

MẠNG NƠ-RON TÍCH CHẬP BA CHIỀU CHO PHÂN LOẠI TỰ ĐỘNG MỨC ĐỘ THOÁI HÓA KHỚP GỐI

THREE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORKS FOR AUTOMATED KNEE OSTEOARTHRITIS GRADING

Trần Doãn Minh^{1*}
Hoàng Văn Hùng¹
Lê Đức Thọ¹
Lê Diệu Linh¹
Nguyễn Đình Thịnh¹

¹ Khoa Kỹ thuật Công nghệ, Trường Đại học Hồng Đức;
*Tác giả liên hệ; Email: trandoanminh@hdu.edu.vn

DOI: <https://doi.org/10.70117/hdujs.3030-4628.90.09.2026.1393>

Ngày nộp bài: 18/05/2026; Ngày gửi phản biện: 22/05/2026; Ngày duyệt đăng: 25/09/2026

Từ khóa:

3D CNN, Knee
Osteoarthritis, 3D
MRI.

TÓM TẮT

Thoái hóa khớp gối là một trong những bệnh lý cơ xương khớp phổ biến, gây ảnh hưởng đáng kể đến vận động và chất lượng cuộc sống. Mặc dù mạng nơ-ron tích chập (CNN) được ứng dụng rộng rãi trong phân tích ảnh y học, các nghiên cứu phân loại thoái hóa khớp gối chủ yếu dựa trên kiến trúc 2D và ảnh X-quang, chưa khai thác đầy đủ thông tin không gian ba chiều của ảnh cộng hưởng từ (MRI). Nghiên cứu này đề xuất sử dụng mạng nơ-ron tích chập ba chiều (3D CNN) để phân tích chuỗi ảnh MRI khớp gối và tự động phân loại mức độ thoái hóa theo thang đo Kellgren–Lawrence từ mức 0 đến mức 4.

Kết quả thực nghiệm cho thấy mô hình 3D CNN có khả năng khai thác đặc trưng không gian giữa các lát cắt MRI và đạt hiệu quả phân loại tốt hơn mô hình CNN 2D truyền thống. Kết quả cho thấy 3D CNN kết hợp MRI có tiềm năng hỗ trợ đánh giá tự động mức độ thoái hóa khớp gối, góp phần nâng cao hiệu quả phân tích hình ảnh y khoa và hỗ trợ bác sĩ trong thực hành lâm sàng.

Keywords:

3D CNN, Knee
Osteoarthritis, 3D
MRI.

ABSTRACT

Knee osteoarthritis (KOA) is one of the most common musculoskeletal disorders, significantly affecting mobility and quality of life. Although convolutional neural networks (CNNs) have been widely applied to medical image analysis, previous studies on KOA severity classification have mainly relied on two-dimensional (2D) CNN architectures and X-ray images, which may not fully exploit the three-dimensional spatial information available in magnetic resonance imaging (MRI). This study proposes a three-

dimensional convolutional neural network (3D CNN) for analyzing knee MRI sequences and automatically classifying KOA severity according to the Kellgren–Lawrence (KL) grading scale from grade 0 to grade 4. Experimental results show that the 3D CNN can effectively extract spatial features across multiple MRI slices and achieves better classification performance than the conventional 2D CNN model. These findings demonstrate the potential of combining MRI with 3D CNNs for the automated assessment of KOA severity, which may improve medical image analysis and support clinicians in clinical practice.

1. MỞ ĐẦU

Theo báo cáo của dịch tễ học ở Hoa Kỳ, thoái hóa khớp (osteoarthritis, OA) là dạng rối loạn khớp phổ biến nhất [1]. Thoái hóa khớp gây đau và là nguyên nhân hàng đầu dẫn đến suy giảm khả năng thực hiện các hoạt động sinh hoạt hằng ngày như đi bộ và leo cầu thang [2]. Ngoài ra, bệnh có mối liên quan chặt chẽ với tuổi tác [3] và được đặc trưng bởi sự suy giảm thể tích ở sụn khớp [4]. Mặc dù có ý nghĩa lớn đối với sức khỏe cộng đồng, hiện nay vẫn chưa có biện pháp can thiệp nào có thể điều chỉnh hiệu quả tiến trình bệnh OA [5]. Nguyên nhân có thể kể tới như sự thiếu hụt các chỉ dấu sinh học hữu ích để phát hiện và theo dõi tiến triển của OA được xem là rào cản công nghệ quan trọng đối với việc phát triển các phương pháp điều trị và phòng ngừa thoái hóa khớp gối.

Chụp X-quang thông thường thường được sử dụng để kiểm tra OA khớp gối định kỳ. Hình ảnh X-quang của một khớp bị thoái hóa khớp thường cho thấy sự thu hẹp khoảng cách giữa các xương của khớp gối nơi sụn đã bị mòn. Tuy nhiên, các triệu chứng của OA khớp gối có thể xuất hiện trước khi tổn thương có thể được nhìn thấy trên phim X-quang tiêu chuẩn. Ví dụ, trong công trình của nhóm tác giả Roemer [6] đã mô tả cách X-quang không thể hiện thị một số kiểu hình cấu trúc của OA và không thể phát hiện một số dấu hiệu bất lợi có thể cho thấy nguy cơ mắc bệnh tiến triển nhanh chóng.

Sự ra đời của chụp cộng hưởng từ (MRI) hứa hẹn giải quyết khoảng trống công nghệ quan trọng bằng cách cho phép định lượng tổn thương cấu trúc ở các khớp. Vì lý do này, các bác sĩ tại các bệnh viện thường sử dụng chụp cộng hưởng từ (MRI) nhạy hơn để phát hiện sớm bệnh thoái hóa khớp (OA). Trong công trình của nhóm tác giả Juras [7] cũng chỉ ra rằng OA cần được phát hiện sớm, và MRI là một phương pháp không xâm lấn để phát hiện các dấu hiệu sinh học sớm.

Hình ảnh MRI 3 chiều sẽ cho phép xem khớp gối như một “cơ quan toàn diện” và mô tả tất cả các mô của khớp. Mặc dù sự thoái hóa khớp gối khác có thể được phát hiện thủ công, nhưng việc xử lý thể tích hình ảnh MRI là công việc rất tốn thời gian. Do đó, tự động hóa các quy trình này bằng các phương pháp học máy là một yêu cầu thiết thực. Ngoài ra, trong khi các phương pháp trước đây sử dụng mạng CNN 2D để phân tích ảnh X-quang [8][9][10][11], trong nghiên cứu này, chúng tôi đề xuất một phương pháp sử dụng mạng 3D CNN và ảnh cộng hưởng từ (MRI).

Bài báo được cấu trúc các phần như sau. Phần đầu cung cấp giới thiệu tổng quan về bài toán phân loại bệnh thoái hoá khớp gối. Kế đến, phần hai chúng tôi trình bày về các nghiên cứu liên quan và đề xuất giải pháp cho phép phân loại mức độ thoái hoá khớp gối. Các kết quả nghiên cứu được trình bày ở phần 3. Cuối cùng là các kết luận và một số hướng nghiên cứu trong tương lai.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Phương pháp tiên tiến hiện nay

Trong thực hành lâm sàng và nghiên cứu dịch tễ học, mức độ thoái hóa khớp gối thường được đánh giá dựa trên hệ thống phân loại Kellgren–Lawrence (KL), một thang đo tiêu chuẩn phản ánh mức độ tổn thương cấu trúc của khớp [12]. Hệ thống này chia thoái hóa khớp gối thành năm mức độ, từ độ 0 đến độ 4. Hình 1 minh họa năm mức độ thoái hóa khớp gối theo thang đo KL. Cụ thể, độ 0 biểu thị khớp gối bình thường, không có dấu hiệu thoái hóa trên hình ảnh chụp. Độ 1 sự xuất hiện nghi ngờ của gai xương và có thể kèm theo hẹp khe khớp rất nhẹ, chưa rõ ràng. Độ 2 được xác định khi gai xương xuất hiện rõ ràng hơn, khe khớp bắt đầu hẹp nhưng chưa ảnh hưởng đáng kể đến trục và hình thái khớp. Độ 3 phản ánh tình trạng thoái hóa mức độ trung bình với hẹp khe khớp rõ rệt, gai xương phát triển, xơ hóa dưới sụn và có thể kèm theo biến dạng nhẹ đầu xương. Cuối cùng, độ 4 biểu hiện thoái hóa nặng, đặc trưng bởi hẹp khe khớp nghiêm trọng, gai xương lớn, xơ hóa dưới sụn rõ rệt và biến dạng khớp đáng kể. Phân loại Kellgren–Lawrence được sử dụng rộng rãi nhờ tính đơn giản, khả năng tái lập và giá trị tham chiếu cao trong đánh giá tiên tri và mức độ nặng của thoái hóa khớp gối.



Hình 1. Minh họa mức độ thoái hoá khớp gối theo thang đo KL

Trong khoảng một thập kỷ trở lại đây, dựa trên phương pháp phân loại theo độ đo KL, các mạng nơ-ron học sâu đã được phát triển để xác định mức độ thoái hóa khớp gối [8][9][10][11].

Nhóm tác giả Anthony và cộng sự đã sử dụng kiến trúc mạng nơ-ron tích chập VGG-16 cổ điển kết hợp với kỹ thuật học chuyển giao trên ảnh X-quang để xác định mức độ nặng của thoái hóa khớp gối [8]. Các ảnh đầu vào được tiền xử lý bằng phương pháp vector hỗ trợ và bộ phát hiện biên Sobel nhằm định vị vùng khớp gối. Nghiên cứu sử dụng dữ liệu ảnh X-quang từ bộ cơ sở dữ liệu OAI, với tổng cộng 4 446 ảnh, tương ứng 8 892 khớp gối. Khi thực hiện phân loại theo năm mức độ KL, phương pháp đạt độ chính xác 59,6%. Trong một nghiên cứu tiếp theo, cùng nhóm tác giả đã cải tiến bước tiền xử lý bằng cách áp dụng mạng nơ-ron tích chập đầy đủ để xác định khung bao của vùng khớp gối. Phương pháp cải tiến cho thấy độ chính xác cao trong việc xác định vùng quan tâm, và khi kết hợp với CNN cho bước phân loại, độ chính xác đạt được là 61,9% [9].

Khác với các bước xử lý hai giai đoạn được đề xuất trong [8][9], một nghiên cứu gần đây [10] đã giới thiệu kiến trúc CNN dạng đầu - cuối (end-to-end) để đánh giá mức độ thoái hóa khớp gối mà không cần sử dụng mạng nơ-ron cho bước tiền xử lý. Phương pháp này tích hợp các nhánh trong CNN, được gọi là các “mô-đun chú ý”, cho phép xác định vùng quan tâm trên ảnh X-quang một cách không giám sát. Một hướng tiếp cận khác đã bổ sung bước phân loại bằng mạng Long Short-Term Memory - LSTM sau các lớp CNN [11]. Do đặc tính của LSTM trong xử lý dữ liệu tuần tự, nhóm tác giả đã tạo thêm các ảnh trong bước tiền xử lý bằng cách cắt một vùng quan tâm cố định và xoay ảnh đã cắt với các góc 5° , 10° , -5° và -10° . Ảnh gốc và các ảnh tăng cường được xếp chồng, tạo ra khoảng 4 600 ảnh dùng cho huấn luyện và khoảng 480 ảnh cho kiểm thử. Nghiên cứu này cũng sử dụng dữ liệu từ OAI và đạt độ chính xác 75,28% trong bài toán phân loại năm mức độ KL.

Qua khảo sát các phương pháp tiên tiến, chúng tôi nhận thấy rằng, việc áp dụng trực tiếp mô hình mạng nơ-ron nhân chập 3 chiều (3D CNN) cho bài toán phân loại mức độ thoái hóa khớp gối vẫn còn rất hạn chế. Phần lớn các nghiên cứu hiện nay tập trung vào khai thác ảnh X-quang hai chiều kết hợp với các kiến trúc 2D CNN, hoặc chỉ sử dụng 3D CNN cho các nhiệm vụ liên quan như phân đoạn hay dự đoán nguy cơ tiến triển bệnh, thay vì phân loại mức độ thoái hóa theo các thang đo lâm sàng. Do đó, vẫn tồn tại một khoảng trống nghiên cứu trong việc tận dụng đầy đủ thông tin không gian ba chiều từ ảnh cộng hưởng từ để đánh giá mức độ nặng của thoái hóa khớp gối bằng 3D CNN.

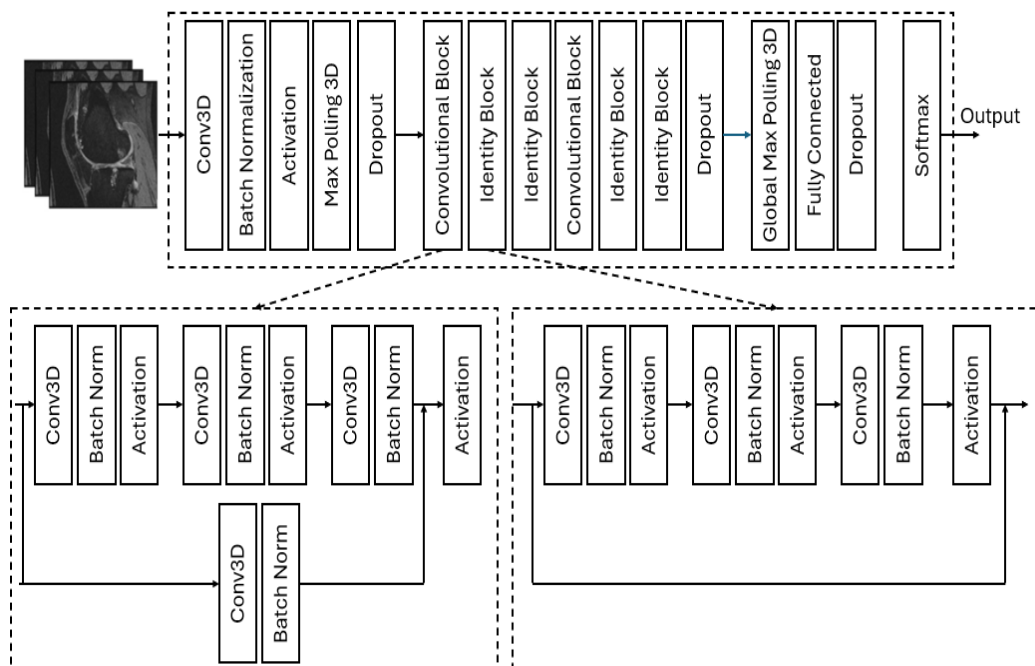
Nhằm giải quyết những thách thức nêu trên, bài báo đề xuất một phương pháp phân loại mức độ thoái hóa khớp gối dựa trên mạng 3D CNN, khai thác hiệu quả đặc trưng thể tích của dữ liệu MRI, qua đó kỳ vọng cải thiện độ chính xác và tính tin cậy so với các phương pháp dựa trên mạng nơ-ron nhân chập và ảnh hai chiều trước đây.

2.2. Giải pháp đề xuất

Ảnh cộng hưởng từ khớp gối tạo ra một biểu diễn ba chiều của cấu trúc khớp thông qua một chuỗi các lát cắt hai chiều được thu nhận theo phương ngang dọc theo khớp gối. Với bản chất dữ liệu mang tính không gian ba chiều của ảnh MRI, việc sử dụng mạng nơ-ron tích chập ba chiều (3D CNN) mang lại nhiều lợi thế trong việc đánh giá toàn bộ chuỗi ảnh như một thể thống nhất. Thông qua việc áp dụng các nhân tích chập ba chiều, thông tin giữa các lát cắt lân cận có thể được tích hợp một cách hiệu quả. Nhờ đó, các đặc trưng không gian ba chiều - những đặc trưng có thể khó hoặc không thể phát hiện bằng các mô hình 2D CNN - có khả năng được khai thác và biểu diễn đầy đủ hơn.

Trong bài báo này, chúng tôi đề xuất một mô hình 3D CNN để phân loại mức độ thoái hóa khớp gối, có tên là OA-Net. Kiến trúc mô hình OA-Net được minh họa trong Hình 2. Mô hình OA-Net được lấy cảm hứng từ mô hình của nhóm tác giả [13]. Sự khác biệt quan trọng nhất giữa mạng 3D CNN và mạng 2D CNN nằm ở cách thức xử lý dữ liệu đầu vào. Cụ thể, 3D CNN sử dụng các nhân tích chập ba chiều để xử lý các khối dữ liệu thể tích của ảnh quét, trong khi 2D CNN chỉ xử lý dữ liệu trên một mặt phẳng giải phẫu đơn lẻ. Các phép nhân tích chập ba chiều cho phép tích hợp thông tin từ các lát cắt

lân cận, nhờ đó có khả năng trích xuất các đặc trưng không gian ba chiều mà các mô hình 2D CNN không thể phát hiện được. Như minh họa trong Hình 2, mô hình 3D CNN được đề xuất bao gồm ba giai đoạn xử lý chính trước khi đi đến lớp phân loại cuối cùng. Các chi tiết cụ thể của mô hình sẽ được trình bày dưới đây.



Hình 2. Kiến trúc mô hình OA-Net

Giai đoạn 1 của mô hình bắt đầu với một lớp tích chập gồm 32 nhân có kích thước $7 \times 7 \times 7$ và bước trượt $2 \times 2 \times 2$. Tiếp theo là lớp chuẩn hóa theo lô (batch normalization) và lớp kích hoạt sử dụng hàm ReLU. Sau đó, một lớp gộp cực đại (max-pooling) được thêm vào với kích thước cửa sổ $2 \times 2 \times 2$ và bước trượt $2 \times 2 \times 2$. Cuối cùng, một lớp dropout được đặt trước khi bắt đầu các khối dư (residual blocks). Trong mô hình 3D CNN này, các lớp dropout được sử dụng ở mỗi giai đoạn nhằm giảm hiện tượng quá khớp (overfitting). Mỗi lớp dropout được thiết lập với tỷ lệ 0,5, nghĩa là mỗi nút có xác suất 50% bị đưa về giá trị 0 trong quá trình huấn luyện.

Giai đoạn thứ hai của mô hình bao gồm một chuỗi sáu khối dư (residual blocks). Mỗi khối dư đều có một kết nối tắt từ đầu vào trực tiếp tới đầu ra. Như minh họa trong Hình 2, mô hình sử dụng hai loại khối dư khác nhau. Khối tích chập (convolutional block) có thêm một lớp tích chập trong nhánh nối tắt, được sử dụng trong trường hợp kích thước của đầu vào thay đổi. Ngược lại, khối đồng nhất (identity block) không bao gồm lớp nào trong nhánh nối tắt và được áp dụng khi kích thước đầu vào và đầu ra là tương đương. Giai đoạn này cũng kết thúc bằng một lớp dropout.

Giai đoạn cuối cùng của mô hình sử dụng lớp gộp cực đại toàn cục (global max pooling), tiếp theo là một lớp kết nối đầy đủ (fully connected) với kích thước 1024 và một lớp dropout. Lớp cuối cùng của mô hình áp dụng hàm kích hoạt softmax, cho ra xác suất mẫu dữ liệu thuộc về từng lớp phân loại.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Bộ dữ liệu và tiền xử lý

Bộ dữ liệu được sử dụng trong nghiên cứu này được lấy từ cơ sở dữ liệu công khai Osteoarthritis Initiative (OAI)¹. Phần lớn các mẫu bệnh nhân trong bộ dữ liệu OAI có kèm theo ảnh X-quang; tuy nhiên, nhiều trường hợp không có chuỗi ảnh cộng hưởng từ (MRI) tương ứng. Trong nghiên cứu này, chúng tôi sử dụng một tập con của dữ liệu OAI gồm 1100 khớp gối, trong đó mỗi khớp gối đều có đầy đủ cả ảnh MRI và ảnh X-quang. Dữ liệu MRI DESS ba chiều của mỗi khớp gối bao gồm một chuỗi 160 ảnh hai chiều, trong khi mỗi bệnh nhân chỉ có một ảnh X-quang chứa đồng thời cả hai khớp gối. Tập dữ liệu được lựa chọn sao cho phân bố cân bằng giữa các mức độ nặng khác nhau của thoái hóa khớp gối (từ 0 đến 4) theo thang đo Kellgren–Lawrence (KL).

Để tiến hành huấn luyện và đánh giá mô hình, bộ dữ liệu được xáo trộn ngẫu nhiên và chia thành ba tập gồm tập huấn luyện, tập xác thực và tập kiểm thử, tương ứng 800, 200 và 100 khớp gối. Việc phân chia được thực hiện trên các mẫu khớp gối và bảo đảm sự cân bằng tương đối giữa năm mức độ Kellgren–Lawrence. Do dữ liệu OAI có thể bao gồm nhiều mẫu liên quan đến cùng một bệnh nhân, khả năng tương quan giữa các mẫu trong quá trình phân chia là một hạn chế cần được lưu ý khi diễn giải kết quả. Các nghiên cứu tiếp theo nên thực hiện phân chia dữ liệu ở cấp bệnh nhân để loại trừ hoàn toàn nguy cơ rò rỉ dữ liệu. Chúng tôi thực hiện cắt ảnh như một bước tiền xử lý nhằm giữ lại vùng sụn khớp trong khi loại bỏ các khu vực ít mang thông tin cho bài toán phân loại thoái hóa khớp gối. Ảnh ban đầu có kích thước 384 x 384 được cắt từ vùng trung tâm với cả cửa sổ hình vuông và hình chữ nhật. Thông qua nhiều thử nghiệm, chúng tôi nhận thấy rằng kích thước cửa sổ 160x160 mang lại giá trị F-measure cao nhất. Vì thế trong các thí nghiệm, chúng tôi báo cáo kết quả với ảnh đầu vào được cắt chỉnh ở kích thước 160x160.

Mỗi mẫu trong cơ sở dữ liệu bao gồm một chuỗi 160 ảnh MRI. Sau khi cắt ảnh, kích thước đầu vào $160 \times 160 \times 160$ vẫn còn rất lớn. Để tiếp tục giảm số chiều của dữ liệu đầu vào, chúng tôi loại bỏ một số lát cắt ở phần đầu, phần giữa và phần cuối của chuỗi ảnh. Việc loại bỏ một số lát cắt ban đầu và kết thúc là do các lát cắt này không chứa thông tin về xương hoặc sụn, nên khó có khả năng mang thông tin liên quan đến thoái hóa khớp. Trong khi đó, các lát cắt ở vùng giữa chuỗi thường có vùng sụn không được xác định rõ ràng và ranh giới xương bị mờ, do quá trình chuyển tiếp giữa phần trong và phần ngoài của xương xảy ra trong khoảng này. Cụ thể, đối với mỗi chuỗi ảnh, chúng tôi loại bỏ 10 lát cắt đầu (1-10), 20 lát cắt ở giữa (71-90) và 10 lát cắt cuối (151-160). 120 lát cắt còn lại (11-70 và 91-150) trong tổng số 160 lát cắt được đưa vào mô hình 3D CNN. Các lát cắt 11-70 và 91-150 được giữ nguyên thứ tự xuất hiện trong chuỗi MRI gốc và được ghép thành tensor đầu vào gồm 120 lát. Việc loại bỏ các lát 71-90 được thực hiện nhằm giảm kích thước dữ liệu và loại bỏ vùng ảnh có ít thông tin liên quan đến cấu trúc xương và sụn. Do đó, lát 70 và lát 91 là hai lát kế tiếp nhau trong tensor đầu vào nhưng không phải là hai lát liên kề trong chuỗi MRI ban đầu. Mô hình 3D CNN xử lý các lát này theo thứ tự trong tensor đã tiền xử lý; đây là một đặc điểm của chiến lược tiền xử lý được sử dụng trong nghiên cứu.

¹ Bộ dữ liệu OAI: <https://nda.nih.gov/oai>

3.2. Thiết lập môi trường

Quá trình huấn luyện và kiểm thử mô hình được thực hiện trên máy trạm trang bị GPU NVIDIA RTX 2080 Ti (12GB VRAM), CPU Intel(R) Xeon(R) W-2133 i6, RAM 32GB, chạy trên Ubuntu 22.04 LTS. Mô hình được triển khai bằng ngôn ngữ lập trình Python với các thư viện PyTorch, OpenCV, NumPy, và scikit-learn. Ngoài ra, mã nguồn được quản lý bằng môi trường ảo conda nhằm đảm bảo tính nhất quán giữa các gói phần mềm, dữ liệu lưu trữ trên ổ cứng SSD để tối ưu tốc độ truy cập. Quá trình huấn luyện và đánh giá được giám sát và ghi lại nhật ký để phục vụ tính toán độ chính xác và giá trị hàm mục tiêu.

Quá trình huấn luyện mô hình OA-Net, chúng tôi sử dụng các siêu tham số gồm: Learning rate = $1e-3$, Batch size = 5, số Epoch = 50, cùng bộ tối ưu Adam với hệ số Weight decay = $5e-4$. Ngoài ra, mô hình áp dụng kỹ thuật Dropout = 0.4 và Early Stopping dựa trên hiệu suất tập xác thực để tránh quá khớp. Việc thiết lập các siêu tham số theo các tiêu chuẩn được công bố trong các công trình trước giúp đảm bảo tính nhất quán, khả năng tái lập và độ tin cậy của kết quả thực nghiệm.

3.3. Kết quả

Trong nghiên cứu này, độ chính xác (Accuracy) được sử dụng làm độ đo chính để đánh giá hiệu năng của các mô hình phân loại. Đối với bài toán phân loại mức độ thoái hóa khớp gối gồm năm mức theo thang đo KL, độ chính xác tổng thể được tính bằng trung bình cộng của độ chính xác đạt được trên từng mức độ thoái hóa. Cách tiếp cận này nhằm đảm bảo rằng mỗi mức KL đều được đóng góp ngang nhau vào kết quả đánh giá cuối cùng, đồng thời hạn chế ảnh hưởng của sự mất cân bằng dữ liệu giữa các lớp. Việc sử dụng độ chính xác trung bình theo từng lớp giúp phản ánh toàn diện hơn khả năng phân loại của mô hình đối với cả các mức độ biên và các mức độ trung gian của thoái hóa khớp gối.

Bảng 1 trình bày hiệu suất của mô hình OA-Net được đề xuất thông qua ma trận nhầm lẫn giữa mức độ thực tế và mức độ dự đoán. Quan sát các kết quả trong Bảng 1 cho thấy hai mức độ biên, cụ thể là KL = 0 và KL = 4, tương đối dễ phân loại với độ chính xác lần lượt đạt 70% và 90%. Ngược lại, các mức độ trung gian gặp nhiều khó khăn hơn trong quá trình phân loại. Đối với mức KL = 1, độ chính xác chỉ đạt 45%; tuy nhiên, phần lớn các trường hợp bị phân loại sai (15 trong tổng số 22 trường hợp) bị nhầm sang mức KL = 0, tức là mức độ nhẹ hơn. Điều này ít nghiêm trọng hơn so với việc bị nhầm sang các mức KL cao hơn, bởi về mặt lâm sàng, cả KL = 0 và KL = 1 đều được xem là chưa mắc thoái hóa khớp. Mức KL = 2 có độ chính xác thấp nhất, chỉ đạt 37%, và đa số các trường hợp bị phân loại sai ở mức này được gán sang KL = 0 hoặc KL = 1. Điều này cho thấy cần có thêm phân tích để làm rõ vì sao mô hình lại xem mức KL = 2 tương đồng hơn với nhóm không thoái hóa, trong khi về mặt lâm sàng, mức này đã được xếp vào nhóm thoái hóa khớp.

Bảng 1. Kết quả đánh giá trên tập xác thực

Thật / Dự đoán	P0	P1	P2	P3	P4	Độ chính xác
A0	28	7	5	0	0	70%
A1	15	18	2	5	0	45%
A2	7	10	15	7	1	37%
A3	0	2	2	24	12	60%
A4	0	0	0	4	36	90%

Để so sánh hiệu suất của mô hình OA-Net với các phương pháp học sâu phổ biến, trong nghiên cứu này chúng tôi lựa chọn hai kiến trúc mạng nơ-ron tích chập hai chiều tiêu biểu là VGG16 và ResNet50. Hai mô hình này được huấn luyện và đánh giá trên dữ liệu ảnh X-quang khớp gối lấy từ cùng tập dữ liệu OAI được sử dụng trong nghiên cứu. Việc sử dụng ảnh X-quang trong cùng một nguồn dữ liệu giúp đảm bảo tính nhất quán về phân bố mẫu, mức độ thoái hóa theo thang đo Kellgren–Lawrence và các yếu tố lâm sàng liên quan. Cách thiết lập này cho phép so sánh một cách khách quan và công bằng giữa các mô hình 2D CNN dựa trên ảnh X-quang và mô hình OA-Net dựa trên mạng 3D CNN khai thác dữ liệu MRI, qua đó làm rõ lợi ích của việc sử dụng thông tin không gian ba chiều trong phân loại mức độ thoái hóa khớp gối.

Bảng 2. So sánh độ chính xác của mô hình đề xuất với các mô hình khác

Mô hình	Tập xác thực	Tập kiểm thử
VGG16	30.2%	28.5%
ResNet50	49.5%	40.0%
OA-Net	60.5%	54.0%

Kết quả trình bày trong Bảng 2 cho thấy sự khác biệt rõ rệt về hiệu năng giữa các mô hình được đánh giá trên cả tập xác thực và tập kiểm thử. Các mô hình 2D CNN truyền thống như VGG16 và ResNet50 cho độ chính xác tương đối thấp, trong đó VGG16 đạt 30,2% trên tập xác thực và 28,5% trên tập kiểm thử, phản ánh hạn chế của việc chỉ khai thác thông tin hai chiều từ dữ liệu ảnh. ResNet50 cho kết quả cải thiện hơn với độ chính xác lần lượt là 49,5% và 40,0%, tuy nhiên hiệu năng vẫn suy giảm đáng kể khi chuyển sang tập kiểm thử, cho thấy khả năng tổng quát hóa còn hạn chế. Ngược lại, mô hình OA-Net do chúng tôi đề xuất, đạt độ chính xác cao nhất trên cả hai tập dữ liệu, lần lượt là 60,5% trên tập xác thực và 54,0% trên tập kiểm thử. Kết quả này cho thấy việc khai thác trực tiếp thông tin không gian ba chiều từ dữ liệu MRI giúp mô hình học được các đặc trưng cấu trúc quan trọng của khớp gối, qua đó cải thiện đáng kể khả năng phân loại mức độ thoái hóa khớp gối so với các mô hình 2D CNN. Điều này cũng khẳng định tính hiệu quả và tiềm năng của hướng tiếp cận dựa trên 3D CNN trong bài toán đánh giá mức độ thoái hóa khớp gối.

4. KẾT LUẬN

Trong bài báo này, chúng tôi đã đề xuất OA-Net, một mô hình mạng nơ-ron tích chập ba chiều (3D CNN) kết hợp với dữ liệu ảnh cộng hưởng từ ba chiều (MRI) nhằm giải quyết bài toán phân loại mức độ thoái hóa khớp gối. Kết quả thực nghiệm cho thấy OA-Net đạt độ chính xác 54,0% trong bài toán phân loại năm mức độ theo thang đo Kellgren-Lawrence và thể hiện sự cải thiện rõ rệt so với các mô hình mạng nơ-ron tích chập hai chiều truyền thống sử dụng ảnh X-quang trên cùng tập dữ liệu bệnh nhân. Những kết quả này khẳng định tiềm năng của việc khai thác dữ liệu MRI kết hợp với mô hình 3D CNN trong việc nâng cao độ chính xác và độ tin cậy của hệ thống hỗ trợ chẩn đoán thoái hóa khớp gối, đồng thời mở ra hướng nghiên cứu mới cho các phương pháp đánh giá dựa trên dữ liệu hình ảnh ba chiều.

TÀI LIỆU THAM KHẢO

- [1] Y. Zhang and J. M. Jordan, "Epidemiology of osteoarthritis," *Clin. Geriatr. Med.*, 26(3):355, 2010, DOI:10.1016/j.cger.2010.03.001.
- [2] M. Fransen, S. McConnell, A. R. Harmer, M. Van der Esch, M. Simic, and K. L. Bennell, "Exercise for osteoarthritis of the knee: a Cochrane systematic review," *Br. J. Sports Med.*, 49(24):1554-1557, 2015, DOI:10.1136/bjsports-2015-095424.
- [3] D. T. Felson, A. Naimark, J. Anderson, L. Kazis, W. Castelli, and R. F. Meenan, "The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study," *Arthritis Rheum. Off. J. Am. Coll. Rheumatol.*, 30(8):914-918, 1987, DOI:10.1002/art.1780300811.
- [4] G. Jones, C. Ding, F. Scott, M. Glisson, and F. Cicuttini, "Early radiographic osteoarthritis is associated with substantial changes in cartilage volume and tibial bone surface area in both males and females," *Osteoarthritis Cartilage*, 12(2): 169-174, 2004, DOI:10.1016/j.joca.2003.08.010.
- [5] D. Bhatia, T. Bejarano, and M. Novo, "Current interventions in the management of knee osteoarthritis," *J. Pharm. Bioallied Sci.*, 5(1):30-38, 2013, DOI:10.4103/0975-7406.106561.
- [6] F. W. Roemer, C. K. Kwok, D. Hayashi, D. T. Felson, and A. Guermazi, "The role of radiography and MRI for eligibility assessment in DMOAD trials of knee OA," *Nat. Rev. Rheumatol.*, 14(6):372-380, 2018, DOI:10.1038/s41584-018-0010-z.
- [7] V. Juras, G. Chang, and R. R. Regatte, "Current status of functional MRI of osteoarthritis for diagnosis and prognosis," *Curr. Opin. Rheumatol.*, 32(1):102-109, 2020, DOI:10.1097/BOR.0000000000000674.
- [8] J. Antony, K. McGuinness, N. E. O'Connor, and K. Moran, "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in 2016 23rd international conference on pattern recognition (ICPR), IEEE, 2016, pp.1195-1200.

- [9] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in International conference on machine learning and data mining in pattern recognition, Springer, 2017, pp.376-390.
- [10] M. Górriz, J. Antony, K. McGuinness, X. Giró-i-Nieto, and N. E. O'Connor, "Assessing knee OA severity with CNN attention-based end-to-end architectures," in International conference on medical imaging with deep learning, PMLR, 2019, pp.197-214.
- [11] R. T. Wahyuningrum, L. Anifah, I. K. E. Purnama, and M. H. Purnomo, "A new approach to classify knee osteoarthritis severity from radiographic images based on CNN-LSTM method," in 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), IEEE, 2019, pp.1-6.
- [12] J. H. Kellgren and J. Lawrence, "Radiological assessment of osteo-arthritis," Ann Rheum Dis, 16(4):494-502, 1957, <https://doi.org/10.1136/ard.16.4.494>.
- [13] T. Wang, K. Leung, K. Cho, G. Chang, and C. M. Deniz, "Total knee replacement prediction using structural MRIs and 3D convolutional neural networks," in International Conference on Medical Imaging with Deep Learning (MIDL) - Extended Abstract Track, 2019, pp.1-5.

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

Thông tin tài trợ (nếu có): Nghiên cứu này không nhận được tài trợ từ bên ngoài.