

## NÂNG CAO HIỆU QUẢ PHÂN LOẠI THOÁI HOÁ KHỚP GỐI BẰNG CNN VỚI DỮ LIỆU MẤT CÂN BẰNG

### IMPROVING KNEE OSTEOARTHRITIS CLASSIFICATION USING CNNs ON IMBALANCED DATA

Lê Thị Thanh Thủy<sup>1</sup>  
Lê Văn Hào<sup>2\*</sup>  
Phạm Thế Anh<sup>3</sup>

<sup>1</sup>Bệnh viện Y Dược Cổ truyền Thanh Hóa;  
<sup>2</sup>Khoa Kỹ thuật Công nghệ, Trường Đại học Hồng Đức;  
<sup>3</sup>Phòng Tổ chức - Hành chính - Quản trị;  
\*Tác giả liên hệ; Email: levanhao@hdu.edu.vn

DOI: <https://doi.org/10.70117/hdujs.3030-4628.90.09.2026.1155>

Ngày nộp bài: 11/02/2026; Ngày gửi phản biện: 13/02/2026; Ngày duyệt đăng: 25/09/2026

#### Từ khóa:

*Mạng nơ-ron tích  
chập, thoái hoá  
khớp gối, dữ liệu  
mất cân bằng.*

#### Keywords:

*CNN, Knee  
Osteoarthritis,  
imbalanced data.*

#### TÓM TẮT

*Bài báo nghiên cứu việc nâng cao hiệu quả phân loại mức độ thoái hoá khớp gối từ ảnh X-quang bằng mạng nơ-ron tích chập (CNN) trong điều kiện dữ liệu mất cân bằng. Các phương pháp học chuyển giao trên những mô hình CNN điển hình gồm VGG16, ResNet50 và EfficientNet-B0 được sử dụng và đánh giá trên bộ dữ liệu chuẩn Osteoarthritis Initiative (OAI). Chiến lược gán trọng số lớp được sử dụng nhằm cải thiện khả năng nhận diện các mức độ thoái hoá trung bình và nặng. Kết quả cho thấy hiệu năng phân loại được cải thiện đáng kể, trong đó EfficientNet-B0 đạt kết quả tốt nhất (macro-F<sub>1</sub> = 68%). Kết quả cho thấy học chuyển giao kết hợp xử lý mất cân bằng dữ liệu có tiềm năng nâng cao khả năng nhận diện các mức độ thoái hoá khớp gối, góp phần hỗ trợ đánh giá X-quang trong thực hành lâm sàng.*

#### ABSTRACT

*This study evaluates several convolutional neural network (CNN) architectures for classifying knee osteoarthritis severity from X-ray images under class-imbalanced conditions. Transfer learning is employed using pre-trained CNN models, including VGG16, ResNet50, and EfficientNet-B0, on the publicly available Osteoarthritis Initiative (OAI) dataset. To address class imbalance, a class-weighted training strategy is adopted to improve the recognition of moderate and severe osteoarthritis cases. Experimental results show improved classification performance when imbalance handling is applied, with EfficientNet-B0 achieving the best performance and a macro-F<sub>1</sub> of 68%. These findings suggest that CNN-based approaches have the potential to support clinical assessment of knee osteoarthritis severity.*

## 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 gối 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]. Một trong những nguyên nhân là sự thiếu hụt các chỉ dấu sinh học đáng tin cậy phục vụ phát hiện và theo dõi tiến triển của OA.

Trong khoảng một thập kỷ trở lại đây, sự ra đời của mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) đã mở ra hướng tiếp cận mạnh mẽ cho bài toán phân loại ảnh y sinh [6]. Tuy nhiên, một thách thức lớn khi áp dụng học sâu cho dữ liệu ảnh y sinh là mất cân bằng lớp: số lượng mẫu cho một số mức độ bệnh (ví dụ mức nặng) thường ít hơn nhiều so với mức nhẹ hoặc bình thường, dẫn đến hiện tượng mô hình bị thiên lệch và giảm hiệu quả phân loại cho các lớp thiểu số [7].

Xuất phát từ các thách thức và kết quả đã đạt được trong các nghiên cứu trước, bài báo này tập trung khảo sát và đánh giá hiệu năng của nhiều kiến trúc CNN khác nhau trong bài toán phân loại mức độ thoái hóa khớp gối từ ảnh X-quang, đồng thời áp dụng chiến lược xử lý mất cân bằng dữ liệu và tăng cường dữ liệu để cải thiện khả năng nhận dạng chính xác các mức độ bệnh khác nhau. Mục tiêu của nghiên cứu là đề xuất một phương pháp tổng thể hiệu quả hơn cho nhiệm vụ phân loại mức độ thoái hóa khớp gối dựa trên mô hình CNN.

## 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 [8]. Hình 1 minh họa các mức độ thoái hóa khớp gối theo thang đo KL. Thang KL chia bệnh thành năm mức độ từ 0 đến 4, tương ứng với mức độ tổn thương khớp tăng dần, từ khớp bình thường (độ 0) đến thoái hóa nặng với hẹp khe khớp nghiêm trọng, gai xương lớn và biến dạng khớp rõ rệt (độ 4).



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

Những tiến bộ gần đây trong học sâu, đặc biệt là mạng CNN, đã cho thấy tiềm năng đáng kể trong phân loại thoái hoá khớp gối. Các tác giả trong [9] thực nghiệm trên sáu mô hình CNN (VGG16, VGG19, ResNet101, MobileNetV2, InceptionResNetV2, và DenseNet121) với kết quả báo cáo độ chính xác đạt 69% cho mô hình ResNet101. Kết quả tương tự cho mô hình ResNet152 với độ chính xác 56.75% cũng được báo cáo trong [10]. Một cách nâng cao hiệu suất phân loại KL bằng việc tổng hợp đặc trưng bởi nhiều mô hình CNN (ResNet-34, VGG-19, DenseNet 121, và DenseNet 161) được đề xuất trong [11]. Qua thực nghiệm [9][10][11], các mô hình nhẹ, mới hơn như ResNet đã chứng minh hiệu quả trong nhiều nhiệm vụ, bao gồm cả phân loại hình ảnh y tế.

Bên cạnh đó, khi được tối ưu hóa đúng cách, các mô hình CNN có thể đạt được hiệu suất tốt hơn so với các mô hình CNN gốc. Các tác giả trong [12], đề xuất một mô hình tự động dò tìm đối tượng và phân loại mức độ thoái hoá khớp gối. Trong đó, một mô-đun kết nối đặc trưng ảnh sử dụng Transformer đã được tích hợp trên bộ khung của mô hình ResNet50, nhằm cải thiện thêm độ chính xác 2.5% so với mô hình ResNet50 gốc.

Tương tự mô hình CNN 2D, tiềm năng của các mô hình 3D CNN trong việc khai thác thông tin thể tích từ dữ liệu ảnh chụp cộng hưởng từ để cải thiện độ chính xác chẩn đoán được khảo sát và đánh giá ở [13]. Tuy nhiên, nghiên cứu cũng chỉ ra rằng việc xây dựng các mô hình 3D CNN có độ chính xác cao và hiệu quả tính toán vẫn là một thách thức và là hướng nghiên cứu mở.

## 2.2. Cơ sở dữ liệu thử nghiệm

Bộ dữ liệu được sử dụng trong nghiên cứu này được lấy từ cơ sở dữ liệu công khai Osteoarthritis Initiative (OAI)<sup>1</sup>. Tất cả các ảnh trong bộ dữ liệu đã được tiền xử lý và chuẩn hóa về kích thước  $224 \times 224$  pixel. Bảng 1 thống kê chi tiết số lượng mẫu tương ứng với mức độ thoái hoá có trong bộ cơ sở dữ liệu này.

**Bảng 1. Phân bố dữ liệu trong cơ sở dữ liệu OAI**

| Mức độ thoái hoá | Huấn luyện | Xác thực | Kiểm thử |
|------------------|------------|----------|----------|
| Mức 0            | 2 286      | 328      | 639      |
| Mức 1            | 1 046      | 153      | 296      |
| Mức 2            | 1 516      | 212      | 447      |
| Mức 3            | 757        | 106      | 223      |
| Mức 4            | 173        | 27       | 51       |
| Tổng             | 5 778      | 826      | 1 656    |

Tuy nhiên, kết quả trong Bảng 1 cho thấy bộ dữ liệu OAI tồn tại hiện tượng mất cân bằng lớp rõ rệt giữa các mức độ thoái hóa khớp gối. Trong tập huấn luyện gồm 5 778 ảnh, các mẫu thuộc mức độ thoái hóa nhẹ và trung bình (mức 0, mức 1 và mức 2) chiếm tỷ lệ áp đảo, với lần lượt 2 286, 1 046 và 1 516 ảnh. Ngược lại, số lượng mẫu ở các mức độ nặng giảm mạnh, đặc biệt là mức 4 chỉ có 173 ảnh, chiếm một tỷ lệ rất nhỏ so với tổng số mẫu. Xu hướng phân bố tương tự cũng được quan sát trong tập xác thực và tập kiểm thử, trong đó số lượng ảnh ở mức 4 lần lượt chỉ là 27 và 51 ảnh.

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

Sự chênh lệch đáng kể giữa các lớp này phản ánh đặc điểm thực tế của dữ liệu lâm sàng, nhưng đồng thời đặt ra thách thức lớn cho quá trình huấn luyện mô hình học sâu. Do đó, việc áp dụng các kỹ thuật xử lý mất cân bằng dữ liệu là cần thiết nhằm nâng cao hiệu quả phân loại trên toàn bộ các mức độ bệnh.

### 2.3. Xử lý mất cân bằng dữ liệu

Để xử lý hiện tượng mất cân bằng dữ liệu trong bộ dữ liệu OAI, chúng tôi sử dụng phương pháp gán trọng số lớp dựa trên phân bố mẫu của tập huấn luyện. Phương pháp này đã được chứng minh tính hiệu quả với các mô hình học sâu như trong tài liệu [14][15].

Trọng số của mỗi lớp được tính theo công thức nghịch đảo tần suất, nhằm đảm bảo các lớp có số lượng mẫu ít hơn sẽ có ảnh hưởng lớn hơn trong quá trình tối ưu hàm mất mát. Cụ thể, trọng số của lớp  $i$  được xác định theo công thức dưới đây:

$$w_i = \frac{N}{K \times n_i}$$

Trong đó:

$N$  là tổng số mẫu trong tập huấn luyện,

$K$  là tổng số lớp (trong nghiên cứu này  $K = 5$ ),

$n_i$  là số lượng mẫu thuộc lớp  $i$ .

**Bảng 2. Bảng trọng số cho từng lớp**

| Mức độ thoái hoá | Số mẫu $n_i$ | Trọng số lớp ( $w_i$ ) |
|------------------|--------------|------------------------|
| Mức 0            | 2 286        | 0.51                   |
| Mức 1            | 1 046        | 1.10                   |
| Mức 2            | 1 516        | 0.76                   |
| Mức 3            | 757          | 1.53                   |
| Mức 4            | 173          | 6.67                   |

Bảng 2 trình bày các trọng số được gán cụ thể cho từng lớp. Có thể nhận thấy rằng lớp mức 4 - lớp có số lượng mẫu ít nhất - được gán trọng số cao hơn 13 và 8 lần so với các lớp chiếm ưu thế như mức 0 và mức 2. Các trọng số này được tích hợp trực tiếp vào hàm mục tiêu trong quá trình huấn luyện mô hình CNN, qua đó làm tăng mức phạt đối với các lỗi phân loại thuộc các lớp thiểu số. Việc áp dụng cơ chế trọng số lớp giúp mô hình giảm xu hướng thiên lệch về các lớp phổ biến, đồng thời cải thiện khả năng học và phân biệt các mức độ thoái hóa trung bình và nặng.

### 2.4. Mô hình CNN và độ đo để đánh giá

Để đảm bảo tính nhất quán trong thiết kế thực nghiệm và cho phép so sánh khách quan giữa các kiến trúc CNN, nghiên cứu này lựa chọn các mô hình tiêu biểu, đã được chứng minh hiệu quả trong phân loại ảnh y sinh. Cụ thể, ba kiến trúc phổ biến gồm VGG [16], ResNet [17] và EfficientNet [18] được sử dụng, đại diện cho các hướng tiếp cận khác nhau trong thiết kế mạng nơ-ron tích chập: từ kiến trúc sâu tuần tự, kiến trúc có kết nối dư, đến kiến trúc tối ưu cân bằng giữa độ chính xác và chi phí tính toán. Các phiên bản được lựa chọn là những cấu hình được sử dụng rộng rãi trong nghiên cứu về OA, nhằm đảm bảo tính đại diện và độ tin cậy cho các thí nghiệm so sánh.

Ngoài ra, phương pháp học chuyển giao được áp dụng cho tất cả mô hình. Các mô hình được khởi tạo với trọng số tiền huấn luyện trên tập dữ liệu ImageNet, sau đó tinh chỉnh các tầng phân loại cuối để phù hợp với bài toán gồm năm mức độ thoái hóa.

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. 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.

Các mô hình được huấn luyện trên bộ dữ liệu OAI có kích thước ảnh  $224 \times 224$  pixel với Batch Size bằng 32. Các kỹ thuật tăng cường dữ liệu đã được áp dụng trên tập huấn luyện gồm xoay ảnh ngẫu nhiên ( $20^\circ$ ), dịch chuyển ngang và dọc (20%), phóng to/thu nhỏ (20%) và lật ngang. Quá trình tối ưu sử dụng bộ tối ưu Adam với tốc độ học ban đầu 0,001 và hàm mục tiêu Categorical Cross-entropy cho bài toán phân loại đa lớp. Các mô hình được huấn luyện tối đa 30 epochs, đồng thời áp dụng cơ chế lưu lại mô hình có giá trị AUC trên tập xác thực (val\_auc) cao nhất. Để kiểm soát hiện tượng quá khớp và tối ưu hóa quá trình hội tụ, các cơ chế Early Stopping và ReduceLROnPlateau được áp dụng dựa trên chỉ số AUC của val\_auc. Cụ thể, quá trình huấn luyện được dừng sớm nếu giá trị val\_auc không được cải thiện sau 10 epoch liên tiếp, trong khi tốc độ học được giảm một nửa khi chỉ số này không cải thiện sau 5 epoch liên tiếp.

Các chỉ số Accuracy, F<sub>1</sub>-score, Precision, Recall và macro-F<sub>1</sub> được sử dụng nhằm đánh giá hiệu năng của các mô hình. Độ chính xác (ACC) và F<sub>1</sub>-score là các chỉ số phổ biến để đánh giá mô hình phân loại. Tuy nhiên, độ chính xác chỉ thực sự có ý nghĩa khi dữ liệu cân bằng, vì với dữ liệu mất cân bằng (như đã phân tích ở Phần 2), giá trị ACC cao có thể chủ yếu đến từ lớp chiếm ưu thế, trong khi các lớp ít mẫu bị xem nhẹ. Do đó, F<sub>1</sub>-score được sử dụng như một thước đo bổ sung quan trọng, vì đây là trung bình điều hòa giữa Precision và Recall. Bên cạnh đó, macro-F<sub>1</sub> được sử dụng để đánh giá hiệu năng tổng thể của mô hình trên tất cả các lớp một cách công bằng hơn trong bối cảnh dữ liệu mất cân bằng.

### **3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN**

Trong mục này, chúng tôi phân tích các kết quả thực nghiệm nhằm đánh giá hiệu quả của phương pháp đề xuất trong bài toán phân loại mức độ thoái hóa khớp gối. Trước hết, ảnh hưởng của hiện tượng mất cân bằng dữ liệu đến hiệu năng mô hình được xem xét. Tiếp theo, kết quả phân loại của các kiến trúc CNN khác nhau được so sánh nhằm làm rõ vai trò của từng mô hình và mức độ cải thiện đạt được khi áp dụng các chiến lược xử lý dữ liệu và học chuyển giao. Cuối cùng, chúng tôi so sánh hiệu suất của phương pháp đề xuất với một số nghiên cứu điển hình gần đây cùng trên bộ dữ liệu OAI.

#### **3.1. Kết quả ảnh hưởng của việc mất cân bằng dữ liệu**

Để đánh giá ảnh hưởng của hiện tượng mất cân bằng dữ liệu, mô hình ResNet50 được lựa chọn là mô hình chính để triển khai với chiến lược học chuyển giao nhằm phân loại mức độ thoái hóa khớp gối từ ảnh X-quang.

Bảng 3 cho thấy kết quả huấn luyện mô hình trên dữ liệu chưa được xử lý mất cân bằng. Có thể nhận thấy rằng hiệu năng phân loại thấp và không đồng đều giữa các lớp. Độ chính xác tổng thể ACC chỉ đạt 0.25, phản ánh khả năng dự đoán kém của mô hình. Đặc biệt, các lớp có số lượng mẫu ít như mức 3 và mức 4 gần như không được mô hình nhận diện, với Recall lần lượt chỉ đạt 0.03 và 0.04, dẫn đến  $F_1$ -score xấp xỉ bằng 0. Điều này cho thấy mô hình có xu hướng thiên lệch nghiêm trọng về một số lớp nhất định và không học được các đặc trưng có ý nghĩa cho các mức độ thoái hóa trung bình và nặng.

**Bảng 3. Hiệu năng của ResNet50 với dữ liệu mất cân bằng**

| Mức độ thoái hoá   | Precision | Recall | $F_1$ -score |
|--------------------|-----------|--------|--------------|
| Mức 0              | 0.34      | 0.21   | 0.26         |
| Mức 1              | 0.30      | 0.01   | 0.02         |
| Mức 2              | 0.27      | 0.57   | 0.37         |
| Mức 3              | 0.08      | 0.03   | 0.04         |
| Mức 4              | 0.00      | 0.04   | 0.01         |
| Độ chính xác (ACC) | 0.25      |        |              |

Ngược lại, khi áp dụng chiến lược gán trọng số lớp, hiệu năng của mô hình được cải thiện rõ rệt trên tất cả các chỉ số đánh giá (kết quả được trình bày trong Bảng 4). Độ chính xác tăng mạnh từ 0.25 lên 0.66, đồng thời Precision, Recall và  $F_1$ -score ở các lớp thiểu số (mức 3 và mức 4) đều đạt giá trị cao, với  $F_1$ -score lần lượt là 0.82 và 0.77. Sự cải thiện đồng đều giữa các mức độ cho thấy mô hình đã học được các đặc trưng phân biệt hiệu quả hơn và giảm đáng kể hiện tượng thiên lệch về các lớp chiếm ưu thế. Kết quả này khẳng định vai trò quan trọng của việc xử lý mất cân bằng dữ liệu bằng trọng số lớp trong các bài toán phân loại ảnh y sinh, đặc biệt đối với các mức độ bệnh có ý nghĩa lâm sàng cao.

**Bảng 4. Hiệu năng của ResNet50 với dữ liệu có trọng số lớp**

| Mức độ thoái hoá   | Precision | Recall | $F_1$ -score |
|--------------------|-----------|--------|--------------|
| Mức 0              | 0.74      | 0.79   | 0.77         |
| Mức 1              | 0.34      | 0.37   | 0.35         |
| Mức 2              | 0.71      | 0.57   | 0.64         |
| Mức 3              | 0.81      | 0.83   | 0.82         |
| Mức 4              | 0.69      | 0.89   | 0.77         |
| Độ chính xác (ACC) | 0.66      |        |              |

Tuy nhiên kết quả Bảng 4 cũng cho thấy ResNet50 gặp nhiều khó khăn trong việc phân loại các mức độ thoái hóa khớp gối ở giai đoạn sớm và trung gian, đặc biệt tại mức 1 với  $F_1$ -score chỉ đạt 0,35. Điều này cho thấy sự chông lán về đặc trưng hình ảnh giữa các mức độ thoái hóa vẫn là thách thức lớn đối với mô hình trong quá trình phân loại.

### 3.2. Kết quả với các mô hình CNN khác

Sau khi áp dụng xử lý mất cân bằng dữ liệu với trọng số lớp trong quá trình huấn luyện, nghiên cứu tiếp hành đánh giá hiệu năng của các mô hình CNN khác nhau nhằm đảm bảo sự so sánh khách quan. Ba kiến trúc tiêu biểu gồm VGG16, ResNet50 và EfficientNet-B0 được lựa chọn để thực nghiệm. Các kết quả được trình bày trong Bảng 5.

**Bảng 5. Hiệu năng của các mô hình CNN**

| Mô hình         | ACC   | Precision | Recall | macro-F <sub>1</sub> |
|-----------------|-------|-----------|--------|----------------------|
| VGG16           | 0.604 | 0.650     | 0.510  | 0.457                |
| ResNet50        | 0.664 | 0.668     | 0.664  | 0.667                |
| EfficientNet-B0 | 0.667 | 0.670     | 0.667  | 0.680                |

Kết quả trong Bảng 5 cho thấy EfficientNet-B0 đạt hiệu năng cao nhất với ACC = 0.667 và macro-F<sub>1</sub> = 0.680, tiếp theo là ResNet50 (ACC = 0.664, macro-F<sub>1</sub> = 0.667), trong khi VGG16 cho kết quả thấp hơn (ACC = 0.604, macro-F<sub>1</sub> = 0.457). Kết quả cho thấy các kiến trúc CNN hiện đại, đặc biệt là EfficientNet-B0 và ResNet50, đạt hiệu năng phân loại cao hơn so với VGG16 trên bộ dữ liệu nghiên cứu.

### 3.3. Kết quả so sánh với các phương pháp khác

Phương pháp đề xuất được so sánh với các nghiên cứu gần đây [9], [10], [12], kết quả được trình bày trong Bảng 6.

**Bảng 6. So sánh hiệu suất với các phương pháp khác**

| Phương pháp (năm) | Mô hình CNN     | ACC  | Precision | Recall | macro-F <sub>1</sub> |
|-------------------|-----------------|------|-----------|--------|----------------------|
| [9] (2023)        | ResNet101       | 0.69 | 0.67      | 0.67   | 0.65                 |
| [10] (2023)       | ResNet101       | 0.57 | -         | -      | -                    |
| [12] (2021)       | YOLO+ResNet50   | 0.69 | -         | -      | -                    |
| Chúng tôi (2026)  | ResNet50        | 0.66 | 0.67      | 0.66   | 0.67                 |
| Chúng tôi (2026)  | EfficientNet-B0 | 0.66 | 0.67      | 0.67   | 0.68                 |

Bảng 6 cho thấy hiệu năng của phương pháp đề xuất đạt hiệu suất tương đương với các nghiên cứu trước đây được báo cáo trên cùng bài toán. Cụ thể, nghiên cứu [9] (2023) sử dụng mô hình ResNet101 đạt ACC 0.69 với Precision và Recall đều đạt 0.67, trong khi nghiên cứu [10] (2023) cũng sử dụng ResNet101 nhưng chỉ đạt ACC 0.57. Phương pháp kết hợp YOLO và ResNet50 trong [12] (2021) đạt độ chính xác 0.69. Trong nghiên cứu này, mô hình EfficientNet-B0 đạt ACC 0.66 cùng với Precision và Recall đều đạt 0.67, và macro-F<sub>1</sub> đạt 0.68. Mặc dù độ chính xác thấp hơn so với một số nghiên cứu trước, phương pháp đề xuất vẫn đạt macro-F<sub>1</sub> cao và duy trì sự cân bằng giữa Precision và Recall.

## 4. KẾT LUẬN

Nghiên cứu đã chứng minh rằng việc áp dụng mạng nơ-ron tích chập kết hợp với chiến lược xử lý mất cân bằng dữ liệu là hướng tiếp cận hiệu quả trong phân loại mức độ thoái hóa khớp gối từ ảnh X-quang. Kết quả thực nghiệm trên bộ dữ liệu OAI cho thấy dữ liệu mất cân bằng ảnh hưởng nghiêm trọng đến hiệu năng mô hình, đặc biệt ở các mức độ thoái hóa trung bình và nặng; tuy nhiên, khi áp dụng phương pháp gán trọng số lớp, hiệu năng phân loại được cải thiện đáng kể và đồng đều hơn giữa các lớp. So sánh các kiến trúc CNN cho thấy ResNet50 và EfficientNet-B0 vượt trội so với VGG16, trong đó EfficientNet-B0 đạt hiệu năng tổng thể tốt nhất với macro-F<sub>1</sub> = 68%. Các kết quả đạt được khẳng định tiềm năng ứng dụng của mô hình CNN trong hỗ trợ chẩn đoán và đánh giá mức độ thoái hóa khớp gối trong thực tiễn lâm sàng.

TÀI LIỆU THAM KHẢO

- [1] Y. Zhang and J. M. Jordan, "Epidemiology of osteoarthritis," *Clinics in Geriatric Medicine*, 26(3):355-369, 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", *British Journal of Sports Medicine*, 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 and Rheumatism*, 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 and 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," *Journal of Pharmacy and Bioallied Sciences*, 5(1):30-38, 2013, DOI:10.4103/0975-7406.106561.
- [6] S. Rani et al., "Deep learning to combat knee osteoarthritis and severity assessment by using CNN-based classification," *BMC Musculoskeletal Disorders*, vol.25, pp.817, 2024, DOI:10.1186/s12891-024-07942-9.
- [7] S. Iqbal, D. Ali, S. Rafique, K. Bagga, A. U. Rahman, and S. Khan, "Comparative analysis of automated knee osteoarthritis severity classification from X-ray images using CNNs and VGG16 architecture," *ICCK Transactions on Sensing, Communication, and Control*, 2(1):36-47, 2025, DOI:10.62762/TSCC.2025.378503.
- [8] J. H. Kellgren and J. Lawrence, "Radiological assessment of osteo-arthrosis", *Annals of the Rheumatic Diseases*, 16(4):494-502, 1957, DOI:10.1136/ard.16.4.494.
- [9] A. S. Mohammed, A. A. Hasanaath, G. Latif, and A. Bashar, "Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed X-ray images," *Diagnostics*, 13(8), pp.1380, 2023, DOI:10.3390/diagnostics13081380.
- [10] A. Rachmad, F. Sonata, J. Hutagalung, D. Hapsari, M. Fuad, and E. M. Sari Rochman, "An automated system for osteoarthritis severity scoring using residual neural networks," *Mathematical Modelling of Engineering Problems*, 10(5):1849-1856, 2023, DOI:10.18280/mmep.100538.
- [11] T. Tariq, Z. Suhail, and Z. Nawaz, "Knee osteoarthritis detection and classification using X-rays," *IEEE Access*, vol.11, pp.48292-48303, 2023, DOI:10.1109/ACCESS.2023.3276810.

- [12] Y. Wang, X. Wang, T. Gao, L. Du, and W. Liu, "An automatic knee osteoarthritis diagnosis method based on deep learning: Data from the Osteoarthritis Initiative," *Journal of Healthcare Engineering*, vol.2021, pp.5586529, DOI:10.1155/2021/5586529.
- [13] P. S. Q. Yeoh et al., "Emergence of deep learning in knee osteoarthritis diagnosis," *Computational Intelligence and Neuroscience*, vol.2021, pp.4931437, DOI:10.1155/2021/4931437.
- [14] J. Byrd and Z. Lipton, "What is the effect of importance weighting in deep learning?," in *International Conference on Machine Learning*, Long Beach, CA, USA, 2019, pp.872-881.
- [15] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, pp.27, 2019, DOI:10.1186/s40537-019-0192-5.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014, DOI:10.48550/arXiv.1409.1556.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp.770-778.
- [18] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, Long Beach, CA, USA, 2019, pp.6105-6114.

**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.