PHÁT HIỆN VÀ PHÂN LOẠI THOÁI HÓA KHỚP SỬ DỤNG VISION TRANSFORMER TRONG MÔI TRƯỜNG PHÂN TÁN

DETECTION AND CLASSIFICATION OF OATHROSIS USING VISION TRANSFORMER IN A DISTRIBUTED ENVIRONMENT

PHAN THƯỢNG CANG[1], PHAN ANH CANG[2], ĐẶNG NHƯ BÁCH[3a]
[1] *Trường Công nghệ Thông tin và Truyền thông - Đại học Cần Thơ*
[2] *Trường Đại học Sư Phạm Kỹ Thuật Vĩnh Long*
[3] *Học viên cao học, Trường Đại học Sư Phạm Kỹ Thuật Vĩnh Long*
[a] *Tác giả liên hệ: bach21041995@gmail.com*

Nhận bài (Received): 25/4/2024; Phản biện (Reviewed):13/5/2024; Chấp nhận (Accepted):16/5/2024

TÓM TẮT

Thoái hóa khớp gối là căn bệnh mãn tính chiếm tỷ lệ không nhỏ trong các bệnh lý xương khớp, gây đau đớn kéo dài, giảm và mất khả năng vận động, có thể gây tàn phế cao nhất hiện nay. Việc phát hiện sớm tình trạng thoái hóa khớp giúp kiểm soát những cơn đau và hạn chế quá trình tiến triển của bệnh khi có biểu hiện thoái hóa khớp gối. Trong bài báo này, chúng tôi đề xuất kỹ thuật phân loại dựa trên kiến trúc mạng Vision Transformer trên môi trường phân tán để phân loại thoái hoá khớp gối dựa trên ảnh X-Quang. Qua kết quả thực nghiệm cho thấy độ chính xác của mô hình đề xuất lên đến 98.85% và cải thiện thời gian huấn luyện gấp đôi so với mô hình huấn luyện trên môi trường cục bộ.

Từ khóa: Vision Transformer, Môi trường phân tán

ABSTRACT

Knee osteoarthritis is a chronic disease that affects a large proportion of joint diseases, causing prolonged pain, reduced and loss of mobility, and can cause the highest freckles today. Early detection of osteoarthritis helps control pain and limit the progression of the disease when knee chemotherapy occurs. In this paper, we propose to publish a classification technique based on Vision Transformer network architecture on a distributed environment for pillow match classification based on X-ray images. Actual results show that the accuracy of the published model is up to 98.85% and improves training time by twice as much as the model trained in the local environment.

Keywords: *Vision Transformer, Distributed environment*

1. MỞ ĐẦU

1.1. Giới thiệu bài toán

Có thể nói nguyên nhân chính gây ra tình trạng thoái hóa khớp gối chính là do tuổi tác. Bệnh thường gặp ở những người trên 55 tuổi. Nguyên nhân do tuổi càng cao thì quá trình tổng hợp của sụn lại càng có xu hướng suy giảm. Vì thế các tế bào gần bị bào mòn và không được tái tạo lại khiến sụn bị thoái hóa nhanh chóng. Tuy nhiên, ngày nay tình trạng thoái hóa khớp

trong đó có khớp gối trẻ ngày càng trẻ hóa. Thống kê cho thấy, chỉ có 10% tình trạng thoái hóa khớp gối xảy ra ở người dưới 26 tuổi. Nhưng từ tuổi 27- 45 là 25,5% và độ tuổi 46 – 60 thì tỉ lệ này lại lên tới 50%. Cần xây dựng một giải pháp để giải quyết bài toán dò tìm và phân loại các dạng thoái hoá khớp gối qua ảnh X-Quang trên môi trường phân tán. Kết quả thu được cho độ chính xác không thay đổi nhiều và cải thiện thời gian chuẩn đoán bệnh, có thể phát triển thành ứng dụng thực tế trong tương lai, hỗ trợ cho Bác sĩ phát hiện sớm bệnh để có giải pháp điều trị kịp thời cho bệnh nhân.

1.2. Những nghiên cứu liên quan

Pingjun Chen và nhóm nghiên cứu [4] lần lượt áp dụng hai mạng lưới thần kinh tích chập sâu (CNN) để tự động đo mức độ nghiêm trọng của viêm khớp gối, được đánh giá bằng hệ thống phân loại Kellgren-Lawrence (KL).

Bochen Guan và nhóm nghiên cứu [6] phát triển và đánh giá các mô hình đánh giá rủi ro học sâu (DL) để dự đoán tiến triển cơn đau ở những đối tượng có hoặc có nguy cơ bị viêm xương khớp đầu gối (OA).

Bin Liu và các cộng sự [7] giới thiệu một mô hình chẩn đoán tự động viêm khớp gối dựa trên phương pháp học sâu end-to-end.

Pauline Shan Qing Yeoh và nhóm nghiên cứu [8] cung cấp một cái nhìn tổng quan bao quát về các phương pháp tiếp cận CNN 2D và 3D hiện tại trong lĩnh vực nghiên cứu viêm khớp.

Yifan Wang và các cộng sự [9] tích hợp mô hình phát hiện đối tượng, YOLO, với bộ biến đổi hình ảnh vào quy trình chẩn đoán, giảm sự can thiệp của con người và cung cấp phương pháp tiếp cận toàn diện để chẩn đoán viêm xương khớp tự động.

Albert Swiecicki và các cộng sự [10] giới thiệu một thuật toán học sâu hoàn toàn tự động phù hợp với hiệu suất của các bác sĩ X quang trong việc đánh giá mức độ nghiêm trọng của viêm xương khớp đầu gối trong ảnh chụp X quang bằng hệ thống phân loại Kellgren-Lawrence.

Simon Olsson và nhóm nghiên cứu [11] đánh giá mức độ AI có thể phân loại mức độ nghiêm trọng của viêm khớp gối, sử dụng toàn bộ loạt hình ảnh và không loại trừ các rối loạn thị giác phổ biến như cấy ghép, bó bột và các bệnh lý không thoái hóa.

S Sheik Abdullah và các cộng sự [12] đã phát triển một công cụ để xác định và phân loại viêm xương khớp đầu gối (OA) từ hình ảnh X-quang kỹ thuật số và minh họa khả năng sử dụng các kỹ thuật học sâu để dự đoán viêm khớp gối theo hệ thống phân loại Kellgren-Lawrence (KL).

Joseph Humberto Cueva và nhóm nghiên cứu [13] đề xuất mô hình CADx bán tự động dựa trên mạng thần kinh tích chập Deep Siamese và ResNet-34 được tinh chỉnh để phát hiện đồng thời các tổn thương viêm khớp ở hai đầu gối theo thang đo KL.

Abdul Sami Mohammed và các nhà nghiên cứu [14] trong miền ML/DL đã sử dụng khả năng của các mô hình mạng thần kinh sâu (DNN) để xác định và phân loại hình ảnh KOA một cách tự động, nhanh hơn và chính xác hơn.

Chen Pingjun và các cộng sự [18] đã tập hợp được Bộ dữ liệu phân loại mức độ nghiêm trọng của bệnh viêm xương khớp đầu gối năm 2018. Tập dữ liệu này chứa dữ liệu X-quang đầu gối để phát hiện khớp gối và phân loại KL đầu gối.

Alexey Dosovitskiy và cộng sự [5] tại Google Research có bài nghiên cứu về "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" từ đó đưa ra Kỹ thuật Vision Transformer (ViT). Mô hình này có thể được huấn luyện trên các

tập dữ liệu lớn như ImageNet và đã cho thấy hiệu suất và độ chính xác cao, vượt qua nhiều mô hình CNN truyền thống trong nhiều tác vụ nhận diện hình ảnh.

Dựa vào các nghiên cứu trên, có thể thấy được các nghiên cứu hiện tại cho độ chính xác phân loại chưa cao và chưa thực sự có được mô hình tối ưu nhất cho việc phát hiện và phân loại bệnh thoái hoá khớp. Những nghiên cứu trên chỉ dừng lại ở việc huấn luyện mô hình ở một máy tính cục bộ, nên việc nghiên cứu phát hiện và phân loại thoái hoá khớp qua ảnh X-Quang trên môi trường phân tán nhiều máy và trên tập dữ liệu lớn hơn, cần đi sâu để tìm hiểu và thực nghiệm, từ đó đưa ra được các đánh giá, so sánh giữa các phương pháp với nhau. Đây là một định hướng nghiên cứu phù hợp với xu thế nghiên cứu chung của thế giới, mang tính cấp thiết cao và có khả năng ứng dụng hiệu quả trong thực tiễn.

1.3. Đặc điểm thoái hoá khớp gối

Trong bài báo này, chúng tôi sử dụng phương pháp phân loại thoái hoá khớp dựa theo Kellgren và Lawrence [14], phương pháp này áp dụng cho đánh giá tổn thương sụn khớp trên phim X-quang. Phương pháp phân loại thoái hoá khớp theo Kellgren và Lawrence là một phương pháp đơn giản và phổ biến được sử dụng để đánh giá mức độ nghiêm trọng của thoái hoá khớp dựa trên hình ảnh chụp X-quang. Phương pháp này đã được giới thiệu lần đầu tiên vào năm 1957 bởi J.H. Kellgren và J.S. Lawrence. Dưới đây là mô tả tổng quát về phương pháp này:

Bình thường (Healthy): không thấy tổn thương khớp.

Giai đoạn 1 (Doubtful): khe khớp gần như bình thường, có thể có gai xương nhỏ.

Giai đoạn 2 (Minimal): khe khớp hẹp nhẹ, có gai xương nhỏ.

Giai đoạn 3 (Moderate): khe khớp hẹp rõ, có nhiều gai xương kích thước vừa, vài chỗ đặc xương dưới sụn, có thể có biến dạng đầu xương.

Giai đoạn 4 (Severe): khe khớp hẹp nhiều, gai xương kích thước lớn, đặc xương dưới sụn, biến dạng rõ đầu xương.

Hình 1. Các mức độ phân loại bệnh

1.4. Các mạng nơ-ron đề xuất

a) Mạng Vision Transformer: là một kiến trúc mạng nơ-ron sâu tiên tiến dùng để xử lý hình ảnh, đặc biệt là trong lĩnh vực thị giác máy tính. Điều đặc biệt về Vision Transformer là nó không sử dụng các kiến trúc mạng nơ-ron tích chập (CNN) truyền thống như ResNet hay DenseNet, mà thay vào đó, nó sử dụng cơ chế transformer được giới thiệu ban đầu cho việc xử lý ngôn ngữ tự nhiên. Đặc điểm quan trọng của mạng Vision Transformer là sự sử dụng các "patch embeddings", trong đó hình ảnh được chia thành các miếng nhỏ (patch) và mỗi patch được biểu diễn bằng một vector embedding. Những embedding này sau đó được đưa vào một cấu trúc transformer để trích xuất thông tin đặc trưng từ hình ảnh. Điều này giúp mô hình nhận biết các đặc trưng không gian trong ảnh và mối quan hệ giữa các patch; quá trình xử lý thông tin được thực hiện dựa trên các phép chuyển đổi tuyến tính và cơ chế tự chú ý mà không yêu cầu các phép tích chập như các mạng

CNN. Đó cũng là lý do chúng tôi chọn mô hình này để thực nghiệm.

b) Mạng ResNet152: là một kiến trúc mạng nơ-ron sâu (deep neural network) thuộc loại mạng học sâu (deep learning). Nó là một biến thể của mạng Residual Network (ResNet), được phát triển bởi các nhà nghiên cứu tại Microsoft Research. Đặc điểm nổi bật của ResNet là sử dụng các khối xây dựng gọi là "residual blocks", trong đó thông tin "còn lại" (residual) từ một lớp được truyền qua một đường đi tắt (shortcut) và được thêm vào lớp đầu ra của khối. Điều này giúp giảm hiện tượng biến mất đạo hàm (vanishing gradient) và cải thiện khả năng học của mạng nơ-ron sâu, cũng chính vì vậy phương pháp này cũng là một lựa chọn của chúng tôi trong mô hình thực nghiệm.

c) Mạng DenseNet121: là một kiến trúc mạng nơ-ron sâu khác, thuộc loại mạng Dense Convolutional Network (DenseNet). DenseNet là một sự tiến bộ từ ResNet, trong đó mỗi lớp nhận đầu vào từ tất cả các lớp trước đó, thay vì chỉ từ các lớp gần nhất như trong ResNet. Đặc điểm chính của DenseNet là sự kết nối mật thiết giữa các lớp, trong đó mỗi lớp nhận đầu vào từ tất cả các lớp trước đó thông qua việc nối chuỗi đầu vào của chúng lại với nhau. Điều này giúp làm tăng sự truyền tải thông tin trong mạng và giảm thiểu số lượng tham số cần học, làm cho mạng trở nên hiệu quả hơn, vì thế chúng tôi chọn DenseNet121 vào mô hình thực nghiệm.

1.5. Xử lý song song và phân tán với Apache Spark và thư viện BigDL

Apache Spark (gọi tắt là Spark) là một công cụ xử lý dữ liệu song song và phân tán có khả năng mở rộng, nhanh so với nhiều khung xử lý dữ liệu khác. Nó bao gồm một số thư viện để giúp xây dựng các ứng dụng cho máy học (MLlib), xử lý luồng (Spark Streaming) và xử lý đồ thị (GraphX),.... Apache Spark dựa trên ý tưởng của mô hình MapReduce nhưng khác biệt chính với Hadoop ở cách tiếp cận xử lý: Apache Spark có thể tính toán và ghi dữ liệu tạm thời ở bộ nhớ trong, trong khi Hadoop phải đọc và ghi vào bộ nhớ ngoài. Do đó, tốc độ xử lý của Apache Spark có thể nhanh hơn gấp 10 đến 100 so với Hahoop. Nó khắc phục được một số hạn chế của Hadoop như: các phép toán đều thực hiện trên đĩa cứng thời gian chậm, không hỗ trợ thời gian thực; hỗ trợ nhiều thư viện MLlib, BigDL,...

BigDL là một thư viện mã nguồn mở được phát triển bởi công ty Intel, nhằm hỗ trợ việc triển khai và huấn luyện mô hình học sâu trên Apache Spark. BigDL cho phép người dùng sử dụng các mô hình học sâu lớn trên cùng một cơ sở hạ tầng phân tán mà Apache Spark cung cấp. BigDL làm cho việc triển khai và huấn luyện mô hình học sâu trên Apache Spark trở nên dễ dàng.

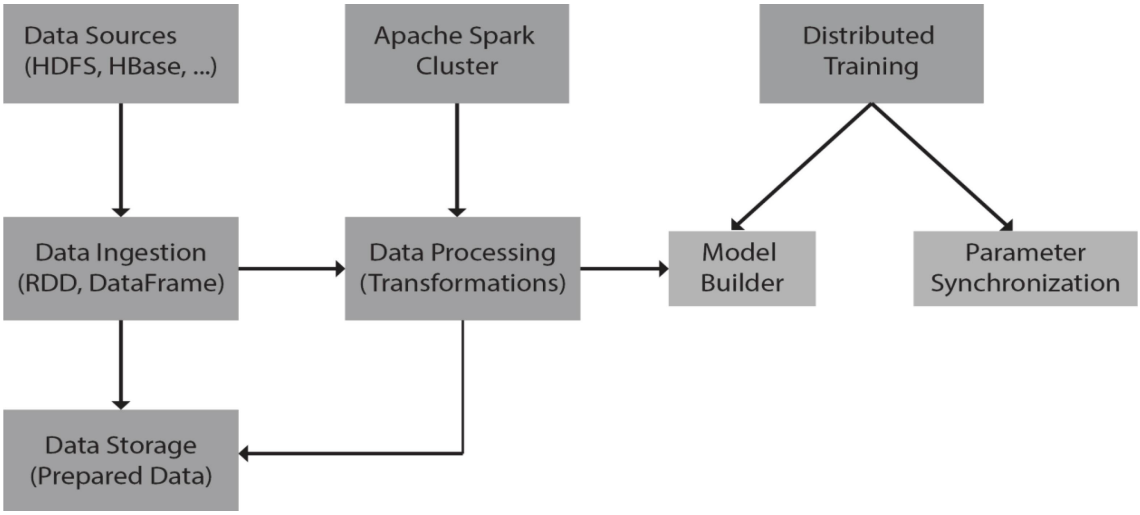
Tận dụng những ưu điểm trên nên trong nghiên cứu này chúng tôi cài đặt và sử dụng thư viện BigDL trên môi trường Apache Spark để huấn luyện, so sánh về thời gian và độ chính xác giữa hai kịch bản đề ra. Để minh họa kiến trúc và mô hình của BigDL trên Apache Spark bằng hình ảnh, chúng ta có thể hình dung các thành phần như sau:

Data Ingestion: Dữ liệu từ các nguồn được chuyển đổi thành các đối tượng RDD hoặc DataFrame của Spark. Đây là bước đầu tiên trong pipeline dữ liệu của Spark.

Apache Spark Cluster: Đảm bảo việc phân phối dữ liệu và tính toán trên các nút trong cluster.

Distributed Training: Quá trình huấn luyện phân tán với sự hỗ trợ của Spark giúp tăng tốc độ và hiệu suất của các tác vụ học sâu.

Data Sources và Data Storage: Đảm bảo dữ liệu được nạp vào và lưu trữ một cách hiệu quả.

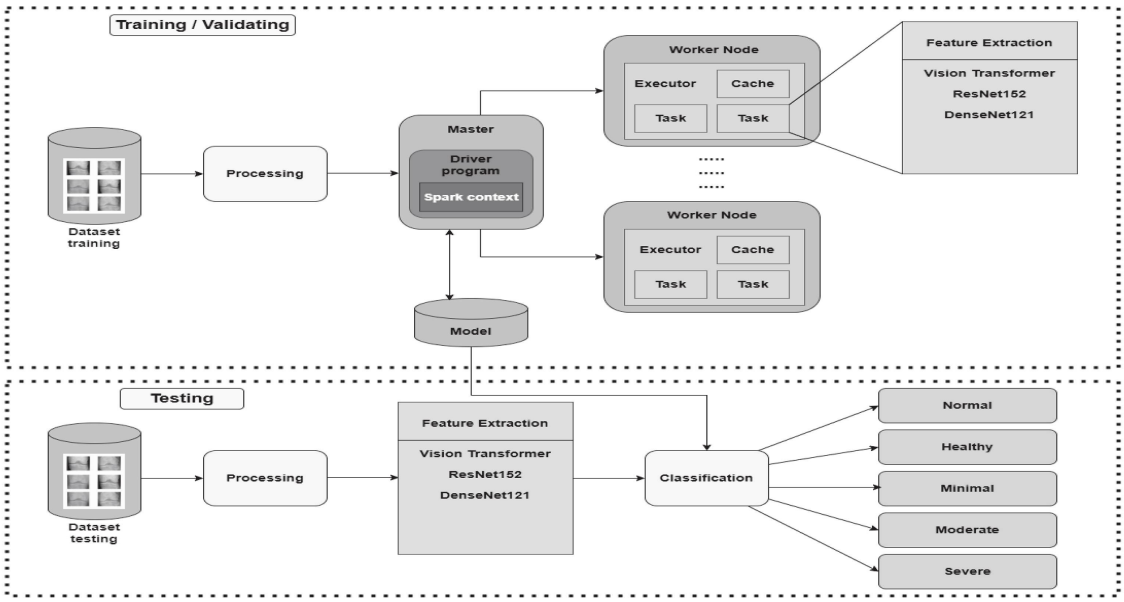


Hình 2. Kiến trúc và mô hình của BigDL trên Apache Spark

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mô hình đề xuất tổng quát

Mô hình tổng quát được đề xuất để giải quyết cho bài toán phát hiện và phân loại gồm có 2 pha, pha huấn luyện và pha kiểm thử. Đối với pha Huấn luyện tập dữ liệu đầu vào được chúng tôi tiền xử lý và đưa vào Spark Master phân chia ngẫu nhiên cho các Worker xử lý rút trích đặc trưng và được Spark Master tổng hợp. Và Model của pha Huấn luyện được pha Kiểm thử sử dụng trong bước Classification để đưa ra kết quả chuẩn đoán.

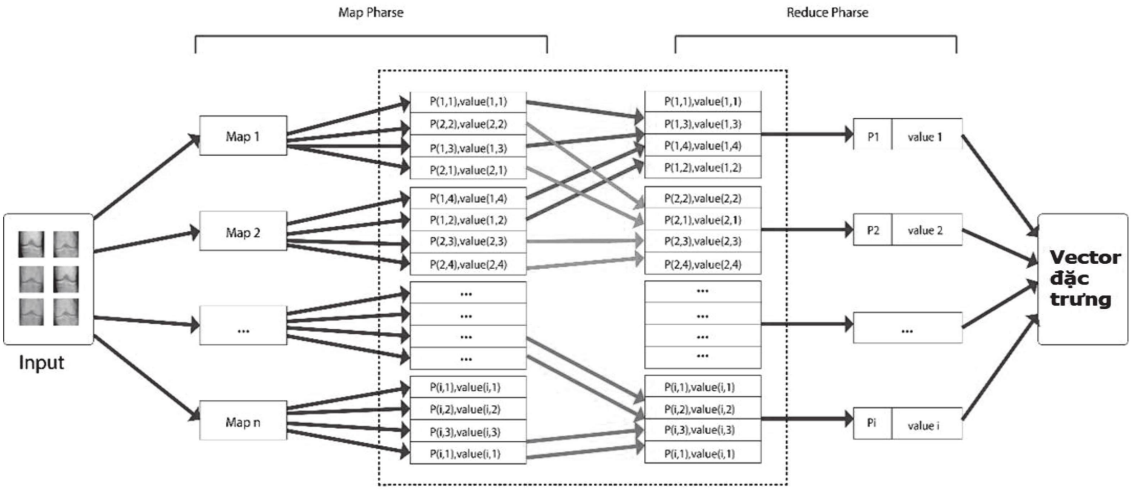


Hình 3. Mô hình tổng quát

2.2. Xử lý giai đoạn rút trích đặc trưng

Apache Spark hỗ trợ một loạt các công cụ và thư viện để thực hiện rút trích đặc trưng trên dữ liệu phân tán một cách hiệu quả trong đó phải kể đến thư viện. Apache

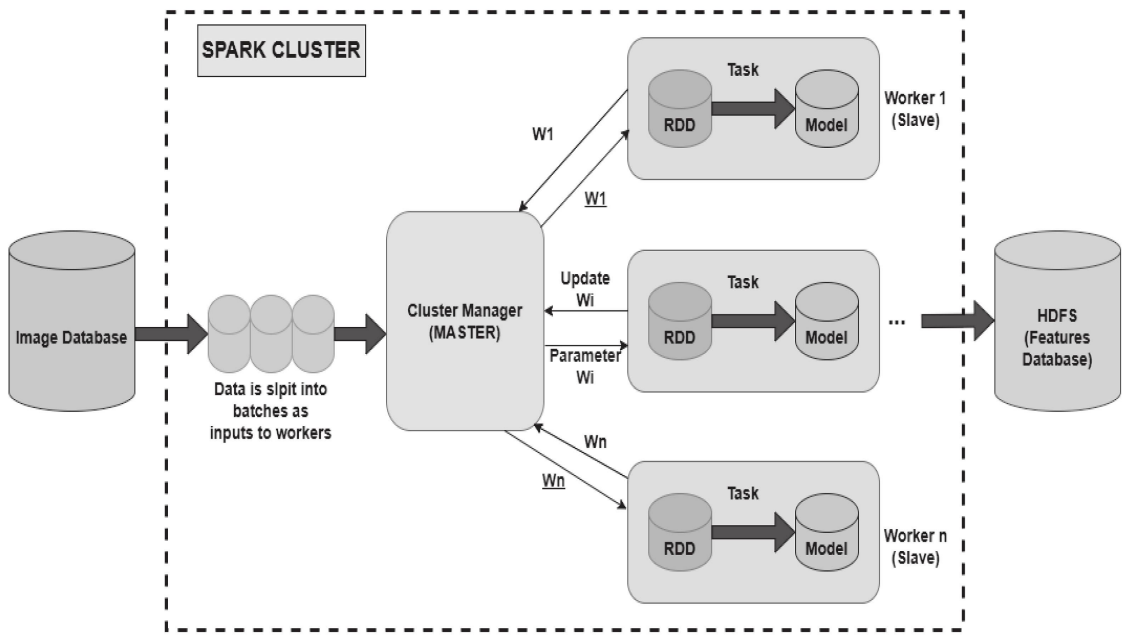
Spark tự động phân tán dữ liệu và phân phối quá trình xử lý và tính toán trên nhiều node trong một cụm máy tính. Điều này giúp tăng tốc quá trình rút trích đặc trưng trên dữ liệu lớn bằng cách sử dụng nhiều tài nguyên tính toán song song.



Hình 4. Mô hình rút trích đặc trưng

2.3. Xử lý giai đoạn huấn luyện

Trong giai đoạn huấn luyện, tập dữ liệu được phân chia ngẫu nhiên cho các worker với W1,...,Wn là các trọng số mạng được liên kết với từng worker 1,..., worker n. W là trung bình trọng số (Global Parameter Vecto W) được thu thập bởi các worker tại máy chủ tham số. Sau khi việc tính trung bình các trọng số được thực hiện các worker 1,..., worker n sẽ được phân bổ W1,..., Wn tương ứng với phần cục bộ mạng. Mỗi worker sẽ thực thi huấn luyện cùng lúc song song nhau. Việc cập nhật trọng số làm giảm thời gian huấn luyện cho mô hình.



Hình 5. Mô hình Apache Spark

3. KẾT QUẢ NGHIÊN CỨU

3.1. Môi trường cài đặt

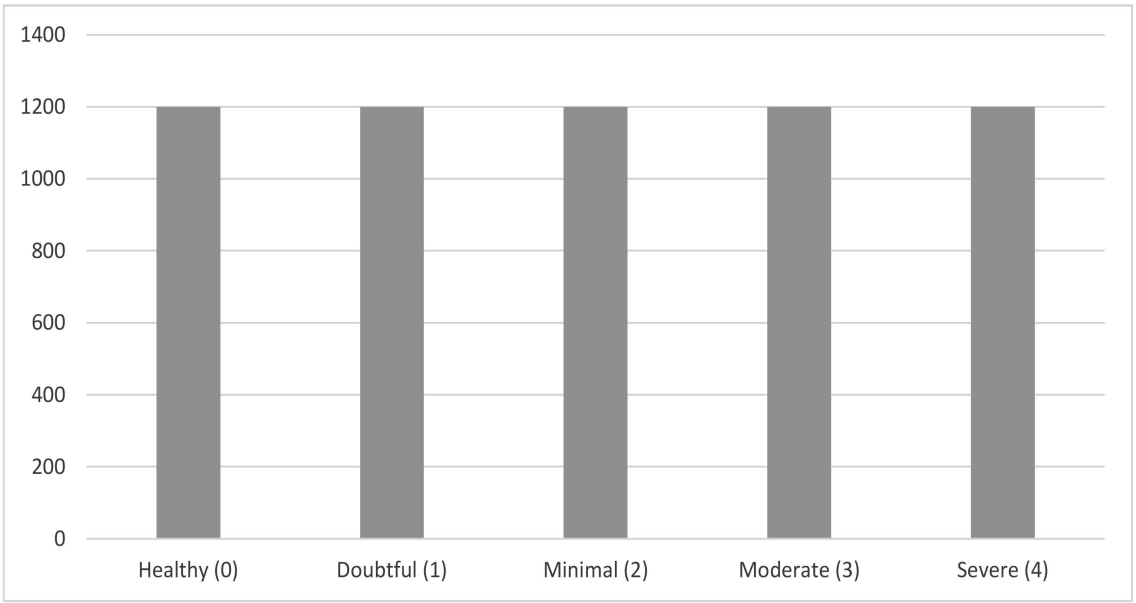
Tiến hành thực hiện cài đặt và huấn luyện trên hai môi trường: cục bộ (Local) ***và*** phân tán (Cluster). Với mỗi máy sẽ có cấu hình RAM 4gb, CPU (4 nhân), hệ điều hành Ubuntu 21.04 LTS x86_64, cùng với các thư viện, và môi trường hỗ trợ: Java JDK 8, python 3.7, tensorflow 2.11, và Apache Spark 2.7. Ở môi trường cục bộ được cấu hình với một máy và môi trường phân tán được cấu hình với số lượng máy là bốn.

Bảng 1. Cấu hình và cài đặt môi trường

Môi trường	Cấu hình	Hệ điều hành	Cài đặt	Số lượng
Cục bộ (Local)	RAM: 4gb CPU: 4 nhân	Ubuntu 21.04 LTS x86_64	Java JDK 8, Python 3.7, Tensorflow 2.11, Apache Spark 2.7	1
Phân tán (Cluster)	RAM: 4gb CPU: 4 nhân	Ubuntu 21.04 LTS x86_64	Java JDK 8, Python 3.7, Tensorflow 2.11 , Apache Spark 2.7	4

3.2. Dữ liệu thực nghiệm

Để có thể thấy được sự khác nhau giữa hai môi trường cục bộ (Local) và phân tán (Cluster) cần phải sử dụng tập dữ liệu có số lượng ảnh lớn. Do đó tập dữ liệu được thực hiện là tập dữ liệu Knee Osteoarthritis Dataset with Severity Grading có 6.000 ảnh đã được tăng cường để mật độ ảnh giữa các lớp là 1.200 ảnh.



Hình 6. Hình ảnh minh hoạ số lượng ảnh giữa các lớp

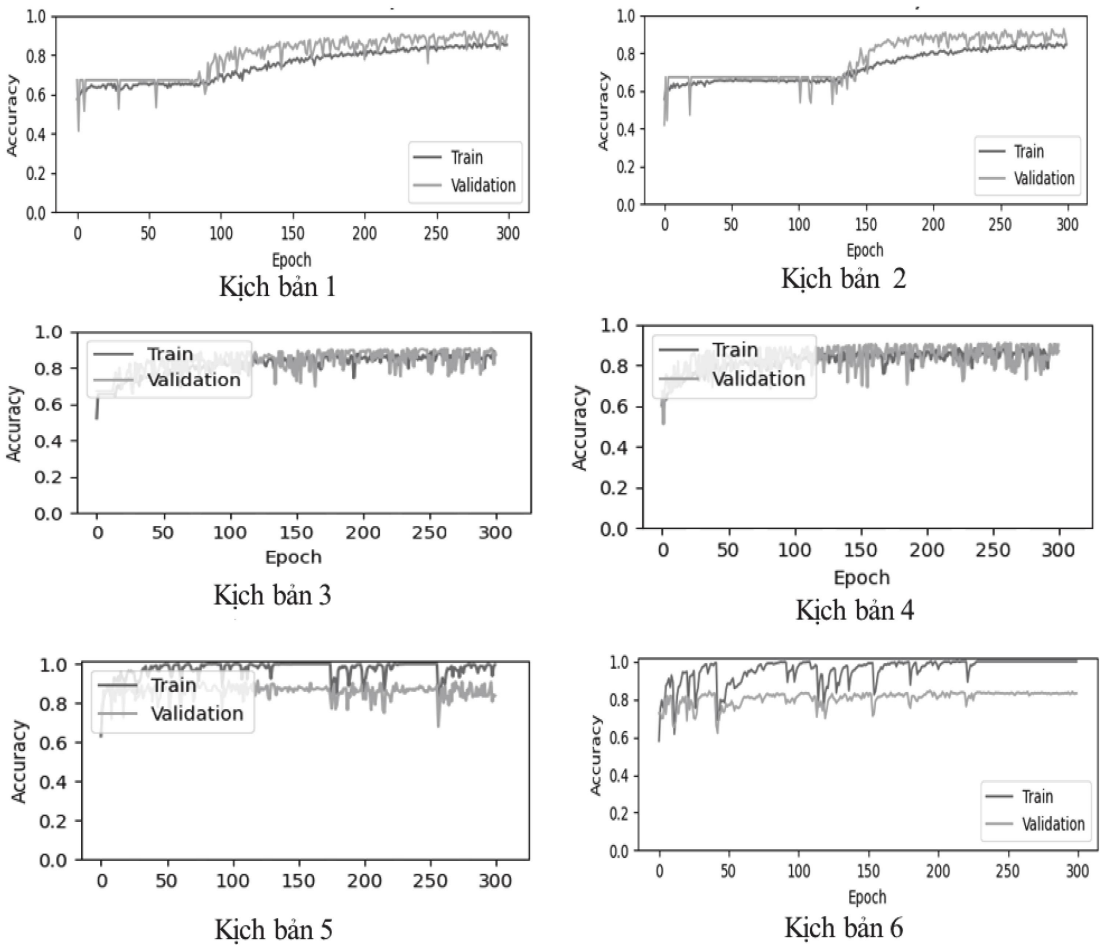
3.3. Kịch bản và kết quả huấn luyện

Thực nghiệm trên năm kịch bản với các tham số huấn luyện như sau: Tỷ lệ học: 0.001; Số lần học: 300; Số lượng ảnh cho một lần học: 64; Kích thước ảnh: 224 x 224.

Bảng 2. Các kịch bản được đề xuất và kết quả huấn luyện

Kịch bản	Kiến trúc mạng	Môi trường	Độ chính xác	Độ mất mát	Thời gian huấn luyện (giờ)
1	Vision Transformer	Local	98.85%	0.028	22.50
2		Cluster	95.17%	0.033	10.13
3	ResNet152	Local	92.01%	0.234	50.54
4		Cluster	88.29%	0.269	33.12
5	DenseNet121	Local	84.93%	0.65	20.12
6		Cluster	83.91%	0.87	11.52

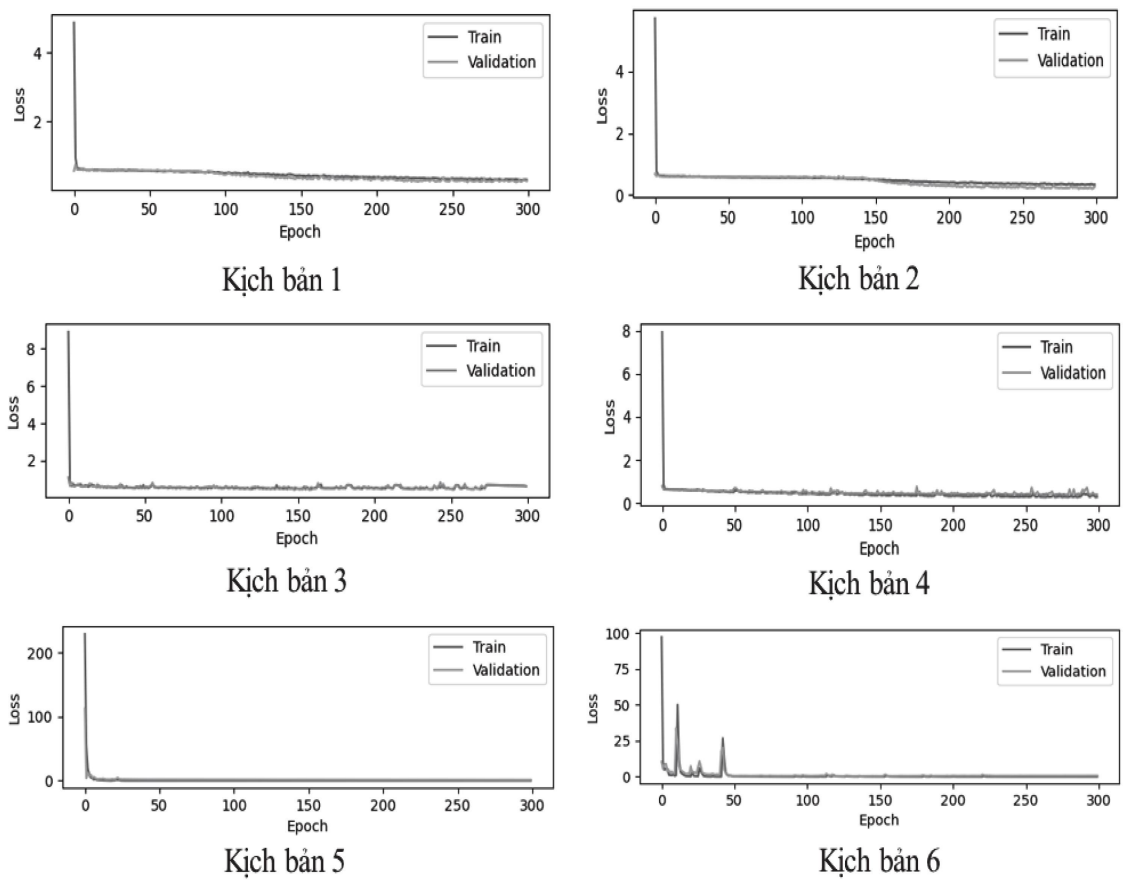
3.4. Độ chính xác



Hình 7. Độ chính xác của các kịch bản

Kịch bản 1 đạt độ chính xác 98.85%, kịch bản 2 có độ chính xác là 95.17%. Kịch bản 3, 4, 5 và 6, độ chính xác thấp hơn hẳn so với hai kịch bản đầu tiên. Mặt khác, kịch bản 5, 6 có giá trị val_accuracy cho kết quả thấp hơn giá trị train accuracy và không ổn định, điều này cho thấy mô hình huấn luyện chưa thật sự hiệu quả và có thể sai lệch trong quá trình dự đoán. Ngược lại, kịch bản 1, 2, 3, 4 có giá trị val_accuracy và train_accuracy chênh lệch ít. Đặc biệt, kịch bản 1 và 2 có giá trị val_accuracy cao hơn so với train_accuracy, điều này cho thấy mô hình trong hai kịch bản 1 và 2 có khả năng dự đoán tốt hơn các mô hình trong bốn kịch bản còn lại. Chúng ta thấy được, trong cùng một tập dữ liệu khi huấn luyện trên mô hình Vision Transformer lại cho kết quả tốt hơn các mô hình CNN, mặc dù các mô hình CNN cũng có khả năng tổng quát hóa tốt, nhưng chúng có thể gặp giới hạn trong việc nắm bắt các mối quan hệ dài hạn và các đặc trưng phức tạp. Ngược lại do cơ chế tự chú ý của Vision Transformer có thể học được các đặc trưng tổng quát và phức tạp hơn, giúp nó tổng quát hóa tốt hơn trên các tập dữ liệu kiểm tra khác nhau, tập dữ liệu càng lớn càng cho thấy sự hiệu quả của Vision Transformer so với các mô hình CNN khác.

3.5. Độ mất mát

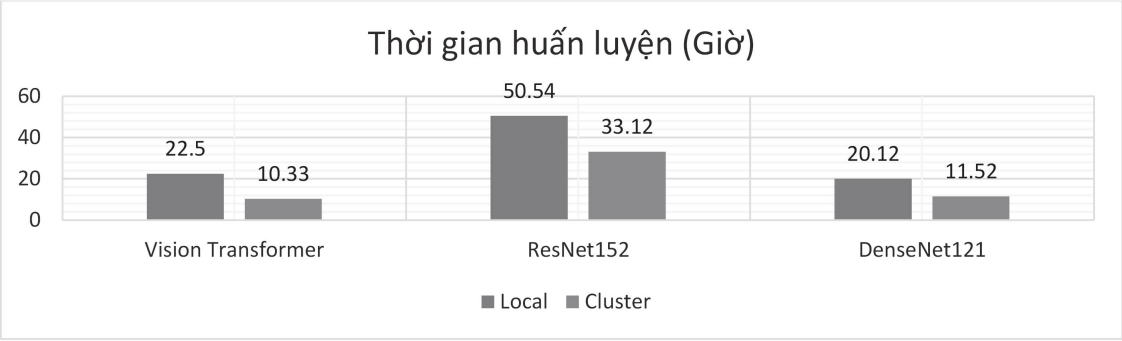


Hình 8. Độ mất mát của các kịch bản

Về độ mất mát, các kịch bản 3 và 4 có giá trị val_loss khá tương đồng nhau với lần lượt 0,234 và 0,269. Tương tự, kịch bản 5 và 6 cũng có giá trị val_loss 0,65 và 0,87. Kịch bản 5 có dấu hiệu underfitting do 2 đường val_loss và train_loss trùng nhau.

Riêng kịch bản 1 và 2 có val_loss thấp nhất lần lượt là 0,028 và 0,033, đặc biệt đường val_loss của hai kịch bản này thấp hơn đường train_loss. Từ đó có thể thấy được mô hình của kịch bản 1 và 2 cho kết quả tốt hơn các mô hình trong các kịch bản còn lại. Tương tự như kết quả độ chính xác, độ mất mát của Vision Transformer cũng thấp hơn so với các mạng CNN còn lại, Vision Transformer chia hình ảnh thành các patch nhỏ và xử lý từng patch như một token, tương tự như cách Transformer xử lý các từ trong NLP. Điều này giúp Vision Transformer học các đặc trưng rất chi tiết và phức tạp từ từng phần của hình ảnh, khả năng này giúp Vision Transformer xây dựng các biểu diễn đặc trưng phong phú và chi tiết hơn, dẫn đến độ mất mát thấp hơn khi so sánh với các mạng CNN truyền thống.

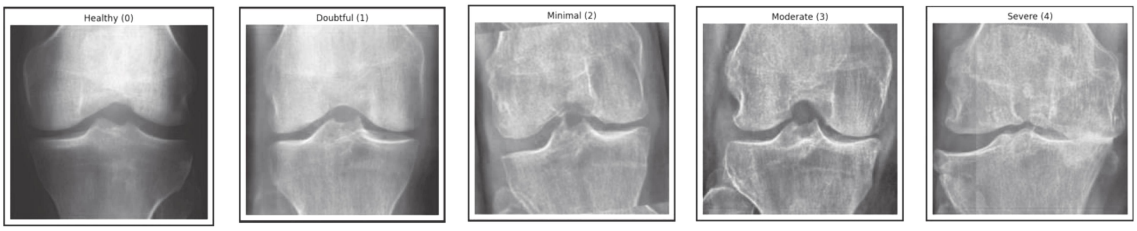
3.6. Thời gian huấn luyện



Hình 9. Thời gian huấn luyện

Về thời gian huấn luyện mô hình, kịch bản 2 cho kết quả tốt nhất với chỉ 10 giờ 13 phút và kịch bản 3 có thời gian huấn luyện lâu nhất với 50 giờ 54 phút. Ta có thể thấy các kịch bản huấn luyện trên môi trường cục bộ có thời gian huấn luyện lâu hơn gấp đôi so với các kịch bản huấn luyện trên môi trường phân tán. Cụ thể với mô hình Vision Transformer ở kịch bản 1 (cục bộ) và kịch bản 2 (phân tán) cho thời gian lần lượt là 22 giờ 5 phút và 10 giờ 33 phút, chênh lệch gấp đôi.

3.7. Một số hình ảnh kiểm thử trên Mô hình Vision Transformer



Hình 10. Kết quả kiểm thử trên mô hình Vision Transformer

Theo kết quả kiểm thử cho thấy, mô hình Vision Transformer cho kết quả phân loại thoái hoá khớp có độ chính xác cao. Cụ thể ảnh kiểm thử mức độ cao nhất (Severe), ta có thể thấy rõ khe khớp trong hình rất hẹp, thậm chí đầu khớp chạm nhau, chứng tỏ sụn khớp không còn nữa, có thể kết luận đây là trường hợp nặng nhất. Ảnh được chuẩn đoán là mức độ 3 (Moderate), độ hẹp khớp là có nhưng không nhiều như hình đầu, trên khớp có các điểm gai nhỏ. Ảnh chuẩn đoán mực độ 2 (Minimal) cũng

tương tự, khớp có gai xương nhỏ và độ hẹp khe khớp không đáng kể, chứng tỏ sụn khớp còn tốt. Và ảnh được chuẩn đoán ở mức độ 0 (Healthy) và mức độ 1 (Doubtful), khớp gối dường như bình thường, có một vài gai xương nhỏ nhưng không đáng kể.

3.8. So sánh kết quả

Qua đánh giá kết quả sau khi huấn luyện cho thấy các mô hình đạt kết quả về độ chính xác cao và ổn định sau các lần huấn luyện ở trong ba mô hình, trong đó Vision Transformer được đánh giá là có độ chính xác cao nhất với độ chính xác huấn luyện đến 98.85% và độ loss đạt 0.028. Do độ phức tạp về cấu trúc mô hình, ảnh hưởng đến thời gian huấn luyện nên thời gian huấn luyện ở mô hình ResNet quá lớn, mô hình Densenet và mô hình Vision Transformer có thời gian huấn luyện tương đương nhau, nhưng Vision Transformer tối ưu hơn về độ chính xác và có thời gian huấn luyện ngắn hơn. So với các bài báo khác cụ thể như " Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss" của Pingjun Chen và các cộng sự năm 2021, phương pháp chúng tôi đề ra có độ chính xác cao hơn rõ rệt. Do các cơ chế tự chú ý, khả năng nắm bắt các mối quan hệ không gian dài hạn và khả năng học các đặc trưng phức tạp hơn từ dữ liệu hình ảnh của Vision Transformer nên kết quả huấn luyện luôn tốt hơn so với các mô hình CNN khác, đặc biệt là trên tập dữ liệu có kích thước lớn. Về thời gian huấn luyện, đối với mô hình ResNet ở môi trường cục bộ có thời gian huấn luyện cao nhất với thời gian lên đến hơn 50 giờ, nhưng trong môi trường phân tán thời gian huấn luyện giảm còn 30 giờ. Sau đó là mô hình Vision Transformer có thời gian huấn luyện ở môi trường cục bộ hơn 20 giờ và ở môi trường phân tán là xấp xỉ 10 giờ. Thời gian huấn luyện ở kịch bản của mô hình DenseNet và của mô hình Vision Transformer ngắn hơn với thời gian huấn luyện huấn luyện ở môi trường cục bộ local từ 20 giờ tới 22 giờ và giảm đi nhiều ở môi trường phân tán từ 10 giờ đến 11 giờ. Có thể thấy được rằng thời gian huấn luyện giảm đi 40 – 50% ở môi trường phân tán so với môi trường cục bộ, nhưng kết quả huấn luyện trên môi trường phân tán lại kém hơn môi trường cục bộ, có thể do một số lý do như trong môi trường phân tán, dữ liệu và thông tin về các tham số mô hình cần được truyền tải và đồng bộ giữa các nút tính toán, có thể gây ra độ trễ và dẫn đến việc cập nhật trọng số không đồng bộ, làm giảm hiệu quả của quá trình huấn luyện. Spark Master có nhiệm vụ tính trung bình các trọng số, việc cập nhật các trọng số làm giảm thời gian huấn luyện nhưng đồng thời cũng làm giảm hiệu quả huấn luyện. Tóm lại, đối với tập dữ liệu càng lớn, khi huấn luyện trên mô hình Vision Transformer và môi trường phân tán càng thấy được sự khác biệt về độ chính xác, độ mất mát và thời gian huấn luyện so với các mô hình khác.

4. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đề xuất một hướng tiếp cận huấn luyện mô hình mạng nơron Vision Transformer thực nghiệm ở môi trường phân tán trên tập dữ liệu ảnh X-Quang khớp gối nhằm rút ngắn thời gian huấn luyện mô hình và hỗ trợ phát hiện và phân loại chính xác thoái hoá khớp gối. Qua thực nghiệm đã thực hiện, một số đóng góp của nghiên cứu này như sau: tổng hợp dữ liệu về tổn thương khớp gối từ nguồn dữ liệu công khai và tiến hành tiền xử lý, thực hiện kiểm tra việc phân

lớp của các mô hình mạng nơron Vision Transformer, ResNet152, DensNet trên môi trường phân tán và cục bộ đối với tập dữ liệu thực nghiệm. Hướng phát triển sắp tới, chúng tôi sẽ tiếp tục cải thiện độ chính xác của mô hình và bổ sung thêm giai đoạn phân đoạn ảnh, chỉ ra và khoanh vùng cụ thể các tổn thương của khớp trên ảnh.

TÀI LIỆU THAM KHẢO

[1]. J. H. Kellgren, J. S. Lawrence, (1957), "*Radiological Assessment of Osteo-Arthrosis*", Ann Rheum Dis 16(4): 494–502 doi: 10.1136/ard.16.4.494.

[2]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, (2016), "*Deep Residual Learning for Image Recognition*", IEEE Conference on Computer Vision and Pattern Recognition.

[3]. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, (2017), "*Densely Connected Convolutional Networks*", IEEE Conference on Computer Vision and Pattern Recognition.

[4]. Pingjun Chen, Linlin Gao, Xiaoshuang Shi, Kyle Allen, Lin Yang, (2019), "*Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss*", Computerized Medical Imaging and Graphics.

[5]. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby, (2020), "*An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*", Computer Vision and Pattern Recognition.

[6]. Bochen Guan, Phương Lưu, Arya Haj Mizaian, Shadpour Demehri, Alexey Samsonov, Ali Guermazi, Richard Kijowski, (2020), "*Deep learning approach to predict pain progression in knee osteoarthritis*", International Journal of Computer Assisted Radiology and Surgery volume 15, pages 457–466.

[7]. Bin Liu, Jianxu Luo, Huan Huang, (2020), "*Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN*", International Journal of Computer Assisted Radiology and Surgery.

[8]. Pauline Shan Qing Yeoh, Khin Wee Lai, Siew Li Goh, Khairunnisa Hasikin, Yan Chai Hum, Yee Kai Tee, Samiappan Dhanalakshmi, (2021), "*Emergence of Deep learning in Knee Osteoarthritis Diagnosis*", Computational Intelligence in Image and Video Analysis.

[9]. Yifan Wang, Xianan Wang, Tianning Gao, Le Du, Wei Liu, (2021), "*An Automatic Knee Osteoarthritis Diagnosis Method Based on Deep learning: Data from the Osteoarthritis Initiative*", Journal of Healthcare Engineering.

[10]. Albert Swiecicki, Nianyi Li, Jonathan O'Donnell, Nicholas Said, Jichen Yang, Richard C Mather, William A Jiranek, Maciej A Mazurowski, (2021), "*Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists*", Computers in Biology and Medicine.

[11]. Simon Olsson, Ehsan Akbarian, Anna Lind, Ali Sharif Razavian, Max Gordon, (2021), "*Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using Deep learning in an unfiltered adult population*", BMC Musculoskeletal Disorders.

[12]. S Sheik Abdullah, M Pallikonda Rajasekaran, (2021), "*Automatic detection and classification of knee osteoarthritis using Deep learning approach*", La radiologia medica.

[13]. Joseph Humberto Cueva, Darwin Castillo, Héctor Espinós-Morató, David Durán, Patricia Díaz, Vasudevan Lakshminarayanan, (2022), "*Detection and Classification of Knee Osteoarthritis*", Machine Learning and Artificial Intelligence in Diagnostics.

[14]. Abdul Sami Mohammed, Ahmed Abul Hasanaath, Ghazanfar Latif, Abul Bashar, (2023), "*Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images*", AI/ML-Based Medical Image Processing and Analysis.

[15]. https://keras.io (20/01/2024)

[16]. https://bigdl.readthedocs.io (20/01/2024)

[17]. https://spark.apache.org (20/01/2024)

[18]. https://data.mendeley.com/datasets/56rmx5bjcr/1 (20/01/2024)