

ỦY BAN NHÂN DÂN TP HỒ CHÍ MINH  
**TRƯỜNG ĐẠI HỌC SÀI GÒN**

~~~~~\*~~~~~

LÊ THỊ THANH NGÂN - 3121411143

**XÂY DỰNG ỨNG DỤNG XÁC ĐỊNH UNG THƯ  
XUÔNG DỰA TRÊN UNET VÀ CƠ CHẾ ATTENTION**

**KHÓA LUẬN TỐT NGHIỆP**

**NGÀNH: CÔNG NGHỆ THÔNG TIN  
TRÌNH ĐỘ ĐÀO TẠO: ĐẠI HỌC**

**TP. HỒ CHÍ MINH, THÁNG 01 NĂM 2025**

ỦY BAN NHÂN DÂN TP HỒ CHÍ MINH  
**TRƯỜNG ĐẠI HỌC SÀI GÒN**

~~~~~\*~~~~~

LÊ THỊ THANH NGÂN - 3121411143

**XÂY DỰNG ỨNG DỤNG XÁC ĐỊNH UNG THƯ  
XUÔNG DỰA TRÊN UNET VÀ CƠ CHẾ ATTENTION**

**KHÓA LUẬN TỐT NGHIỆP**

**NGÀNH: CÔNG NGHỆ THÔNG TIN  
TRÌNH ĐỘ ĐÀO TẠO: ĐẠI HỌC**

**Người hướng dẫn: TS. Đỗ Như Tài**

**TP. HỒ CHÍ MINH, THÁNG 01 NĂM 2025**

## **LỜI CAM ĐOAN**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi, các số liệu và kết quả nghiên cứu nêu trong khóa luận là trung thực và chưa được công bố trong một công trình nào khác.

Tác giả khóa luận

**Lê Thị Thanh Ngân**

## LỜI CẢM ƠN

Tôi xin chân thành gửi lời cảm ơn sâu sắc nhất đến các quý thầy/cô trong Khoa Công nghệ thông tin Trường Đại học Sài Gòn đã ân cần giảng dạy và chia sẻ kiến thức quý báu, đồng thời đã tạo điều kiện thuận lợi về cơ sở vật chất và môi trường học tập, nghiên cứu trong suốt những năm tháng tôi theo học tại trường.

Đặc biệt, tôi xin bày tỏ lòng biết ơn đến giảng viên hướng dẫn của tôi - tiến sĩ Đỗ Như Tài, người đã tận tình hướng dẫn, chỉ bảo và động viên tôi trong suốt quá trình thực hiện đề tài nghiên cứu này. Sự hỗ trợ về kiến thức chuyên môn, những góp ý quý báu và định hướng kịp thời của thầy là yếu tố then chốt giúp tôi có thể hoàn thành tốt công việc nghiên cứu của mình.

Cuối cùng, tôi xin dành lời cảm ơn chân thành nhất đến bạn bè và gia đình, những người đã luôn là điểm tựa tinh thần vững chắc, động viên và tạo mọi điều kiện tốt nhất để tôi có thể tập trung vào học tập và nghiên cứu.

Mặc dù đã có nhiều cố gắng, nhưng do hạn chế về thời gian và kiến thức, đề tài này chắc chắn không tránh khỏi những thiếu sót. Tôi rất mong nhận được những ý kiến đóng góp quý báu từ quý thầy/cô và các bạn để nghiên cứu này ngày càng hoàn thiện hơn.

Tôi xin chân thành cảm ơn!

## MỤC LỤC

|                                            |      |
|--------------------------------------------|------|
| LỜI CAM ĐOAN .....                         | i    |
| LỜI CẢM ƠN .....                           | ii   |
| MỤC LỤC .....                              | iii  |
| DANH MỤC CÁC CỤM TỪ VIẾT TẮT .....         | vii  |
| DANH MỤC BẢNG BIỂU .....                   | viii |
| DANH MỤC CÁC SƠ ĐỒ .....                   | ix   |
| DANH MỤC HÌNH ẢNH .....                    | x    |
| LỜI MỞ ĐẦU .....                           | xi   |
| CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI .....        | 1    |
| 1.1. Đặt vấn đề .....                      | 1    |
| 1.2. Mục tiêu và nhiệm vụ nghiên cứu ..... | 2    |
| 1.2.1. Mục tiêu .....                      | 2    |
| 1.2.2. Nhiệm vụ .....                      | 2    |
| 1.3. Đối tượng và phạm vi nghiên cứu ..... | 2    |
| 1.3.1. Đối tượng nghiên cứu .....          | 2    |
| 1.3.2. Phạm vi nghiên cứu .....            | 3    |
| 1.4. Phương pháp nghiên cứu .....          | 3    |
| 1.5. Tính mới và đóng góp của đề tài ..... | 4    |
| 1.5.1. Về mặt khoa học .....               | 4    |
| 1.5.2. Về mặt thực tiễn .....              | 4    |
| 1.6. Tiến độ và kế hoạch nghiên cứu .....  | 5    |

|                                                                                       |          |
|---------------------------------------------------------------------------------------|----------|
| 1.7. Cấu trúc của luận văn .....                                                      | 6        |
| <b>CHƯƠNG 2: CƠ SỞ LÝ THUYẾT .....</b>                                                | <b>7</b> |
| 2.1. Tổng quan về ung thư xương.....                                                  | 7        |
| 2.1.1. Định nghĩa và phân loại ung thư xương.....                                     | 7        |
| 2.1.2. Nguyên nhân và tỉ lệ mắc ung thư xương .....                                   | 7        |
| 2.1.2.1. Nguyên nhân gây ra ung thư xương.....                                        | 7        |
| 2.1.2.2. Tỷ lệ mắc bệnh ung thư xương .....                                           | 8        |
| 2.2. Vai trò của học sâu trong chẩn đoán hình ảnh y tế.....                           | 9        |
| 2.2.1. Tổng quan về học sâu.....                                                      | 9        |
| 2.2.2. Nguyên lý cơ bản của học sâu .....                                             | 10       |
| 2.2.3. Mạng nơ-ron trong học sâu .....                                                | 10       |
| 2.2.4. Các kiến trúc mạng nơ-ron tích chập – CNN.....                                 | 11       |
| 2.3. Vai trò và ứng dụng của học sâu trong y học và chuẩn hình ảnh y tế.....          | 12       |
| 2.3.1. Ứng dụng chung của học sâu trong y học .....                                   | 12       |
| 2.3.2. Học sâu trong chuẩn đoán hình ảnh y tế .....                                   | 13       |
| 2.4. Các kỹ thuật học sâu trong phân đoạn hình ảnh.....                               | 13       |
| 2.5. U-Net trong phân đoạn hình ảnh y tế .....                                        | 14       |
| 2.5.1. Kiến trúc và nguyên lý hoạt động của U-net .....                               | 14       |
| 2.5.2. Lợi thế của U-net trong phân đoạn 2D và 3D .....                               | 15       |
| 2.6. Cơ chế Attention trong học sâu .....                                             | 16       |
| 2.6.1. Khái niệm về cơ chế Attention .....                                            | 16       |
| 2.6.2. Tích hợp cơ chế Attention kết hợp với U-net trong phân đoạn hình ảnh y tế..... | 16       |

|                                                                  |           |
|------------------------------------------------------------------|-----------|
| 2.6.3. Ứng dụng thực tiễn của U-net và Attention trong y tế..... | 17        |
| 2.7. Các công nghệ và framework .....                            | 18        |
| 2.7.1. Giới thiệu về PyTorch.....                                | 18        |
| 2.7.2. Giới thiệu về MONAI .....                                 | 19        |
| 2.7.3. PyTorch Lightning.....                                    | 19        |
| 2.8. Các nghiên cứu liên quan.....                               | 20        |
| 2.8.1. Dự án BTCV .....                                          | 20        |
| 2.8.2. Dự án Spleen (U-Net 3D).....                              | 22        |
| 2.8.3. Kết luận.....                                             | 23        |
| <b>CHƯƠNG 3: DỮ LIỆU VÀ CÁC PHƯƠNG PHÁP ĐỀ XUẤT .....</b>        | <b>25</b> |
| 3.1. Phân tích và chuẩn bị dữ liệu .....                         | 25        |
| 3.1.1. Nguồn dữ liệu.....                                        | 25        |
| 3.1.2. Đặc điểm dữ liệu .....                                    | 27        |
| 3.2. Quy trình huấn luyện mô hình học sâu .....                  | 32        |
| 3.3. Tiền xử lý dữ liệu .....                                    | 33        |
| 3.3.1. Chuẩn hóa kích thước ảnh .....                            | 33        |
| 3.3.2. Chia tập dữ liệu .....                                    | 38        |
| 3.4. Phương pháp đề xuất.....                                    | 40        |
| 3.4.1. Mã hóa (Encoder).....                                     | 43        |
| 3.4.2. Giải mã (Decoder) .....                                   | 43        |
| 3.4.3. Đề xuất hàm mất mát.....                                  | 43        |
| 3.4.4. Phương pháp huấn luyện.....                               | 45        |
| <b>CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM.....</b>                        | <b>47</b> |

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>4.1. Kết quả thực nghiệm .....</b>                                                              | <b>47</b> |
| <b>4.1.1. Mô tả tập dữ liệu thực nghiệm .....</b>                                                  | <b>47</b> |
| <b>4.1.2. Môi trường thực nghiệm và cấu hình huấn luyện.....</b>                                   | <b>47</b> |
| <b>4.1.3. Kết quả đánh giá trên tập test.....</b>                                                  | <b>48</b> |
| <b>4.1.4. Xây dựng ứng dụng xác định ảnh ung thư xương dựa trên Unet và cơ chế Attention .....</b> | <b>51</b> |
| <b>4.1. Phân tích kết quả .....</b>                                                                | <b>53</b> |
| <b>4.2. Ứng dụng thực tiễn.....</b>                                                                | <b>54</b> |
| <b>CHƯƠNG 5: KẾT LUẬN VÀ KHUYẾN NGHỊ .....</b>                                                     | <b>56</b> |
| <b>5.1. Kết luận.....</b>                                                                          | <b>56</b> |
| <b>5.1.1. Ưu điểm .....</b>                                                                        | <b>56</b> |
| <b>5.1.2. Khó khăn, thách thức.....</b>                                                            | <b>57</b> |
| <b>5.2. Hướng phát triển .....</b>                                                                 | <b>57</b> |
| <b>DANH MỤC TÀI LIỆU THAM KHẢO.....</b>                                                            | <b>59</b> |



## DANH MỤC CÁC CỤM TỪ VIẾT TẮT

|       |                                                  |
|-------|--------------------------------------------------|
| MONAI | Medical Open Network for Artificial Intelligence |
| BTCV  | Brain Tumor Classification and Segmentation      |
| CPU   | Central Processing Unit                          |
| GPU   | Graphics Processing Unit                         |
| ReLU  | Rectified Linear Unit                            |
| IoU   | Intersection over Union                          |
| CNN   | Convolutional Neural Network                     |

## **DANH MỤC BẢNG BIỂU**

|                                                                                                            |    |
|------------------------------------------------------------------------------------------------------------|----|
| Bảng 1. 1. Bảng tiến độ công việc .....                                                                    | 5  |
| Bảng 3. 1. Đặt điểm cơ bản của tập dữ liệu BTXRD [12] .....                                                | 26 |
| Bảng 3. 2. Bảng thống kê số lượng ảnh ung thư ở các vị trí xương khác nhau có trong bộ dữ liệu BTXRD ..... | 29 |
| Bảng 3. 3. Bảng thống kê số lượng các loại ung thư có trong bộ dữ liệu BTXRD .....                         | 31 |
| Bảng 3. 4. Bảng kết quả chia dữ liệu lần 1 .....                                                           | 38 |
| Bảng 3. 5. Bảng kết quả chia dữ liệu lần 2 .....                                                           | 39 |
| Bảng 4. 1. Bảng kết quả sau khi đánh giá trên tập kiểm thử.....                                            | 48 |

## **DANH MỤC CÁC SƠ ĐỒ**

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| Sơ đồ 3. 1. Sơ đồ kiến trúc của Attention Gate được sử dụng trong mô hình. .... | 41 |
|---------------------------------------------------------------------------------|----|

## DANH MỤC HÌNH ẢNH

|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| Hình 3. 1. Quy trình thiết kế tập dữ liệu BTXRD.....                                           | 26 |
| Hình 3. 2. Biểu đồ biểu diễn số lượng hình ảnh xương bình thường, lành tính, ác tính.<br>..... | 28 |
| Hình 3. 3. Hình ảnh 10 dòng dữ liệu đầu trong file classification.xlsx .....                   | 28 |
| Hình 3. 4. Biểu đồ biểu diễn số lượng ảnh ung thư ở nam và nữ. ....                            | 29 |
| Hình 3. 5. Biểu đồ cột biểu diễn số lượng ảnh tại các vị trí xương khác nhau .....             | 30 |
| Hình 3. 6. Biểu đồ cột biểu diễn số lượng ảnh trong từng loại ung thư.....                     | 32 |
| Hình 3. 7. Biểu đồ biểu diễn kích thước ảnh trước khi xử lý.....                               | 34 |
| Hình 3. 8. Ba mươi ảnh ngẫu nhiên có khoanh vùng khối u trước khi xử lý dữ liệu. ...           | 35 |
| Hình 3. 9. Ba mươi ảnh ngẫu nhiên sau khi đồng nhất kích thước ảnh. ....                       | 37 |
| Hình 3. 10. Mô hình học sâu SegUnet .....                                                      | 40 |
| Hình 4. 1. Trực quan hóa kết quả ngẫu nhiên phân đoạn của mô hình từ tập kiểm thử .            | 50 |
| Hình 4. 2. Giao diện ứng dụng web mang tên "Bone Detect - AI Prediction" .....                 | 51 |
| Hình 4. 3. Giao diện trước khi upload ảnh cần nhận diện khối u .....                           | 52 |
| Hình 4. 4. Giao diện khi upload ảnh khối u .....                                               | 52 |
| Hình 4. 5. Giao diện kết quả phân đoạn vùng ung thư .....                                      | 53 |

## LỜI MỞ ĐẦU

Ung thư xương tuy không phải là căn bệnh phổ biến, nhưng là bệnh lý ác tính nghiêm trọng, ảnh hưởng lớn đến chất lượng sống của bệnh nhân. Việc phát hiện sớm và chẩn đoán chính xác đóng vai trò then chốt trong điều trị hiệu quả và cải thiện tiên lượng. Tuy nhiên, các phương pháp chẩn đoán hình ảnh như X-quang, CT, MRI truyền thống lại đòi hỏi chuyên môn cao, mất thời gian và dễ bị ảnh hưởng bởi yếu tố chủ quan.

Trong bối cảnh đó, trí tuệ nhân tạo, đặc biệt là học sâu, đang mở ra những cơ hội mới trong phân tích ảnh y tế. Mô hình U-Net đã chứng minh hiệu quả vượt trội trong phân đoạn ảnh y sinh. Việc tích hợp cơ chế Attention vào U-Net giúp mô hình tập trung tốt hơn vào các vùng đặc trưng quan trọng, từ đó nâng cao độ chính xác trong nhận diện tổn thương.

Xuất phát từ tiềm năng đó, đề tài "Xây dựng ứng dụng xác định ung thư xương dựa trên U-Net và cơ chế Attention" được thực hiện nhằm xây dựng hệ thống tự động phân đoạn vùng nghi ngờ ung thư xương từ ảnh X-quang. Nghiên cứu tập trung phát triển mô hình kết hợp U-Net và Attention, huấn luyện trên tập dữ liệu BTXRD và đánh giá hiệu quả qua các chỉ số định lượng và định tính.

## CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

Ung thư xương là một căn bệnh hiểm gặp nhưng có ảnh hưởng nghiêm trọng đến sức khỏe và chất lượng cuộc sống của bệnh nhân. Đây là một thách thức lớn đối với ngành y tế, không chỉ vì mức độ phức tạp trong chẩn đoán mà còn do tính chất tiến triển nhanh của bệnh. Trong những năm gần đây, với sự phát triển vượt bậc của trí tuệ nhân tạo (AI) và học sâu (Deep Learning), các công cụ hỗ trợ chẩn đoán hình ảnh y tế đã và đang tạo nên những bước tiến quan trọng. Đặc biệt, mô hình U-Net kết hợp cơ chế Attention là một giải pháp mang tính đột phá, được chứng minh là rất hiệu quả trong việc phân đoạn các vùng tổn thương từ hình ảnh y tế, như ảnh X-quang hoặc CT. Việc xây dựng một ứng dụng chuyên biệt dựa trên các công nghệ tiên tiến này không chỉ giúp cải thiện khả năng phát hiện sớm ung thư xương mà còn tạo ra cơ hội nâng cao hiệu quả điều trị.

### 1.1. Đặt vấn đề

Trong lĩnh vực y tế, việc ứng dụng công nghệ học sâu vào phân tích hình ảnh y tế đã trở thành một hướng nghiên cứu quan trọng và thu hút. Tuy nhiên, tại Việt Nam, các nghiên cứu tập trung vào ung thư xương, đặc biệt là phân đoạn hình ảnh 2D, 3D vẫn còn khá hạn chế. Đây là một thực tế đáng quan ngại khi ung thư xương là căn bệnh thường được phát hiện muộn, dẫn đến hiệu quả điều trị thấp và chi phí cao.

Với đề tài “Xây dựng ứng dụng xác định ung thư xương dựa trên U-Net và cơ chế Attention”, tôi kỳ vọng sẽ tạo ra một hệ thống có khả năng tự động phân tích hình ảnh y tế, xác định chính xác các vùng tổn thương trong xương, từ đó hỗ trợ bác sĩ chẩn đoán nhanh hơn và hiệu quả hơn.

Ngoài ra, sự kết hợp giữa U-Net và cơ chế Attention không chỉ giúp tăng độ chính xác mà còn giải quyết được những thách thức trong phân đoạn hình ảnh y tế, chẳng hạn như xác định các vùng nhỏ, không rõ ràng. Điều này là tiềm năng lớn trong việc áp dụng

công nghệ tiên tiến vào thực tiễn, đặc biệt trong các tình huống khẩn cấp hoặc thiếu nguồn lực y tế.

## **1.2. Mục tiêu và nhiệm vụ nghiên cứu**

### **1.2.1. Mục tiêu**

Mục tiêu chính của đề tài là xây dựng một ứng dụng tự động và chính xác để hỗ trợ bác sĩ phát hiện ung thư xương từ các hình ảnh y tế 2D. Hệ thống này không chỉ cung cấp kết quả phân đoạn với độ chính xác cao mà còn tối ưu hóa thời gian và nguồn lực trong quá trình chẩn đoán.

### **1.2.2. Nhiệm vụ**

Để đạt được mục tiêu trên, đề tài đặt ra các nhiệm vụ cụ thể bao gồm:

- Nghiên cứu các mô hình học sâu hiện đại, đặc biệt là U-Net và cơ chế Attention, để ứng dụng vào bài toán phân đoạn hình ảnh y tế.
- Xây dựng quy trình tiền xử lý dữ liệu như chuẩn hóa hình ảnh, cắt bỏ vùng không liên quan và cân chỉnh kích thước ảnh nhằm đảm bảo dữ liệu đầu vào đạt chất lượng cao nhất và kích thước nhất định.
- Phát triển và huấn luyện mô hình trên bộ dữ liệu y tế thực tế có đáp ứng yêu cầu về tính minh bạch và bảo mật, sử dụng các thuật toán tối ưu hóa như Dice Loss và Focal Loss để cải thiện hiệu suất phân đoạn.
- Đánh giá hiệu quả mô hình thông qua các chỉ số chuyên môn, đồng thời so sánh với các phương pháp đã có để khẳng định tính vượt trội của ứng dụng.
- Đề xuất cải tiến, mở rộng khả năng ứng dụng của hệ thống.

## **1.3. Đối tượng và phạm vi nghiên cứu**

### **1.3.1. Đối tượng nghiên cứu**

Đối tượng nghiên cứu của đề tài là các hình ảnh y tế 2D, cụ thể là các ảnh X-quang hoặc CT của bệnh nhân bị nghi ngờ hoặc chẩn đoán mắc ung thư xương. Những hình

ảnh này chứa thông tin quan trọng về cấu trúc xương và các vùng tổn thương, cung cấp một nền tảng dữ liệu cần thiết để phát triển hệ thống.

### **1.3.2. Phạm vi nghiên cứu**

Nghiên cứu sẽ tập trung vào việc triển khai và tối ưu hóa mô hình U-Net, một kiến trúc mạng nơ-ron tích chập (CNN) phổ biến và hiệu quả cho các tác vụ phân đoạn ảnh y tế. Các thử nghiệm có thể bao gồm việc tích hợp cơ chế Attention và U-Net vào các vùng đặc trưng quan trọng của ảnh để đánh giá khả năng cải thiện hiệu suất phân đoạn.

Không những thế, để tối ưu hóa hiệu suất phân đoạn và giải quyết vấn đề mất cân bằng lớp, nghiên cứu sẽ khảo sát việc sử dụng hàm mất mát kết hợp giữa Dice Loss và Focal Loss. Quá trình huấn luyện mô hình sẽ được thực hiện bằng thuật toán Adam, kết hợp các kỹ thuật điều chuẩn như L2 regularization, Dropout và tăng cường dữ liệu cơ bản bao gồm lật, xoay, điều chỉnh sáng/tối, tỷ lệ... để cải thiện khả năng tổng quát hóa.

Hiệu suất của mô hình sẽ được đánh giá định lượng trên tập kiểm thử thông qua các chỉ số phổ biến như Dice Coefficient, IoU, Precision và Recall, đặc biệt tập trung vào lớp khối u. Toàn bộ quá trình xây dựng, huấn luyện và đánh giá mô hình sẽ được triển khai bằng TensorFlow và Keras API, kết hợp với nền tảng Weights & Biases (WandB) để theo dõi, ghi lại và trực quan hóa các thực nghiệm.

Đề tài không chỉ dừng lại ở việc phát triển mô hình mà còn mở rộng sang việc thử nghiệm và đánh giá trên bộ dữ liệu thực tế, với kỳ vọng đưa ra được một hệ thống khả thi trong môi trường y tế tại Việt Nam.

Ngoài ra, đề tài cũng tìm hiểu các thách thức đặc thù của hình ảnh xương, như độ phức tạp của cấu trúc 2D, sự tương đồng giữa các vùng tổn thương và mô lành, để đảm bảo mô hình có thể ứng dụng hiệu quả trong thực tiễn.

## **1.4. Phương pháp nghiên cứu**

Đề tài sử dụng các phương pháp nghiên cứu chính bao gồm:



- Thu thập và tiền xử lý dữ liệu: Bộ dữ liệu bao gồm các ảnh X-quang hoặc CT của bệnh nhân ung thư xương sẽ được chuẩn hóa, cân chỉnh kích thước và loại bỏ các vùng không liên quan. Sau đó, đưa qua tiền xử lý dữ liệu để đảm bảo chất lượng và tính nhất quán của đầu vào.
- Phát triển mô hình: Đề tài triển khai mô hình U-Net kết hợp với cơ chế Attention để tận dụng cả thông tin tổng quát và chi tiết trong hình ảnh. Các thuật toán tối ưu như Dice Loss hoặc Focal Loss cũng được áp dụng trong quá trình huấn luyện để cải thiện hiệu quả phân đoạn.
- Huấn luyện và đánh giá: Mô hình sẽ được huấn luyện trên tập dữ liệu thực tế, sử dụng kỹ thuật tăng cường dữ liệu (Data Augmentation) để giảm thiểu hiện tượng overfitting. Hiệu quả mô hình được đánh giá thông qua các chỉ số chuyên môn như Dice Coefficient và IoU.
- Thử nghiệm và cải tiến: Sau khi hoàn thiện mô hình, hệ thống sẽ được thử nghiệm trên các bộ dữ liệu khác nhau để đánh giá tính ổn định và khả năng mở rộng. Các cải tiến sẽ được đề xuất để nâng cao hiệu quả và khả năng ứng dụng thực tế.

## **1.5. Tính mới và đóng góp của đề tài**

### **1.5.1. Về mặt khoa học**

Đề tài góp phần mở rộng trong việc ứng dụng kết hợp U-Net và cơ chế Attention vào bài toán xác định ung thư xương tại Việt Nam. Đây là một hướng đi mới, không chỉ giải quyết được những hạn chế trong phân đoạn hình ảnh y tế hiện tại mà còn mở rộng khả năng ứng dụng của học sâu vào các lĩnh vực y tế khác.

### **1.5.2. Về mặt thực tiễn**

Về thực tiễn, đề tài này có thể hỗ trợ bác sĩ chẩn đoán nhanh chóng và chính xác, giảm tải khối lượng công việc trong các cơ sở y tế. Đồng thời, hệ thống còn tiết kiệm chi phí cho bệnh nhân thông qua việc giảm thời gian chẩn đoán và các xét nghiệm không

cần thiết. Đây là một giải pháp thiết thực, góp phần nâng cao chất lượng dịch vụ y tế và thúc đẩy xu hướng số hóa y tế tại Việt Nam.

### 1.6. Tiến độ và kế hoạch nghiên cứu

Đề tài này được thực hiện trong 15 tuần, từ ngày 27/12/2024 đến khoảng ngày 22/05/2025.

*Bảng 1. 1. Bảng tiến độ công việc*

| <b>Tuần</b> | <b>Công việc</b>                                                                      |
|-------------|---------------------------------------------------------------------------------------|
| 1 – 2       | Tìm hiểu và nghiên cứu các tài liệu liên quan đến MONAI, Unet và Attention Mechanism. |
| 3 – 5       | Tìm hiểu, nghiên cứu và chạy thử một số dự án liên quan.                              |
| 6           | Thu thập dữ liệu.                                                                     |
| 6 – 10      | Tiền xử lý dữ liệu và huấn luyện mô hình.                                             |
| 11          | Thiết kế giao diện ứng dụng.                                                          |
| 12          | Xây dựng ứng dụng và tích hợp mô hình vào ứng dụng                                    |
| 12 – 13     | Đánh giá kết quả và tối ưu hóa.                                                       |
| 14          | Kết luận và khuyến nghị                                                               |

|         |                    |
|---------|--------------------|
| 14 – 15 | Hoàn thành báo cáo |
|---------|--------------------|

### **1.7. Cấu trúc của luận văn**

Ngoài phần mở đầu, kết luận và danh mục tài liệu tham khảo luận văn gồm có 5 chương:

Chương 1: TỔNG QUAN VỀ ĐỀ TÀI

Chương 2: CƠ SỞ LÝ THUYẾT

Chương 3: DỮ LIỆU VÀ PHƯƠNG PHÁP ĐỀ XUẤT

Chương 4: KẾT QUẢ THỰC NGHIỆM

Chương 5: KẾT LUẬN VÀ KHUYẾN NGHỊ

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

### 2.1. Tổng quan về ung thư xương

#### 2.1.1. Định nghĩa và phân loại ung thư xương

Ung thư xương là một thuật ngữ chung mô tả sự phát triển bất thường và không kiểm soát của các tế bào ác tính bắt nguồn từ mô xương hoặc sụn. Các tế bào ung thư này có khả năng xâm lấn và phá hủy các mô xương khỏe mạnh xung quanh, đồng thời có thể di căn đến các bộ phận khác của cơ thể [1].

Ung thư xương được phân thành hai nhóm chính:

- Ung thư xương nguyên phát: khi khối u khởi phát trực tiếp từ các tế bào xương hoặc sụn. Ví dụ như Osteosarcoma (ung thư xương tạo xương), Chondrosarcoma (ung thư sụn), Ewing sarcoma (u xương Ewing), Fibrosarcoma và Malignant Fibrous Histiocytoma (sarcoma sợi và u sợi bào ác tính).
- Ung thư xương thứ phát (hay còn gọi là di căn xương): khi các tế bào ung thư từ một cơ quan khác (như phổi, vú, tuyến tiền liệt) lan đến xương và hình thành khối u mới.

Việc phân loại chính xác loại ung thư xương là vô cùng quan trọng, vì nó ảnh hưởng trực tiếp đến tiên lượng và lựa chọn phương pháp điều trị cho bệnh nhân.

#### 2.1.2. Nguyên nhân và tỉ lệ mắc ung thư xương

##### 2.1.2.1. Nguyên nhân gây ra ung thư xương

Cho đến nay, nguyên nhân chính gây ra hầu hết các trường hợp ung thư xương vẫn chưa được xác định rõ ràng. Tuy nhiên, một số yếu tố nguy cơ đã được ghi nhận có thể làm tăng khả năng mắc bệnh, bao gồm:

- Yếu tố di truyền (ví dụ: một số hội chứng di truyền như hội chứng Li-Fraumeni, retinoblastoma di truyền).
- Do ảnh hưởng hoặc tiếp xúc với môi trường hóa chất và bức xạ.
- Tiền sử xạ trị hoặc hóa trị liệu cho các bệnh ung thư khác.
- Một số bệnh lý xương lành tính như bệnh Paget xương và chấn thương xương mạn tính (mặc dù mối liên hệ này còn gây tranh cãi).

Ung thư xương nguyên phát là một loại ung thư tương đối hiếm gặp, chiếm một tỷ lệ nhỏ trong tổng số các ca ung thư được chẩn đoán hàng năm. Tỷ lệ mắc bệnh có thể thay đổi tùy theo độ tuổi, giới tính và chủng tộc, với một số loại như Osteosarcoma thường gặp hơn ở trẻ em và thanh thiếu niên, trong khi Chondrosarcoma lại phổ biến hơn ở người lớn tuổi. Ngược lại, ung thư xương thứ phát (di căn xương) lại phổ biến hơn nhiều, do nhiều loại ung thư ở các cơ quan khác có xu hướng di căn đến xương.

#### **2.1.2.2. Tỷ lệ mắc bệnh ung thư xương**

Ung thư xương là một bệnh hiếm gặp, chiếm một tỷ lệ nhỏ trong tổng số các ca ung thư được chẩn đoán hàng năm. Tỷ lệ mắc bệnh có thể thay đổi tùy theo độ tuổi, giới tính và chủng tộc. Tỷ lệ mắc ung thư xương nguyên phát trên toàn cầu dao động từ 1-3 ca trên mỗi triệu người mỗi năm, trong đó phổ biến nhất là các dạng như osteosarcoma, chondrosarcoma, và Ewing sarcoma. Một số loại như Osteosarcoma thường bắt gặp ở trẻ em và thanh thiếu niên ở nhóm tuổi từ 10-20, trong khi Chondrosarcoma lại phổ biến hơn ở người lớn tuổi [1, 2].

Ngược lại, ung thư xương thứ phát (di căn xương) lại phổ biến hơn nhiều, do nhiều loại ung thư ở các cơ quan khác có xu hướng di căn đến xương. Đặc biệt, theo thống kê từ tạp chí Cancer 2009, nam giới có nguy cơ mắc bệnh cao hơn nữ giới, với tỷ lệ nam:nữ khoảng 1,35:1 [3].

## **2.2. Vai trò của học sâu trong chẩn đoán hình ảnh y tế**

Chẩn đoán hình ảnh y tế, bao gồm các kỹ thuật như X-quang, CT, MRI, đóng vai trò không thể thiếu trong y học hiện đại, cung cấp thông tin trực quan quý giá cho việc phát hiện, chẩn đoán, theo dõi và lập kế hoạch điều trị nhiều loại bệnh lý, đặc biệt là ung thư. Tuy nhiên, do khối lượng dữ liệu hình ảnh ngày càng tăng cùng với sự phức tạp của các dấu hiệu bệnh lý đã đặt ra thách thức lớn cho các chuyên gia y tế về thời gian và độ chính xác. Để giải quyết những thách thức này, học sâu (Deep Learning) là cộng sự đầy tiềm năng.

### **2.2.1. Tổng quan về học sâu**

Học sâu hay còn gọi là Deep Learning là một nhánh thuộc lĩnh vực trí tuệ nhân tạo, phát triển từ Machine Learning với nền tảng là các mạng nơ-ron nhân tạo (Artificial Neural Networks - ANN). Đây là phương pháp học máy có khả năng học biểu diễn dữ liệu thông qua nhiều lớp trừu tượng hóa, cho phép máy tính tự động trích xuất đặc trưng mà không cần đến các kỹ thuật thủ công trong thiết kế đặc trưng như các phương pháp truyền thống.

Điểm đặc biệt của Deep Learning là khả năng xử lý dữ liệu lớn (Big Data) nhờ vào sự hỗ trợ của phần cứng mạnh mẽ như GPU, đặc biệt là sau sự xuất hiện của dòng GPU GTX 10 series từ NVIDIA. Điều này giúp tăng tốc độ huấn luyện và mở rộng khả năng ứng dụng của Deep Learning từ phòng thí nghiệm ra thực tiễn. Sự thành công vang dội của mô hình AlexNet năm 2012 tại cuộc thi ILSVRC (ImageNet Large Scale Visual Recognition Challenge) đã đánh dấu bước ngoặt cho Deep Learning khi lần đầu tiên mô hình mạng nơ-ron nhân tạo đạt được kết quả vượt trội so với các phương pháp truyền thống trong nhận diện hình ảnh [4].

Deep Learning đã chứng minh tính ứng dụng rộng rãi của mình trong nhiều lĩnh vực như thị giác máy tính (computer vision), xử lý ngôn ngữ tự nhiên (natural language processing), nhận diện giọng nói, y tế, giao thông, và nhiều lĩnh vực khác trong cuộc

sống. Và trong đề tài này, Deep Learning cũng đóng một vai trò quan trọng, bằng cách tự động học các đặc trưng phức tạp từ dữ liệu hình ảnh y tế, các mô hình học sâu có thể hỗ trợ đắc lực cho bác sĩ trong việc phát hiện sớm các tổn thương nhỏ, phân đoạn chính xác các vùng quan tâm, phân loại bệnh lý và thậm chí dự đoán tiên lượng, qua đó hứa hẹn cải thiện đáng kể chất lượng chăm sóc sức khỏe và hiệu quả chẩn đoán.

### **2.2.2. Nguyên lý cơ bản của học sâu**

Học sâu là một lĩnh vực con của học máy, tập trung vào việc xây dựng và huấn luyện các mạng nơ-ron nhân tạo (Artificial Neural Networks - ANNs) với nhiều lớp ẩn để học các biểu diễn dữ liệu phức tạp và thực hiện các tác vụ dự đoán hoặc ra quyết định.

Khác với các phương pháp học máy truyền thống thường đòi hỏi quá trình thiết kế và trích xuất đặc trưng thủ công tốn nhiều công sức bởi các chuyên gia, học sâu có khả năng tự động học các đặc trưng phân cấp trực tiếp từ dữ liệu thô. Các lớp gần đầu vào của mạng ANN học các đặc trưng ở mức độ thấp (ví dụ: cạnh, góc trong hình ảnh), trong khi các lớp sâu hơn kết hợp các đặc trưng này để học các khái niệm trừu tượng và phức tạp hơn.

Nguyên lý hoạt động cốt lõi dựa trên việc tối ưu hóa các tham số như trọng số và độ lệch của mạng thông qua một thuật toán, thường là lan truyền ngược (backpropagation) và các biến thể của gradient descent, nhằm giảm thiểu sự khác biệt giữa đầu ra dự đoán và giá trị thực tế của dữ liệu huấn luyện, được đo lường bằng một hàm mất mát (loss function).

### **2.2.3. Mạng nơ-ron trong học sâu**

Mạng nơ-ron (ANN) là cấu trúc tính toán nền tảng của học sâu, được lấy cảm hứng từ cách thức hoạt động của não bộ sinh học, bao gồm các đơn vị xử lý cơ bản gọi là nơ-ron (neurons) hay nút (nodes) được kết nối với nhau. Mỗi nơ-ron nhận đầu vào từ các nơ-ron khác hoặc từ dữ liệu gốc, thực hiện một phép tính, thường là tổng có trọng số của các đầu vào cộng với một độ lệch – bias. Sau đó truyền kết quả qua một hàm kích

hoạt (activation function) phi tuyến để tạo ra đầu ra. Các hàm kích hoạt phổ biến bao gồm Sigmoid, Tanh, ReLU và các biến thể của nó, đóng vai trò quan trọng trong việc cho phép mạng học các mối quan hệ phức tạp. Các nơ-ron được tổ chức thành các lớp: lớp đầu vào nhận dữ liệu, một hoặc nhiều lớp ẩn (hidden layers) thực hiện các phép biến đổi và tính toán trung gian và lớp đầu ra đưa ra dự đoán cuối cùng. Độ "sâu" của một mạng học sâu chính là số lượng các lớp ẩn này.

Trong quá trình huấn luyện, thông tin được lan truyền xuôi (feedforward) từ lớp đầu vào qua các lớp ẩn đến lớp đầu ra để tạo ra dự đoán. Sau đó, sự khác biệt giữa dự đoán và giá trị thực tế được tính toán bằng hàm mất mát. Thuật toán lan truyền ngược (backpropagation) được sử dụng để tính toán gradient của hàm mất mát theo từng trọng số và độ lệch trong mạng. Các gradient này sau đó được sử dụng bởi một thuật toán tối ưu hóa, ví dụ: Stochastic Gradient Descent - SGD, Adam, để cập nhật các trọng số và độ lệch theo hướng giảm thiểu lỗi. Quá trình này được lặp lại qua nhiều lượt (epochs) trên tập dữ liệu huấn luyện cho đến khi mạng đạt được hiệu suất mong muốn. Mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi quy (RNN) và các kiến trúc phức tạp hơn như Transformers đều là các dạng cụ thể của mạng nơ-ron được thiết kế cho các loại dữ liệu và tác vụ khác nhau. Trong bài nghiên cứu này chủ yếu sử dụng CNN để thiết kế mô hình.

#### **2.2.4. Các kiến trúc mạng nơ-ron tích chập – CNN**

Mạng nơ-ron tích chập hay còn gọi là Convolutional Neural Network là một lớp kiến trúc mạng nơ-ron sâu chuyên biệt, đạt được thành công vang dội trong các bài toán liên quan đến dữ liệu có cấu trúc dạng lưới, đặc biệt là hình ảnh. Đặc điểm nổi bật của CNN là việc sử dụng các lớp tích chập - convolutional layers để tự động học các bộ lọc filters hoặc kernels có khả năng phát hiện các đặc trưng không gian cục bộ trong ảnh, chẳng hạn như cạnh, góc, kết cấu ở các lớp nông và các đối tượng hay bộ phận phức tạp hơn ở các lớp sâu hơn. Các lớp tích chập thường đi kèm với pooling layers, ví dụ: max pooling giúp giảm kích thước không gian của bản đồ đặc trưng, tăng tính bất biến với



các thay đổi nhỏ về vị trí và giảm số lượng tham số tính toán. Cuối cùng, các đặc trưng được học thường được đưa qua một hoặc nhiều lớp kết nối đầy đủ (fully connected layers) để thực hiện nhiệm vụ phân loại hoặc hồi quy.

Một số kiến trúc CNN mang tính bước ngoặt đã định hình sự phát triển của lĩnh vực này. LeNet-5 [5] là một trong những CNN đầu tiên thành công trong nhận dạng chữ số viết tay. AlexNet [6] với kiến trúc sâu hơn và việc sử dụng GPU đã giành chiến thắng thuyết phục trong cuộc thi ImageNet, đánh dấu sự trỗi dậy mạnh mẽ của học sâu.

Sau đó, các kiến trúc như VGGNet [7] với các bộ lọc nhỏ 3x3 xếp chồng lên nhau, GoogLeNet [8] với module “Inception” và ResNet [9] với các kết nối tắt - residual connections cho phép huấn luyện các mạng rất sâu đã liên tục đẩy lùi giới hạn về độ chính xác trong các tác vụ thị giác máy tính. Mặc dù nghiên cứu này tập trung vào U-Net, việc hiểu các nguyên lý thiết kế từ những kiến trúc nền tảng này là rất hữu ích.

## **2.3. Vai trò và ứng dụng của học sâu trong y học và chuẩn hình ảnh y tế**

### **2.3.1. Ứng dụng chung của học sâu trong y học**

Học sâu đã và đang tạo ra những thay đổi mang tính cách mạng trong nhiều khía cạnh của y học và chăm sóc sức khỏe, vượt ra ngoài lĩnh vực phân tích hình ảnh. Một trong những ứng dụng nổi bật là trong lĩnh vực gen và proteomics, nơi các mô hình học sâu được sử dụng để dự đoán cấu trúc protein, xác định các biến thể gen liên quan đến bệnh tật, và phân tích dữ liệu biểu hiện gen để hiểu rõ hơn về cơ chế bệnh sinh. Trong phát triển thuốc, học sâu hỗ trợ việc sàng lọc các hợp chất tiềm năng, dự đoán độc tính và hiệu quả của thuốc, và thậm chí thiết kế các phân tử thuốc mới, qua đó rút ngắn đáng kể thời gian và chi phí nghiên cứu.

Ngoài ra, học sâu còn được ứng dụng trong việc phân tích hồ sơ bệnh án điện tử (EHR) để dự đoán nguy cơ mắc bệnh, phát hiện sớm các đợt bệnh cấp tính, tối ưu hóa phác đồ điều trị và hỗ trợ các quyết định lâm sàng dựa trên bằng chứng từ lượng lớn dữ liệu bệnh nhân. Các ứng dụng khác bao gồm robot phẫu thuật được tăng cường bởi AI,

hệ thống theo dõi sức khỏe cá nhân thông minh và phát hiện các mẫu bất thường trong tín hiệu sinh lý như điện tâm đồ (ECG) hay điện não đồ (EEG).

### **2.3.2. Học sâu trong chuẩn đoán hình ảnh y tế**

Lĩnh vực chẩn đoán hình ảnh y tế có lẽ là nơi học sâu thể hiện rõ rệt nhất tiềm năng biến đổi của mình. Việc phân tích hình ảnh y tế truyền thống thường dựa vào kinh nghiệm và khả năng nhận diện mẫu của các bác sĩ x-quang, một quá trình có thể bị ảnh hưởng bởi sự mệt mỏi, khối lượng công việc lớn và tính chủ quan nhất định, dẫn đến nguy cơ bỏ sót hoặc chẩn đoán không nhất quán. Các mô hình học sâu, đặc biệt là CNN, có khả năng tự động học các đặc trưng hình ảnh phức tạp liên quan đến bệnh lý trực tiếp từ dữ liệu, giúp giải quyết nhiều thách thức này.

Trong phát hiện bệnh, các thuật toán học sâu có thể được huấn luyện để xác định sự hiện diện của các dấu hiệu bất thường như khối u, nốt phổi, xuất huyết não, hay gãy xương với độ nhạy và độ đặc hiệu cao, đôi khi tương đương hoặc vượt trội so với chuyên gia con người trong một số tác vụ cụ thể.

Trong phân loại bệnh lý, học sâu hỗ trợ việc phân biệt giữa các tình trạng khác nhau dựa trên đặc điểm hình ảnh, ví dụ như phân biệt khối u lành tính và ác tính, hoặc xác định các giai đoạn khác nhau của bệnh. Đối với bài toán phân đoạn ảnh, một trọng tâm của nghiên cứu này, các mô hình học sâu như U-Net có khả năng khoanh vùng chính xác các cơ quan, mô hoặc tổn thương bệnh lý, cung cấp thông tin định lượng về kích thước, hình dạng, thể tích, điều này rất quan trọng cho việc lập kế hoạch phẫu thuật, xạ trị, và theo dõi đáp ứng điều trị. Có thể thấy, học sâu có khả năng xử lý lượng lớn hình ảnh một cách nhanh chóng và nhất quán không chỉ giúp giảm tải cho các chuyên gia y tế mà còn hứa hẹn cải thiện khả năng tiếp cận các dịch vụ chẩn đoán chất lượng cao, đặc biệt ở những khu vực thiếu hụt nhân lực y tế.

## **2.4. Các kỹ thuật học sâu trong phân đoạn hình ảnh**

Phân đoạn ảnh y tế là quá trình phân chia một hình ảnh y tế thành nhiều vùng hoặc đối tượng có ý nghĩa, trong đó mỗi pixel trong ảnh được gán cho một nhãn lớp cụ thể. Mục tiêu của phân đoạn có thể là để xác định ranh giới chính xác của các cơ quan giải phẫu, ví dụ như não, tim, gan..., các cấu trúc mô bệnh học, ví dụ như tế bào ung thư, hoặc các vùng tổn thương bệnh lý, ví dụ như khối u...

Kết quả phân đoạn cung cấp thông tin định lượng và không gian quan trọng, hỗ trợ đắc lực cho nhiều ứng dụng lâm sàng như chẩn đoán bệnh, đánh giá mức độ nghiêm trọng, lập kế hoạch can thiệp phẫu thuật hoặc xạ trị, theo dõi sự phát triển của bệnh hoặc đáp ứng với điều trị. Ví dụ, trong nghiên cứu ung thư, việc phân đoạn chính xác khối u cho phép đo lường thể tích và theo dõi sự thay đổi kích thước của nó theo thời gian. Đây là một chỉ số quan trọng để đánh giá hiệu quả của phác đồ điều trị.

Tuy nhiên, phân đoạn ảnh y tế là một bài toán đầy thách thức do nhiều yếu tố, chẳng hạn như hình ảnh y tế thường có độ tương phản thấp giữa các mô khác nhau, ranh giới giữa các cấu trúc có thể mờ nhạt hoặc không rõ ràng, sự hiện diện của nhiễu và các tạo ảnh (artifacts) do quá trình thu nhận hình ảnh, cũng như sự đa dạng lớn về hình dạng, kích thước và vị trí của các cấu trúc giải phẫu hoặc tổn thương giữa các bệnh nhân khác nhau. Hơn nữa, việc thu thập dữ liệu gán nhãn ở cấp độ pixel (ground truth masks) cho việc huấn luyện các mô hình phân đoạn đòi hỏi nhiều thời gian và công sức từ các chuyên gia y tế. Các phương pháp phân đoạn truyền thống dựa trên ngưỡng, phát hiện biên, hay phát triển vùng thường gặp khó khăn với những thách thức này.

Do đó, sự ra đời của các mô hình học sâu, đặc biệt là các kiến trúc dựa trên mạng nơ-ron tích chập, đã mang lại những tiến bộ đáng kể trong việc giải quyết bài toán phân đoạn ảnh y tế một cách tự động và chính xác hơn.

## **2.5. U-Net trong phân đoạn hình ảnh y tế**

### **2.5.1. Kiến trúc và nguyên lý hoạt động của U-net**

U-Net, được giới thiệu bởi Ronneberger và cộng sự vào năm 2015 [10], là một kiến trúc mạng nơ-ron tích chập được thiết kế chuyên biệt cho các tác vụ phân đoạn ảnh y sinh và đã nhanh chóng trở thành một trong những kiến trúc nền tảng và được sử dụng rộng rãi nhất trong lĩnh vực này. Kiến trúc của U-Net có dạng đối xứng hình chữ U, bao gồm hai thành phần chính: đường đi xuống – contracting path hoặc encoder và đường đi lên – expanding path hoặc decoder. Đường đi xuống có cấu trúc tương tự như một mạng CNN phân loại điển hình, bao gồm các khối lặp lại của hai lớp tích chập  $3 \times 3$  không đệm – unpadded convolutions, theo sau bởi một hàm kích hoạt ReLU và một lớp gộp cực đại – max pooling  $2 \times 2$  với sai bước 2. Mục đích của đường đi xuống là để giảm dần kích thước không gian của bản đồ đặc trưng (feature map) trong khi tăng số lượng kênh đặc trưng. Qua đó giúp nắm bắt thông tin ngữ cảnh và các đặc trưng trừu tượng của ảnh đầu vào.

Đường đi lên – decoder, thực hiện nhiệm vụ ngược lại: tăng dần độ phân giải không gian của bản đồ đặc trưng để tạo ra một mặt nạ phân đoạn có cùng kích thước với ảnh đầu vào. Điều này thường đạt được thông qua việc sử dụng các lớp giải tích chập – transposed convolution hoặc up-convolution  $2 \times 2$ , làm tăng gấp đôi kích thước không gian và giảm một nửa số kênh đặc trưng. Một đặc điểm then chốt và mang tính đột phá của U-Net là skip connections. Các kết nối này sao chép và ghép nối – concatenate các bản đồ đặc trưng có độ phân giải cao từ đường đi xuống với các bản đồ đặc trưng tương ứng đã được giải tích chập ở đường đi lên. Nhờ các kết nối tắt này, thông tin không gian chi tiết bị mất trong quá trình gộp ở đường đi xuống có thể được phục hồi và kết hợp với thông tin ngữ cảnh cấp cao được học ở các lớp sâu hơn, cho phép mô hình định vị chính xác ranh giới của các đối tượng cần phân đoạn. Lớp cuối cùng của mạng thường là một lớp tích chập  $1 \times 1$  sử dụng hàm kích hoạt Sigmoid hoặc Softmax để tạo ra bản đồ xác suất cho từng pixel thuộc về mỗi lớp.

### **2.5.2. Lợi thế của U-net trong phân đoạn 2D và 3D**

Kiến trúc U-Net ban đầu được thiết kế và chứng minh hiệu quả cho các tác vụ phân đoạn trên ảnh y tế 2D, nơi nó có khả năng phân tích từng lát cắt riêng lẻ của dữ liệu. Một trong những lợi thế lớn của U-Net là khả năng hoạt động tốt ngay cả khi tập dữ liệu huấn luyện có kích thước hạn chế. Đây là một trong những tình huống thường gặp trong y tế. Skip connections cho phép truyền tải hiệu quả thông tin chi tiết, giúp U-Net phân đoạn chính xác các cấu trúc phức tạp và có ranh giới tinh vi.

Hơn thế nữa, U-Net có thể được huấn luyện từ đầu – end-to-end và tương đối dễ triển khai so với một số kiến trúc phức tạp khác. Khả năng xử lý ảnh 2D của U-Net phù hợp với nhiều loại dữ liệu y tế như ảnh X-quang, lát cắt CT.

## **2.6. Cơ chế Attention trong học sâu**

### **2.6.1. Khái niệm về cơ chế Attention**

Attention Mechanism trong học sâu là một kỹ thuật lấy cảm hứng từ cách hệ thống thị giác của con người tập trung vào các phần quan trọng của một cảnh phức tạp trong khi bỏ qua các chi tiết không liên quan. Thay vì xử lý tất cả các phần của đầu vào với tầm quan trọng như nhau, cơ chế Attention cho phép mô hình tự động học cách gán các "trọng số" khác nhau cho các phần khác nhau của dữ liệu đầu vào, tùy thuộc vào mức độ liên quan của chúng đối với tác vụ hiện tại. Điều này giúp mô hình tập trung tài nguyên tính toán vào những thông tin hữu ích nhất, cải thiện hiệu suất và khả năng diễn giải [10, 11].

Về cơ bản, cơ chế chú ý tính toán một vector ngữ cảnh – context vector như một tổng có trọng số của các vector biểu diễn đầu vào, trong đó các trọng số được xác định dựa trên trạng thái hiện tại của mô hình hoặc một truy vấn – query cụ thể.

### **2.6.2. Tích hợp cơ chế Attention kết hợp với U-net trong phân đoạn hình ảnh y tế**

Attention U-Net là một kiến trúc cải tiến từ U-Net gốc, được đề xuất bởi Oktay và cộng sự vào năm 2018 [11], nhằm tăng cường khả năng của mô hình trong việc phân

đoạn các cấu trúc y tế bằng cách tích hợp các Attention Gates - AGs. AGs này được thiết kế để tự động học cách làm nổi bật các đặc trưng liên quan đến đối tượng mục tiêu và làm mờ đi các phản hồi không liên quan từ các vùng nền hoặc các cấu trúc giải phẫu khác không phải là đối tượng quan tâm.

Trong kiến trúc Attention U-Net, AGs được tích hợp vào skip connections trước khi các đặc trưng từ đường đi xuống – encoder được ghép nối với các đặc trưng từ đường đi lên – decoder.

Mỗi AGs nhận hai đầu vào: một tín hiệu từ lớp trước đó ở đường đi lên (gating signal), mang thông tin ngữ cảnh cấp cao về vị trí và đặc điểm của đối tượng cần tìm, và một tín hiệu từ lớp tương ứng ở đường đi xuống, mang thông tin không gian chi tiết. AG sử dụng gating signal để tạo ra một attention coefficient map có cùng kích thước với đặc trưng từ đường đi xuống. Bản đồ này sau đó được nhân theo từng phần tử element-wise multiplication với đặc trưng từ đường đi xuống, có tác dụng "lọc" và chỉ giữ lại những đặc trưng được xem là quan trọng dựa trên ngữ cảnh hiện tại [11].

Kết quả là các đặc trưng không liên quan sẽ bị triệt tiêu, trong khi các đặc trưng nổi bật liên quan đến đối tượng mục tiêu sẽ được khuếch đại. Việc này giúp cải thiện độ nhạy và độ đặc hiệu của mô hình, đặc biệt hiệu quả trong việc phân đoạn các đối tượng nhỏ, có hình dạng phức tạp, hoặc khi có sự hiện diện của nhiều đối tượng tương tự nhau trong ảnh.

### **2.6.3. Ứng dụng thực tiễn của U-net và Attention trong y tế**

Kiến trúc U-Net kết hợp cơ chế Attention đã tạo ra công cụ phân tích ảnh y tế mạnh mẽ, được ứng dụng rộng rãi trong nhiều chuyên khoa và với nhiều loại hình ảnh khác nhau. Trong lĩnh vực ung thư học, các mô hình này được sử dụng để phân đoạn chính xác các khối u trên ảnh MRI, CT, siêu âm hoặc nhũ ảnh, chẳng hạn như khối u não, ung thư vú, ung thư phổi, gan... Những ứng dụng này không chỉ hỗ trợ chẩn đoán ban đầu mà còn giúp định lượng thể tích khối u, theo dõi sự thay đổi theo thời gian nhằm đánh

giá đáp ứng điều trị, cũng như hỗ trợ lập kế hoạch xạ trị bằng cách xác định chính xác vùng chiếu xạ và các cơ quan lân cận cần bảo vệ.

Trong thần kinh học, các mô hình này được sử dụng để phân đoạn các cấu trúc não bộ như chất trắng, chất xám hoặc các tổn thương liên quan đến đột quỵ và bệnh đa xơ cứng. Trong nhãn khoa, chúng có khả năng phân đoạn mạch máu võng mạc và các tổn thương do bệnh võng mạc tiểu đường.

Khả năng phân đoạn chính xác và tự động của mô hình này giúp giảm đáng kể khối lượng công việc thủ công của bác sĩ, tăng tính nhất quán trong đánh giá, và mở ra tiềm năng cho các hệ thống hỗ trợ chẩn đoán thông minh, góp phần nâng cao chất lượng chăm sóc y tế.

## **2.7. Các công nghệ và framework**

### **2.7.1. Giới thiệu về PyTorch**

PyTorch là một thư viện học máy mã nguồn mở dựa trên Torch, được phát triển bởi phòng nghiên cứu AI của Meta. Kể từ khi ra mắt năm 2016, PyTorch nhanh chóng trở nên phổ biến nhờ tính linh hoạt, giao diện Python trực quan và khả năng gỡ lỗi hiệu quả. Một đặc điểm nổi bật của PyTorch là việc sử dụng đồ thị tính toán động "define-by-run", cho phép thay đổi cấu trúc mạng trong quá trình thực thi – rất hữu ích cho các mô hình phức tạp như RNN.

PyTorch có thư viện cung cấp các tensor tính toán trên GPU, module autograd để tự động tính đạo hàm, cùng các công cụ xây dựng và huấn luyện mạng nơ-ron như `torch.nn`, `torch.optim`. PyTorch cũng có hệ sinh thái phong phú gồm `torchvision`, `torchaudio`, `torchtext`, và hỗ trợ triển khai mô hình qua TorchScript hoặc ONNX.

Chính nhờ sự linh hoạt và hỗ trợ mạnh mẽ này, PyTorch trở thành nền tảng lý tưởng để xây dựng, huấn luyện và đánh giá các mô hình phân đoạn ảnh y tế như U-Net và Attention U-Net trong nghiên cứu này.

### 2.7.2. Giới thiệu về MONAI

MONAI hay còn gọi là Medical Open Network for AI là một framework mã nguồn mở dựa trên PyTorch, được phát triển nhằm thúc đẩy ứng dụng trí tuệ nhân tạo trong lĩnh vực ảnh y tế. MONAI cung cấp một bộ công cụ toàn diện và chuẩn hóa cho toàn bộ quy trình phát triển mô hình AI y tế – từ tải và tiền xử lý dữ liệu, tăng cường dữ liệu, xây dựng và huấn luyện mô hình, đến đánh giá và triển khai.

Framework này đặc biệt phù hợp với dữ liệu y tế nhờ hỗ trợ nhiều định dạng như DICOM, NIfTI, dữ liệu đa chiều từ 2D, 3D đến 4D và các phép biến đổi chuyên biệt như biến đổi hình học bảo toàn cấu trúc hoặc điều chỉnh cường độ theo từng modality. MONAI tích hợp sẵn nhiều kiến trúc mạng phổ biến (U-Net, V-Net, DenseNet3D), hàm mất mát, chỉ số đánh giá, cùng các pipeline huấn luyện mẫu tối ưu hóa sẵn.

Nhờ sự kết hợp giữa tính linh hoạt của PyTorch và các tính năng chuyên biệt cho y tế, MONAI giúp rút ngắn thời gian phát triển, tăng tính tái lập và hỗ trợ chia sẻ, hợp tác hiệu quả trong nghiên cứu AI y tế.

### 2.7.3. PyTorch Lightning

PyTorch Lightning là một thư viện mã nguồn mở cấp cao, bao bọc PyTorch nhằm chuẩn hóa quy trình huấn luyện mô hình học sâu. Thư viện này giúp tách biệt rõ ràng phần logic mô hình bao gồm kiến trúc mạng, hàm mất mát, thuật toán tối ưu, tính toán metrics, tách khỏi các thao tác kỹ thuật như vòng lặp huấn luyện, phân phối tính toán hay logging. Tất cả đều được đóng gói trong lớp LightningModule.

Lightning tự động xử lý nhiều khâu trong huấn luyện như di chuyển dữ liệu lên GPU, lan truyền ngược, cập nhật trọng số, hỗ trợ huấn luyện với mixed precision, phân tán trên nhiều GPU hoặc máy và tích hợp logging với TensorBoard hoặc Weights & Biases. Nhờ đó, người dùng có thể tập trung phát triển mô hình, giám lặp lại code, tăng tính tái sử dụng và mở rộng quy mô dễ dàng.



Với thiết kế rõ ràng và hiệu quả, PyTorch Lightning là công cụ hữu ích trong nghiên cứu và triển khai các mô hình học sâu phức tạp, giúp tăng năng suất và đảm bảo tính tái lập trong các thí nghiệm.

## **2.8. Các nghiên cứu liên quan**

### **2.8.1. Dự án BTCV**

Dự án BTCV - 3D Multi-organ Segmentation with UNETR (BTCV Challenge) của MONAI Consortium tập trung vào phân đoạn khối u não trong hình ảnh y tế 3D, đặc biệt là các hình ảnh CT và MRI. Mô hình UNETR kết hợp U-Net với Transformer để tăng cường khả năng học các đặc trưng toàn cục, mang lại sự cải thiện vượt trội trong phân đoạn khối u so với các mô hình truyền thống.

Bộ dữ liệu BTCV cung cấp tập hợp ảnh CT y tế với nhãn phân đoạn đa lớp cho 13 cơ quan trong cơ thể người. Các nhãn này cho phép mô hình học cách xác định vị trí, hình dạng và kích thước của từng cơ quan, trong đó có thể áp dụng cho nhận diện tổn thương hoặc khối u liên quan. Việc sử dụng dữ liệu 3D đòi hỏi mô hình không chỉ xử lý từng lớp ảnh riêng biệt mà còn phải duy trì được mối quan hệ không gian giữa các lát cắt, điều mà UNETR giải quyết hiệu quả bằng cách mã hóa toàn bộ ảnh đầu vào trước khi đưa qua các khối Transformer.

### **Bước 1: Tiền xử lý dữ liệu trong dự án BTCV**

Trước khi đưa dữ liệu vào mô hình UNETR, các bước tiền xử lý đóng vai trò quan trọng để đảm bảo tính đồng nhất và hiệu quả huấn luyện. Dữ liệu ảnh CT đầu vào, vốn có định dạng NIfTI hoặc DICOM và thường chứa nhiều voxel có giá trị không đồng đều, sẽ được chuẩn hóa cường độ sáng về một dải cố định – chẳng hạn  $[0, 1]$  – thông qua các kỹ thuật như `ScaleIntensityRanged`. Điều này giúp giảm nhiễu và giúp mô hình học hiệu quả hơn.

Sau khi chuyển đổi, các ảnh được cắt vùng chứa thông tin quan trọng (crop foreground) để loại bỏ các vùng nền trống rỗng, từ đó giảm thiểu chi phí tính toán và tập

trung vào các khu vực cần phân đoạn. Ngoài ra, một loạt kỹ thuật augmentation 3D cũng được áp dụng như xoay ngẫu nhiên, lật trục, co giãn, phóng to/thu nhỏ hoặc biến đổi gamma, giúp mô hình học được các đặc trưng ở nhiều góc nhìn và hình thái khác nhau. Những thao tác này được xây dựng bằng pipeline Compose trong MONAI.

## **Bước 2: Xây dựng mô hình**

Dự án sử dụng mô hình UNETR (UNet with Transformers) là kiến trúc hiện đại, kết hợp self-attention từ Transformer với pipeline encoder-decoder như U-Net. Ảnh đầu vào 3D được chia thành các patch nhỏ, qua lớp token embedding và được đưa vào bộ encoder Transformer. Các đặc trưng từ các mức độ khác nhau được đưa vào decoder thông qua skip connections, và cuối cùng khôi phục kích thước bằng ConvTranspose3d. Mô hình này khai thác tốt mối quan hệ không gian toàn cục trong ảnh 3D.

## **Bước 3: Huấn luyện mô hình**

Quá trình huấn luyện diễn ra trên GPU với cấu hình gồm batch size = 2, learning rate =  $1e-4$ , số epoch = 150. Bộ tối ưu hóa AdamW được sử dụng cùng với hàm mất mát DiceCELoss, là sự kết hợp giữa Dice Loss (để tăng độ chồng khớp) và CrossEntropy (phân biệt nhiều lớp rõ ràng).

Trong mỗi epoch, dữ liệu được lấy từ DataLoader sử dụng CacheDataset để tăng tốc độ đọc dữ liệu. Các callback như ModelCheckpoint, EarlyStopping, LRMonitor được sử dụng giúp theo dõi tiến trình, lưu lại mô hình tốt nhất và dừng khi không còn cải thiện nữa. Việc huấn luyện cũng được giám sát chặt chẽ bằng Dice Score và IoU trên tập validation để đảm bảo hiệu suất mô hình.

## **Bước 4: Đánh giá và tinh chỉnh**

Sau khi hoàn thành quá trình huấn luyện, mô hình được đánh giá trên tập test của BTCV với cùng quy trình tiền xử lý. Mô hình UNETR đạt kết quả Dice trung bình trên 13 cơ quan  $> 0.87$ , nổi bật ở các cơ quan nhỏ hoặc có hình dạng phức tạp. Cơ chế

attention từ Transformer giúp mô hình duy trì mối liên kết không gian giữa các lát cắt CT và cải thiện khả năng nhận diện ở những vùng có độ tương phản thấp.

Từ kết quả này, mô hình UNETR đã chứng minh được khả năng tổng quát hóa tốt và phù hợp cho các ứng dụng thực tế như hỗ trợ chẩn đoán đa cơ quan hoặc phát hiện tổn thương phức tạp trong ảnh y tế.

### **2.8.2. Dự án Spleen (U-Net 3D)**

Dự án Spleen nhằm đến phân đoạn cơ quan lá lách từ hình vùng lá lách từ các ảnh CT/MRI.

#### **Bước 1: Tiền xử lý dữ liệu**

Dữ liệu của dự án này đơn giản hơn dự án BTCV, tập trung vào một cơ quan duy nhất là lá lách. Đầu tiên, các ảnh CT 3D cũng được chuyển sang định dạng NIfTI, sau đó chuẩn hóa cường độ voxel trong khoảng  $[-57, 164]$  – đây là ngưỡng xác định phù hợp với vùng mô mềm có chứa lá lách. Ảnh được resample về spacing đồng nhất ( $1.5\text{mm} \times 1.5\text{mm} \times 2.0\text{mm}$ ), sau đó cắt vùng chứa cơ quan chính bằng cách crop ảnh theo nhãn.

Tác giả cũng sử dụng kỹ thuật augmentation để tránh overfitting do số lượng ảnh hạn chế. Một số kỹ thuật được áp dụng gồm xoay ngẫu nhiên trong không gian 3D, lật ảnh, thay đổi độ sáng, contrast và co giãn hình học. MONAI hỗ trợ chuỗi pipeline augmentation này qua các lớp như RandRotate90d, RandAffined, RandAdjustContrastd.

#### **Bước 2: Xây dựng mô hình**

Dự án sử dụng mô hình U-Net 3D, được cấu trúc gồm 4 khối encoder và 4 khối decoder, mỗi khối gồm các lớp Conv3D, BatchNorm, ReLU, kết hợp với skip connection từ encoder sang decoder. Việc xây dựng mô hình được thực hiện hoàn toàn bằng MONAI với module UNet, trong đó các tham số như số lớp, số filters, kích thước kernel, stride đều được định nghĩa linh hoạt.

Đồng thời tích hợp PyTorch Lightning vào LightningModule, giúp việc tổ chức huấn luyện, đánh giá, lưu mô hình trở nên chuẩn hóa và dễ dàng hơn. Logger của Lightning giúp ghi lại các chỉ số trong quá trình huấn luyện và cùng lúc đó visualization của segmentation mask cũng được tích hợp sẵn.

### **Bước 3: Huấn luyện**

Mô hình được huấn luyện trong 300 epochs với optimizer Adam, learning rate =  $2e-4$ , sử dụng hàm mất mát DiceLoss. Batch size được đặt là 2 do kích thước lớn của ảnh 3D. Việc theo dõi Dice score, Precision và Recall trên tập validation giúp đánh giá mức độ ổn định và độ chính xác của mô hình trong từng epoch.

PyTorch Lightning hỗ trợ callbacks EarlyStopping để dừng huấn luyện khi mô hình không cải thiện và ModelCheckpoint để lưu lại mô hình tốt nhất theo Dice score. Các metric này cũng được tính toán trên GPU giúp tiết kiệm thời gian đáng kể.

### **Bước 4: Đánh giá và tinh chỉnh**

Kết quả cuối cùng của dự án Spleen (U-Net 3D) trên tập test cho thấy Dice score đạt trung bình xấp xỉ 0.90, cho thấy hiệu quả cao trong việc phân đoạn lá lách. Khả năng học các đặc trưng không gian 3D giúp mô hình phân biệt được lá lách trong nhiều tư thế, độ lệch hoặc trường hợp bị che khuất bởi mô lân cận.

So với dự án BTCV, dự án Spleen có cấu trúc đơn giản hơn nhưng vẫn hiệu quả với bài toán đơn cơ quan.

#### **2.8.3. Kết luận**

Các dự án phân đoạn hình ảnh y tế 3D trên đã cung cấp một cái nhìn tổng quát, một cơ sở mạnh mẽ cho việc phát triển ứng dụng xác định ung thư xương sử dụng mô hình U-Net kết hợp với cơ chế Attention. Mô hình này cho phép tối ưu hóa việc phân đoạn các khối u bằng cách học được các đặc trưng chi tiết từ các hình ảnh y tế 3D, từ đó đạt được sự chính xác cao hơn trong việc phát hiện các bất thường.

Cơ chế Attention giúp mô hình tập trung vào các khu vực quan trọng trong hình ảnh, đặc biệt là các vùng có khả năng chứa khối u. Việc kết hợp U-Net với Attention không chỉ tối ưu hóa khả năng học các đặc trưng không gian và toàn cục mà còn giúp nâng cao khả năng phát hiện ung thư xương từ các hình ảnh y tế 3D. Điều này góp phần cải thiện chất lượng và độ tin cậy của quá trình chẩn đoán, đồng thời hỗ trợ các bác sĩ trong việc phát hiện và điều trị ung thư xương một cách hiệu quả hơn.

## CHƯƠNG 3: DỮ LIỆU VÀ CÁC PHƯƠNG PHÁP ĐỀ XUẤT

### 3.1. Phân tích và chuẩn bị dữ liệu

#### 3.1.1. Nguồn dữ liệu

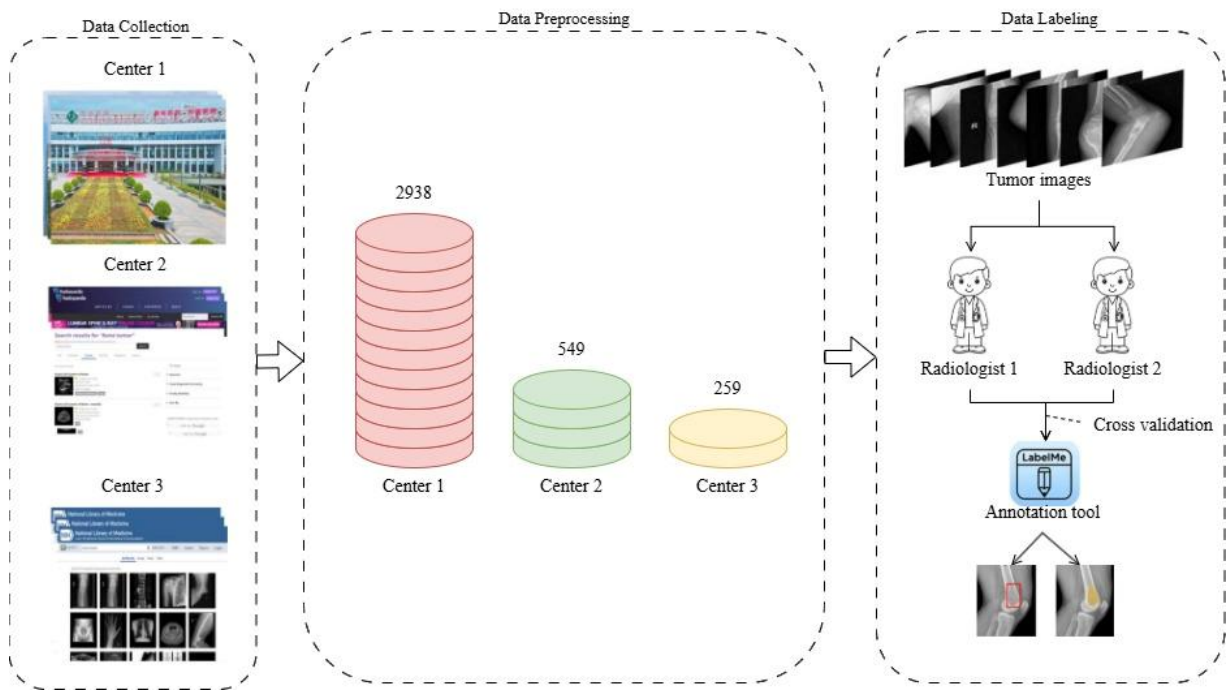
Bài nghiên cứu này sử dụng bộ dữ liệu BTXRD. BTXRD hay còn có tên gọi khác là Bone Tumor X-quang Radiograph Dataset) là một bộ dữ liệu công khai được xây dựng và công bố với mục tiêu hỗ trợ cộng đồng nghiên cứu phát triển các hệ thống học sâu trong chẩn đoán và phân tích hình ảnh y tế trên fgshare.com [12]. Đây là một đóng góp đáng kể trong bối cảnh dữ liệu y tế, đặc biệt là ảnh chụp X-quang về khối u xương.

([https://figshare.com/articles/dataset/A\\_Radiograph\\_Dataset\\_for\\_the\\_Classification\\_on\\_Localization\\_and\\_Segmentation\\_of\\_Primary\\_Bone\\_Tumors/27865398?file=50653575](https://figshare.com/articles/dataset/A_Radiograph_Dataset_for_the_Classification_on_Localization_and_Segmentation_of_Primary_Bone_Tumors/27865398?file=50653575))

Tổng kích thước của tập dữ liệu BTXRD là 801 MB, có thể tải xuống mà không cần đăng ký. Tập tin tải xuống ở định dạng zip, bao gồm các thư mục con “images”, “Annotations” và một tập tin “dataset.csv”. Thư mục “images” chứa tất cả các hình ảnh X-quang, với mỗi tên hình ảnh bắt đầu bằng “IMG” theo sau là một giá trị sáu chữ số đệm số không duy nhất và kết thúc bằng “.jpeg”. Thư mục “Annotations” chứa các tệp định dạng “.json” với thông tin chú thích để phân đoạn và đặt bounding boxes cho từng hình ảnh khối u xương. Tệp “dataset.csv” chứa bản tóm tắt thông tin dữ liệu cho tất cả 3.746 hình. Nhưng trong bài nghiên cứu này, tệp “dataset.csv” không được sử dụng mà thay vào đó sẽ sử dụng tệp “classification.xlsx” tự thiết kế chứa thông tin lâm sàng cơ bản của ảnh như id ảnh, trung tâm thu thập ảnh, tuổi, giới tính của bệnh nhân, nhãn loại tổn thương (lành tính, ác tính, hoặc không ung thư), vị trí khối u xương, góc chụp, nhóm chi, tạo độ vùng ung thư và các loại ung thư...

Bộ dữ liệu BTXRD được lấy từ ba trung tâm lưu trữ hình ảnh y tế. Trung tâm 1 bao gồm ba bệnh viện ở Trung Quốc: Bệnh viện liên kết đầu tiên của Đại học Y Guangxi,

bệnh viện liên kết của cao đẳng Y khoa Youjiang và Bệnh viện nhân dân Baise. Trung tâm 2 là Radiopaedia - nền tảng trực tuyến org. Và trung tâm 3 là cơ sở dữ liệu công cộng MedPix. Mỗi hình ảnh trong tập dữ liệu là duy nhất, nhưng một số có thể đến từ cùng một tổn thương, vì hình ảnh đã được thu thập từ các góc chụp khác nhau (góc nhìn chính diện, góc nhìn bên và góc nhìn xiên) cho mỗi tổn thương. Quy trình làm việc của nghiên cứu này được minh họa trong Hình 3. 1 và Bảng 3. 1, với các bước chính, bao gồm thu thập dữ liệu, tiền xử lý dữ liệu và chú thích dữ liệu.



Hình 3. 1. Quy trình thiết kế tập dữ liệu BTXRD

Bảng 3. 1. Đặt điểm cơ bản của tập dữ liệu BTXRD [12]

| BTXRD    | Age (y)           | Gender (M/F) | Normal | Benign tumor | Malignant tumor | Total |
|----------|-------------------|--------------|--------|--------------|-----------------|-------|
| Center 1 | 40.36 $\pm$ 19.23 | 1600/1338    | 1593   | 1110         | 235             | 2938  |
| Center 2 | 22.32 $\pm$ 14.57 | 326/223      | 259    | 214          | 76              | 549   |
| Center 3 | 22.58 $\pm$ 13.75 | 172/87       | 27     | 201          | 31              | 259   |

|       |                   |           |      |      |     |      |
|-------|-------------------|-----------|------|------|-----|------|
| Total | 37.53 $\pm$ 19.58 | 2098/1648 | 1879 | 1525 | 342 | 3746 |
|-------|-------------------|-----------|------|------|-----|------|

Từ Hình 3. 1 và Bảng 3. 1, có thể thấy tập dữ liệu BTXRD được thu thập từ 3 trung tâm, toàn bộ ảnh trong bộ dữ liệu đều được xử lý và ẩn danh theo đúng quy trình trong nghiên cứu y sinh học. Các dữ liệu trước khi được đưa vào tiền xử lý (data preprocessing) phải đảm bảo được các tiêu chí như:

- U xương nguyên phát được xác nhận bằng xét nghiệm mô bệnh hoặc bởi chuyên gia có kinh nghiệm.
- Bệnh nhân được chụp X-quang.
- Có thể biết tuổi và giới tính bệnh nhân.
- Loại trừ các ảnh có chất lượng kém, bị nhiễu hoặc không thấy khối u trên phim X-quang và khối u xuất hiện ở đầu, cổ, cột sống, ngực và bụng.
- Các ảnh xương bình thường cũng được thu thập.

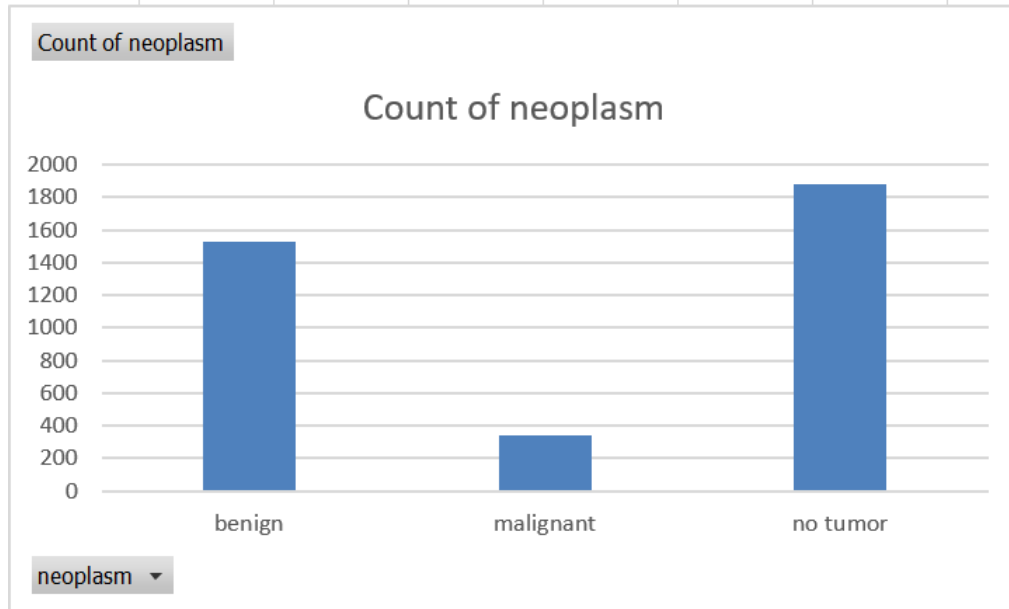
Dữ liệu thu thập cần phải được loại bỏ các mẫu khối u xương nguyên phát từ vùng đầu, cổ, ngực và bụng vì số lượng của chúng bị hạn chế và có nhiều hiện vật. Do đó, tập dữ liệu BTXRD chỉ bao gồm hình ảnh khối u xương từ các chi trên (bàn tay, xương trụ, xương quay, xương cánh tay, khớp cổ tay, khớp khuỷu tay và khớp vai), chi dưới (bàn chân, xương chày, xương ống chân, xương đùi, khớp mắt cá chân và khớp gối) và xương chậu (hông, khớp hông).

Tiếp đến, tiền xử lý dữ liệu sẽ giúp chuyển đổi hình ảnh DICOM sang định dạng ảnh “jpeg”, xóa các thông tin liên quan đến bệnh nhân có trong ảnh, lưu và thống kê lại độ tuổi, giới tính của bệnh nhân và góc chụp, vị trí chụp ảnh và vị trí xương và phân nhóm khối u xương trên tập dữ liệu. Cuối cùng, các ảnh khối u được chú thích dưới sự trợ giúp của hai chuyên gia về X-quang để sử dụng trong nhiều ứng dụng học sâu khác nhau sau này.

### 3.1.2. Đặc điểm dữ liệu



Tập dữ liệu này bao gồm 3.746 hình ảnh, với 1.879 hình ảnh xương bình thường, 1.525 hình ảnh khối u xương lành tính và 342 hình ảnh khối u xương ác tính được mô tả như Hình 3. 2.



Hình 3. 2. Biểu đồ biểu diễn số lượng hình ảnh xương bình thường, lành tính, ác tính.

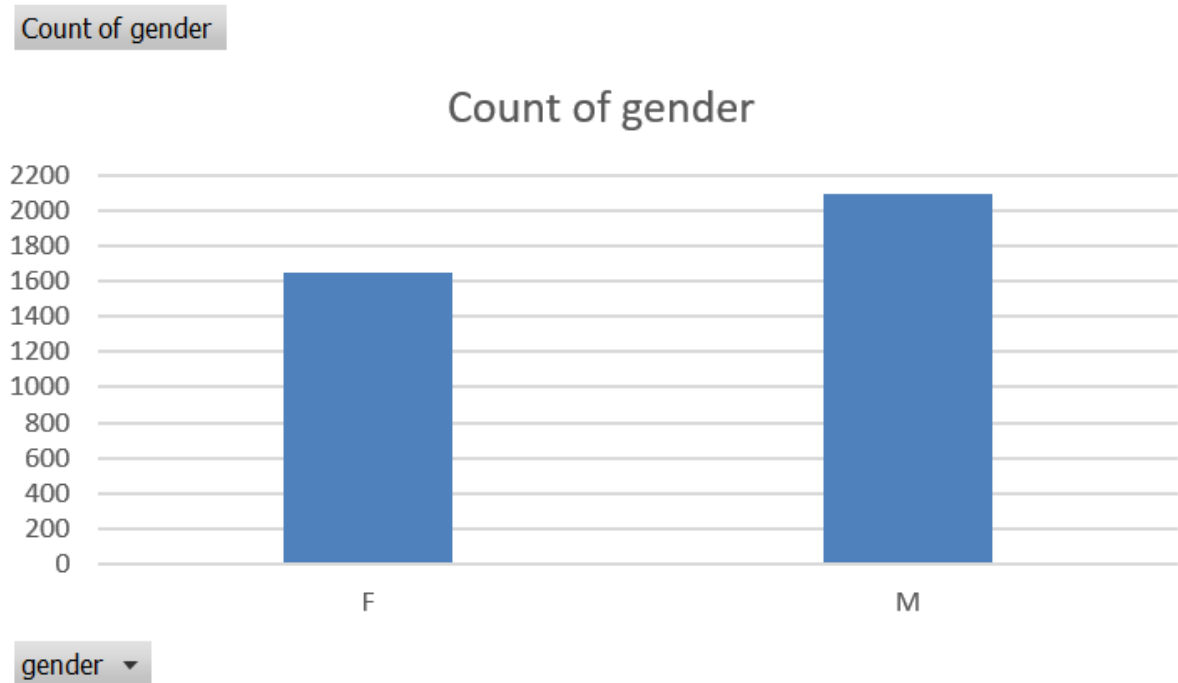
Từ Hình 3. 2, có thể thấy tập dữ liệu BTXRD cung cấp một nguồn dữ liệu lớn đủ để huấn luyện, kiểm thử và đánh giá các mô hình học sâu trong nhiều bài toán khác nhau: từ phân loại (classification), xác định vị trí (localization), đến phân đoạn khối u (segmentation).

Từ bộ dữ liệu BTXRD, tôi đã thống kê file excel classification.xlsx chứa các thông tin như Hình 3. 3.

| image_id | center | age | gender | neoplasm  | tumor_type  | img_type | skeletal_type | bones_type | points      | area   | image_area | tumor_ratio | image_filename  | type  |
|----------|--------|-----|--------|-----------|-------------|----------|---------------|------------|-------------|--------|------------|-------------|-----------------|-------|
| IMG00000 | 1      | 48  | F      | malignant | other mt    | frontal  | pelvis        | hip bone   | [[[2027.92  | 294967 | 7701561    | 3.83        | IMG0000001.jpeg | train |
| IMG00000 | 1      | 12  | M      | malignant | osteosarcom | frontal  | lower limb    | tibia      | [[[895.675  | 131041 | 4972480    | 2.64        | IMG0000002.jpeg | train |
| IMG00000 | 1      | 12  | M      | malignant | osteosarcom | lateral  | lower limb    | tibia      | [[[811.711  | 120650 | 6900832    | 1.75        | IMG0000003.jpeg | train |
| IMG00000 | 1      | 52  | M      | malignant | osteosarcom | frontal  | lower limb    | foot       | [[[1302.66  | 91065  | 7214337    | 1.26        | IMG0000004.jpeg | train |
| IMG00000 | 1      | 52  | M      | malignant | osteosarcom | oblique  | lower limb    | foot       | [[[1294.24  | 88469  | 7017152    | 1.26        | IMG0000005.jpeg | train |
| IMG00000 | 1      | 62  | M      | malignant | osteosarcom | frontal  | lower limb    | femur      | [[[591.913  | 806550 | 5627824    | 14.33       | IMG0000006.jpeg | train |
| IMG00000 | 1      | 62  | M      | malignant | osteosarcom | lateral  | lower limb    | femur      | [[[868.363  | 592080 | 5172384    | 11.45       | IMG0000007.jpeg | train |
| IMG00000 | 1      | 6   | M      | malignant | osteosarcom | frontal  | upper limb    | humerus    | [[[585.0, 9 | 61063  | 2650000    | 2.3         | IMG0000008.jpeg | train |
| IMG00000 | 1      | 6   | M      | malignant | osteosarcom | frontal  | upper limb    | humerus    | [[[671.621  | 49362  | 3085000    | 1.6         | IMG0000009.jpeg | val   |

Hình 3. 3. Hình ảnh 10 dòng dữ liệu đầu trong file classification.xlsx

Hình 3. 3 cho thấy một cái nhìn tổng quan về dữ liệu mà ảnh trong bộ BTXRD sẽ chứa. File giúp cung cấp thêm các thông tin lâm sàng cơ bản của ảnh như id ảnh, trung tâm thu thập ảnh, tuổi, giới tính của bệnh nhân, nhãn loại tổn thương (lành tính, ác tính, hoặc không ung thư), vị trí khối u xương, góc chụp, nhóm chi, tạo độ vùng ung thư và các loại ung thư... Để hiểu rõ hơn, hãy theo dõi các Hình 3. 4, Hình 3. 5, Hình 3. 6 và Bảng 3. 2 và Bảng 3. 3.



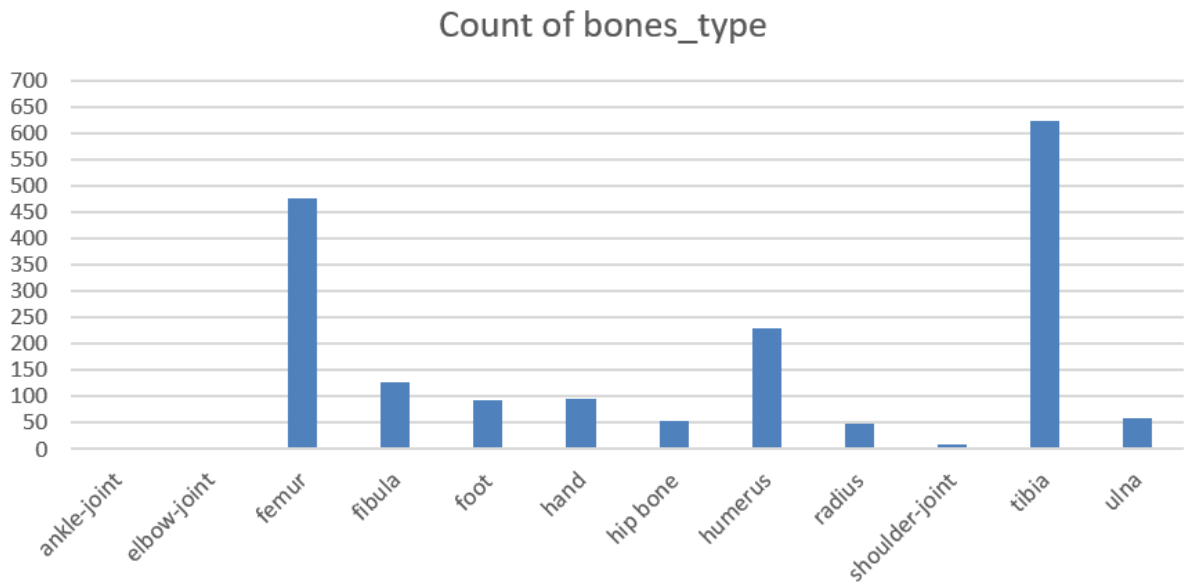
Hình 3. 4. Biểu đồ biểu diễn số lượng ảnh ung thư ở nam và nữ.

Bảng 3. 2. Bảng thống kê số lượng ảnh ung thư ở các vị trí xương khác nhau có trong bộ dữ liệu BTXRD

| <b>bones_type</b> | <b>Count of bones_type</b> |
|-------------------|----------------------------|
| ankle-joint       | 3                          |
| elbow-joint       | 5                          |
| ftôiu             | 477                        |
| fibula            | 127                        |

|                |     |
|----------------|-----|
| foot           | 93  |
| hand           | 97  |
| hip bone       | 55  |
| humerus        | 230 |
| radius         | 49  |
| shoulder-joint | 8   |
| tibia          | 624 |
| ulna           | 60  |

Count of bones\_type



bones\_type ▼

Hình 3. 5. Biểu đồ cột biểu diễn số lượng ảnh tại các vị trí xương khác nhau

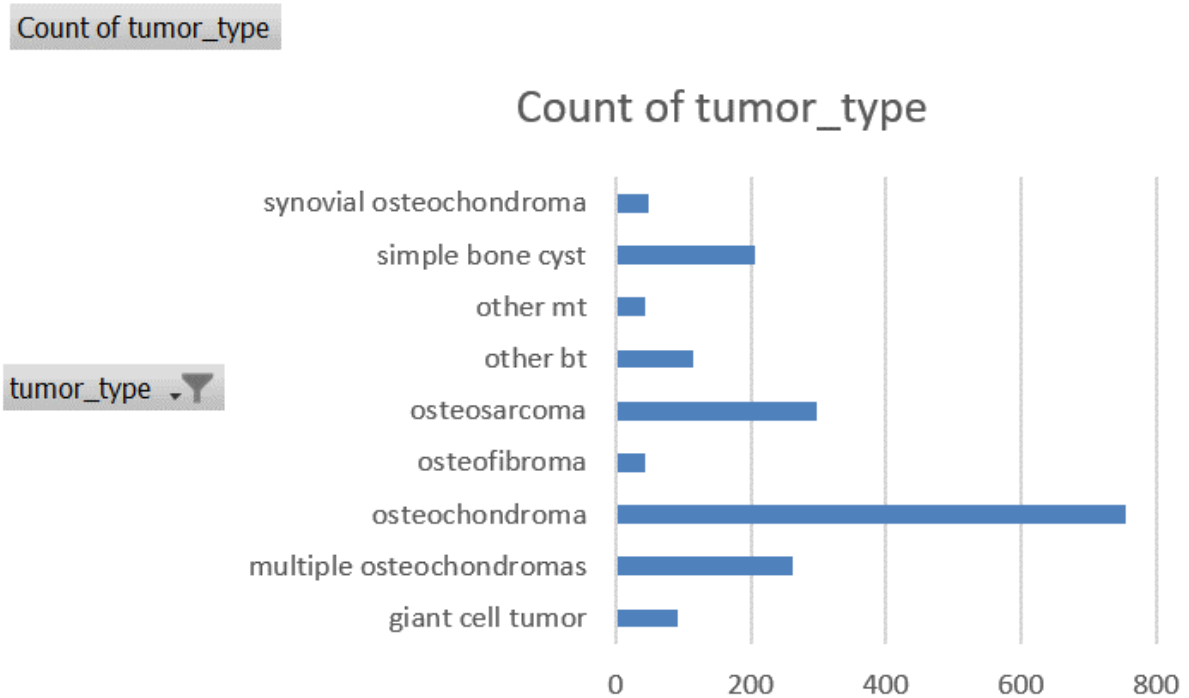
Bảng 3. 2 và Hình 3. 5 đã cung cấp cái nhìn tổng quan về số lượng ảnh ung thư tại các vị trí xương khác nhau trong bộ dữ liệu BTXRD. Có thể thấy rằng một số loại xương có số lượng ảnh ung thư cao hơn đáng kể so với các loại khác. Ví dụ, femur, tibia và humerus là ba loại xương có số lượng ảnh ung thư nhiều nhất, lần lượt là 477, 624 và

230 ảnh. Ngược lại, ankle-joint, elbow-joint và shoulder-joint có số lượng ảnh ung thư thấp hơn rõ rệt, với chỉ 3, 5 và 8 ảnh.

Không dừng lại ở đó, bộ dữ liệu BTXRD còn thể hiện tính đa dạng của nhiều loại ung thư thông qua Bảng 3. 3 và Hình 3. 6

*Bảng 3. 3. Bảng thống kê số lượng các loại ung thư có trong bộ dữ liệu BTXRD*

| tumor_type               | Count of tumor_type |
|--------------------------|---------------------|
| giant cell tumor         | 93                  |
| multiple osteochondromas | 263                 |
| osteochondroma           | 754                 |
| osteofibroma             | 44                  |
| osteosarcoma             | 297                 |
| other bt                 | 115                 |
| other mt                 | 45                  |
| simple bone cyst         | 206                 |
| synovial osteochondroma  | 50                  |



Hình 3. 6. Biểu đồ cột biểu diễn số lượng ảnh trong từng loại ung thư

Bảng 3. 3 kết hợp với Hình 3. 6 giúp tạo một cái nhìn bao quát hơn về số lượng các loại khối u xương có trong bộ dữ liệu. Có thể thấy osteochondroma là loại ung thư phổ biến nhất với số lượng vượt trội 754 ảnh, tiếp theo là osteosarcoma với số lượng 297 ảnh và multiple osteochondromas 263 ảnh. Ngược lại, osteofibroma và other mt có số lượng thấp nhất, chỉ tầm 44, 45 ảnh.

Sự chênh lệch rõ rệt này có thể phản ánh mức độ phổ biến của ung thư xương ở các vị trí khác nhau, giúp định hướng nghiên cứu y học tập trung vào những khu vực xương có nguy cơ cao hơn.

### 3.2. Quy trình huấn luyện mô hình học sâu

Một mô hình Deep Learning thường trải qua các giai đoạn sau trong quy trình huấn luyện bao gồm 4 bước:

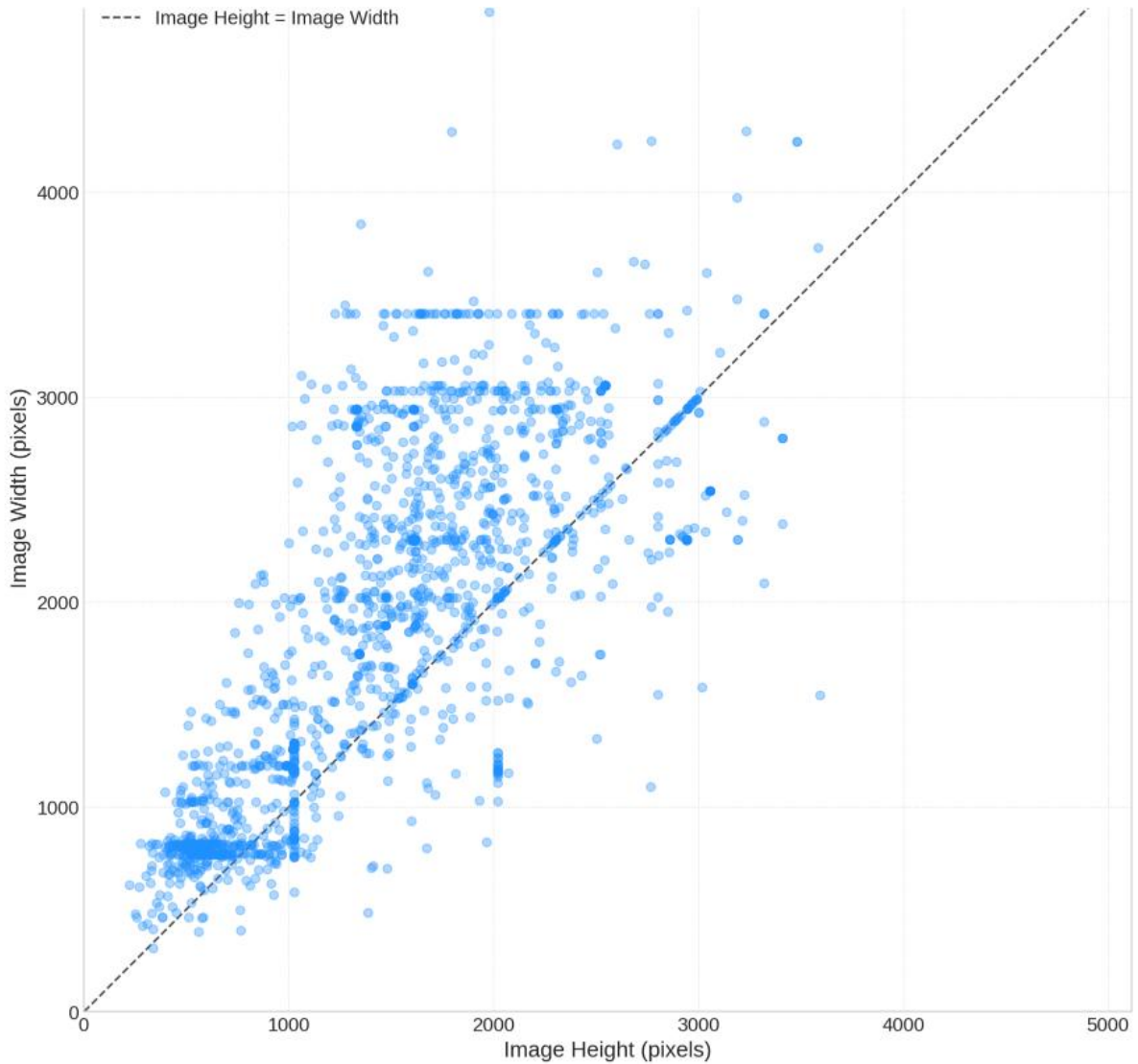
- Bước 1: Tiền xử lý dữ liệu. Dữ liệu được chuẩn hóa về kích thước, cường độ sáng, hoặc chuyển đổi định dạng (chẳng hạn từ RGB sang grayscale). Các kỹ thuật augmentation như xoay, lật, zoom, làm mờ,... được áp dụng để tăng cường độ đa dạng của dữ liệu và tránh overfitting.
- Bước 2: Xây dựng mô hình. Việc chọn kiến trúc phù hợp (MLP, CNN, RNN...) phụ thuộc vào loại bài toán. Thư viện Keras hoặc TensorFlow thường được sử dụng để xây dựng mô hình với các lớp phù hợp (Conv2D, Dense, Flatten...).
- Bước 3: Huấn luyện. Mô hình được tối ưu hóa bằng các thuật toán như SGD, Adam. Hàm mất mát phù hợp sẽ được chọn như cross-entropy cho phân loại, dice loss cho segmentation. Quá trình huấn luyện được theo dõi bằng các chỉ số như accuracy, loss.
- Bước 4: Đánh giá và tinh chỉnh. Sau khi huấn luyện, mô hình được kiểm thử với tập dữ liệu mới để đánh giá khả năng tổng quát hóa. Callbacks như EarlyStopping, ReduceLROnPlateau giúp dừng huấn luyện sớm khi mô hình không còn cải thiện.

### 3.3. Tiền xử lý dữ liệu

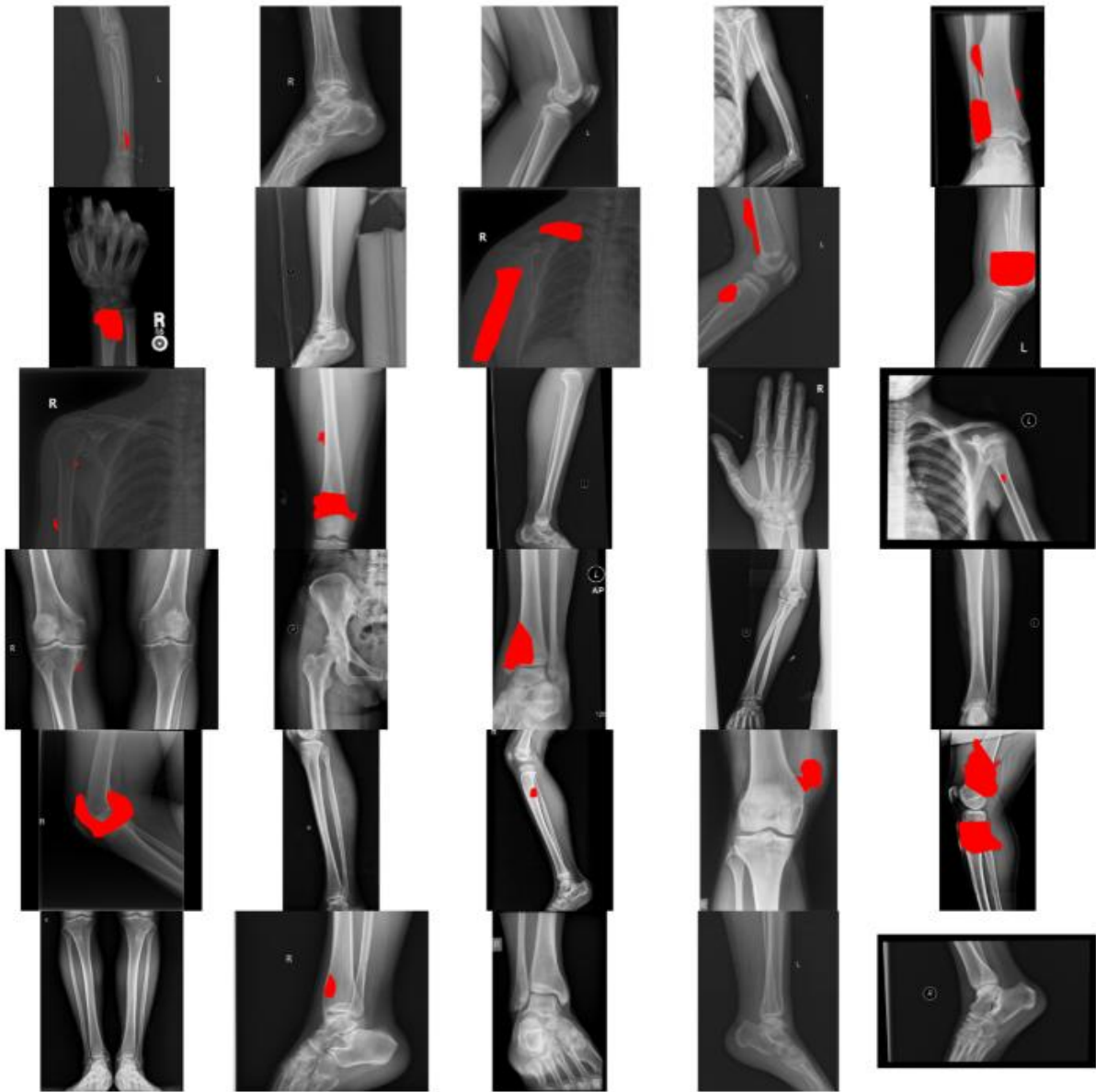
Trong giai đoạn đầu, tập dữ liệu BTXRD bao gồm tổng cộng 10.089 hình ảnh. Sau khi làm sạch dữ liệu, 3.746 hình ảnh đã được giữ lại. Trong số đó, 2.938 hình ảnh gồm 1.593 hình ảnh bình thường, 1.110 hình ảnh khối u xương lành tính và 235 hình ảnh khối u xương ác tính đến từ trung tâm 1; 549 hình ảnh gồm 259 hình ảnh bình thường, 214 hình ảnh khối u xương lành tính và 76 hình ảnh khối u xương ác tính từ trung tâm 2; và 259 hình ảnh gồm 27 hình ảnh bình thường, 201 hình ảnh khối u xương lành tính và 31 hình ảnh khối u xương ác tính từ trung tâm 3 [12].

#### 3.3.1. Chuẩn hóa kích thước ảnh

Do tổng hợp dữ liệu ảnh từ nhiều nguồn nên không thể tránh khỏi việc ảnh có độ phân giải và tỉ lệ khác nhau, bao gồm cả ảnh hình chữ nhật ngang và dọc, với kích thước dao động từ nhỏ (vài trăm pixel) đến lớn (trên 4.000 pixel) như Hình 3. 7 và Hình 3. 8 sau:



*Hình 3. 7. Biểu đồ biểu diễn kích thước ảnh trước khi xử lý.*



*Hình 3. 8. Ba mươi ảnh ngẫu nhiên có khoanh vùng khối u trước khi xử lý dữ liệu.*

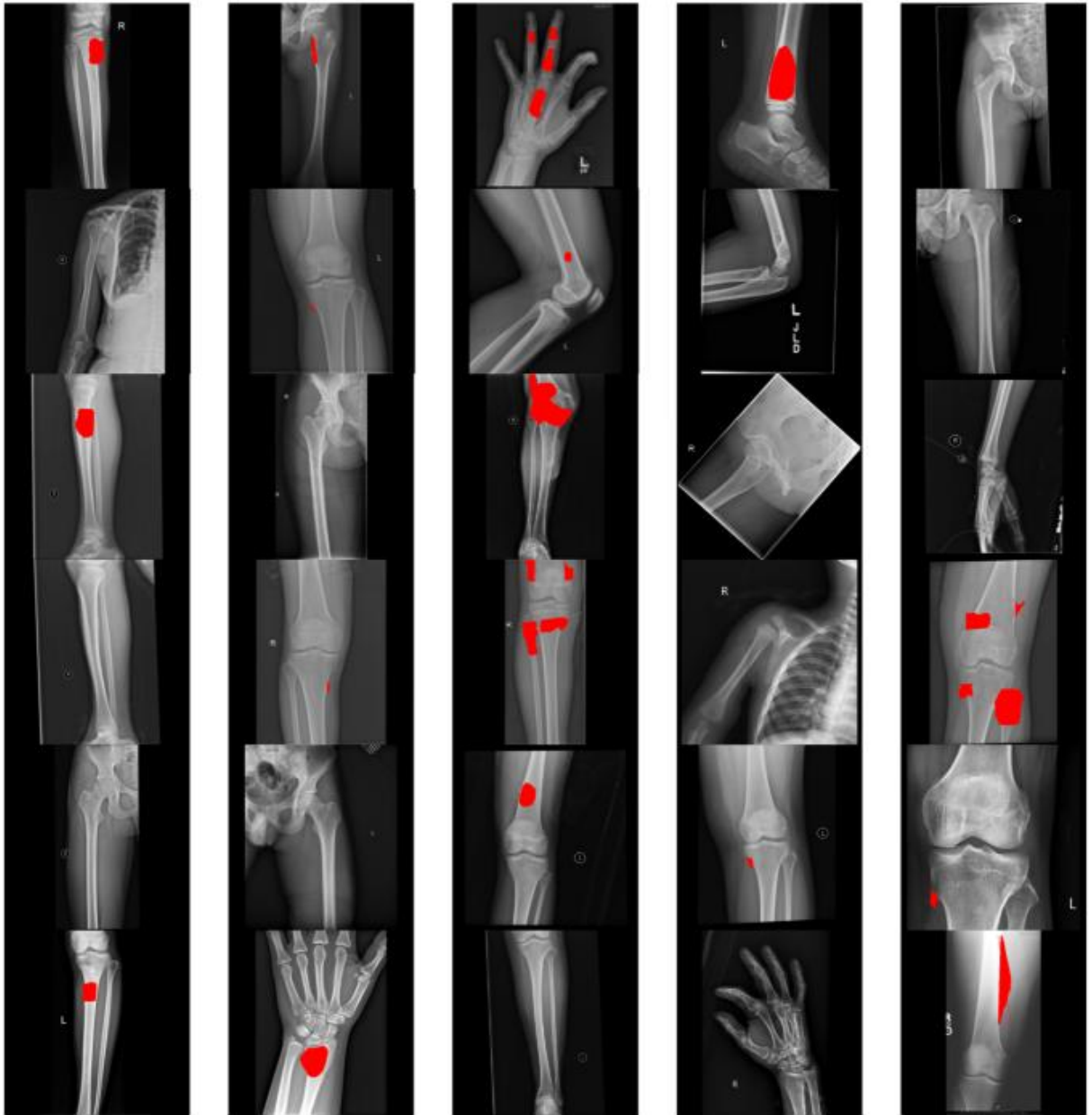
Biểu đồ phân tán trong Hình 3. 7 trên thể hiện mối quan hệ giữa chiều cao và chiều rộng của hình ảnh. Đường nét đứt biểu diễn cho “Image Height = Image Width” giúp xác định những hình ảnh có tỷ lệ cân đối, trong khi các điểm dữ liệu xung quanh phản ánh sự phân bố kích thước thực tế. Kết hợp với Hình 3. 8 có thể thấy rằng nhiều hình ảnh có chiều cao và chiều rộng tương đối đồng đều, nhưng cũng có những trường hợp kích thước chênh lệch đáng kể. Sự không đồng nhất này gây khó khăn khi đưa vào mô



hình học sâu, đặc biệt là các kiến trúc như U-Net vốn yêu cầu kích thước đầu vào cố định.

Để xử lý vấn đề này, tất cả ảnh cần được đưa về kích thước chuẩn  $512 \times 512$  pixel bằng cách kết hợp hai bước: resize và padding.

Đầu tiên, ảnh được resize sao cho chiều dài lớn nhất bằng 512 pixel mà vẫn đảm bảo giữ nguyên tỉ lệ. Sau đó, ảnh được padding (đệm thêm viền) để đạt kích thước vuông chuẩn. Padding được thực hiện bằng cách thêm các pixel màu đen có giá trị  $= 0$  vào cạnh còn thiếu. Sau khi chuẩn hóa ảnh, các ảnh sẽ có kích thước đồng nhất như Hình 3. 9.



*Hình 3. 9. Ba mươi ảnh ngẫu nhiên sau khi đồng nhất kích thước ảnh.*

Từ Hình 3. 9, có thể thấy việc chuẩn hóa ảnh không chỉ là một bước quan trọng để đảm bảo đầu vào đồng nhất cho mô hình, mà còn là bước giúp giảm bộ nhớ GPU tiêu thụ trong quá trình huấn luyện, đồng thời giữ được tỷ lệ vùng tổn thương tương đối trong ảnh.

### 3.3.2. Chia tập dữ liệu

Trước khi bắt đầu đi vào huấn luyện mô hình, việc phân chia bộ dữ liệu thành các tập con riêng biệt cũng là một bước không thể thiếu trong quy trình phát triển và đánh giá mô hình học máy, nhằm đảm bảo tính khách quan của mô hình.

Trong nghiên cứu này, bộ dữ liệu được phân chia thành ba tập con riêng biệt: tập huấn luyện (train), tập kiểm định (val) và tập kiểm thử (test) với tỉ lệ tương ứng là 72%, 18% và 10%. Trong đó:

- Tập huấn luyện được sử dụng để huấn luyện các tham số của mô hình.
- Tập kiểm định dùng để theo dõi hiệu suất của mô hình trong quá trình huấn luyện và tinh chỉnh tham số.
- Tập kiểm thử được giữ lại hoàn toàn và chỉ sử dụng một lần duy nhất để đánh giá khách quan hiệu suất cuối cùng của mô hình đã huấn luyện.

Để đảm bảo ảnh của từng loại khối u đều xuất hiện trong ba tập con, bài nghiên cứu sử dụng tham số stratify. Tham số này cho phép đảm bảo rằng tỷ lệ phân phối các lớp tumor\_type trong các tập train, val và test không bị mất cân bằng, nhằm tránh việc một số lớp hiếm chỉ xuất hiện trong một tập và không xuất hiện ở tập khác. Cụ thể, việc chia tập dữ liệu được tách thành 2 lần:

- Lần 1, dữ liệu được chia thành tập train+val và tập test. Tại bước này nếu dữ liệu chỉ có một lớp hoặc có số lượng ảnh của một loại ung thư nào đó có ít hơn 2 mẫu sẽ thực hiện chia ngẫu nhiên để tránh lỗi, kết quả chia lần 1 được thể hiện trong Bảng 3. 4 sau:

*Bảng 3. 4. Bảng kết quả chia dữ liệu lần 1*

|                       | <i>train+val</i> | <i>test</i> | <i>Total</i> |
|-----------------------|------------------|-------------|--------------|
| <i>osteochondroma</i> | 678              | 76          | 754          |
| <i>osteosarcoma</i>   | 267              | 30          | 297          |

|                                 |      |     |      |
|---------------------------------|------|-----|------|
| <i>multiple osteochondromas</i> | 237  | 26  | 263  |
| <i>simple bone cyst</i>         | 185  | 21  | 206  |
| <i>other bt</i>                 | 103  | 12  | 115  |
| <i>giant cell tumor</i>         | 84   | 9   | 93   |
| <i>synovial osteochondroma</i>  | 85   | 5   | 50   |
| <i>other mt</i>                 | 41   | 4   | 45   |
| <i>osteofibroma</i>             | 40   | 4   | 44   |
| <b>Total</b>                    | 1680 | 187 | 1867 |

- Lần 2, tiếp tục chia tập train+val thành train và val như Bảng 3. 5. Tương tự như lần 1, stratify tiếp tục được sử dụng để duy trì tỷ lệ lớp hợp lý trong tập val. Nếu sau chia lần 1, dữ liệu chỉ còn một lớp hoặc có lớp có số mẫu quá ít, stratify cũng sẽ bị bỏ qua để đảm bảo chia không lỗi.

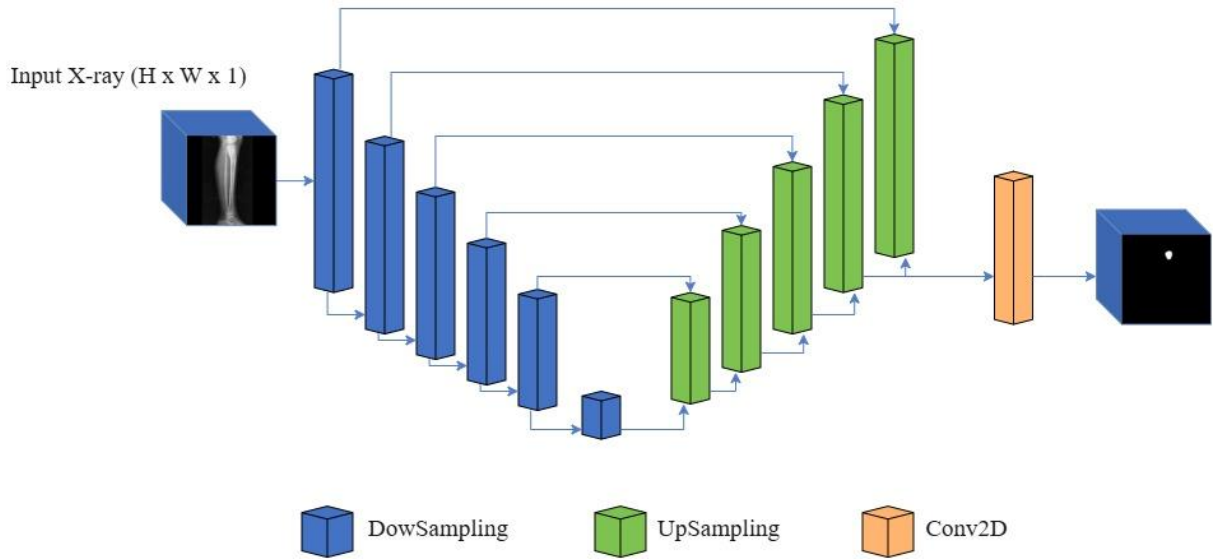
*Bảng 3. 5. Bảng kết quả chia dữ liệu lần 2*

|                                 | <i>train</i> | <i>val</i> | <i>Train + val</i> |
|---------------------------------|--------------|------------|--------------------|
| <i>osteochondroma</i>           | 542          | 136        | 678                |
| <i>osteosarcoma</i>             | 214          | 53         | 267                |
| <i>multiple osteochondromas</i> | 190          | 47         | 237                |
| <i>simple bone cyst</i>         | 148          | 37         | 185                |
| <i>other bt</i>                 | 82           | 21         | 103                |
| <i>giant cell tumor</i>         | 67           | 17         | 84                 |
| <i>synovial osteochondroma</i>  | 36           | 9          | 85                 |
| <i>other mt</i>                 | 33           | 8          | 41                 |
| <i>osteofibroma</i>             | 32           | 8          | 40                 |
| <b>Total</b>                    | 1344         | 336        | 1680               |

Sau khi chia tập dữ liệu như Bảng 3. 4, Bảng 3. 5, số lượng mẫu cụ thể trong mỗi tập được xác định là 1344 ảnh cho tập huấn luyện (train), 336 ảnh cho tập kiểm định (val) và 187 ảnh cho tập kiểm thử (test). Việc phân chia này đảm bảo rằng mô hình được đánh giá trên những dữ liệu mà nó chưa từng thấy trước đó, phản ánh đúng khả năng khái quát hóa của nó.

### 3.4. Phương pháp đề xuất

Trong nghiên cứu này, tôi đề xuất một phương pháp học sâu để giải quyết bài toán phân đoạn xác định ung thư xương dựa trên U-Net và cơ chế Attention trên ảnh X-quang. Mô hình đề xuất được gọi là SegUNet, dựa trên kiến trúc U-Net nổi tiếng [10] với cấu trúc đối xứng gồm một đường mã hóa (encoder) và một đường giải mã (decoder) được kết nối bằng skip connections. Hình 3. 10 sau minh họa tổng quan kiến trúc mô hình của này.

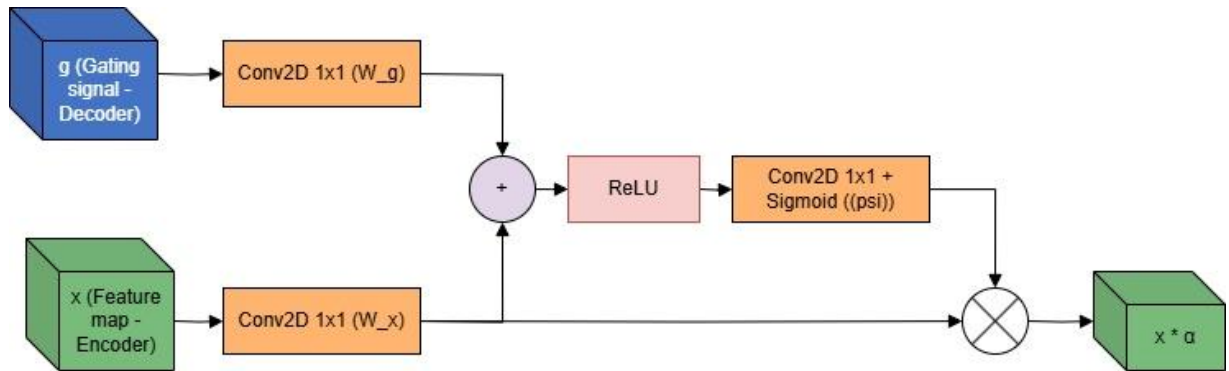


Hình 3. 10. Mô hình học sâu SegUnet

Trong Hình 3. 10, mô hình SegUNet bao gồm khối DownSampling, khối UpSampling và Conv2D. Mục tiêu chính của mô hình là phân loại chính xác từng pixel  $p = (x, y)$  từ ảnh X-quang đầu vào  $I \in \mathbb{R}^{W \times H \times 1}$  (trong đó W là chiều rộng, H là chiều

cao, và 1 là số kênh ảnh xám) thành một trong hai lớp: lớp 0 - nền (background) hoặc lớp 1 - vùng chứa khối u (tumor region).

Để giải quyết vấn đề này, tôi đã thiết kế và triển khai một mô hình dựa trên kiến trúc Unet và cơ chế Attention, có sơ đồ kiến trúc như Sơ đồ 3. 1.



Sơ đồ 3. 1. Sơ đồ kiến trúc của Attention Gate được sử dụng trong mô hình.

Sơ đồ 3. 1 minh họa chi tiết kiến trúc của Attention Gate - AG được tích hợp vào mô hình U-Net đề xuất trong bài nghiên cứu này. AG được thiết kế để tăng cường hiệu quả của skip connections bằng cách cho phép mô hình học cách tập trung có chọn lọc vào các vùng đặc trưng quan trọng và làm giảm ảnh hưởng của các đặc trưng nhiễu hoặc không liên quan từ nền.

Attention Gate nhận hai đầu vào:

- Tín hiệu cổng (g - Gating signal): Đây là bản đồ đặc trưng được lấy từ tầng giải mã (decoder) trước đó và đã được tăng kích thước (upsampled). Tín hiệu này chứa thông tin ngữ cảnh từ các tầng sâu hơn của mạng.
- Feature map đầu vào (x - Feature map): Đây là bản đồ đặc trưng được truyền từ tầng mã hóa (encoder) tương ứng thông qua skip connections. Nó chứa thông tin chi tiết về không gian và các đặc trưng cấp thấp hơn.

Quá trình xử lý bên trong Attention Gate diễn ra như sau:

- Cả tín hiệu cổng  $g$  và feature map  $x$  đầu tiên được chiếu qua các lớp tích chập 2D riêng biệt với kích thước bộ lọc  $1 \times 1$  (ký hiệu tương ứng lần lượt là  $W_g$  và  $W_x$ ). Các phép chiếu này nhằm mục đích đưa  $g$  và  $x$  về cùng một không gian đặc trưng và giảm số kênh nếu cần, tạo ra  $g'$  và  $x'$ .
- Sau đó,  $g'$  và  $x'$  được cộng theo từng phần tử (element-wise addition). Kết quả của phép cộng này được truyền qua một hàm kích hoạt ReLU để loại bỏ các giá trị âm và tăng tính phi tuyến.
- Tiếp đến, kết quả feature map tiếp tục đẩy qua một lớp tích chập 2D  $1 \times 1$  khác (ký hiệu là  $\psi$  - psi), theo sau bởi một hàm kích hoạt Sigmoid. Hàm Sigmoid đảm bảo rằng đầu ra của bước này, gọi là các hệ số chú ý (attention coefficients -  $\alpha$ ). Các hệ số này có giá trị nằm trong khoảng từ 0 đến 1. Mỗi giá trị trong  $\alpha$  biểu thị mức độ quan trọng của pixel tương ứng trong bản đồ đặc trưng  $x$ .
- Cuối cùng, các hệ số chú ý  $\alpha$  được nhân theo từng phần tử (element-wise multiplication) với feature map đầu vào ban đầu  $x$ . Kết quả là một feature map mới  $x * \alpha$ , trong đó các vùng đặc trưng quan trọng trong  $x$  có trọng số cao hơn bởi  $\alpha$  được "nhấn mạnh" (\*), trong khi các vùng ít liên quan hơn, có trọng số thấp hơn đã bị "làm mờ".

Feature map  $x * \alpha$  (\*) này sau đó sẽ được nối (concatenate) với bản đồ đặc trưng từ tầng giải mã, đã qua Conv2Dtranspose, để tiếp tục quá trình giải mã, giúp mô hình tái tạo lại bản đồ phân đoạn một cách chính xác hơn.

Ngoài việc tích hợp cơ chế Attention, bài nghiên cứu này còn cải tiến bằng cách sử dụng một hàm mất mát Combined Loss Function nhằm tăng cường khả năng phân đoạn chính xác, đặc biệt trong bối cảnh dữ liệu mất cân bằng thường gặp trong ảnh y tế.

Có thể thấy mô hình đề xuất SegUNet được xây dựng dựa trên kiến trúc U-Net kinh điển [10]. Kiến trúc này bao gồm hai phần chính: một đường mã hóa (Encoder Path)

để trích xuất các đặc trưng đa cấp độ từ ảnh đầu vào và một đường giải mã (Decoder Path) để tái tạo lại bản đồ phân đoạn với độ phân giải ban đầu. Tiếp sau đây, bài nghiên cứu sẽ đi sâu hơn vào encoder, decoder và hàm mất mát Combined Loss trước khi đi vào phương pháp huấn luyện.

#### **3.4.1. Mã hóa (Encoder)**

Phần mã hóa bao gồm một chuỗi các khối tích chập (Convolutional Blocks) lặp lại đi lặp lại. Mỗi khối tích chập bao gồm hai lớp tích chập 2D (Conv2D) với kích thước kernel 3x3, theo sau bởi một lớp chuẩn hóa theo lô (Batch Normalization) và một hàm kích hoạt ReLU. Sau mỗi khối tích chập, một lớp Max Pooling với kích thước 2x2 được sử dụng để giảm chiều không gian của feature map xuống một nửa và tăng receptive field.

Trong mỗi khối tích chập, tôi cũng tùy chọn sử dụng kỹ thuật Dropout với một tỷ lệ nhất định để giảm thiểu hiện tượng overfitting. Số lượng filter trong các lớp Conv2D tăng dần theo độ sâu của mạng (ví dụ: 32, 64, 128, 256, 512) để học các đặc trưng từ đơn giản đến phức tạp.

#### **3.4.2. Giải mã (Decoder)**

Phần giải mã (Decoder) cũng bao gồm một chuỗi các khối. Mỗi khối trong Decoder bắt đầu bằng một lớp Conv2DTranspose nhằm tăng kích thước (upsampling) feature map.

Để giữ lại thông tin chi tiết từ các tầng nông hơn của bộ mã hóa và hỗ trợ việc định vị chính xác, Skip Connections được thiết lập để nối các feature map từ encoder với các feature map tương ứng ở decoder.

Sau phép nối, một khối tích chập tương tự như trong encoder được áp dụng để tinh chỉnh các đặc trưng như hai lớp Conv2D 3x3, Batch Normalization, ReLU...

#### **3.4.3. Đề xuất hàm mất mát**



Để tránh tình trạng số lượng pixel nền và pixel khối u trong dữ liệu ảnh y tế mất cân bằng, việc lựa chọn hàm mất mát phù hợp là rất quan trọng. Bài nghiên cứu này đề xuất sử dụng một hàm mất mát kết hợp (Combined Loss Function) là sự tổng hợp giữa Dice Loss và Focal Loss.

### **Dice Loss:**

Dice loss chủ yếu tập trung vào việc tối đa hóa sự chồng chéo (overlap) giữa vùng dự đoán và vùng thực tế. Điều này giúp mô hình quan tâm nhiều hơn đến sự trùng khớp của các vùng khối u, đặc biệt là khối u có kích thước nhỏ, hạn chế bị ảnh hưởng bởi nền hoặc vùng lớn.

Công thức Dice Coefficient (Soft Dice Coefficient) [13, 14]:

$$DiceLoss = 1 - \frac{2 \sum_i P_i G_i}{\sum_i P_i + \sum_i G_i + \epsilon} \quad (1)$$

Trong đó:

- $p_i$  : giá trị xác suất (soft prediction) ở pixel  $i$  sau sigmoid/softmax (0 đến 1).
- $g_i$  : nhãn thực tế tại pixel  $i$  (0 hoặc 1).
- $\epsilon$ : hằng số rất nhỏ để tránh chia cho 0.

### **Focal Loss**

Focal Loss được thiết kế để giải quyết vấn đề mất cân bằng lớp bằng cách giảm trọng số của các mẫu dễ phân loại như pixel nền và tập trung hơn vào các mẫu khó thường là pixel khối u.

Công thức Focal Loss [15]:

$$FL = -\frac{1}{N} \sum_{i=1}^N \alpha_t (1 - p_i)^\gamma \log(p_i) \quad (2)$$

Trong đó:

- $p_i$ : Xác suất dự đoán tại pixel  $i$ .
- $N$ : Tổng số pixel.
- $\alpha_t$ : Hệ số cân bằng.
- $\gamma$ : Tham số điều chỉnh độ tập trung.

### Combined Loss Function

Từ công thức (1) và (2) suy ra công thức Combined Loss Function như sau [13]:

$$\text{Combined Loss} = \omega_{Dice} \times L_{Dice} + (1 - \omega_{Dice}) \times L_{FL} \quad (3)$$

Trong đó:

- $\omega_{Dice}$ : Trọng số của Dice Loss (thường trong khoảng 0.5 – 0.8)
- $L_{Dice}$ : Kết quả công thức Dice Loss (1).
- $L_{FL}$ : Kết quả công thức Focal Loss (2).

#### 3.4.4. Phương pháp huấn luyện

Trong bài nghiên cứu này, mô hình được huấn luyện bằng cách sử dụng optimizer Adam với tốc độ học ban đầu là 1e-4. Bài nghiên cứu sử dụng kỹ thuật giảm tốc độ học ReduceLROnPlateau dựa trên val\_dice\_coef\_metric\_tumor. Để giải quyết vấn đề mất cân bằng ở mức độ dữ liệu, kỹ thuật tính trọng số lớp (class weights) cũng được sử dụng trong quá trình huấn luyện để gán trọng số cao hơn cho lớp khối u.

Không những thế, kỹ thuật Early Stopping cũng được sử dụng để tránh quá khớp, dựa trên cùng một metric. Kích thước batch được đặt bằng 4 và mô hình được huấn luyện trong tối đa 300 epochs. Dữ liệu đầu vào được chuẩn hóa bằng cách trừ đi giá trị trung bình và chia cho độ lệch chuẩn tính toán trên tập huấn luyện. Các phép tăng cường dữ liệu (data augmentation) bao gồm lật ngẫu nhiên, xoay ngẫu nhiên, thay đổi độ

sáng/tương phản, zoom ngẫu nhiên,... cũng được áp dụng trong quá trình huấn luyện để tăng tính đa dạng của dữ liệu và cải thiện khả năng tổng quát hóa của mô hình.

## CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM

Chương này trình bày chi tiết các kết quả thực nghiệm thu được từ việc huấn luyện và đánh giá mô hình SegUNet trên tập dữ liệu ảnh x-quang xương đã được chuẩn bị. Các chỉ số đánh giá định lượng và ví dụ phân đoạn trực quan sẽ được phân tích để làm rõ hiệu năng của phương pháp đề xuất.

### 4.1. Kết quả thực nghiệm

#### 4.1.1. Mô tả tập dữ liệu thực nghiệm

Nghiên cứu này sử dụng bộ dữ liệu BTXRD, một tập hợp các hình ảnh X-quang hai chiều cùng với các chú thích chi tiết về vị trí khối u được cung cấp dưới dạng tập JSON. Sau quá trình phân chia dữ liệu, tập train bao gồm 1344 mẫu, tập validation chứa 336 mẫu và tập test được sử dụng để đánh giá cuối cùng hiệu suất của mô hình bao gồm 187 mẫu.

Các mẫu trong tập test được lựa chọn ngẫu nhiên, đảm bảo bao gồm tất cả các loại khối u và đảm bảo không trùng lặp với dữ liệu đã được sử dụng trong quá trình train hay validation, nhằm cung cấp một đánh giá khách quan về khả năng tổng quát hóa của mô hình trên dữ liệu mới.

Tập thử nghiệm test thể hiện sự đa dạng về kích thước, hình dạng và vị trí của khối u, cũng như các biến thể về chất lượng hình ảnh và sự hiện diện của các cấu trúc giải phẫu khác có thể gây nhiễu. Việc đánh giá trên một tập test đại diện như vậy là rất quan trọng để khẳng định tính hiệu quả và độ tin cậy của mô hình phân đoạn được đề xuất trong các tình huống gần với thực tế lâm sàng.

#### 4.1.2. Môi trường thực nghiệm và cấu hình huấn luyện

Toàn bộ quá trình thực nghiệm, bao gồm huấn luyện – train và đánh giá mô hình – test, được thực hiện trên một hệ thống máy tính với cấu hình phần cứng bao gồm Intel(R) Xeon(R) CPU, 16GB RAM và một card đồ họa GPU P100 với 16384MiB VRAM, cho

phép tăng tốc đáng kể các phép toán học sâu. Về phần mềm, nghiên cứu sử dụng môi trường Python phiên bản 3.10.12, cùng với thư viện TensorFlow phiên bản 2.17.1 và Keras API phiên bản 3.5.0 để xây dựng, huấn luyện và đánh giá mô hình. Các thư viện hỗ trợ tính toán khoa học như NumPy và xử lý ảnh như OpenCV, Pillow cũng được tận dụng. Nền tảng Weights & Biases (WandB) phiên bản 0.19.1 được tích hợp để theo dõi và ghi lại chi tiết quá trình huấn luyện cũng như các kết quả thực nghiệm.

Để đánh giá trên tập test, mô hình SegUNet đã được huấn luyện với các siêu tham số tối ưu bao gồm tốc độ học ban đầu là  $1e-4$  cho thuật toán Adam, kích thước batchsize là 4 và hàm mất mát kết hợp Dice Loss và Focal Loss. Quá trình huấn luyện kéo dài 300 epochs với các kỹ thuật điều chuẩn như L2 regularization =  $1e-5$  và Dropout = 0.3. Dữ liệu huấn luyện cũng được tăng cường bằng các phép biến đổi hình học và cường độ điểm ảnh để cải thiện khả năng tổng quát hóa.

#### 4.1.3. Kết quả đánh giá trên tập test

Sau khi hoàn tất quá trình huấn luyện và lựa chọn được mô hình có hiệu suất tốt nhất dựa trên tập validation, mô hình này đã được sử dụng để đánh giá trên tập test. Kết quả đánh giá được tổng hợp và trình bày trong Bảng 4. 1 dưới đây, bao gồm các chỉ số chính thường được sử dụng để đo lường hiệu suất của các thuật toán phân đoạn ảnh y tế. Các chỉ số này cung cấp một cái nhìn toàn diện về khả năng của mô hình trong việc xác định chính xác vùng khối u cũng như độ chính xác tổng thể trên toàn bộ ảnh.

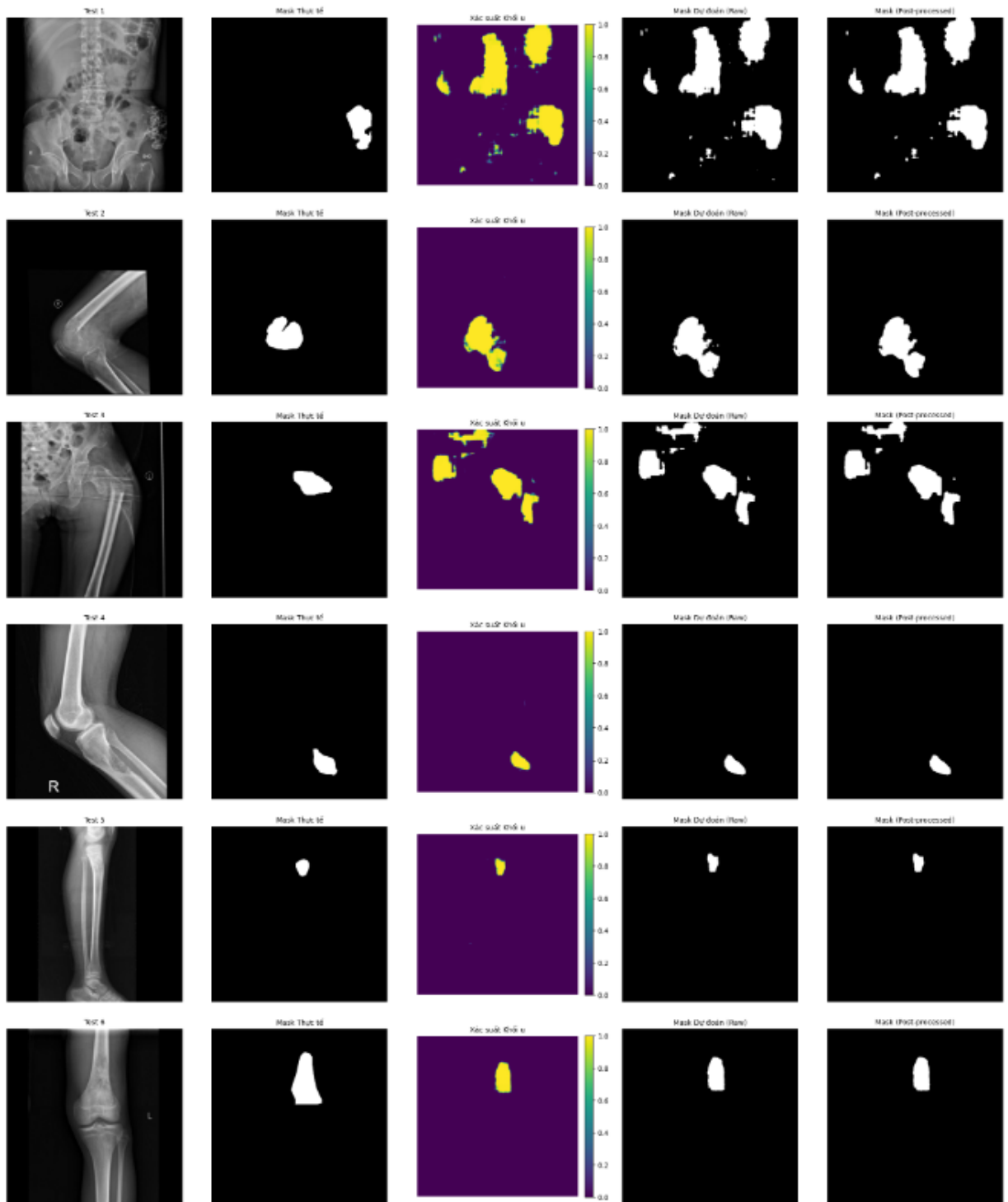
*Bảng 4. 1. Bảng kết quả sau khi đánh giá trên tập kiểm thử*

| ĐẦU RA         | CHỈ SỐ |
|----------------|--------|
| TEST LOSS      | 0.3079 |
| TEST ACCURACY  | 0.7921 |
| TEST MEAN IOU  | 0.4901 |
| TEST TUMOR IOU | 0.4596 |
| TEST DICE COEF | 0.5633 |

|                       |        |
|-----------------------|--------|
| <b>TEST PRECISION</b> | 0.6105 |
| <b>TEST RECALL</b>    | 0.5792 |
| <b>TEST F1-SCORE</b>  | 0.5948 |

Các giá trị trong Bảng 4. 1 đã phản ánh hiệu suất cuối cùng của mô hình trên dữ liệu mới, là cơ sở quan trọng để phân tích và đưa ra các kết luận về tính hiệu quả của phương pháp đề xuất.

Bên cạnh các chỉ số định lượng, việc kiểm tra trực quan kết quả phân đoạn của mô hình trên các mẫu cụ thể từ tập thử nghiệm cũng đóng vai trò quan trọng trong việc hiểu rõ hơn về điểm mạnh và điểm yếu của nó. **Error! Reference source not found.** dưới đây minh họa một số trường hợp dự đoán điển hình, bao gồm cả những trường hợp mô hình hoạt động tốt và những trường hợp gặp khó khăn. Mỗi hàng trong hình ảnh hiển thị ảnh X-quang gốc, mặt nạ thực tế (ground truth mask) được chú thích bởi chuyên gia, và mặt nạ dự đoán (predicted mask) được tạo ra bởi mô hình SegUNet đã huấn luyện.

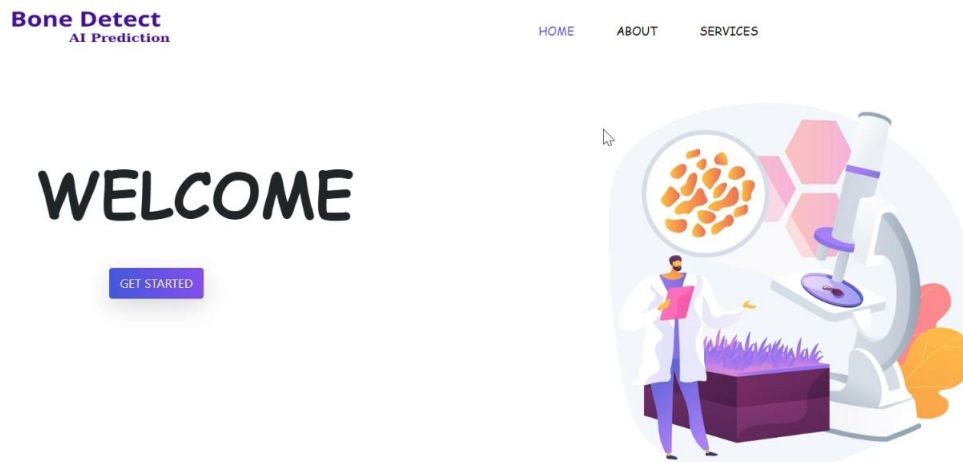


Hình 4. 1. Trực quan hóa kết quả ngẫu nhiên phân đoạn của mô hình từ tập kiểm thử

Hình 4. 1 cho thấy ảnh X-quang gốc, mask thực tế (ground truth), bản đồ xác suất khối u do mô hình dự đoán, kết quả phân đoạn thô (raw prediction) và kết quả sau khi áp dụng xử lý hậu kỳ (post-processed prediction). Tuy nhiên do chỉ số accuracy chưa cao nên vẫn còn một số ảnh không xác định được.

#### 4.1.4. Xây dựng ứng dụng xác định ảnh ung thư xương dựa trên Unet và cơ chế Attention

Sau khi mô hình SegUNet được huấn luyện và đánh giá định lượng trên tập thử nghiệm với các chỉ số hiệu suất đã được trình bày ở mục 4.1.3. Bước quan trọng tiếp theo là trực quan hóa khả năng hoạt động của mô hình trong một kịch bản gần với ứng dụng thực tế. Để hiện thực hóa điều này, một ứng dụng web mang tên "Bone Detect - AI Prediction" đã được phát triển với giao diện như Hình 4. 2.



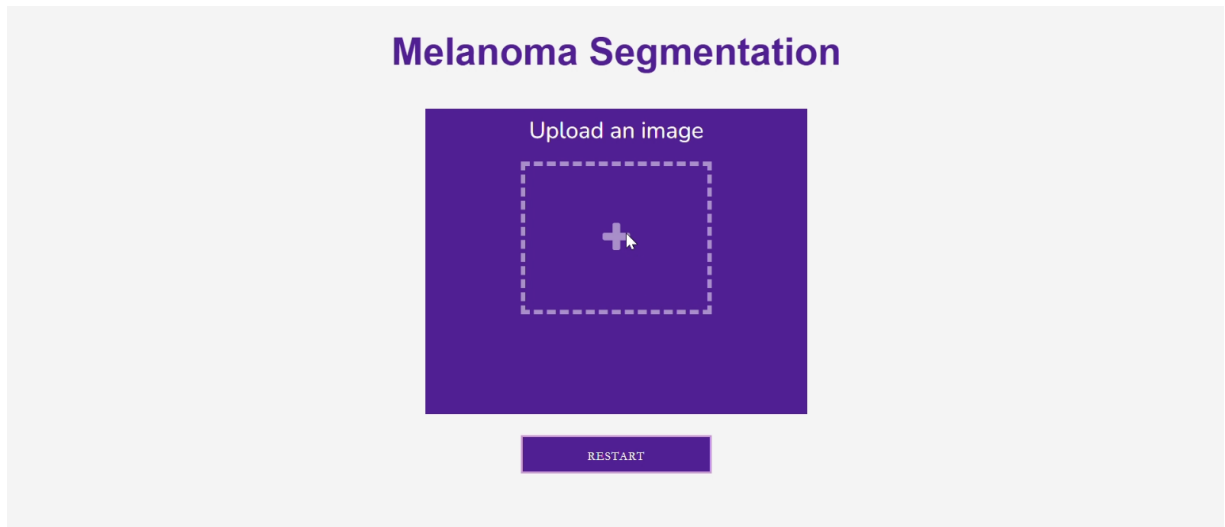
Hình 4. 2. Giao diện ứng dụng web mang tên "Bone Detect - AI Prediction"

Ứng dụng web Hình 4. 2 này tích hợp mô hình đã huấn luyện, cho phép người dùng tải lên hình ảnh X-quang xương và nhận lại kết quả phân đoạn các vùng nghi ngờ là tổn thương ung thư do mô hình dự đoán. Việc xây dựng ứng dụng web như Hình 4. 2 không chỉ nhằm mục đích minh họa khả năng hoạt động của mô hình một cách trực quan, mà

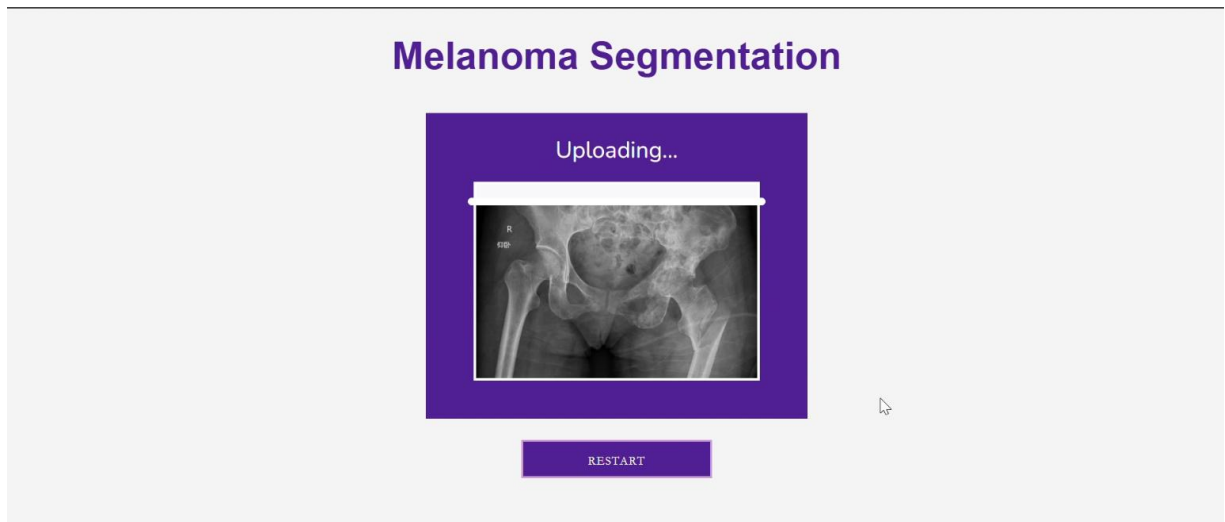


còn là một bước đầu trong việc đánh giá tính khả dụng và tiềm năng triển khai của giải pháp AI trong thực tế.

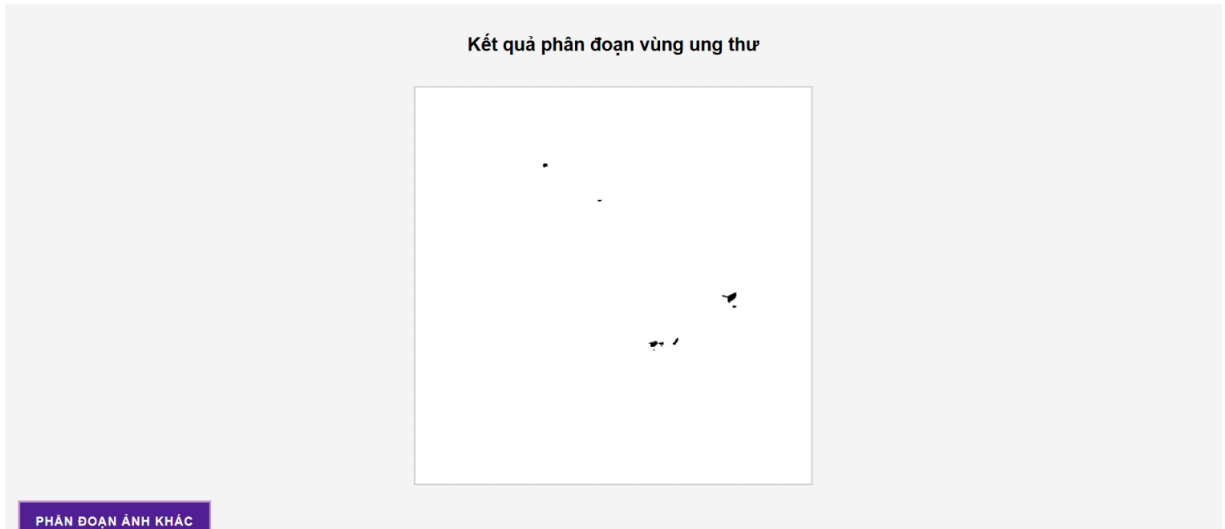
Người dùng có thể dễ dàng tương tác với hệ thống thông qua các bước đơn giản: chọn và tải lên tệp ảnh X-quang như Hình 4. 3 và Hình 4. 4, sau đó hệ thống sẽ tự động xử lý ảnh bằng mô hình học sâu đã được huấn luyện và hiển thị kết quả phân đoạn trực tiếp trên giao diện và hiển thị kết quả như Hình 4. 5.



*Hình 4. 3. Giao diện trước khi upload ảnh cần nhận diện khối u*



*Hình 4. 4. Giao diện khi upload ảnh khối u*



*Hình 4. 5. Giao diện kết quả phân đoạn vùng ung thư*

Hình 4. 5 đã minh họa kết quả đầu ra của ứng dụng "Bone Detect" khi xử lý một số hình ảnh X-quang cụ thể từ Hình 4. 4.

#### **4.1. Phân tích kết quả**

Kết quả định lượng thu được từ việc đánh giá tập test như Bảng 4. 1 cho thấy mô hình SegUNet đạt được hiệu suất trung bình trong bài toán phân đoạn khối u từ ảnh X-quang. Cụ thể, hệ số Dice cho lớp khối u (Test Dice Coef) là 0.5633 và chỉ số Intersection over Union cho lớp khối u (Test Tumor IoU) là 0.4596. Các giá trị này phản ánh mức độ tương đồng giữa vùng dự đoán và vùng thực tế của khối u, với giá trị càng gần 1 thì độ chính xác càng cao. Chỉ số Accuracy tổng thể đạt 0.7921, cho thấy mô hình có khả năng phân loại đúng phần lớn các pixel trong ảnh. Tuy nhiên, do chỉ số này chưa cao nên vẫn còn một số ảnh không xác định được và có thể bị ảnh hưởng lớn bởi sự mất cân bằng lớp nếu vùng nền chiếm đa số.

Từ các ví dụ dự đoán trực quan được trình bày trong Hình 4. 1, có thể thấy mô hình hoạt động hiệu quả trong việc xác định các khối u có ranh giới tương đối rõ ràng và kích thước đủ lớn. Trong những trường hợp này, mặt nạ dự đoán có độ tương đồng cao với

mặt nạ thực tế, cho thấy mô hình đã học được các đặc trưng hình thái quan trọng của khối u. Tuy nhiên, mô hình cũng bộc lộ một số hạn chế ở các trường hợp khó hơn. Ví dụ, đối với các khối u có kích thước rất nhỏ hoặc có hình dạng không điển hình, mô hình có thể bỏ sót hoặc chỉ phân đoạn được một phần. Tương tự, khi ranh giới khối u mờ nhạt hoặc bị che lấp bởi các cấu trúc giải phẫu khác có cường độ điểm ảnh tương tự, độ chính xác của việc xác định đường biên có thể giảm sút.

Dựa trên các kết quả định lượng được trình bày trong Bảng 4. 1, mô hình SegUNet đã đạt được mức độ chính xác chấp nhận được. Cụ thể, chỉ số Tumor IoU là và Dice Coefficient cho lớp khối u lần lượt là 0.4596 và 0.5633 cho thấy khả năng của mô hình trong việc xác định vùng tổn thương.

Giá trị Accuracy cao cho thấy mô hình đã hội tụ tốt trên tập kiểm thử. Tuy nhiên cần được xét cẩn trọng do sự mất cân bằng giữa lớp nền và lớp khối u. Các chỉ số tập trung vào lớp khối u như Tumor IoU, Dice, Precision, Recall và F1-Score cung cấp cái nhìn sâu sắc hơn về hiệu năng thực sự của mô hình đối với đối tượng quan tâm.

Qua các ví dụ trực quan hóa trong Hình 4. 5, có thể thấy mô hình hoạt động tạm ổn trong việc xác định các khối u có kích thước từ trung bình đến lớn. Bản đồ xác suất khối u cho thấy mô hình có khả năng tập trung vào đúng vùng nghi ngờ, và kết quả sau xử lý hậu kỳ đã loại bỏ được nhiều nhiễu và làm mượt ranh giới.

## 4.2. Ứng dụng thực tiễn

Kết quả nghiên cứu cho thấy mô hình SegUNet có tiềm năng ứng dụng trong thực tiễn lâm sàng như một công cụ hỗ trợ các bác sĩ chẩn đoán hình ảnh trong việc phát hiện và khoanh vùng ung thư xương trên ảnh X-quang. Việc tự động xác định vùng nghi ngờ có thể giúp giảm tải công việc, tăng tính khách quan và đẩy nhanh quá trình chẩn đoán, đặc biệt trong các cơ sở y tế có khối lượng công việc lớn.

Mô hình này có thể được tích hợp vào một hệ thống phần mềm hoàn chỉnh, cung cấp giao diện cho phép bác sĩ tải ảnh lên, nhận kết quả phân đoạn trực quan và các thông tin định lượng về vùng tổn thương, ví dụ: kích thước, vị trí ước tính...

Tuy nhiên, để có thể triển khai rộng rãi trong thực tế, mô hình cần được kiểm định thêm trên các tập dữ liệu lớn hơn, đa dạng hơn từ nhiều nguồn khác nhau và được đánh giá bởi các chuyên gia y tế. Các nghiên cứu sâu hơn về việc tối ưu hóa mô hình, giảm thiểu các trường hợp dự đoán sai và cải thiện khả năng diễn giải của mô hình cũng là những hướng phát triển quan trọng.

## CHƯƠNG 5: KẾT LUẬN VÀ KHUYẾN NGHỊ

### 5.1. Kết luận

Nghiên cứu này đã thành công trong việc xây dựng và đánh giá một mô hình học sâu dựa trên kiến trúc U-Net tích hợp cơ chế Attention, để giải quyết bài toán phân đoạn vùng khối u xương trên ảnh X-quang. Mô hình được huấn luyện và kiểm thử trên tập dữ liệu BTXRD, cho thấy khả năng xác định các vùng tổn thương với độ chính xác khá cao, nhưng vẫn còn hạn chế ở một số mặt.

Cụ thể, mô hình SegUNet đã đạt được các chỉ số đánh giá chính trên tập kiểm thử như sau: Test Tumor IoU là 0.4596, Dice score là 0.5633, recall là 0.5792. Những kết quả này minh chứng cho tiềm năng của việc ứng dụng học sâu, đặc biệt là kiến trúc U-Net, trong việc tự động hóa quá trình phân tích ảnh y tế, hỗ trợ chẩn đoán ung thư xương. Việc sử dụng hàm mất mát kết hợp giữa Dice Loss và Focal Loss cùng với kỹ thuật cân bằng lớp đã góp phần cải thiện khả năng xử lý dữ liệu mất cân bằng của mô hình.

#### 5.1.1. Ưu điểm

Mô hình đề xuất sở hữu nhiều ưu điểm nổi bật, đặc biệt là khả năng tự động hóa trong việc xác định và khoanh vùng khối u xương trên ảnh X-quang, giúp giảm thiểu đáng kể thời gian và công sức cho các bác sĩ so với phương pháp phân tích thủ công truyền thống.

Dựa trên kiến trúc U-Net đã được kiểm chứng rộng rãi trong lĩnh vực phân đoạn ảnh y tế, mô hình đảm bảo khả năng học các đặc trưng không gian và tái tạo chi tiết tốt. Hơn nữa, khi tích hợp thêm cơ chế Attention, mô hình còn cho thấy khả năng cải thiện việc tập trung vào các vùng đặc trưng của khối u, giúp tăng nhẹ độ chính xác phân đoạn so với phiên bản U-Net cơ bản.

Ngoài ra, mô hình cũng được trang bị các kỹ thuật xử lý mất cân bằng lớp hiệu quả thông qua việc áp dụng hàm mất mát kết hợp Dice và Focal Loss cùng với kỹ thuật class

weights, giúp giải quyết tốt thách thức phổ biến về sự chênh lệch lớn giữa lớp nền và lớp khối u trong dữ liệu y tế.

Những kết quả ban đầu khả quan cho thấy mô hình có tiềm năng phát triển thành công cụ hỗ trợ chẩn đoán, góp phần sàng lọc và đánh giá nhanh các trường hợp nghi ngờ ung thư xương.

Cuối cùng, kiến trúc mô hình cũng cho thấy tính linh hoạt và khả năng mở rộng cao khi có thể dễ dàng điều chỉnh hoặc thử nghiệm thêm các biến thể khối tích chập, cơ chế attention phức tạp hơn để phù hợp với các tập dữ liệu hoặc yêu cầu đặc thù khác.

### **5.1.2. Khó khăn, thách thức**

Mặc dù vậy, mô hình vẫn đối mặt với nhiều thách thức, đặc biệt là hạn chế về chất lượng và số lượng dữ liệu huấn luyện, trong khi việc thu thập và chú thích ảnh X-quang chất lượng cao rất tốn kém và đòi hỏi chuyên môn. Độ phức tạp và đa dạng của khối u xương với nhiều hình dạng, kích thước, vị trí và độ tương phản khác nhau cũng gây khó khăn cho mô hình trong việc phân đoạn chính xác, nhất là ở các vùng có ranh giới mờ hoặc dễ nhầm lẫn với cấu trúc giải phẫu bình thường.

Bên cạnh đó, khả năng nhận diện nhiều khối u hoặc các khối u nhỏ vẫn còn hạn chế và việc huấn luyện và vận hành mô hình cũng đòi hỏi tài nguyên tính toán lớn.

## **5.2. Hướng phát triển**

Để vượt qua các thách thức trên, nghiên cứu cần mở rộng và đa dạng hóa tập dữ liệu bằng cách thu thập thêm từ nhiều nguồn, cải thiện quy trình chú thích dữ liệu với sự hỗ trợ của công cụ và chuyên gia. Về mô hình, nên thử nghiệm các biến thể U-Net tiên tiến hơn như U-Net++, UNet 3+, hoặc kết hợp Transformer với CNN (TransUNet) để tăng cường khả năng học ngữ cảnh. Đồng thời, nghiên cứu các cơ chế attention phức tạp hơn, áp dụng deep supervision, cũng như thử nghiệm các hàm mất mát mới như Lovasz-

Softmax, Tversky loss và tối ưu hóa siêu tham số với các kỹ thuật tự động hóa. Việc áp dụng augmentation nâng cao với các phép biến đổi mạnh cũng là một hướng khả thi.

## DANH MỤC TÀI LIỆU THAM KHẢO

- [1] B. v. N. T. Phương. "Ung thư xương và những điều cần biết." <https://suckhoedoisong.vn/ung-thu-xuong-nguyen-nhan-trieu-chung-phong-ngua-va-dieu-tri-169240707001743615.htm> (accessed 05/01/2025).
- [2] S. k. Đ. sống. "Ung thư xương: Nguyên nhân, triệu chứng, phòng ngừa và điều trị." <https://suckhoedoisong.vn/ung-thu-xuong-nguyen-nhan-trieu-chung-phong-ngua-va-dieu-tri-169240707001743615.htm> (accessed 05/01/2025).
- [3] T. R. Mirabello L, Savage SA., "Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program," *Cancer* vol. 115, no. 7, pp. 1531–1543, 2009, doi: 10.1002/cncr.24121.
- [4] N. T. Tuấn, *Deep Learning cơ bản*. 2020.
- [5] L. B. Y. LeCun, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
- [6] I. S. A. Krizhevsky, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks,," *Communications of the ACM*, vol. 60, pp. 84–90, 2017, doi: 10.1145/3065386.
- [7] K. S. a. A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>.
- [8] W. L. C. Szegedy, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, và A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.



- [9] X. Z. K. He, S. Ren, và J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
- [10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," presented at the Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.
- [11] O. Oktay, J. Schlemper, and e. al, "Attention U-Net: Learning Where to Look for the Pancreas.," *Medical Imaging with Deep Learning 2018*, 2018.
- [12] Y. H. ShunhanYao, XiaoyuWang, Yiwen Zhang, IanCosta Paixao, ZhikangWang, Charla LuChai, HongtaoWang, Dinggui Lu, Geoffrey IWebb, Shanshan Li, YumingGuo, QingfengChen & Jiangning Son, "A Radiograph Dataset for the Classification, Localization, and Segmentation of Primary Bone Tumors," *Scientific Data*, vol. 43, pp. 312-324, 2024, doi: 10.1038/s41597-024-04311-y.
- [13] W. L. Carole H Sudre, Tom Vercauteren, Sébastien Ourselin, M. Jorge Cardoso, "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," *Proceedings of 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565-571, 2017, doi: 10.1109/3DV.2016.79.
- [14] C. M. Daniel Bell, "Dice Similarity Coefficient (DSC)," *Ecology*, vol. 26, pp. 297-302, 1945, doi: 10.2307/1932409.
- [15] P. G. Tsung-Yi Lin, Ross Girshick, Kaiming He, Piotr Dollár, "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017, doi: 10.1109/ICCV.2017.324.