



Bài báo

Các mô hình học sâu để theo dõi tư thế Yoga

Deborah Swain ¹ , Santosh Satapathy ¹, Biswaranjan Acharya ^{2,*} , Madhu Shukla ³,
4,* Vassilis C. Gerogiannis ⁴ , Andreas Kanavos ^{5*} và Dimitris Giakovis ⁶

- ¹ Khoa Khoa học Máy tính và Kỹ thuật, Đại học Năng lực Pandit Deendayal, Gandhi Nagar 382007, Ấn Độ
 - ² Khoa Kỹ thuật Máy tính-AI, Đại học Marwadi, Rajkot 360003, Ấn Độ
 - ³ Khoa Kỹ thuật Máy tính-AI và Phân tích Dữ liệu Lớn, Đại học Marwadi, Rajkot 360003, Ấn Độ
 - ⁴ Khoa Hệ thống Kỹ thuật số, Đại học Thessaly, 41500 Larissa, Hy Lạp
 - ⁵ Khoa Tin học, Đại học Ionian, 49100 Corfu, Hy Lạp
 - ⁶ Trụ sở Thực nghiệm Larissa, Bộ Giáo dục, 41334 Larissa, Hy Lạp
- * Thư từ: biswaranjanacharya2020@gmail.com (BA); ; vgerogian@uth.gr (VCG); akanavos@ionio.gr (AK)

Tóm tắt: Nhận dạng hoạt động là quá trình theo dõi liên tục hoạt động và chuyển động của một người. Nhận dạng tư thế con người có thể được sử dụng để lắp ráp một khung thực hành tự hướng dẫn cho phép các cá nhân học và thực hành chính xác các tư thế yoga mà không cần sự trợ giúp từ ai khác. Với việc sử dụng các thuật toán học sâu, chúng tôi đề xuất một cách tiếp cận hiệu quả phát hiện và công nhận các tư thế yoga khác nhau. Bộ dữ liệu được chọn bao gồm 85 video với 6 tư thế yoga được thực hiện bởi 15 người tham gia, trong đó các điểm chính của người dùng được trích xuất bằng cách sử dụng Thư viện mediapipe. Sự kết hợp giữa Mạng nơ-ron tích chập (CNN) và Ngăn chặn dài hạn Bộ nhớ (LSTM) đã được sử dụng để nhận dạng tư thế yoga thông qua các video được giám sát theo thời gian thực như một mô hình học sâu. Cụ thể, lớp CNN được sử dụng để trích xuất các tính năng từ các điểm chính và lớp LSTM sau đây hiểu sự xuất hiện của chuỗi khung cho dự đoán để thực hiện. Sau đây, các tư thế được phân loại là đúng hay sai; nếu đúng tư thế được xác định, sau đó hệ thống sẽ cung cấp cho người dùng phản hồi tư vấn ứng thông qua văn bản/lời nói. Bài viết này kết hợp nền tảng học máy với cấu trúc dữ liệu như là sức mạnh tổng hợp giữa hai lĩnh vực này có thể được thiết lập theo nghĩa là các kỹ thuật học máy và đặc biệt là học sâu học có thể nhận ra các lược đồ dữ liệu một cách hiệu quả và làm cho chúng có thể tương tác với nhau.

Từ khóa: tư thế yoga; học máy; học kỹ càng; cấu trúc dữ liệu; asana; CNN; LSTM; ống mMedia; Đặt ra dự đoán



Trích dẫn: Swain, D.; thần kinh, S.; Acharya, B.; Shukla, M.; Gerogiannis, V.C.; Kanavos, A.; Giakovis, D. Sâu Mô hình học tập cho tư thế Yoga. *Giám sát. Thuật toán* 2022, 15, 403. <https://doi.org/10.3390/a15110403>

Biên tập học thuật: Alwin Poullose và Melania Susi

Nhận: ngày 18 tháng 9 năm 2022

Chấp nhận: 28 tháng 10 năm 2022

Đã xuất bản: 31 tháng 10 năm 2022

Ghi chú của nhà xuất bản: MDPI vẫn trung lập liên quan đến các khiếu nại về thẩm quyền trong bản đồ được xuất bản và mối quan hệ thể chế iation.



Bản quyền: © 2022 thuộc về các tác giả.

Người dùng được cấp phép MDPI, Basel, Thụy Sĩ.

Bài viết này là một bài báo truy cập mở phân phối theo các điều khoản và điều kiện của Creative Commons

Giấy phép Ghi công (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

1. Giới thiệu

Nhận biết hoạt động của con người thông qua thị giác máy tính là một lĩnh vực được các nhà nghiên cứu quan tâm hàng đầu trong nhiều năm, do khả năng ứng dụng sâu rộng của nó [1]. Nhân loại nhận dạng hoạt động có khả năng tác động đến các lĩnh vực khác nhau như robot, con người-tư vấn tác máy tính, chơi game, giám sát video, xác minh sinh trắc học, phát hiện hỗn loạn, theo dõi thể thao và theo dõi sức khỏe trong số nhiều thứ khác [2]. Trong khi nhiều cơ hội để thực hiện các hệ thống nhận dạng đã được khám phá trong suốt nhiều năm, ứng dụng trong việc phát hiện tư thế yoga được coi là một ứng dụng tư vấn đối mới và chưa được nghiên cứu. cánh đồng. Cụ thể, nó nắm giữ khả năng to lớn để thúc đẩy sức khỏe cộng đồng và phúc lợi thông qua việc thúc đẩy thực hành Ấn Độ cổ đại này, bên cạnh việc đóng vai trò quan trọng cột mốc quan trọng cho các nhà nghiên cứu trong sự tiến bộ trong lĩnh vực nhận dạng hoạt động này [3].

Nhận dạng hoạt động là vấn đề dự đoán chuyển động của một người, dựa trên dữ liệu cảm biến hoặc chuỗi video. Đây là một vấn đề thị giác máy tính đầy thách thức đòi hỏi một nhiều sự quan tâm và cải tiến [1,4]. Nó có nhiều ứng dụng thực tế như giao diện người dùng thiết kế, học robot, giám sát, v.v. Nó cũng gặp nhiều thách thức do sự khác biệt

về chất lượng video, hình thức một phần, độ sáng, độ gần, tiếng ồn xung quanh, góc và điểm nhìn, v.v. Nhận dạng hoạt động của con người là mối quan tâm của nhận dạng tư thế yoga [3].

Yoga là một hình thức tập thể dục cực kỳ hữu ích trong việc giảm căng thẳng tinh thần và giữ cho một người khỏe mạnh về thể chất và tinh thần. Hơn nữa, nó cải thiện sự tập trung, tập trung, bình tĩnh và lưu thông máu [5]. Yoga có nguồn gốc từ Ấn Độ cổ đại và hiện đã được các nước trên thế giới áp dụng. Yoke đề cập đến yuj trong tiếng Phạn, một thiết bị vật lý được sử dụng để gia súc. Những thiết bị này rất to lớn, thô bạo và rắn chắc và đã từng đeo ách cho ngựa chiến từ lâu. Yoga vừa là phương pháp vừa là công cụ để trấn an ngựa để bạn có thể khiến chúng tập trung và thể hiện tốt trong chiến tranh [6]. Nghiên cứu cho thấy những người tập yoga hàng ngày có tư duy tích cực, cảm giác được cải thiện năng lượng để sống một cuộc sống trọn vẹn và kiểm soát tốt hơn thời giờ quá độ. Cuốn Kinh Yoga cổ điển do Patanjali viết lần đầu tiên mô tả triết lý yoga và các thực hành của nó [7].

Yoga, một bộ môn cổ xưa có nguồn gốc từ Ấn Độ và từng là địa phương của quốc gia, hiện đang trở nên phổ biến trên toàn thế giới nhờ những lợi ích khác nhau về thể chất, tinh thần và tâm linh [5]. Ý nghĩa ngày càng tăng của yoga trong y học có thể là do tác dụng chữa bệnh phi thường của nó trong nhiều tình trạng ảnh hưởng đến cơ thể con người như các vấn đề về hô hấp, bệnh tim, các vấn đề về cơ xương và ứng dụng học sâu trong chăm sóc sức khỏe [6-8]. Tuy nhiên, tồn tại một khoảng cách nhất định giữa thể hệ mới với sự hiểu biết và nhận thức của họ về lợi ích của yoga, dẫn đến một loạt các vấn đề sức khỏe liên quan đến lối sống vội vã ngày nay có thể dễ dàng hạn chế bằng cách áp dụng yoga như một phần của thói quen hàng ngày. Một trong những yếu tố chính dẫn đến những quan niệm sai lầm xung quanh yoga, cuối cùng góp phần khiến mọi người không muốn kết hợp nó vào cuộc sống của họ, là không có hướng dẫn phù hợp. Tuy nhiên, sự đổi mới ngày càng tăng trong công nghệ mang đến khả năng giải quyết vấn đề không thể tiếp cận với việc dạy kèm phù hợp thông qua một công cụ hỗ trợ tự học theo thời gian thực, có khả năng phát hiện các tư thế yoga khác nhau bằng các công cụ nhận dạng hoạt động, cho phép nó phục vụ như một phương tiện hướng dẫn thuận tiện, cần thiết để giúp phổ biến hình thức theo cách mong muốn.

Học máy có thể được coi là nghiên cứu về các thuật toán cho phép các chương trình máy tính tự động cải thiện thông qua kinh nghiệm. Cụ thể, các thuật toán học máy xây dựng một mô hình dựa trên dữ liệu đào tạo để đưa ra dự đoán hoặc quyết định mà không được lập trình rõ ràng để làm như vậy. Trong những năm gần đây, một lĩnh vực phụ của máy học gần đây đã phát triển đáng chú ý, được gọi là học sâu. Truồng này đề cập đến việc sử dụng các kiến trúc khác nhau dựa trên mạng thần kinh nhân tạo, thông qua hệ thống phân cấp các lớp với các đơn vị xử lý phi tuyến tính, tìm hiểu các khái niệm trừu tượng cấp cao cho dữ liệu. Học máy cũng như các kỹ thuật học sâu đã được sử dụng rộng rãi để khám phá mối quan hệ nhân quả và mối tương quan giữa dữ liệu lớn.

Một trong những đối tượng nghiên cứu chính trong các lĩnh vực khoa học về học sâu và thị giác máy tính là kỹ năng của con người để xác định và diễn giải hành động của người khác. Nhận dạng hoạt động đề cập đến chủ đề nghiên cứu rộng lớn tập trung vào việc phát hiện các hoạt động và mục tiêu của một hoặc nhiều cá nhân từ một chuỗi các quan sát về hành động của cá nhân và môi trường xung quanh. Nhận dạng hoạt động của con người là tập hợp con liên quan đến mô hình cơ thể con người để nhận dạng chuyển động và hành vi độc đáo của một người thông qua dữ liệu nhận được từ cảm biến hoặc kỹ thuật dựa trên tầm nhìn. Các chuyển động được xác định có thể bao gồm các hoạt động phổ biến như ngồi, đứng, trò chuyện hoặc đi bộ hoặc có thể bao gồm các cử chỉ rất cụ thể và phức tạp như các tư thế yoga sống động hoặc các tư thế thể thao cụ thể [9]. Việc phân loại chính xác các tư thế của con người vẫn là một thách thức đối với cộng đồng khoa học, kích thích các công việc tiếp theo trong lĩnh vực này.

Thông qua các phương pháp mô tả đặc tính khác nhau phổ biến, cách tiếp cận hai mặt đối với vấn đề phân loại vẫn thống nhất: thứ nhất, phần nhận dạng và thứ hai, vấn đề bản địa hóa, cho biết mô tả hành động được phát hiện và nhận dạng phần thực. -thời gian chuyển động có chứa đối tượng quan tâm tương ứng [1,10]. Toàn bộ quá trình thường bao gồm ba giai đoạn về mặt biểu diễn: công nghệ lõi cấp thấp, phát hiện hành vi cấp trung và triển khai cấp cao [2]. Trong bước đầu tiên của giai đoạn đầu tiên, phân đoạn đối tượng là

được tiến hành trên từng khung hình của video để phân biệt và cô lập đối tượng mục tiêu khỏi môi trường của nó. Cấp độ thứ hai liên quan đến việc truy xuất các thuộc tính của các đối tượng riêng biệt như màu sắc, hình dạng, đường viền và chuyển động và biểu diễn tiếp theo của chúng thành các tính năng, có thể được phân loại thành bốn loại chính: biến đổi tần số, định danh cục bộ, thống kê không gian-thời gian, và dữ liệu mô hình cơ thể. Các kỹ thuật phân loại và phát hiện sự kiện được sử dụng để phát hiện các tư thế và chuyển động khác nhau của con người dựa trên các đặc điểm thu được từ bước trước đó bao gồm giai đoạn thứ ba của thao tác.

Nó liên quan đến những khó khăn gặp phải ở mỗi cấp độ phụ này sau vô số

các yếu tố như ánh sáng kém, thay đổi chất lượng video, khả năng hiển thị một phần, khoảng cách gần, nhiều nền, góc liên quan, v.v., khiến cho việc nhận dạng hoạt động của con người trở thành một nhiệm vụ khá khó khăn [11].

Các nghiên cứu chuyên sâu tiếp tục được tiến hành trong lĩnh vực này, với mục đích cải thiện các kết quả hiện có đã đạt được. Gần đây, các phương pháp dựa trên học sâu đã trở nên phổ biến do sử dụng các thuật toán học biểu diễn có thể trích xuất các tính năng phù hợp theo bản năng từ dữ liệu đầu vào do các cảm biến cung cấp mà không cần sự can thiệp của con người và các mẫu cơ bản duy nhất [12,13]. Trong bài báo cụ thể này, chúng tôi đề xuất một mô hình học sâu để nhận biết và điều chỉnh các tư thế yoga trong thời gian thực, với mục đích giải quyết tình trạng thiếu đào tạo phù hợp và làm cho việc giáo dục đó trở nên hợp lý và phù hợp hơn với mọi người. Điều này được thực hiện bằng cách cố gắng khai thác cơ hội tự luyện tập tại nhà thông qua một ứng dụng có thể phân biệt chính xác giữa các tư thế được huấn luyện viên chuyên nghiệp thực hành với tư thế của người mới bắt đầu, dẫn đến các khuyến nghị đáng tin cậy cho người học về những điều chỉnh cần thiết. Với việc sử dụng các mô hình CNN và LSTM, một hệ thống giám sát tư thế yoga thông minh đã được tạo ra để lấy đầu vào là video của các đối tượng mẫu, sau đó chuyển đổi chúng thành một chuỗi các điểm chính, là tọa độ của các điểm đối tượng chính trên một người, sau đó thực hiện nhiệm vụ phát hiện và cuối cùng, cung cấp phản hồi cần thiết cho người dùng dựa trên sự tương đồng ở dạng văn bản/lời nói [9]. Các phần sắp tới sẽ làm sáng tỏ các phương pháp chính xác được sử dụng và kết quả đạt được, trước khi làm nổi bật phạm vi cải tiến hơn nữa.

Trong nghiên cứu này, một hệ thống có thể nhận dạng và sửa các tư thế yoga trong thời gian thực bằng cách sử dụng các kỹ thuật học sâu khác nhau được đề xuất. Bằng cách sử dụng một hệ thống như vậy, toàn bộ bầu không khí của lớp học yoga có thể được tạo ra tại nhà của người dùng nơi hệ thống phát hiện và tự động sửa tư thế yoga. Tìm một lớp học yoga có giá cả phải chăng và phù hợp với lịch trình của người dùng là điều khó khăn. Bài báo này cố gắng lấp đầy khoảng trống đó của các lớp yoga theo yêu cầu bằng phản hồi về thói quen tập yoga trong thời gian thực. Ý tưởng chính là tạo ra một mô hình học sâu có thể phân loại chính xác tư thế yoga của người dùng bằng cách huấn luyện tư thế đó trên bộ dữ liệu video yoga.

Phần còn lại của bài báo được tổ chức như sau. Tài liệu cơ sở được giới thiệu trong Phần 2. Phần 3 tổng quan về các khái niệm và phương pháp cơ bản được sử dụng trong bài báo này cùng với bộ dữ liệu được sử dụng. Phần 4 trình bày các kết quả nghiên cứu, và cuối cùng, Phần 5 trình bày các kết luận và định hướng cho công việc trong tương lai.

2. Công việc liên quan

Các phương pháp và kỹ thuật khác nhau cho nhiều ứng dụng đã được đề xuất và sử dụng trong lĩnh vực giám sát con người theo thời gian thực. Nghiên cứu được đề xuất nhằm mục đích xem xét các tài liệu hiện có và đóng góp hơn nữa vào việc nâng cao các kết quả hiện có đạt được trong lĩnh vực thị giác máy tính để ước tính tư thế con người [14].

Hơn nữa, hiệu quả của mạng thần kinh LSTM và đóng góp quan trọng của nó cho các nhiệm vụ cụ thể đã được chứng minh [15,16]. Một nỗ lực đã được phát triển nhằm tạo ra một hệ thống dựa trên học sâu để phát hiện hiệu quả các tư thế yoga và thực tế hoạt động như một sự thay thế cho huấn luyện viên bằng cách đưa ra phản hồi chính xác cho người dùng.

Một nghiên cứu so sánh giữa học máy và kỹ thuật học sâu để ước tính tư thế yoga đã được thực hiện trong [3] liên quan đến cả hai công nghệ và phân tích hiệu suất của khung bằng cách sử dụng Máy vectơ hỗ trợ (SVM), Mạng thần kinh chuyển đổi (CNN) và Mạng thần kinh chuyển đổi Mạng cùng với dài ngắn hạn

Bộ nhớ (LSTM). Nghiên cứu kết luận rằng hiệu suất của mô hình lai CNN-LSTM dẫn đến số lượng phân loại sai ít nhất. Một hệ thống mà bộ xử lý của người dùng ban đầu được phát hiện bằng cách sử dụng thuật toán gọi là ước lượng tf-pose, đã được giới thiệu trong [17]. Sau đây, sáu mô hình học máy khác nhau, cụ thể là Cây quyết định, Rừng ngẫu nhiên, Hồi quy logistic, Naive Bayes, SVM và KNN đã được thử nghiệm và so sánh. Bộ dữ liệu đã sử dụng chứa 10 tư thế yoga khác nhau với khoảng 5500 hình ảnh và hiệu suất tối ưu được phân loại Random Forest thu được với độ chính xác 99,04%.

Một mô hình có tên ExNet, là một Mạng thần kinh chuyển đổi đa lớp, đã được đề xuất trong [18]. Các tác giả đã sử dụng một bộ dữ liệu hình ảnh chứa 2000 hình ảnh về các tư thế tập thể dục của con người được chia thành năm loại được dán nhãn là chống đẩy, kéo lên, đạp xe, uốn cong gân kheo và đi bộ. Mô hình đã sử dụng trình tối ưu hóa Adam và phương pháp giảm Tỷ lệ học tập tự động. Sau 50 kỷ nguyên, ExNET có độ chính xác 82,68% khi phân loại tư thế tập thể dục của con người 2D từ bộ dữ liệu. Tuy nhiên, mô hình cần điều chỉnh siêu tham số tốt hơn vì nó phải đối mặt với vấn đề trang bị quá mức. Các tác giả trong [19] đã quản lý để thiết kế chính xác mô hình học sâu thời gian thực bằng OpenPose để trích xuất các điểm chính từ video được cung cấp bằng cách sử dụng bộ dữ liệu bao gồm sáu asana, tức là Trikonasana, Padmasana, Vrikshasana, Bhujangasana, Tadasana và Shavasana. Mặc dù hệ thống có khả năng đạt độ chính xác 99,04% với đầu vào được lấy theo khung, nhưng việc phát hiện trên cơ sở khảo sát 45 cạnh đặt trước mang lại thước đo hiệu suất là 99,38%. Công việc đã được thử nghiệm thêm trên mười hai cá nhân và được chứng minh là có hiệu quả cao với độ chính xác 1

Các tác giả trong [20] đã ghi lại thông tin video bằng Microsoft Kinect. Kinect có một công cụ theo dõi bộ xử lý có thể thừa nhận 20 khớp trên cơ thể một người. Mười điểm khớp đã chọn được sử dụng để tính toán và cấu trúc tham chiếu cho từng tư thế yoga được thiết kế bằng cách thu thập thông tin về các điểm khớp từ tư thế của con người. Sau đây, độ đồng dạng cosin của các vectơ được tính bằng cách tìm các góc giữa tất cả các vectơ nối hai điểm chung bất kỳ. Cụ thể, nếu độ lệch được tính toán lớn hơn giá trị ngưỡng đã đặt, thì tư thế được phân loại là không chính xác. Một hệ thống có khả năng phát hiện các tư thế liên quan đến bốn loại hoạt động bình thường của con người và xác định hành động của con người ở cả hai chiều thứ hai và thứ ba, đã được trình bày trong [21]. Nghiên cứu được thực hiện trên dữ liệu lấy từ bốn nguồn khác nhau, tức là Human 3.6M, NTU, Penn Action và MPII bằng cách sử dụng Mạng thần kinh chuyển đổi đa nhiệm [22].

Trong một phương pháp được sử dụng khác, khi người dùng tạo tư thế trước máy ảnh của mình, thư viện Mediapipe sẽ thực hiện phân tích hình học dựa trên các khung thu được từ máy ảnh [23]. Phân tích hình học tạo ra đầu ra dựa trên các góc giữa các khớp khác nhau. Các góc được tính toán sau đó được so sánh với các góc chính xác được lưu trong cơ sở dữ liệu cho tư thế yoga cụ thể. Nếu sự khác biệt giữa hai góc này lớn hơn ngưỡng đã đặt, thì phản hồi tương ứng sẽ được cung cấp cho người dùng, ở dạng văn bản hoặc thông qua thiết bị âm thanh. Một trong những công trình gần đây nhất thể hiện hệ thống dựa trên LSTM-CNN để phân loại [24]. Cụ thể, nhiệm vụ phân loại đã được cải thiện khi phương pháp được đề xuất giảm thời gian thực hiện theo các giá trị nằm trong khoảng từ 30% đến 42%.

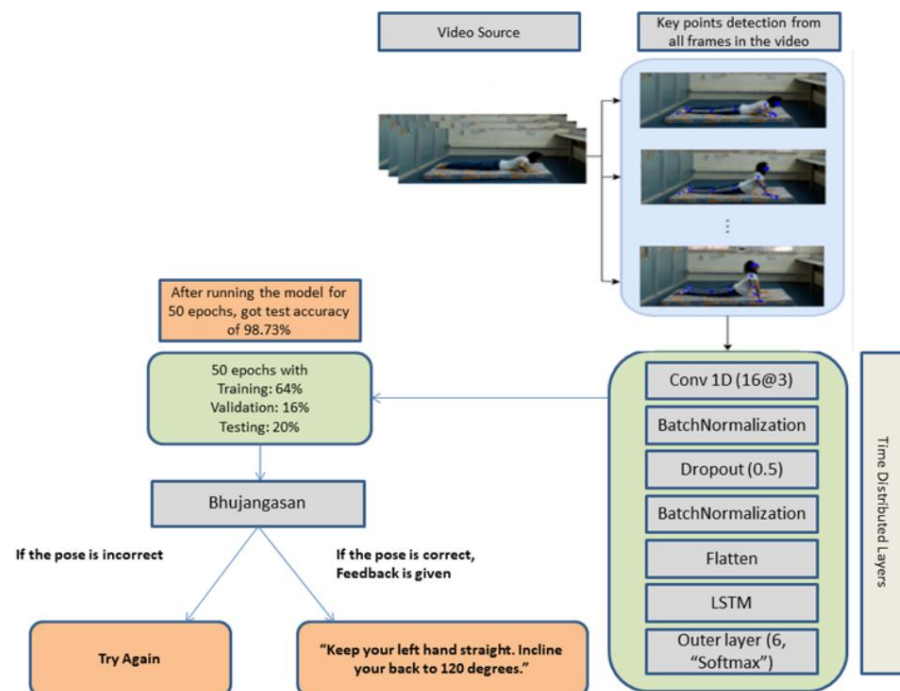
Một huấn luyện viên thể dục thông minh về các khái niệm phát hiện và ước tính tư thế bằng các phương pháp học sâu, đã được sử dụng trong [25], trong đó khung ước tính tư thế nhiều người trong khu vực đã được sử dụng. Ngoài ra, AlphaPose đã được sử dụng để xác định các điểm chính, là các khớp của con người trong mọi khung hình. Con số thu được khi kết nối chúng sau đó được so sánh với các tư thế lý tưởng, dẫn đến gợi ý cho người dùng về cách thực hành các bài tập một cách chính xác. Một mô hình khác để điều chỉnh tư thế trong khi tập thể dục đã được đề xuất trong [26], trong đó việc đào tạo mô hình được thực hiện bằng cách sử dụng OpenPose và 18 điểm chính được xác định để phát hiện chính xác sự khác biệt trong các tư thế. Bộ dữ liệu được sử dụng chứa hơn 100 video về các bài tập được thực hiện đúng và sai. Nhiệm vụ được thực hiện trên 4 bài tập khác nhau, cụ thể là nâng cao phía trước, nhấn vai đứng, uốn cong bắp tay và nhún vai.

Cuối cùng, một kỹ thuật học sâu khác để huấn luyện người mẫu nhận biết các tư thế yoga, đã được áp dụng trong [27]. Sau khi trích xuất các tính năng thông qua ước tính tư thế đa mục đích của Keras,

quá trình phân loại được sử dụng thành một trong sáu tư thế: tư thế xác chết, tư thế cái cây, tư thế ngọn núi, tư thế tam giác, tư thế hoa sen và tư thế rắn hổ mang. Quá trình này được thực hiện theo các góc giữa mười hai điểm chính bằng cách sử dụng Multilayer Perceptron, dẫn đến độ chính xác là 99,58%.

3. Hệ thống được đề

Hệ thống được đề xuất lấy các khung chuỗi video làm đầu vào trong thời gian thực. Đầu ra sẽ là tư thế yoga được dự đoán cùng với phản hồi tương ứng về cả góc độ và hiệu chỉnh tư thế. Hệ thống bao gồm ba giai đoạn chính, cụ thể là trích xuất điểm chính, dự đoán tư thế cũng như hiệu chỉnh tư thế. Giai đoạn trích xuất điểm chính nhằm mục đích phát hiện và trích xuất vị trí của các điểm chính quan trọng dựa trên vị trí của người dùng [14]. Giai đoạn dự đoán tư thế xác định kiến trúc mô hình và phân loại xem tư thế đó có đúng hay không. Giai đoạn cuối cùng là chỉnh sửa tư thế trong đó người dùng được cung cấp thêm phản hồi để chỉnh sửa tư thế và cũng mô tả tỷ lệ phần trăm tương tự so với tư thế thực tế. Hình 1 minh họa kiến trúc hệ thống được đề xuất cùng với ba giai đoạn trên.

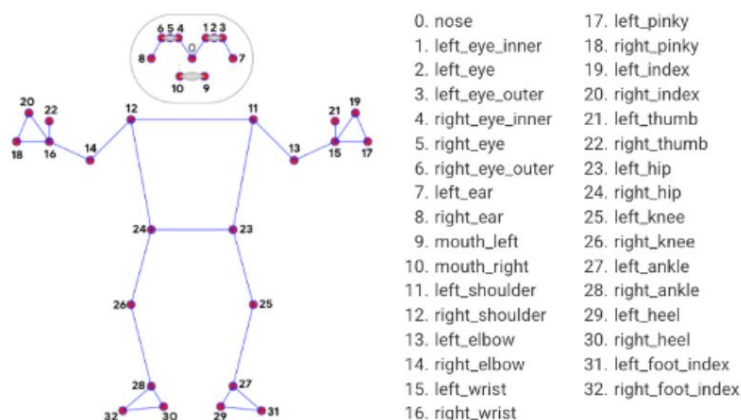


Hình 1. Lưu đồ của Hệ thống Đề xuất.

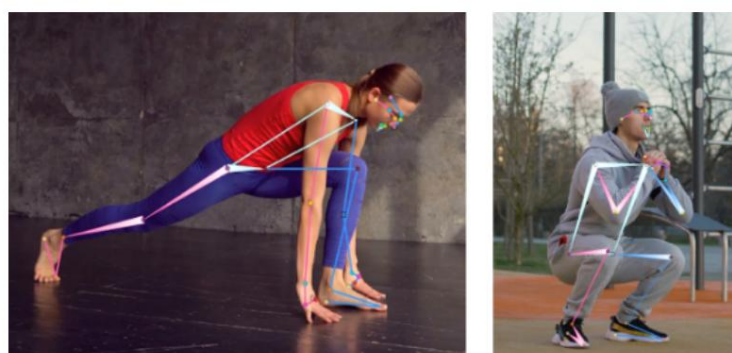
3.1. Khai thác điểm chính

Giai đoạn đầu tiên bao gồm việc trích xuất các điểm chính từ tất cả các khung hình của video và sau đó lưu trữ nó ở định dạng JSON. Các điểm chính bao gồm các điểm khác nhau ở một người rất quan trọng trong việc hình thành một tư thế yoga; các ví dụ bao gồm vai, khuỷu tay, cổ tay, đầu gối, v.v. Chúng tôi đã sử dụng thư viện MediaPipe để trích xuất điểm chính, đây là thư viện đa nền tảng do Google phát triển cung cấp các giải pháp ML sẵn sàng sử dụng tuyệt vời cho các tác vụ thị giác máy tính.

Giai đoạn này sử dụng mô hình CNN được đào tạo trước được tối ưu hóa cao được sử dụng để theo dõi tư thế cơ thể với độ chính xác cao, suy ra 33 mốc 3D và mặt nạ phân đoạn nền trên toàn bộ cơ thể từ các khung video RGB. Thư viện MediaPipe tạo ra 3 tọa độ (X, Y và Z), trong đó Z chỉ độ sâu của tọa độ 2D [10]. Hình 2 trình bày 33 điểm chính do thư viện MediaPipe cung cấp, trong khi Hình 3 minh họa kết quả sau khi trích xuất các điểm chính bằng cách sử dụng thư viện MediaPipe.



Hình 2. Điểm chính MediaPipe.



Hình 3. Các điểm chính của MediaPipe trên con người thực tế.

Sau khi chuyển đổi video sang định dạng JSON, quá trình phân chia thành tập dữ liệu đào tạo, xác thực và thử nghiệm đang diễn ra. Tỷ lệ phân chia được sử dụng là 64:16:20 và mỗi tập dữ liệu thử nghiệm bao gồm một chuỗi 45 khung với kích thước chồng lấp là 36 khung, chứa tọa độ của tất cả 33 điểm chính. Do đó, hình dạng đầu vào của một tập dữ liệu thử nghiệm đơn lẻ có thể được đưa ra là (45, 33, 2). Tổng số mẫu đào tạo, xác nhận và thử nghiệm là 7063, 1832 và 2202.

Giai đoạn này sử dụng các kỹ thuật máy học để thiết kế cấu trúc dữ liệu một cách hiệu quả trong ngữ cảnh của ứng dụng cụ thể. Sự đóng góp nằm ở sự phát triển hệ thống thành công cũng như việc sử dụng các hoạt động đánh giá.

3.2. Dự đoán tư thế

Giai đoạn thứ hai liên quan đến việc tạo ra một mô hình học sâu nhằm phân loại chính xác bất kỳ video thời gian thực nào theo một trong sáu tư thế được đưa ra trong bộ dữ liệu. Ở đây, một mô hình lai được kết hợp, là sự kết hợp giữa CNN và LSTM, trong đó CNN được sử dụng để trích xuất đặc trưng [13]. Cụ thể, CNN là Mạng nơ-ron nhân tạo nhiều lớp (ANN) được tạo riêng để hoạt động trên hình ảnh và được sử dụng cho các tác vụ như nhận dạng đối tượng và phân loại hình ảnh [28]. Mặt khác, LSTM rất hữu ích để hiểu chuỗi các khung xảy ra trong một tư thế yoga cụ thể. Cụ thể, LSTM là một loại RNN được trang bị để học và ghi nhớ các phụ thuộc cực kỳ dài hạn đối với chuỗi dài dữ liệu đầu vào [29].

Trong công việc này, một lớp Phân phối Thời gian cùng với CNN, đặc biệt hữu ích khi làm việc với các khung hình video hoặc dữ liệu chuỗi thời gian [30], được sử dụng. Ngoài ra, một lớp Softmax, sử dụng hàm lũy thừa chuẩn hóa, được sử dụng để tìm xác suất của từng tư thế yoga. Tư thế có xác suất cao nhất được dự đoán là đầu ra.

Cụ thể hơn, Lớp đầu tiên được sử dụng trong dự đoán tư thế là lớp CNN với 16 bộ lọc và kích thước cửa sổ bằng 3. Hàm kích hoạt được sử dụng trong lớp này là ReLU, đây là một hàm tuyến tính theo từng phần cung cấp đầu ra bằng 0 khi đầu vào

miễn là nhỏ hơn 0, nếu không, nó cho đầu ra là đầu vào đã cho [31,32]. Phương trình (1) trình bày chức năng kích hoạt của ReLU:

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

trong đó $x \in \mathbb{R}$.

Lớp thứ hai được sử dụng là lớp Chuẩn hóa hàng loạt giải quyết vấn đề dịch chuyển đồng biến bên trong và giúp luồng dữ liệu dễ dàng hơn qua các lớp khác nhau [33].

Lớp tiếp theo được sử dụng là lớp Dropout, đây là một kỹ thuật chính quy hóa để ngăn chặn việc trang bị quá mức thông tin ở tỷ lệ 0,5 [34]. Sau đó, một lớp Chuẩn hóa hàng loạt khác được sử dụng. Đầu ra thu được từ lớp này sau đó được chuyển đến lớp Làm phẳng để chuyển đổi dữ liệu thành mảng một chiều. Lớp tiếp theo được sử dụng là lớp LSTM có kích thước là 20 đơn vị với độ lệch quên được đặt thành True để trả về đầu ra của mọi nút. Đầu ra của LSTM thứ 0 được tạo thông qua một loạt các hoạt động có Cổng, như Cổng Quên, Cổng Đầu vào và Cổng Đầu ra. Các phương trình toán học cho một LSTM được trình bày trong các Phương trình (2)-(7) sau đây.

Thông tin trạng thái của ô hiện tại và ô trước đó được lưu trữ trong C. Đầu ra của hàm sigmoid trong cổng quên cho biết thông tin nào cần giữ lại hoặc quên. Đầu ra chủ yếu phụ thuộc vào vectơ đầu ra trạng thái trước đó h_{t-1} và vectơ đầu vào trạng thái hiện tại x_t . Nếu có sự khác biệt thì hàm σ không cho phép giữ lại thông tin. Ngoài ra, xu hướng quên làm thay đổi đầu ra theo hướng 1 (giữ lại) hoặc 0 (quên); mô hình hoạt động tốt hơn với giá trị sai lệch quên bằng 0,5. Trong cổng đầu vào, hàm σ quyết định giá trị nào cần cập nhật và hàm tanh thêm giá trị mới cho trạng thái. Đối với lớp đầu ra, hàm σ xác định giá trị được chọn làm đầu ra h_t .

$$l_r = \sigma(W_l \cdot [h_{t-1}, x_t] + b_l) \quad (2)$$

$$m_r = \sigma(W_m \cdot [h_{t-1}, x_t] + b_m) \quad (3)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$C_r = l_r \cdot C_{t-1} + i_r \cdot C_t \quad (5)$$

$$\text{hoặc } = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$g_i = \text{hoặc } \tanh(C_r) \quad (7)$$

Ở đây

- l_r , m_r hoặc h_t lần lượt là đầu ra của cổng quên, cổng đầu vào, cổng đầu ra và trạng thái hiện tại;
- W_l , W_m và W_o lần lượt là các trọng số của cổng quên, cổng đầu vào và cổng đầu ra;
- x_t là đầu vào trạng thái
- hiện tại; b_l , b_m và b_o lần lượt là độ lệch của cổng quên, cổng đầu vào và cổng đầu ra; m_r là đầu ra của cổng đầu vào;
- C_t và C_r là trạng thái ô hiện tại và trước đó.

Lớp cuối cùng được sử dụng là lớp Dense sử dụng Softmax làm chức năng kích hoạt, gán xác suất của các từ thể khác nhau dựa trên đầu vào đã cho hiện tại [35]. Phương trình toán học của hàm kích hoạt Softmax được trình bày trong Công thức (8).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (8)$$

trong đó σ là softmax, z là vectơ đầu vào, e^{z_i} là hàm mũ tiêu chuẩn cho vectơ đầu vào, K là số lớp trong trình phân loại nhiều lớp và e^{z_j} là hàm mũ tiêu chuẩn cho vectơ đầu ra.

Đầu ra thu được sau lớp này được thăm dò trên 45 khung hình để có được dự đoán cuối cùng. Trình tối ưu hóa Adam [36] được sử dụng với tốc độ học bằng 0,0001; trình tối ưu hóa này

giúp mô hình hội tụ nhanh bằng cách bổ sung số hạng động lượng và số hạng tỉ lệ như đã chứng minh trong phương trình (9). Nó kết hợp ý tưởng về động lượng và trình tối ưu hóa RMSprop, đồng thời giúp tránh sự suy giảm theo cấp số nhân của vấn đề tốc độ học tập.

$$\theta_{\text{mới}} = \theta_{\text{cũ}} - \eta \cdot m^{\wedge} \cdot \frac{\phi}{\sqrt{s^{\wedge}}} + \quad (9)$$

trong đó $\theta_{\text{mới}}$ và θ_{old} lần lượt là giá trị trọng số mới và cũ, η là tốc độ học, m^{\wedge} là số hạng động lượng, s^{\wedge} là số hạng tỷ lệ, là số hạng làm mịn để tránh sai số chia bằng 0 và ϕ là phép chia thành phần.

Hàm mất mát được sử dụng là entropy chéo phân loại, rất phổ biến cho các nhiệm vụ phân loại nhiều lớp [37]. Phương trình (10) mô tả phương trình toán học được sử dụng trong hàm mất entropy chéo phân loại.

$$ETC = - \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N \text{piclog}(y_{ic}) \quad (10)$$

trong đó ECC là entropy chéo phân loại, N là số cặp có sẵn trong tập huấn luyện, C là số loại, pic là hàm chỉ báo nhị phân phát hiện xem loại thứ i và yic có phải là phân phối xác suất dự đoán hay không. ~~đó là độ chính xác của dự đoán thuộc lớp c.~~

Số liệu được sử dụng để đánh giá hiệu suất của mô hình là độ chính xác [38], trong đó mô hình cụ thể được đào tạo trong tổng số 50 kỷ nguyên. Ban đầu, sự tăng trưởng theo cấp số nhân và trở nên ổn định sau một vài kỷ nguyên. Cụ thể, sau mỗi kỷ nguyên, chúng tôi kiểm tra xem độ chính xác có được cải thiện và trở nên tốt hơn so với độ chính xác tốt nhất đạt được hay không. Nếu nó tốt hơn độ chính xác tốt nhất, thì độ chính xác tốt nhất được thay thế bằng độ chính xác hiện tại. Tất cả các tham số được sử dụng trong mô hình đã được điều chỉnh hoàn hảo bằng cách sử dụng điều chỉnh siêu tham số để thu được kết quả tối ưu nhất [39,40].

3.3. chỉnh sửa tư thế

Sau khi tư thế dự đoán được phân loại là chính xác, đối với tư thế đã chọn, người dùng được cung cấp phản hồi thích hợp và sau đó, tỷ lệ phần trăm tư thế tự (sử dụng độ tự tư cosine) được tính để hiển thị cho người dùng.

Đối với tất cả sáu tư thế yoga có trong bộ dữ liệu sẽ được giới thiệu bên dưới, các góc quan trọng và quan trọng đã được xác định và các quy tắc đã được xây dựng cho từng tư thế. Đối với mỗi quy tắc, một người dùng được đặt, tạo thành độ lệch tối đa được phép cho người dùng so với tư thế chuẩn. Nếu người dùng vượt quá giá trị người dùng này, một phản hồi tư thế ứng sẽ được đưa ra tư thế ứng dưới dạng văn bản và lời nói. Có thể tìm ra góc giữa hai điểm chính bằng cách tính nghịch đảo tiếp tuyến của độ dốc với trục X dự định. Phương trình (11) mô tả công thức tìm góc, cho hai tọa độ của các điểm chính:

$$\theta = \tan^{-1} \left(\frac{y_2 - y_1}{x_2 - x_1} \right) \quad (11)$$

trong đó (x_1, y_1) và (x_2, y_2) là tọa độ của hai điểm chính.

Phản hồi ban đầu thu được ở dạng văn bản, sau đó được chuyển đổi thành lời nói sử dụng thư viện Pyttsx3 [41]; nó là một công cụ chuyển đổi văn bản thành giọng nói và cũng hoạt động ngoại tuyến.

Độ tư thế tự cosin, là phép đo so sánh hai vectơ bằng cách tính cosin của các góc giữa chúng [42], cũng được báo cáo cho người dùng. Giá trị của số liệu này thay đổi từ -1 đến +1. Nếu điểm tư thế tự nằm trong khoảng từ -1 đến 0 thì nó được nhân với -1 để biến nó thành mặt tích cực. Sau đó, độ tư thế tự được tính theo điểm trong khoảng từ 0 đến +1. Công thức toán học của độ tư thế tự cosin được đưa ra trong phương trình (12):

$$\cos \theta = \frac{A \cdot B}{|A| \cdot |B|} \quad (12)$$

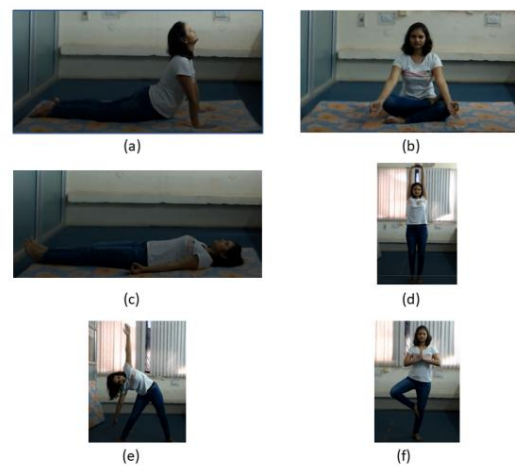
trong đó, A và B là hai vectơ trong không gian nhiều chiều.

Trong công việc này, độ tương tự cosine được tính toán giữa các điểm chính của tư thế người dùng và tư thế chuẩn. Bằng cách này, biện pháp này cho thấy mức độ gần gũi với tư thế thực tế.

Vì khoảng cách của những người dùng khác nhau có thể thay đổi dựa trên vị trí của họ so với máy ảnh, nên tất cả các điểm chính ban đầu được chuẩn hóa để xác định tỷ lệ tương tự cho tất cả chúng.

3.4. tập dữ liệu

Bộ dữ liệu bao gồm tổng cộng 85 video với chất lượng HD 1080p, được ghi lại bởi 15 hãng khác nhau. cá nhân. Nó chứa các video về 6 tư thế yoga khác nhau, cụ thể là Bhujangasana, Padmasana, Shavasana, Tadasana, Trikonasana và Vrikshasana như minh họa trong Hình 4. Các video được ghi bằng GPU NVIDIA TITAN X và bộ xử lý Intel Xeon với RAM 32 GB với tốc độ khung hình bằng 30 khung hình mỗi giây. Người dùng đã cố gắng thực hiện khác nhau asana với nhiều biến thể. Các đặc điểm của tập dữ liệu tương ứng được trình bày trong Bảng 1.



Hình 4. Các tư thế Yoga: (a) Bhujangasana, (b) Padmasana, (c) Shavasana, (d) Tadasana, (e) Trikonasana, (f) Vrikshasana.

Bảng 1. Tập dữ liệu.

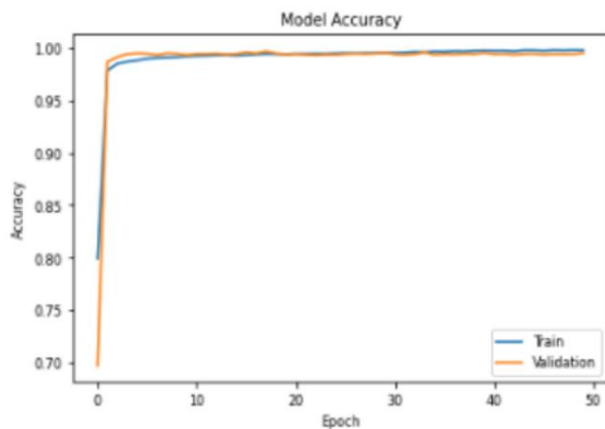
KHÔNG.	Tên tư thế	#người	#Video
1	Bhujangasana	15	16
2	Liên Hoa Sinh	14	14
3	Shavasana	15	15
4	tadasana	15	15
5	Trikonasana	13	13
6	Vrikshasana	12	12

4. Kết quả

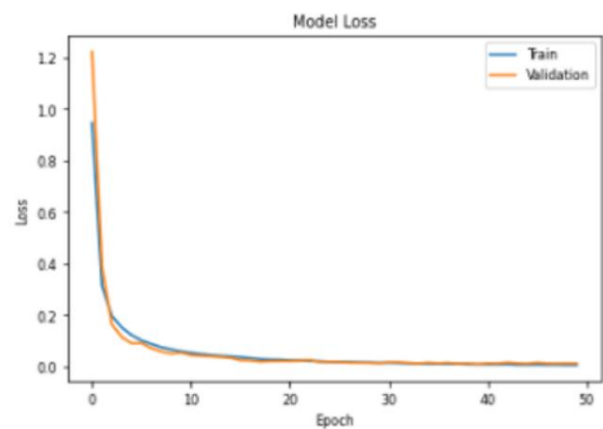
Dự đoán về tư thế yoga cho chuỗi 45 khung hình được triển khai bằng cách sử dụng tính năng bỏ phiếu, trong đó chế độ của tất cả các dự đoán tư thế yoga được coi là dự đoán cuối cùng.

4.1. Độ chính xác của mô hình và mất mô hình

Sau khi chạy mô hình trong 50 kỷ nguyên, độ chính xác đào tạo thu được là 99,49%, trong khi độ chính xác thực là 99,70%. Hình 5 minh họa biểu diễn đồ họa của cả độ chính xác của mô hình và sự mất mát của mô hình.



(Môc)



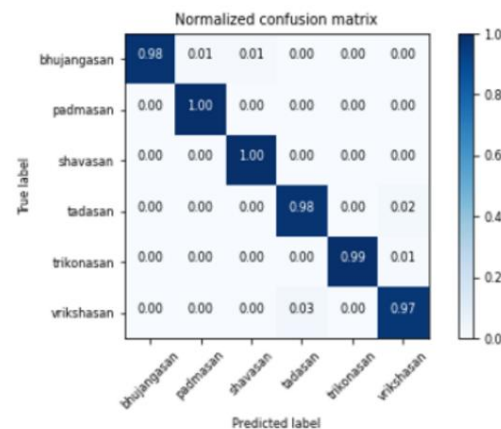
(b)

Hình 5. (a) Độ chính xác của mô hình, (b) Mất mô hình.

4.2. Ma trận nhầm lẫn Ma

trận nhầm lẫn được coi là thước đo hiệu suất quan trọng được sử dụng để tính toán các phân tích khác như thu hồi, độ chính xác, độ đặc hiệu, độ nhạy, v.v. Nó cũng được sử dụng cho các vấn đề phân loại để tóm tắt kết quả dự đoán [43]. Hơn nữa, nó bao gồm bốn giá trị chính là dự đoán tính thật, âm tính thật, dự đoán tính giả và âm tính giả [44].

Độ chính xác của bài kiểm tra thu được là 99,53% và Hình 6 trình bày ma trận nhầm lẫn giữa các asana yoga được dự đoán và các asana thực tế. Màu xanh trên đường chéo càng đậm thì mối tương quan giữa giá trị dự đoán và giá trị thực càng cao. Vì hầu hết các giá trị đường chéo gần bằng 1, ma trận biểu thị mối tương quan cao và điều này tương ứng với một mô hình khá chính xác.



Hình 6. Ma trận nhầm lẫn.

4.3. Độ chính xác

Độ chính xác là một số liệu được sử dụng để đánh giá hiệu suất của một mô hình. Cụ thể, nó được định nghĩa là tỷ lệ mẫu được phân loại chính xác là dự đoán tính với tất cả các mẫu thực sự dự đoán tính [21]. Độ chính xác thu được từ mô hình là 0,9866. Công thức toán học của nó được đưa ra trong phương trình (13):

$$\text{Độ chính xác} = \frac{TP}{TP + FP} \quad (13)$$

trong đó TP là dự đoán tính thực và FP dự đoán tính giả.

4.4. Nhớ lại

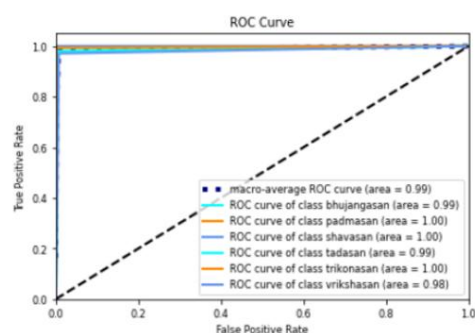
Nhớ lại là một số liệu khác được sử dụng để đánh giá hiệu suất của một mô hình. Nó được định nghĩa là tỷ lệ mẫu được phân loại chính xác là dự đoán tính với tất cả các mẫu được dự đoán là dự đoán tính [45]. Nhớ lại thu được từ mô hình là 0,9869. Công thức toán học của nó được đưa ra trong phương trình (14):

$$\text{Nhớ lại} = \frac{TP}{TP + FN} \quad (14)$$

trong đó TP là giá trị dự đoán thực và FN là giá trị âm sai.

4.5. Dự đoán công đặc tính hoạt động của máy thu (ROC)

Hơn nữa, chúng tôi sử dụng dự đoán công ROC, là biểu diễn đồ họa cho thấy hiệu suất của mô hình bằng cách vẽ biểu đồ tỷ lệ dự đoán tính giả so với tỷ lệ dự đoán tính thực [46]. Một mô hình được cho là lý tưởng khi diện tích dưới đường dự đoán công ROC có xu hướng bằng 1. Hình 7 minh họa dự đoán công ROC của tất cả các asana được sử dụng cho các thí nghiệm của chúng tôi. Diện tích thu được dưới đường dự đoán công là 0,99, một giá trị có thể được coi là đặc biệt cao.

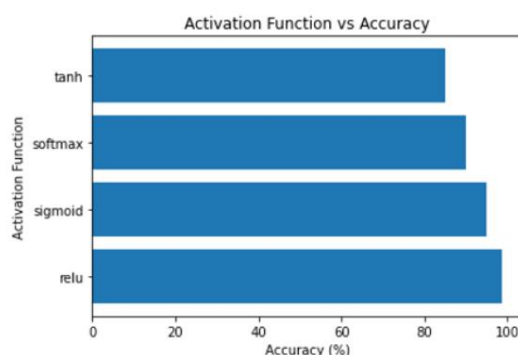


Hình 7. Dự đoán công ROC.

4.6. Phân tích so sánh các hàm kích hoạt được sử dụng Các hàm

Kích hoạt rất cần thiết đối với bất kỳ mô hình học sâu nào trong việc giúp khung được thiết kế nhân tạo tự động học các mẫu phức tạp từ dữ liệu được cung cấp cho nó. Còn được gọi là hàm truyền, nó thường nằm ở cuối, với nhiệm vụ được giao là thực hiện một số hoạt động phi tuyến tính trên đầu vào, trước khi hoàn thiện nó thành đầu ra và gửi nó đến lớp tế bào thần kinh kế tiếp.

Bốn chức năng kích hoạt khác nhau đã được sử dụng và thử nghiệm trong nghiên cứu này, đó là ReLU, Sigmoid, Softmax và Tanh, theo thứ tự bằng chữ cái. Như có thể được mô tả trong Hình 8, ReLU được chứng minh là chính xác nhất, tiếp theo là Sigmoid, Softmax và Tanh.



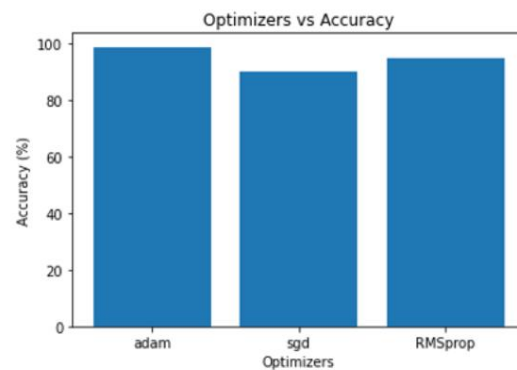
Hình 8. Các Chức năng Kích hoạt được sử dụng dựa trên Độ chính xác tương ứng của chúng.

4.7. Phân tích so sánh các trình tối ưu hóa được sử dụng

Các trình tối ưu hóa trong mạng thần kinh đóng một vai trò quan trọng trong việc giảm tổn thất mô hình và đồng thời cải thiện độ chính xác. Đây là những thuật toán chịu trách nhiệm cho

sửa đổi các khía cạnh nhất định của mạng, chẳng hạn như trọng số và tốc độ học tập, để mang lại kết quả tối ưu.

Ba công cụ tối ưu hóa khác nhau đã được sử dụng và thử nghiệm trong nghiên cứu này, đó là Adam, RMSprop và Stochastic Gradient Descent (SGD), theo thứ tự bảng chữ cái. Như có thể được mô tả trong Hình 9, Adam được chứng minh là chính xác nhất, tiếp theo là RMSprop và SGD. Cụ thể, các giá trị độ chính xác thu được cho Adam, RMSprop và SGD lần lượt là 98,72%, 94,66% và 89,16%.



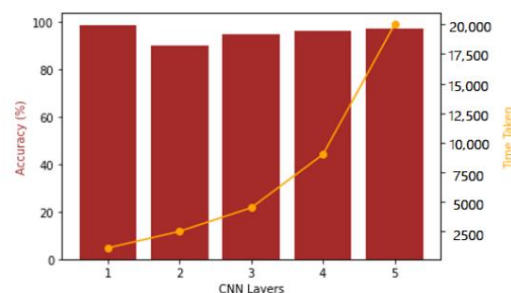
Hình 9. Trình tối ưu hóa được sử dụng và Độ chính xác tương ứng của chúng.

4.8. Tác động của số lượng lớp CNN đến hiệu suất mô hình

Số lượng các lớp ẩn trong mô hình CNN có thể ảnh hưởng đáng kể đến độ chính xác phân loại của nó. Mặc dù số lượng lớp tăng lên thường được quy cho độ chính xác tổng thể tăng lên, nhưng đây không phải là quan sát chung và thường phụ thuộc vào mức độ phức tạp của nhiệm vụ liên quan. Trong trường hợp không có tập huấn luyện đủ lớn, việc thêm các lớp bổ sung có thể mang lại sự tăng cường đáng kể cho một mạng rất lớn và phức tạp, dễ bị trang bị quá mức và do đó làm giảm độ chính xác thu được trên dữ liệu thử nghiệm.

Mặt khác, việc bổ sung thêm các lớp cũng có thể góp phần kéo dài thời gian thực hiện, như được minh họa qua biểu đồ đường trong Hình 10. Đối với mô hình cụ thể này, kết quả tốt nhất thu được khi sử dụng một lớp tích chập duy nhất.

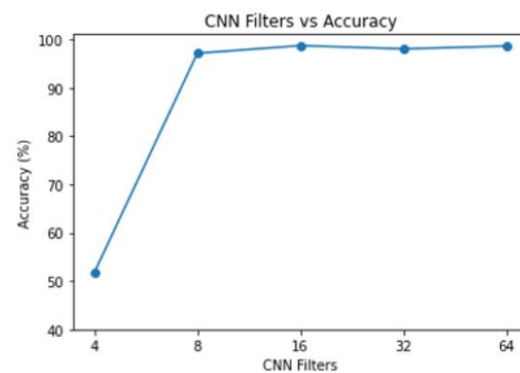
Thời gian cần thiết của mô hình do bao gồm các lớp từ 1 đến 5 lần lượt là 1100 ms, 2500 ms, 4600 ms, 9100 ms và 19900 ms. Điều này chứng tỏ rằng việc bổ sung các lớp làm tăng thời gian xử lý mô hình về mặt dự đoán.



Hình 10. Độ chính xác liên quan đến số lượng Lớp CNN khác nhau.

4.9. Tác động của số lượng bộ lọc CNN đến hiệu suất mô hình

Có thể xác định mối tương quan chặt chẽ giữa số lượng bộ lọc CNN được sử dụng và độ chính xác tương ứng như có thể xem trong Hình 11. Mỗi lớp bộ lọc chịu trách nhiệm trích xuất một bộ tính năng nhất định từ dữ liệu đầu vào. Khi các mẫu tiếp tục trở nên phức tạp hơn, số lượng bộ lọc ngày càng tăng được yêu cầu để trích xuất chính xác một số tổ hợp tính năng nổi bật khác nhau từ dữ liệu nhận được. Cụ thể, đối với ít nhất 8 bộ lọc CNN, độ chính xác có giá trị bằng 100%.

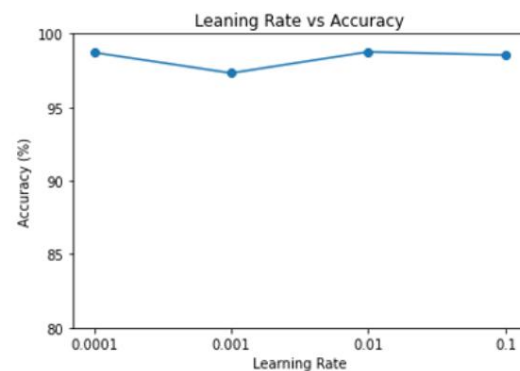


Hình 11. Độ chính xác liên quan đến số lượng Bộ lọc CNN khác nhau.

4.10. Tác động của tỷ lệ học tập đối với hiệu suất

của mô hình Tỷ lệ học tập đề cập đến mức độ điều chỉnh được phép trong mô hình để đáp ứng với lỗi quan sát được mỗi khi trọng số của mô hình được thay đổi. Việc quyết định tốc độ học tập tối ưu có thể khó khăn, vì cả hai thái cực đều không được mong muốn. Mặc dù tỷ lệ học tập thấp hơn có thể giúp xây dựng một mô hình chính xác hơn, nhưng nó có thể dẫn đến quá trình đào tạo cực kỳ chậm. Ngược lại, một giá trị rất cao có thể giảm đáng kể thời gian học, nhưng chi phí cho độ chính xác có thể tăng thêm.

Hình 12 trình bày sự phụ thuộc của độ chính xác vào tỷ lệ học tập. Cụ thể, tỷ lệ học tập nhận các giá trị trong khoảng từ 0,0010 đến 0,1 trong khi độ chính xác ít nhất là 97%.

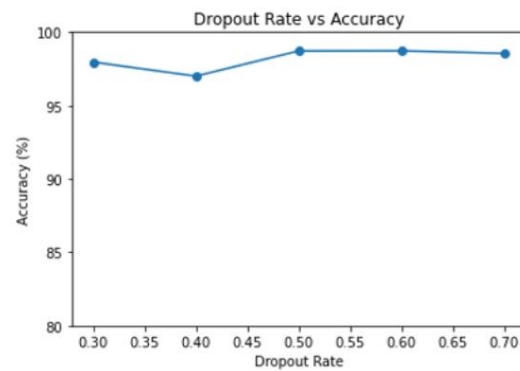


Hình 12. Tác động của Tỷ lệ Học đối với Độ chính xác.

4.11. Tác động của tỷ lệ bỏ học đối với hiệu suất của

mô hình Ở mọi giai đoạn của quá trình đào tạo, một số lượng tế bào thần kinh nhất định sẽ bị loại bỏ khỏi mỗi lớp. Phần lớn -ron này, có giá trị bị vô hiệu hóa, làm tăng tỷ lệ bỏ học. Quan trọng hơn, việc bỏ học là cần thiết để ngăn chặn việc trang bị quá mức trong các mạng thần kinh và cũng được sử dụng như một phương tiện để xử lý các sự đồng thích ứng phức tạp do các tế bào thần kinh có trọng số kết nối tương tự gây ra trong quá trình đào tạo.

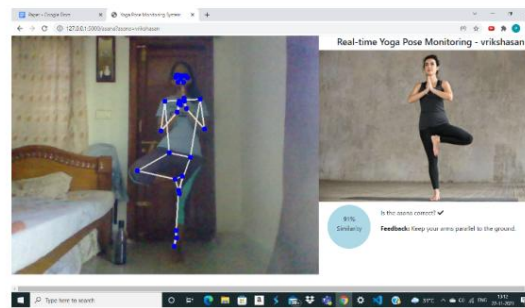
Hình 13 trình bày tỷ lệ bỏ học lý tưởng đạt được cho mô hình này. Cụ thể, tỷ lệ bỏ học lấy các giá trị trong khoảng từ 0,30 đến 0,70 trong khi độ chính xác ít nhất là 97%.



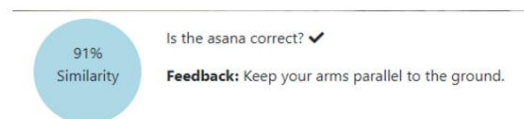
Hình 13. Tác động của tỷ lệ bỏ học đối với độ chính xác.

4.12. Kết quả chung cuộc

Sau khi kết hợp ba giai đoạn nói trên, một giao diện người dùng đáng tin cậy đã được tạo để theo dõi người dùng trong thời gian thực để dự đoán và chỉnh sửa tư thế. Hình 14 minh họa chế độ xem tổng thể của giao diện người dùng trong khi Hình 15 trình bày kết quả được cung cấp cho người dùng sau khi dự đoán và chỉnh sửa tư thế.



Hình 14. Giao diện người dùng.



Hình 15. Dự đoán và Hiệu chỉnh.

4.13. Cuộc thảo luận

Công trình đề xuất đã thực hiện một bước tiến so với các công trình khác có kiến trúc tương tự [19,27]. Cụ thể, Bảng 2 mô tả những khác biệt cơ bản của công việc đề xuất của chúng tôi so với những điều trên.

Việc sử dụng nhiều điểm chính hơn trong công việc được đề xuất dẫn đến chất lượng dự đoán tốt hơn và đáng tin cậy hơn. Ngoài ra, lớp Chuẩn hóa hàng loạt (BN) được sử dụng sau lớp Bỏ học để hội tụ nhanh hơn; việc giới thiệu lớp BN này đảm bảo rằng các pixel đầu vào có phân phối tương tự và giúp đào tạo nhanh hơn và tốt hơn. Hơn nữa, trình tối ưu hóa Adadelta được sử dụng trong [27] chậm hơn và không hiệu quả vì nó xem xét sự hội tụ từng bước.

Bảng 2. Sự khác biệt với các tác phẩm khác.

Tính năng	[19]	[27]	Công việc được đề xuất
Thư viện dùng để giải nén	Tư thế mở	-	phương tiện truyền thông
Những điểm chính			
Số điểm chính			33
Kỷ nguyên hội tụ	18 100	13 1000	50
trình tối ưu hóa	-	Adadelta	Adam
lớp	bỏ học và làm phẳng	LSTM và CNN	bỏ học và Chuẩn hóa hàng loạt
Phản hồi sau tư thế Yoga Nhận biết	KHÔNG	Tìm góc giữa các khớp	Tìm sự giống nhau với tư thế yoga đúng

5. Kết luận và công việc trong tương lai

Trong bài báo này, một hệ thống hiệu quả để theo dõi yoga theo thời gian thực đã được đề xuất. Nó ban đầu xác định các điểm chính của một người dùng cụ thể bằng cách sử dụng thư viện MediaPipe, trong đó tọa độ quan trọng được ghi lại và lưu trữ ở định dạng JSON. Sau đó, chúng tôi tạo trong thời gian thực một chuỗi gồm 45 khung hình và chuyển nó cho mô hình. Theo sau, mô hình, khi sử dụng kết hợp CNN và LSTM, tìm các tính năng hữu ích khi sử dụng CNN và quan sát sự xuất hiện của chuỗi khung sử dụng LSTM. Lớp Softmax cuối cùng tìm thấy xác suất của mỗi tư thế yoga cho chuỗi khung hình hiện tại và cho kết quả cao nhất asana xác suất. Hơn nữa, đầu ra của mỗi khung hình được thăm dò trên 45 khung hình trong đó chế độ được tính toán và cung cấp cho người dùng dưới dạng đầu ra.

Hệ thống đạt được độ chính xác tuyệt vời 99,53% trên tập dữ liệu thử nghiệm. Nếu tư thế được phân loại là chính xác, phản hồi tiếp theo sẽ được cung cấp cho người dùng dựa trên ngưỡng đã đặt. Cụ thể, ngưỡng được đặt theo cách không làm cho hệ thống trở nên quá nghiêm ngặt và đồng thời đảm bảo tư thế và góc độ chính xác do người dùng tạo ra. Cuối cùng, sự giống nhau phần trăm được hiển thị cho người dùng khi so sánh với tư thế tiêu chuẩn.

Về công việc trong tương lai, các biến thể và sự kết hợp của bộ mô hình được đề xuất được trình bày trong tác phẩm này rất đáng để thử, để nghiên cứu xem liệu có thể tiếp tục cải thiện độ chính xác. Hơn nữa, các bộ phân loại hiện có có thể được kiểm tra trong các bộ dữ liệu lớn hơn để xác minh mức độ chính xác cao đạt được trong phát hiện tình cảm. Ngoài các khối lượng lớn hơn của tập dữ liệu, điều quan trọng là phải thêm nhiều tính năng số hơn những tính năng đó chứa trong bộ được sử dụng cho mục đích của công việc này. Hơn nữa, sự kém hiệu quả của các mô hình đơn lẻ có thể được giải quyết bằng cách áp dụng một số kỹ thuật kết hợp, điều này sẽ dẫn đến cho kết quả chính xác hơn như trong [47].

Đóng góp của tác giả: Quản lý dữ liệu, DS, SS, BA, MS, VCG và AK; Phương pháp luận, DS, SS, BA, MS, VCG và AK; Viết–bản thảo gốc, DS, SS, BA, MS, VCG và AK; Viết–đánh giá và chỉnh sửa, DS, SS, BA, MS, VCG, AK và DG Tất cả các tác giả đã đọc và đồng ý với phiên bản xuất bản của bản thảo.

Kinh phí: Nghiên cứu này không nhận được tài trợ từ bên ngoài.

Tuyên bố về tính khả dụng của dữ liệu: Không áp dụng.

Xung đột lợi ích: Các tác giả tuyên bố không có xung đột lợi ích.

Người giới thiệu

1. Vrigkas, M.; Nikou, C.; Kakadiaris, IA Đánh giá về các phương pháp nhận biết hoạt động của con người. *Đăng trực. người máy. AI* 2015, 2, 28. [\[Tham khảo chéo\]](#)

2. Kế, S.; Hoàng, LUT; Lee, Y.; Hwang, J.; Yoo, J.; Choi, K. Đánh giá về Nhận dạng hoạt động của con người dựa trên video. *máy vi tính* 2013, 2, 88–131. [\[Tham khảo chéo\]](#)

3. Kothari, S. Phân loại tư thế Yoga bằng cách học sâu. *Bằng tiến sĩ. Luận án*, Đại học Bang San Jose, San Jose, CA, Hoa Kỳ, 2020.

4. Acharya, BR; Gantayat, PK Công nhận Hoạt động Bất thường của Con người trong Video Giám sát. *Quốc tế J. Res. Khoa học. đổi mới. (IJRSI)* 2015, 2, 18–23.

5. Stephens, I. *Liệu pháp Yoga Y học*. Trẻ em 2017, 4, 12. [\[CrossRef\]](#)

6. Newcombe, S. Sự phát triển của Yoga hiện đại: Khảo sát thực địa. tôn giáo. La bàn 2009, 3, 986-1002. [\[Tham khảo chéo\]](#)

7. Woodyard, C. Khám phá tác dụng trị liệu của Yoga và khả năng nâng cao chất lượng cuộc sống của nó. quốc tế J. Yoga 2011, 4, 49. [\[Tham khảo chéo\]](#)

8. Dấu gạch ngang, S.; Acharya, BR; Mittal, M.; Áp-ra-ham, A.; Kelemen, A. Kỹ thuật học sâu cho tin học y sinh và sức khỏe; Springer: Berlin/Heidelberg, Đức, 2020.

9. Brownlee, J. Các mô hình học sâu để nhận biết hoạt động của con người. 2018. Có sẵn trực tuyến: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1480070&dsid=-5372> (truy cập ngày 10 tháng 9 năm 2022).

10. Zhang, F.; Bazarevsky, V.; Vakunov, A.; Tkachenka, A.; Sung, G.; Chang, C.; Grundmann, M. MediaPipe Hands: Theo dõi tay trong thời gian thực trên thiết bị. arXiv 2020, arXiv:2006.10214.

11. Alzahrani, M.; Kammoun, S. Nhận biết hoạt động của con người: Các thách thức và các giai đoạn của quá trình. quốc tế J. Đổi mới. độ phân giải Điện toán. cộng đồng Tiếng Anh (IJIRCCE) 2016, 5, 1111-1118.

12. Gupta, S. Nhận dạng hoạt động của con người dựa trên học sâu (HAR) bằng cách sử dụng dữ liệu cảm biến có thể đeo được. quốc tế J. Inf. quản lý. Thông tin chi tiết về dữ liệu 2021, 1, 100046. [\[CrossRef\]](#)

13. Jogin, M.; Madhulika, MS; Divya, GD; Meghana, Vư ơ ng quốc Anh; Apoorva, S. Trích xuất tính năng bằng cách sử dụng Mạng thần kinh chuyển đổi (CNN) và Học sâu. Trong Kỳ yếu của Hội nghị Quốc tế IEEE lần thứ 3 về các Xu hướng Gần đây trong Công nghệ Điện tử, Thông tin & Truyền thông (RTEICT), Bangalore, Ấn Độ, 18-19 tháng 5 năm 2018; trang 2319-2323.

14. Josyula, R.; Ostadabbas, S. Đánh giá về ước tính tư thế con người. arXiv 2021, arXiv:2110.06877. Có sẵn trực tuyến: <https://arxiv.org/pdf/2110.06877.pdf> (truy cập ngày 10 tháng 9 năm 2022).

15. Kanavos, A.; Kounellis, F.; Iliadis, L.; Makris, C. Các mô hình học sâu để dự báo chuỗi thời gian nhu cầu hàng không. thần kinh Điện toán. ứng dụng 2021, 33, 16329-16343. [\[Tham khảo chéo\]](#)

16. Lyras, A.; Vernikou, S.; Kanavos, A.; Sioutas, S.; Mylonas, P. Mô hình hóa Độ tin cậy trong Dữ liệu lớn xã hội bằng Mạng thần kinh LSTM. Trong Kỳ yếu Hội nghị Quốc tế lần thứ 17 về Hệ thống Thông tin và Công nghệ Web (WEBIST), Trực tuyến, ngày 26-28 tháng 10 năm 2021; trang 599-606.

17. Agrawal, Y.; Shah, Y.; Sharma, A. Triển khai Kỹ thuật Máy học để Xác định Tư thế Yoga. Trong Kỳ yếu của Hội nghị Quốc tế IEEE lần thứ 9 về Hệ thống Truyền thông và Công nghệ Mạng (CSNT), Gwalior, Ấn Độ, ngày 10-12 tháng 4 năm 2020; trang 40-43.

18. Haque, S.; Rabby, ASA; Laboni, MA; Neehal, N.; Hossain, SA ExNET: Mạng thần kinh sâu để phát hiện tư thế tập thể dục. Trong Kỳ yếu của Hội nghị quốc tế lần thứ 2 về các xu hướng gần đây trong xử lý hình ảnh và nhận dạng mẫu (RTIP2R), Solapur, Ấn Độ, ngày 21-22 tháng 12 năm 2018; Truyền thông trong Khoa học Máy tính và Thông tin; Springer: Berlin/Heidelberg, Đức, 2018; Tập 1035, trang 186-193.

19. Yadav, SK; Singh, A.; Gupta, A.; Raheja, JL Nhận dạng Yoga trong thời gian thực bằng cách sử dụng Deep Learning. Máy tính thần kinh. ứng dụng năm 2019, 31, 9349-9361. [\[Tham khảo chéo\]](#)

20. Kadbhane, SG; Datir, KD; Jagdale, TS; Dhongade, SS; Jagtap, Nhận dạng tư thế GR Yoga. quốc tế J. Adv. độ phân giải Điện toán. cộng đồng Tiếng Anh (IJARCCE) 2021, 10, 143-147.

21. Buckland, MK; Gey, FC Mối quan hệ giữa Thu hồi và Độ chính xác. Mút. Sóc. thông tin liên lạc Khoa học. (JASIS) 1994, 45, 12-19. [\[Tham khảo chéo\]](#)

22. Luvizon, DC; Picard, D.; Tabia, H. 2D/3D Pose Estimation và Action Recognition bằng Multitask Deep Learning. Trong Kỳ yếu của Hội nghị IEEE về Tầm nhìn Máy tính và Nhận dạng Mẫu (CVPR), Thành phố Salt Lake, UT, Hoa Kỳ, ngày 18-22 tháng 6 năm 2018; trang 5137-5146.

23. Anilkumar, A.; Athulya, K.; Sajan, S.; Sreeja, K. Pose Hệ thống theo dõi Yoga ước tính. Trong Kỳ yếu của Hội nghị quốc tế về Hệ thống thông minh & Mạng điều khiển dựa trên IoT (ICICNIS), Kottayam, Ấn Độ, ngày 28-29 tháng 6 năm 2021.

24. Savvopoulos, A.; Kanavos, A.; Mylonas, P.; Sioutas, S. Máy gia tốc LSTM để Nhận dạng Đối tượng Tích chập. thuật toán 2018, 11, 157. [\[CrossRef\]](#)

25. Zou, J.; Lý, B.; Vư ơ ng, L.; Lý, Y.; Lý, X.; Lôi, R.; Sun, S. Hệ thống Huấn luyện viên Thể hình Thông minh Dựa trên Ước tính Tư thế Con người. Trong Kỳ yếu của Hội nghị Quốc tế về Xử lý Tín hiệu và Thông tin, Mạng và Máy tính (ICSINC), Ngọc Lâm, Trung Quốc, 29 tháng 11-1 tháng 12 năm 2018; trang 593-599.

26. Trần, S.; Yang, RR Pose Trainer: Điều chỉnh tư thế tập thể dục bằng cách sử dụng Ước tính tư thế. arXiv 2020, arXiv:2006.11718.

27. Thoutam, VA; Srivastava, A.; Badal, T.; Mishra, VK; Sinha, GR; Sakalle, A.; Bhardwaj, H.; Raj, M. Yoga Ước tính tư thế và tạo phản hồi bằng cách sử dụng Deep Learning. Điện toán. thông minh. Thần kinh học. 2022. Có sẵn trực tuyến: <https://www.hindawi.com/tap chí/cin/2022/4311350/> (truy cập ngày 10 tháng 9 năm 2022).

28. Al-Saffar, AAM; Tao, H.; Talab, MA Đánh giá về Mạng thần kinh tích chập sâu trong phân loại hình ảnh. Trong Kỳ yếu của Hội nghị Quốc tế IEEE về Ăng-ten Radar, Lò vi sóng, Điện tử và Viễn thông (ICRAMET), Jakarta, Indonesia, 23-24 tháng 10 năm 2017; trang 26-31.

29. Shiranthika, C.; Premakumara, N.; chiu, HL; Samani, H.; Shyalika, C.; Yang, CY Nhận dạng hoạt động của con người bằng CNN & LSTM. Trong Kỳ yếu của Hội nghị Quốc tế IEEE lần thứ 5 về Nghiên cứu Công nghệ Thông tin (ICITR), Moratuwa, Sri Lanka, ngày 2-4 tháng 12 năm 2020; trang 1-6.

30. Ullah, A.; Ahmad, J.; Muhammad, K.; Sajjad, M.; Baik, SW Nhận dạng hành động trong chuỗi video bằng Deep Bi-Directional LSTM Với các tính năng của CNN. Truy cập IEEE 2018, 6, 1155-1166. [\[Tham khảo chéo\]](#)

31. Agarap, AF Deep Learning sử dụng các đơn vị tuyến tính được chỉnh sửa (ReLU). arXiv 2018, arXiv:1803.08375.

32. Swain, D.; Pani, SK; Swain, D. Chẩn đoán bệnh động mạch vành bằng mạng thần kinh tích chập 1-D. 2019. Có sẵn trực tuyến: <https://www.ijrte.org/wp-content/uploads/papers/v8i2/B2693078219.pdf> (truy cập ngày 10 tháng 9 năm 2022).
33. Thakkar, V.; Tewary, S.; Chakraborty, C. Batch Normalization in Convolutional Neural Networks—Một nghiên cứu so sánh với dữ liệu CIFAR-10. Trong Kỷ yếu của Hội nghị Quốc tế IEEE lần thứ 5 về Ứng dụng Công nghệ Thông tin Mới nổi (EAIT), Kolkata, Ấn Độ, ngày 12–13 tháng 1 năm 2018; trang 1–5.
34. Srivastava, N.; Hinton, GE; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Bỏ học: Một cách đơn giản để ngăn chặn mạng nơ-ron trang bị quá mức. J. Mach. Học hỏi. độ phân giải 2014, 15, 1929–1958.
35. Szandala, T. Đánh giá và so sánh các chức năng kích hoạt thường được sử dụng cho mạng lưu ý thần kinh sâu. Trong cảm hứng sinh học Điện toán thần kinh; Springer: Singapore, 2021.
36. Kingma, DP; Ba, J. Adam: Phương pháp tối ưu hóa ngẫu nhiên. arXiv 2014, arXiv:1412.6980.
37. Rusiecki, A. Entropy chéo phân loại được cắt bớt để học sâu với nhiều nhãn. điện tử. Hãy để. 2019, 55, 319–320. [\[Tham khảo chéo\]](#)
38. Fatourehchi, M.; Phurong, RK; Thọ, N.; SG; Huggins, JE; Schlögl, A.; Birch, GE So sánh các chỉ số đánh giá trong các ứng dụng phân loại với bộ dữ liệu không cân bằng. Trong Kỷ yếu của Hội nghị Quốc tế lần thứ 7 về Máy học và Ứng dụng (ICMLA), San Diego, CA, USA, 11–15 tháng 12 năm 2008; trang 777–782.
39. Diaz, GI; Fokoue-Nkoutche, A.; Nannicini, G.; Samulowitz, H. Một thuật toán hiệu quả để tối ưu hóa siêu tham số của mạng thần kinh. IBM J. Res. nhà phát triển 2017, 61, 1–11. [\[Tham khảo chéo\]](#)
40. Swain, D.; Pani, SK; Swain, D. Một hệ thống hiệu quả để dự đoán bệnh động mạch vành bằng cách sử dụng mạng lưu ý thần kinh dày đặc với Điều chỉnh siêu thông số. quốc tế J. Đổi mới. công nghệ. thám hiểm. Tiếng Anh (IJITEE) 2019, 8, 689–695.
41. Subhash, S.; Siddesh, S.; Srivatsa, PN; Ullas, A.; Santhosh, B. Phát triển Giao diện Ngủ để dùng Đồ họa cho Trí tuệ Nhân tạo Trợ lý giọng nói dựa trên. quốc tế J. Organ. Sự u. tầm. thông minh. (IJOCI) 2021, 11, 49–67. [\[Tham khảo chéo\]](#)
42. Lahitani, AR; Permanasari, AE; Setiawan, NA Cosine Độ tương tự để Xác định Độ đo Độ tương tự: Trùng hợp Nghiên cứu trong Đánh giá Bài luận Trực tuyến. Trong Kỷ yếu của Hội nghị Quốc tế IEEE lần thứ 4 về Quản lý Dịch vụ CNTT và Mạng, Bandung, Indonesia, 26–27 tháng 4 năm 2016; trang 1–6.
43. Hasnain, M.; Pasha, MF; Ghani, I.; Imran, M.; Alzahrani, C. A. T. I.; Budiarto, R. Đánh giá dự báo lòng tin và các biện pháp ma trận nhằm lẫn để xếp hạng dịch vụ web. Truy cập IEEE 2020, 8, 90847–90861. [\[Tham khảo chéo\]](#)
44. Swain, D.; Ballal, P.; Dolase, V.; Dấu gạch ngang, B.; Santhappan, J. Một hệ thống dự đoán bệnh tim hiệu quả bằng cách sử dụng máy học. máy móc. Học hỏi. thông tin liên lạc Quá trình. 2020, 1101, 39–50.
45. Bradley, AP Việc sử dụng vùng dưới đường cong ROC trong đánh giá các thuật toán học máy. Nhận dạng mẫu. 1997, 30, 1145–1159. [\[Tham khảo chéo\]](#)
46. Swain, D.; Pani, SK; Swain, D. Một cuộc điều tra ẩn dụ về dự đoán bệnh tim bằng cách sử dụng máy học. Trong Kỷ yếu của Hội nghị Quốc tế IEEE về Điện toán Tiên tiến và Viễn thông (ICACAT), Bhopal, Ấn Độ, ngày 28–29 tháng 12 năm 2018; trang 1–6.
47. Vernikou, S.; Lyras, A.; Kanavos, A. Phân tích tình cảm đa lớp trên các Tweet liên quan đến COVID-19 bằng cách sử dụng Mô hình học sâu. Máy tính thần kinh. ứng dụng 2022, 34, 19615–19627. [\[Tham khảo chéo\]](#)