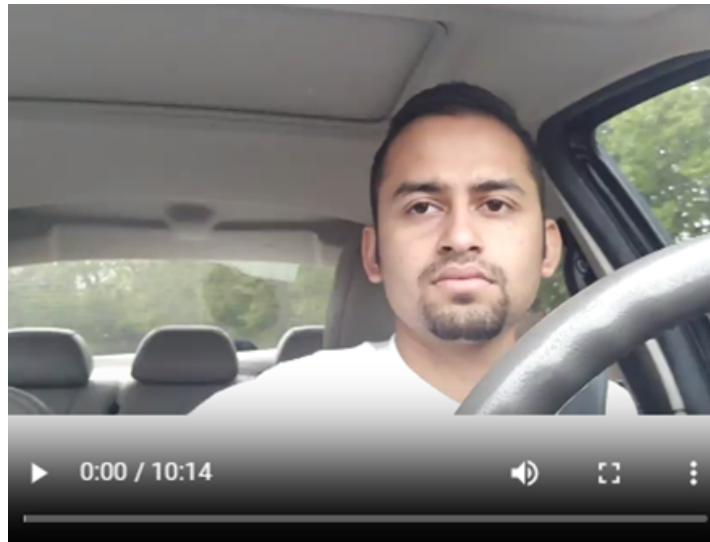
2.1 Bộ dữ liệu

2.1.1 The Real-Life Drowsiness Dataset (RLDD)

Bộ dữ liệu RLDD được tạo ra cho nhiệm vụ phát hiện mệt mỏi nhiều mức độ, nhằm đến không chỉ các trường hợp nghiêm trọng và dễ nhìn thấy, mà còn các trường hợp mệt mỏi khó phát hiện. Việc phát hiện những trường hợp mệt mỏi khó phát hiện này có thể quan trọng để phát hiện mệt mỏi ở giai đoạn đầu, nhằm kích hoạt các cơ chế ngăn ngừa mệt mỏi. Bộ dữ liệu RLDD là bộ dữ liệu mô phỏng mệt mỏi lớn nhất cho đến nay.

Bộ dữ liệu RLDD bao gồm khoảng 30 giờ video RGB của 60 người tham gia khỏe mạnh. Đối với mỗi người tham gia, thu thập một video cho mỗi trong ba lớp khác nhau: tỉnh táo, giảm sự chú ý và mệt mỏi, tổng cộng 180 video. Người tham gia là sinh viên đại học hoặc sau đại học và các thành viên nhân viên tham gia tự nguyện hoặc nhận tín chỉ bổ sung trong một khóa học. Tất cả các người tham gia đều trên 18 tuổi. Có 51 nam và 9 nữ, thuộc các dân tộc và độ tuổi khác nhau (10 người da trắng, 5 người da Mỹ Latinh không trắng, 30 người da Ấn-Âu và Dravidian, 8 người Trung Đông và 7 người Đông Á), (từ 20 đến 59 tuổi với trung bình 25 và độ lệch chuẩn 6). Có 21 video trong tổng số 180 video,

người tham gia đeo kính, và có râu ria đáng kể trong 72 video trong tổng số 180 video. Video được quay từ các góc độ khác nhau trong các môi trường và phong nền thực tế khác nhau. Mỗi video được người tham gia tự ghi lại bằng điện thoại di động hoặc webcam. Tốc độ khung hình luôn nhỏ hơn 30 fps, đại diện cho tốc độ khung hình dự kiến của các máy ảnh thông thường được sử dụng bởi công chúng chung.



Hình 2.1: Video trong trạng thái tỉnh táo.



Hình 2.2: Video trong trạng thái buồn ngủ.

2.1.2 The Closed Eyes in Wild (CEW)

Bộ dữ liệu CEW là bộ dữ liệu có quy mô lớn này bao gồm cả hình ảnh mắt đóng và mở của con người, chủ yếu được sử dụng cho việc phát hiện mắt và có thể được mở rộng để phát hiện mặt mới.

Hình ảnh trong bộ dữ liệu được thu thập dưới nhiều điều kiện khác nhau, bao gồm độ sáng, khoảng cách, độ phân giải, góc mặt và góc mắt. Những thông số này giúp đạt được kết quả tốt với khả năng chính xác cao nhất.

Bộ dữ liệu này có nhiều tập tin. Tập tin 1 có 10.000 hình ảnh được chia thành 5.000 hình ảnh cho mắt đóng và 5.000 hình ảnh cho mắt mở. Tập tin 2 được chia thành 2.500 hình ảnh cho mắt đóng và 2.500 hình ảnh cho mắt mở. Tập tin 3 có một bộ sưu tập khác gồm 10.000 hình ảnh, được chia thành 5.000 hình ảnh cho mắt đóng và 5.000 hình ảnh cho mắt mở. Tập tin 4 của bộ dữ liệu, tức tập tin 4, có tổng cộng 4.000 hình ảnh, gồm 2.000 hình ảnh cho mắt đóng và 2.000 hình ảnh cho mắt mở.



Hình 2.3: Ảnh nhắm mắt.

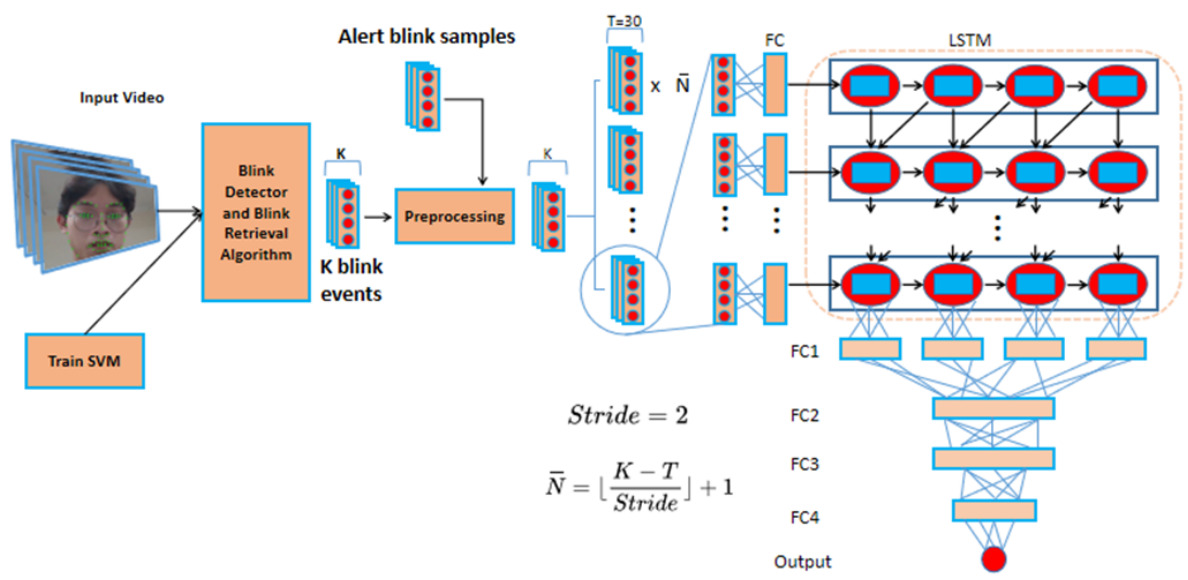


Hình 2.4: Ảnh mở mắt.

2.2 Phương pháp đề xuất

Phần này chỉ ra các thành phần trong quy trình hoạt động của hệ thống phát hiện buồn ngủ. Đầu tiên, hệ thống sử dụng thuật toán phát hiện nháy mắt và trích xuất các đặc trưng

nháy mắt từ video đầu vào. Đầu ra của quá trình này là dữ liệu về các nháy mắt. Sau đó, hệ thống thực hiện tiền xử lý dữ liệu làm đầu vào cho mô hình. Cuối cùng là sử dụng mô hình LSTM nhiều lớp để dự đoán nhãn phân loại cho mỗi đoạn video



Hình 2.5: Quy trình hoạt động.

2.2.1 Trích xuất đặc trưng

2.2.1.1 Phát hiện nháy mắt

Báo cáo này sử dụng một công cụ phát hiện khuôn mặt đã được huấn luyện trước của dlib, dựa trên một phiên bản sửa đổi của phương pháp Histogram of Oriented Gradients + Linear SVM tiêu chuẩn cho việc phát hiện đối tượng.

2.2.1.2 Trích xuất các đặc trưng nháy mắt

Mục đích đằng sau việc sử dụng các đặc trưng liên quan đến nháy mắt như thời gian nháy mắt, biên độ, tần số và tốc độ mở mắt là để bắt những khoảnh khắc xuất hiện tự nhiên trong mắt người và có thể bị bỏ qua bởi các bộ đặc trưng không gian như CNNs.



Hình 2.6: Dự đoán 68 điểm trên khuôn mặt theo góc thẳng và góc nghiêng.

Trong các thực nghiệm, các phương pháp xác định các chớp mắt nhanh liên tiếp như một chớp mắt duy nhất gây ra vấn đề khi phát hiện tình trạng buồn ngủ, vì nhiều chớp mắt liên tiếp thường là một dấu hiệu của tình trạng buồn ngủ. Vì vậy chúng tôi đã cải thiện thuật toán để giúp xác định các chớp mắt liên tiếp.

Đầu vào của thuật toán phát hiện nháy mắt là toàn bộ video (có độ dài khoảng mười phút trong tập dữ liệu). Đầu ra của thuật toán phát hiện nháy mắt là một chuỗi các sự kiện nháy mắt $blink_1, \dots, blink_K$. Mỗi blinki là một vector bốn chiều chứa bốn đặc trưng mô tả nháy mắt: thời gian, biên độ, tốc độ mở mắt và tần số. Đối với mỗi sự kiện nháy mắt blinki, xác định starti, bottomi và endi là các điểm "bắt đầu", "đáy" và "kết thúc" (frames) trong nháy mắt đó (được giải thích trong phần phụ lục "Blink Retrieval Algorithm"). Ngoài ra, đối với mỗi frame k , ký hiệu là:

$$EAR[k] = \frac{\|\vec{p}_2 - \vec{p}_6\| + \|\vec{p}_3 - \vec{p}_5\|}{\|\vec{p}_1 - \vec{p}_4\|} \quad (1)$$

Trong đó \vec{p}_i là vị trí 2D của một điểm định vị trên khuôn mặt trong vùng mắt (hình 2.7).



Hình 2.7: 68 điểm trên khuôn mặt

Khi tính toán EAR thì hệ thống lấy trung bình cộng của mắt phải và mắt trái. Nhưng có một nhược điểm phương pháp này là nếu ở góc nghiêng thì diện tích vùng mắt của một mắt sẽ bị nhỏ hơn mắt còn lại.

Do đó, chúng tôi đã cải thiện thuật toán của Reza Ghoddoosian, Marnim Galib và Vassilis Athitsos để có thể phát hiện nháy mắt ở góc nghiêng. Để nhận biết góc mặt nghiêng sang bên nào, ta tính toán khoảng cách $d_1 = D(point_{49}, point_5)$ và $d_2 = D(point_{55}, point_{13})$ rồi so sánh d_1, d_2 .

Nếu $d_1 > 2d_2$ thì nghĩa là mặt người đang quay sang phải theo hướng camera thì khi đó:

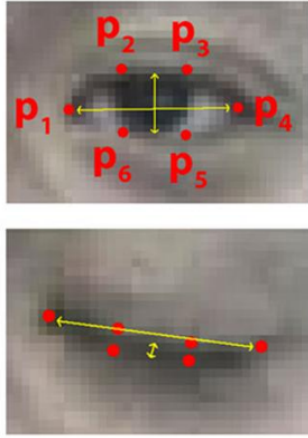
$$EAR = EAR_{left} \quad (2)$$

Ngược lại $d_2 > 2d_1$ thì nghĩa là mặt người đang quay sang trái theo hướng camera thì khi đó:

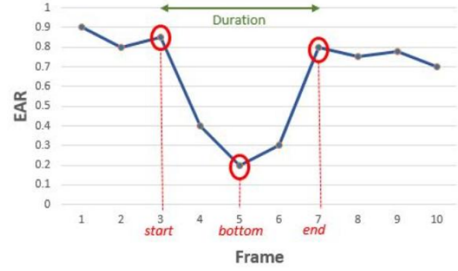
$$EAR = EAR_{right} \quad (3)$$

Hệ thống xác định các điểm $start_i, bottom_i, end_i$ và sử dụng chỉ số EAR để trích xuất bốn

đặc trưng chính từ nháy mắt.



Hình 2.8



Hình 2.9

Đây là các đặc trưng được sử dụng cho phương pháp phát hiện buồn ngủ sớm:

- Tần suất chớp mắt là số lần chớp mắt trong một khoảng thời gian nhất định.

$$Frequency_i = \frac{Number\ of\ blinks\ up\ to\ blink_i}{Number\ of\ frames\ up\ to\ end_i} \times 100 \quad (4)$$

- Thời gian chớp mắt là khoảng thời gian mà mí mắt nhắm lại.

$$Duration_i = end_i - start_i + 1 \quad (5)$$

- Biên độ chớp mắt là mức độ mở rộng của mí mắt khi chớp mắt.

$$Amplitude_i = \frac{EAR[start_i] - 2EAR[bottom_i] + EAR[end_i]}{2} \quad (6)$$

- Tốc độ chớp mắt là tốc độ mà mí mắt mở và nhắm lại khi chớp mắt.

$$Velocity_i = \frac{EAR[start_i] - 2EAR[bottom_i] + EAR[end_i]}{end_i - bottom_i} \quad (7)$$

Khi buồn ngủ, thời gian của một chớp mắt có xu hướng tăng còn tần suất, biên độ, tốc

độ của một chớp mắt có xu hướng giảm do mí mắt trở nên nặng nề và khó mở.

2.2.2 Xây dựng mạng LSTM phát hiện buồn ngủ

Tiền xử lý dữ liệu: Bước tiền xử lý dữ liệu là một bước rất quan trọng. Ở báo cáo này, các đặc trưng được chuẩn hóa qua từng người tham gia đã giúp giải quyết một trong những thách thức lớn của việc áp dụng vào thực tế. Đó là sự khác biệt trong mẫu nháy mắt của mỗi người là khác nhau. Sự khác nhau đó có thể do cấu tạo mắt từng người khác nhau hoặc tình trạng mắt lúc bình thường và khi bị thương khác nhau.

Để giải quyết thách thức này, phương pháp sử dụng một phần đầu tiên thời gian của mỗi cá nhân ở trạng thái tỉnh táo để tính trung bình và độ lệch chuẩn của mỗi đặc trưng nháy mắt cho mỗi cá nhân, sau đó sử dụng công thức bên dưới để chuẩn hóa phần nháy mắt còn lại của cùng một người (m) và đặc trưng (n):

$$normalized_Feature_{n,m} = \frac{Feature_{n,m} - \mu_{n,m}}{\sigma_{n,m}} \quad (8)$$

Ở đây, $\mu_{Feature}$ và $\sigma_{Feature}$ là trung bình và độ lệch chuẩn của đặc trưng n trong ba phút đầu tiên trong video trạng thái tỉnh táo của một người (m).

Quá trình chuẩn hóa này được thực hiện cho cả dữ liệu huấn luyện và kiểm tra của tất cả các đặc trưng. Quá trình chuẩn hóa này là một ràng buộc thực tế: khi tài xế bắt đầu lái chiếc xe mới hoặc một người lao động bắt đầu làm việc, camera có thể sử dụng vài phút đầu tiên (trong đó người đó được kỳ vọng tỉnh táo) để tính trung bình và phương sai, và hiệu chỉnh hệ thống.

Lớp biến đổi đặc trưng: Thay vì xác định một số lượng lớn đặc trưng ban đầu và sau đó lựa chọn những đặc trưng quan trọng nhất. Tiếp theo, mạng học sâu sử dụng bốn đặc trưng chính của nháy mắt và học cách ánh xạ chúng vào một không gian đặc trưng có số chiều cao hơn để giảm thiểu hàm mất mát. Mục tiêu của lớp kết nối đầy đủ trước module LSTM là nhận mỗi vector đặc trưng 4D tại mỗi khoảng thời gian làm đầu vào và biến đổi

nó thành một không gian L chiều với trọng số chung ($\mathbf{W} \in \mathbb{R}^{4 \times L}$) và độ lệch ($\mathbf{b} \in \mathbb{R}^{1 \times L}$) qua các khoảng thời gian. Định nghĩa T là số bước thời gian được sử dụng cho Mạng LSTM và $\mathbf{f}_i \in \mathbb{R}^{1 \times L}$ cho mỗi nháy mắt tại mỗi khoảng thời gian i , sao cho:

$$F = \text{ReLU}(BW + \vec{a}) \quad (9)$$

Mạng LSTM: nhận mỗi hàng của F làm đầu vào tại mỗi khoảng thời gian và xuất ra trạng thái ẩn $h_l \in \mathbb{R}^{1 \times H}$ chỉ tại khoảng thời gian cuối cùng cho mỗi tầng l . H là số trạng thái ẩn cho mỗi tầng.

Lớp kết nối đầy đủ: sử dụng một lớp kết nối đầy đủ (với $W_{1,l} \in \mathbb{R}^{H \times L_1}$ là ma trận trọng số và $b_{1,l} \in \mathbb{R}^{1 \times L_1}$ làm độ lệch) vào đầu ra của mỗi tầng l với L_1 đơn vị để thu thập kết quả từ mạng LSTM từ các quan điểm phân cấp khác nhau một cách riêng biệt. Đặt $e_{1l} \in \mathbb{R}^{1 \times L_1}$ cho mỗi tầng, sao cho:

$$e_{1l} = \text{ReLU}(h_l W_{1,l} + b_{1,l}) \quad (10)$$

Sau đó, nối các vector $e_{1l} \forall l \in \{1, 2, \dots, L\}$ để tạo thành $e_1 = [e_{11}, e_{12}, \dots, e_{1L}]$, trong đó $e_1 \in \mathbb{R}^{1 \times (L_1 \times L)}$ và L là số lớp.

Tương tự, e_1 được đưa vào các lớp kết nối đầy đủ khác (với hàm kích hoạt ReLU) trong $FC2$, $FC3$ và $FC4$, kết quả là $e_4 \in \mathbb{R}^{1 \times L_4}$, trong đó L_4 là số đơn vị trong $FC4$.

Đơn vị hồi quy: Một nút duy nhất ở cuối mạng này xác định mức độ buồn ngủ bằng cách đưa ra một số thực từ 0 đến 10, phụ thuộc vào mức độ tỉnh táo hoặc buồn ngủ của các nháy mắt đầu vào (11). Thang điểm từ 0 đến 10 này giúp mạng mô hình hóa sự chuyển đổi tự nhiên từ sự tỉnh táo đến buồn ngủ, khác với các công trình trước đây, nơi đầu vào được phân loại trực tiếp thành các lớp khác nhau một cách rời rạc.

$$\text{out} = 10 \times \text{Sigmoid}(e_4 W_o + b_o) \quad (11)$$

Ở đây, $W_o \in \mathbb{R}^{L_4 \times 1}$ và $b_o \in \mathbb{R}^{1 \times 1}$ là các tham số hồi quy, và $\text{out} \in \mathbb{R}^{1 \times 1}$ là đầu ra hồi

quy cuối cùng.

Phân cấp các mức độ buồn ngủ: Khi một người buồn ngủ, điều này không có nghĩa là tất cả các nháy mắt của họ sẽ đại diện cho sự buồn ngủ. Do đó, việc phân loại mức độ buồn ngủ của mỗi đoạn video như trạng thái chiếm ưu thế nhất được dự báo từ tất cả các chuỗi nháy mắt trong đoạn video đó là quan trọng. Sử dụng (12) để phân chia đầu ra hồi quy thành từng lớp đã được xác định trước.

$$class(out) = \begin{cases} \text{Alert}, & \text{if } 0 \leq out \leq 5 \\ \text{Drowsy}, & \text{if } 5 < out \leq 10 \end{cases}$$

Giả sử có K nháy mắt trong video V . Sử dụng cửa sổ trượt có độ dài T , T nháy mắt liên tiếp tạo thành một chuỗi nháy mắt được đưa vào mạng (9), dẫn đến có thể có nhiều chuỗi nháy mắt. Lớp được dự báo phổ biến nhất từ những chuỗi này sẽ là kết quả phân loại cuối cùng của video V .

Chương 3

KẾT QUẢ - ĐÁNH GIÁ

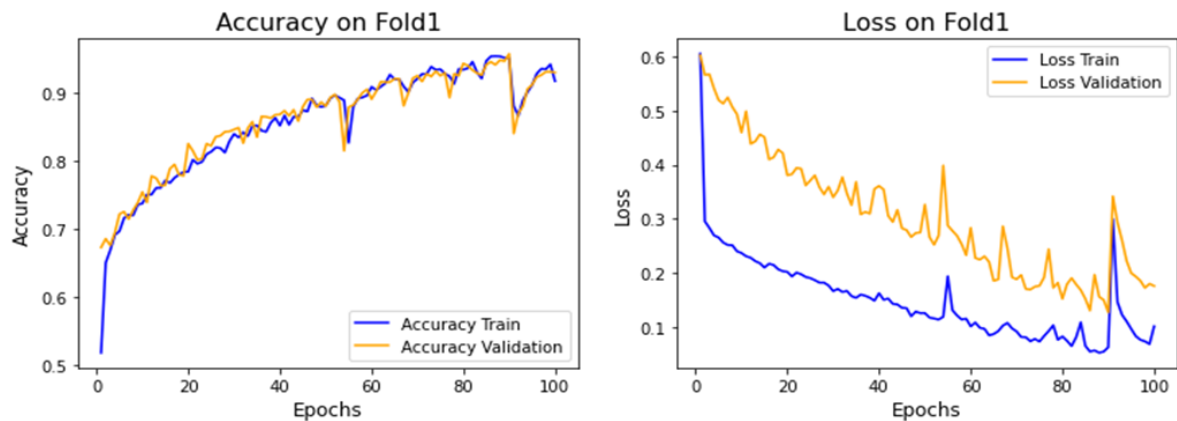
3.1 Triển khai

Hệ thống sử dụng một phần của bộ dữ liệu RLDD làm tập kiểm tra (test set), và bốn phần còn lại để huấn luyện (train set). Sau khi lặp lại quy trình này cho mỗi Fold, kết quả được tính trung bình qua năm Fold. Với tham số T đã được xác định trước đó, số lượng nhảy mắt liên tiếp cần đưa vào mạng là 30. Chuỗi nhảy mắt được tạo ra bằng cách sử dụng một cửa sổ trượt có độ dài 30 nhảy mắt trên mỗi video, với bước nhảy là hai. Đối với các video có ít hơn 30 nhảy mắt, cần thêm các giá trị 0 để đủ 30 nhảy mắt.

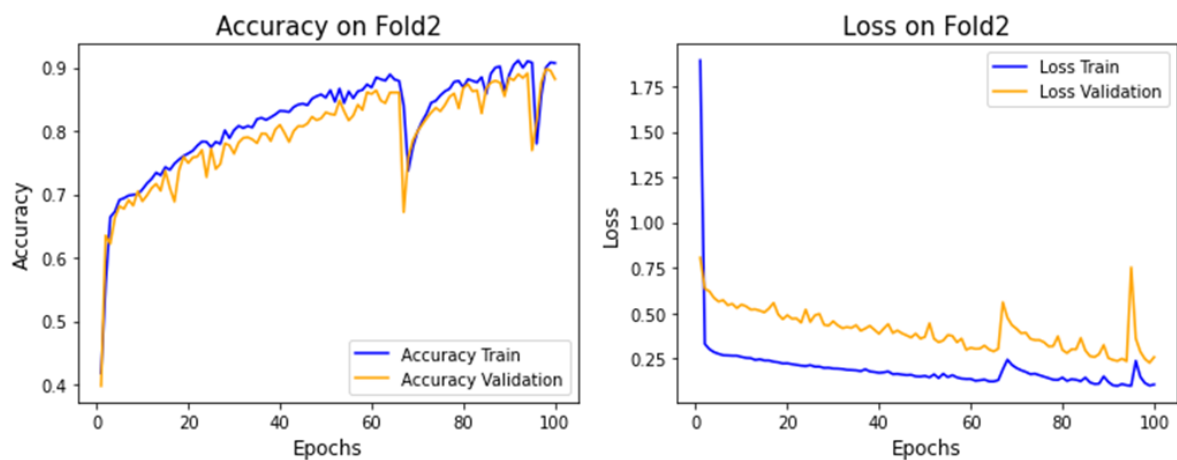
Batch size là 128 được sử dụng trong 100 epoch cho cả năm phần, kết hợp với chuẩn hóa batch (batch normalization) và chuẩn hóa L2 với hệ số (λ) là 0.1. Mô hình LSTM có bốn lớp, mỗi lớp có 32 trạng thái ẩn. Quá trình này được thực hiện trên một máy laptop với 8GB RAM và bộ xử lý 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz.

3.2 Thực nghiệm

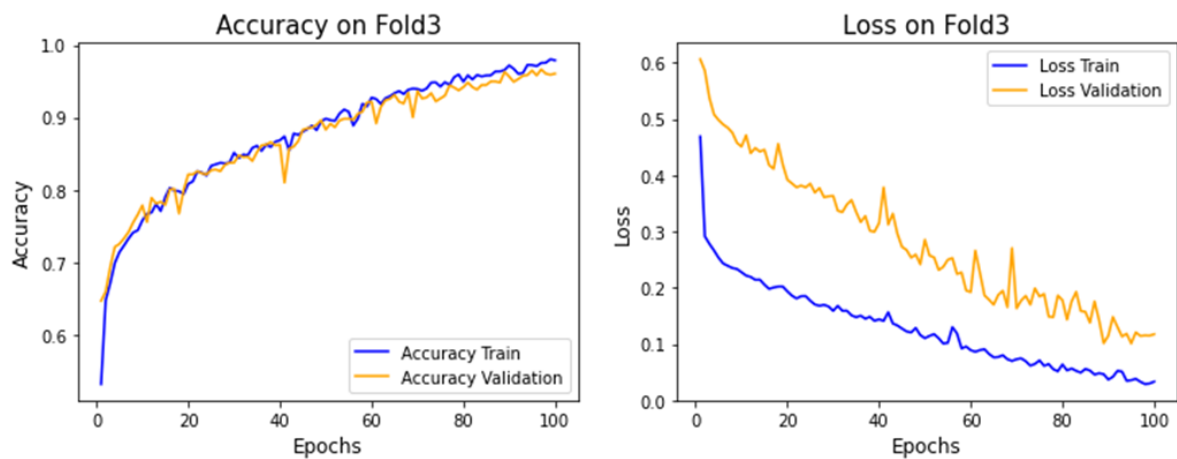
Trong phần này, chúng tôi đánh giá phương pháp cơ sở của chúng tôi với tiêu chí đánh giá: **accuracy** và **loss function**.



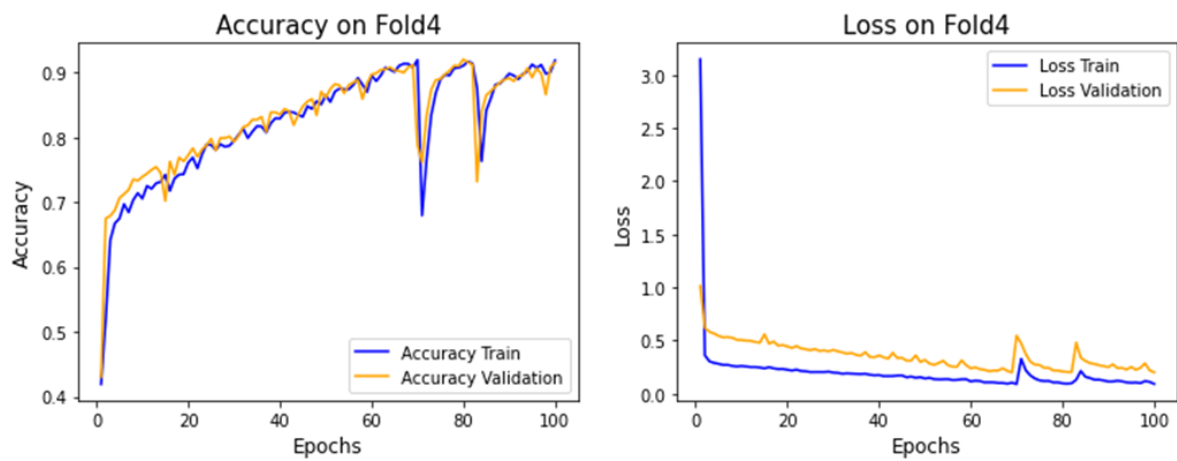
Hình 3.1: Kết quả thực nghiệm trên fold 1.



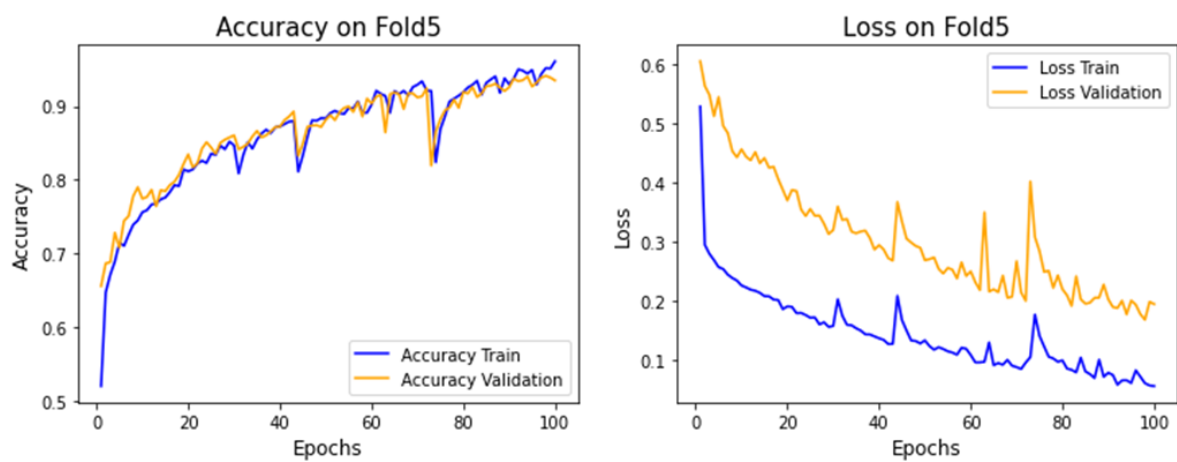
Hình 3.2: Kết quả thực nghiệm trên fold 2.



Hình 3.3: Kết quả thực nghiệm trên fold 3.

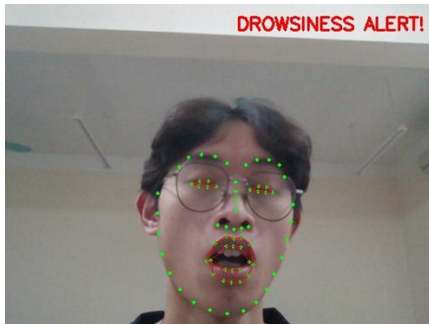


Hình 3.4: Kết quả thực nghiệm trên fold 4.

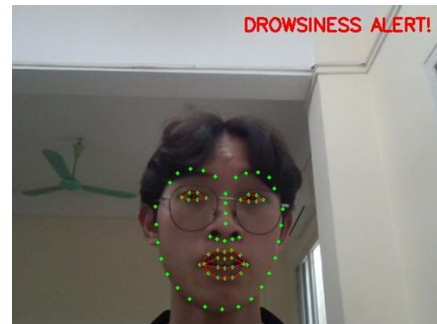


Hình 3.5: Kết quả thực nghiệm trên fold 5.

Mô tả hệ thống dự báo



Hình 3.6: Kết quả dự báo đúng.



Hình 3.7: Kết quả dự báo sai.

3.3 Đánh giá

Thuật toán của Reza Ghoddoosian, Marnim Galib và Vassilis Athitsos đã được cải tiến để có thể trích xuất đặc trưng nháy mắt ở các góc nghiêng, và kết quả thu được rất khả quan. Mô hình triển khai với khoảng 14.000 tham số đã phân loại các đối tượng tỉnh táo và buồn ngủ với độ chính xác trên 90%, và hiếm khi nhầm lẫn giữa sự tỉnh táo và buồn ngủ. Việc lưu trữ các tham số này không chiếm nhiều không gian bộ nhớ, nên mô hình có thể dễ dàng lưu trữ trên cả điện thoại di động cấp thấp. Tuy nhiên, khi triển khai kiểm tra thực tế với khuôn mặt, mô hình vẫn gặp một số lần nhận diện sai.

Chương 4

KẾT LUẬN - HƯỚNG PHÁT TRIỂN

Báo cáo này đã trình bày một phương pháp sử dụng mối quan hệ của các lần nháy mắt theo thời gian và một bộ dữ liệu về tình trạng buồn ngủ trong đời sống thực (RLDD). Phương pháp được đề xuất có yêu cầu tính toán và lưu trữ thấp. Kết quả đã chứng minh rằng phương pháp có độ chính xác cao và có thể ứng dụng trong thực tế.