ỦY BAN NHÂN DÂN TP HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

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

SINH VIÊN THỰC HIỆN: LÊ THỊ THANH NGÂN

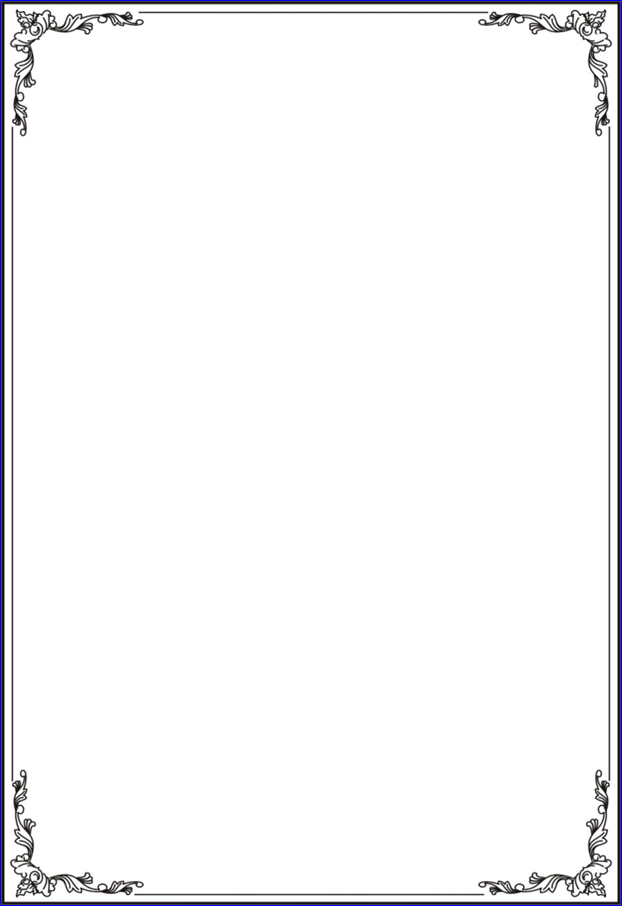
**ĐỀ TÀI: XÂY DỰNG ỨNG DỤNG XÁC ĐỊNH UNG THƯ XƯƠNG DỰA TRÊN UNET VÀ CƠ CHẾ ATTENTION**

**BÁO CÁO TIẾN ĐỘ 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**

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

SINH VIÊN THỰC HIỆN: LÊ THỊ THANH NGÂN

**ĐỀ TÀI: XÂY DỰNG ỨNG DỤNG XÁC ĐỊNH UNG THƯ XƯƠNG DỰA TRÊN UNET VÀ CƠ CHẾ ATTENTION**

**BÁO CÁO TIẾN ĐỘ 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 cửa 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**

**MỤC LỤC**

[**MỞ ĐẦU** 1](#_Toc195902829)

[**1.** **Lý do chọn đề tài** 1](#_Toc195902830)

[**2.** **Mục đích và nhiệm vụ nghiên cứu** 2](#_Toc195902831)

[**3.** **Đối tượng và phạm vi nghiên cứu** 2](#_Toc195902832)

[**4.** **Phương pháp nghiên cứu** 3](#_Toc195902833)

[**5.** **Tính mới và đóng góp của đề tài** 4](#_Toc195902834)

[**6.** **Tiến độ và kế hoạch nghiên cứu** 4](#_Toc195902835)

[**7.** **Cấu trúc của luận văn** 5](#_Toc195902836)

[**NỘI DUNG** 6](#_Toc195902837)

[**CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI** 6](#_Toc195902838)

[**1.** **Tổng quan về ung thư xương** 6](#_Toc195902839)

[**1.1.** **Định nghĩa và phân loại ung thư xương** 6](#_Toc195902840)

[**1.2.** **Nguyên nhân và tỉ lệ mắc ung thư xương** 6](#_Toc195902841)

[**1.2.1.** **Nguyên nhân gây ra ung thư xương** 6](#_Toc195902842)

[**1.2.2.** **Tỷ lệ mắc bệnh ung thư xương** 7](#_Toc195902843)

[**2.** **Học sâu và ứng dụng trong y học** 7](#_Toc195902844)

[**3.** **Quy trình huấn luyện mô hình học sâu (Deep Learning)** 7](#_Toc195902845)

[**4.** **Các nghiên cứu liên quan** 8](#_Toc195902846)

[**4.1.** **Dự án BTCV** 8](#_Toc195902847)

[**4.2.** **Dự án Spleen (U-Net 3D)** 10](#_Toc195902848)

[**4.3.** **Kết luận** 12](#_Toc195902849)

[**CHƯƠNG 2: CƠ SỞ LÝ THUYẾT** 13](#_Toc195902850)

[**1.** **Giới thiệu về Deep Learning** 13](#_Toc195902851)

[**2.** **Mạng nơ-ron trong Deep Learning** 13](#_Toc195902852)

[**3.** **Deep Learning trong xử lý ảnh y tế** 14](#_Toc195902853)

[**3.1.** **U-net trong phân đoạn hình ảnh y tế** 15](#_Toc195902854)

[**3.1.1.** **Kiến trúc và nguyên lý hoạt động của U-net** 15](#_Toc195902855)

[**3.1.2.** **Lợi thế của U-net trong phân đoạn 2D và 3D** 16](#_Toc195902856)

[**3.2.** **Cơ chế Attention** 17](#_Toc195902857)

[**3.2.1.** **Khái niệm về Attention trong học sâu** 17](#_Toc195902858)

[**3.2.2.** **Attention kết hợp với U-net trong phân đoạn hình ảnh y tế** 17](#_Toc195902859)

[**3.3.** **Ứng dụng thực tiễn của U-net và Attention trong y tế** 18](#_Toc195902860)

[**4.** **Công nghệ PyTorch, MONAI và PyTorch Lightning** 18](#_Toc195902861)

[**CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU** 19](#_Toc195902862)

[**1.** **Thu thập và tiền xử lý dữ liệu** 19](#_Toc195902863)

[**2.** **Thiết kế mô hình** 19](#_Toc195902864)

[**3.** **Huấn luyện mô hình** 19](#_Toc195902865)

[**4.** **Đánh giá và tối ưu hóa** 19](#_Toc195902866)

[**CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM** 20](#_Toc195902867)

[**1.** **Kết quả thực nghiệm** 20](#_Toc195902868)

[**2.** **Phân tích kết quả** 20](#_Toc195902869)

[**3.** **Ứng dụng thực tiễn** 20](#_Toc195902870)

[**KẾT LUẬN VÀ KHUYẾN NGHỊ** 21](#_Toc195902871)

[**1.** **Kết luận** 21](#_Toc195902872)

[**1.1.** **Ưu điểm** 21](#_Toc195902873)

[**1.2.** **Khó khăn, thách thức** 21](#_Toc195902874)

[**2.** **Khuyến nghị / Hướng phát triển** 21](#_Toc195902875)

[**TÀI LIỆU THAM KHẢO** 22](#_Toc195902876)

**DANH MỤC CÁC TỪ VIẾT TẮT**

|  |  |
| --- | --- |
| MONAI | Medical Open Network for AI |
| BTCV | Brain Tumor Classification and Segmentation |
| CPU | Central Processing Unit |
| GPU | Graphics Processing Unit |
|  |  |
|  |  |

**DANH MỤC BẢNG BIỂU**

**DANH MỤC HÌNH ẢNH**

**No table of figures entries found.**

# **MỞ ĐẦU**

Ung thư xương là một căn bệnh hiếm 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 CT hoặc MRI. 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 mở ra cơ hội nâng cao hiệu quả điều trị và quản lý bệnh.

1. **Lý do chọn đề tài**

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 3D để hỗ trợ chẩn đoán vẫn còn 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”, chúng 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 Attention Mechanism 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 cho thấy 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. **Mục đích và nhiệm vụ nghiên cứu**
   1. **Mục đích**

Mục đích 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 3D. 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. **Nhiệm vụ**

Để đạt được mục tiêu đó, 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, cắt bỏ vùng không liên quan và cân chỉnh kích thước ảnh nhằm đảm bảo dữ liệu đầu vào đạt chất lượng cao nhất.
* Phát triển và huấn luyện mô hình trên bộ dữ liệu y tế thực tế, sử dụng các thuật toán tối ưu hóa như Dice Loss hoặc Cross-Entropy 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. **Đối tượng và phạm vi nghiên cứu**
   1. **Đối tượng**

Đối tượng nghiên cứu của đề tài là các hình ảnh y tế 3D, cụ thể là các ảnh CT hoặc MRI 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 nền tảng dữ liệu cần thiết để phát triển hệ thống.

* 1. **Phạm vi nghiên cứu**

Phạm vi nghiên cứu tập trung vào việc phát triển mô hình phân đoạn hình ảnh, nhằm xác định chính xác các vùng tổn thương hoặc khối u trong xương. 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 3D, 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. **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 CT hoặc MRI của bệnh nhân ung thư xương sẽ được chuẩn hóa, cân chỉnh kích thước và loại bỏ các vùng không liên quan. Tiền xử lý dữ liệu là bước quan trọng 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 Attention Mechanism để tận dụng cả thông tin tổng quát và chi tiết trong hình ảnh. Các thuật toán tối ưu như Dice Loss hoặc Cross-Entropy Loss được áp dụng 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 Similarity Coefficient và Intersection over Union (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. **Tính mới và đóng góp của đề tài**
   1. **Về mặt khoa học**

Đề tài tiên phong 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. **Về mặt thực tiễn**

Về thực tiễn, ứng dụng 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. **Tiến độ và kế hoạch nghiên cứu**

Đề tài này được thực hiện từ ngày 27/12/2024 đến khoảng ngày 18/05/2025.

* 27/12/2024: Tra cứu và tìm hiểu về một số nghiên cứu liên quan đến chuẩn đoán ung thư bằng hình ảnh dự trên U-net.
* 01/01/2025: Tìm hiểu và chạy thử các MONAI Tutorial.

1. **Cấu trúc của luận văn**

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

Chuơng 1: TỔNG QUAN VỀ ĐỀ TÀI

Chuơng 2: PHÂN TÍCH MỘT SỐ DỰ ÁN LIÊN QUAN

Chuơng 3: PHÂN TÍCH VÀ THIẾT KẾ ĐỀ TÀI

Chuơng 4: KẾT QUẢ THỰC NGHIỆM

# **NỘI DUNG**

**CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI**

1. **Tổng quan về ung thư xương**
   1. **Định nghĩa và phân loại ung thư xương**

Ung thư xương là một loại bệnh ung thư bắt nguồn từ các tế bào trong xương. Bệnh này có thể xảy ra ở bất kỳ xương nào trong cơ thể, nhưng phổ biến nhất là ở xương dài như xương đùi, xương chày (ở chân) hoặc xương cánh tay.

Ung thư xương được chia làm hai loại chính:

* Ung thư xương nguyên phát: Đây là dạng ung thư xuất phát từ các tế bào của mô xương, mô sụn hoặc mô liên kết trong xương; 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 (di căn): Là loại ung thư xuất hiện do các tế bào ung thư từ một cơ quan khác như phổi, vú, thận,… di căn đến xương.
  1. **Nguyên nhân và tỉ lệ mắc ung thư xương**
     1. **Nguyên nhân gây ra ung thư xương**

Cho đến nay, nguyên nhân chính gây ra ung thư xương vẫn chưa được xác định. Tuy nhiên, các nhà khoa học đã xác định được một số yếu tố có thể làm tăng nguy cơ phát triển ung thư xương, bao gồm:

* Các yếu tố di truyền và sinh học (như đột biến gen).
* Do tiếp xúc với liều cao bức xạ ion hóa.
* Ảnh hưởng từ môi trường và hóa chất.
* Tiền sử gia đình mắc ung thư.
* Bệnh lý xương từ trước.
  + 1. **Tỷ lệ mắc bệnh ung thư xương**

Ung thư xương là một bệnh hiếm gặp, chiếm chưa đến 1% tổng số các loại ung thư. 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. Bệnh thường xuất hiện ở trẻ em, thanh thiếu niên và người cao tuổi, với osteosarcoma chiếm ưu thế ở nhóm tuổi từ 10-20. Nam giới có nguy cơ mắc bệnh cao hơn nữ giới, với tỷ lệ nam/nữ khoảng 1,5:1 [1, 2].

1. **Học sâu và ứng dụng trong y học**

Sự phát triển của Deep Learning đã mở ra một kỷ nguyên mới trong lĩnh vực xử lý ảnh y tế, nơi yêu cầu độ chính xác cao, khả năng tổng quát hóa mạnh và tốc độ xử lý nhanh. Trong các hệ thống truyền thống, việc phân tích ảnh y tế phụ thuộc rất nhiều vào chuyên gia và quy trình thủ công, dẫn đến độ chính xác không đồng đều và tốn thời gian. Deep Learning, đặc biệt là các mạng nơ-ron tích chập (CNN), đã giúp tự động hóa và nâng cao độ tin cậy trong phân tích ảnh y tế.

Các mô hình Deep Learning không chỉ hỗ trợ bác sĩ trong việc chẩn đoán mà còn có thể phát hiện ra các mẫu dữ liệu tiềm ẩn mà mắt người khó nhận thấy. Chúng có thể được huấn luyện để phát hiện tổn thương trong ảnh X-quang, phân vùng các khối u trong ảnh MRI hoặc CT scan, và nhận diện các tế bào ung thư trong ảnh sinh thiết với độ chính xác cao.

Tuy nhiên, việc ứng dụng Deep Learning trong ảnh y tế vẫn có nhiều thách thức như: sự thiếu hụt dữ liệu gán nhãn chất lượng cao, sự chênh lệch giữa ảnh từ các thiết bị khác nhau. Do đó, việc nghiên cứu kỹ các kiến trúc mạng, quy trình tiền xử lý và đánh giá mô hình một cách bài bản là điều kiện tiên quyết để phát triển các hệ thống phân tích ảnh y tế đáng tin cậy.

1. **Quy trình huấn luyện mô hình học sâu (Deep Learning)**

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.

1. **Các nghiên cứu liên quan**
   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 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.

* 1. **Dự án Spleen (U-Net 3D)**

Dự án Spleen nhắm 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 ánh 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 (1.5mm × 1.5mm × 2.0mm), 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.

* 1. **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 2: CƠ SỞ LÝ THUYẾT**

1. **Giới thiệu về Deep Learning**

Deep Learning (học sâu) là một nhánh thuộc lĩnh vực trí tuệ nhân tạo (AI), phát triển từ Machine Learning (học máy) 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 .

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.

1. **Mạng nơ-ron trong Deep Learning**

Mạng nơ-ron trong Deep Learning gồm nhiều kiến trúc khác nhau, mỗi loại được thiết kế để phù hợp với từng dạng dữ liệu và bài toán đặc thù.

* MLP (Multi-Layer Perceptron): Là dạng mạng nơ-ron cơ bản nhất gồm các lớp fully-connected. MLP được ứng dụng trong những bài toán phân loại đơn giản với dữ liệu đã được làm phẳng (flatten).
* CNN (Convolutional Neural Network): Là kiến trúc chủ đạo trong thị giác máy tính, CNN sử dụng các lớp tích chập (convolutional layers) để tự động trích xuất các đặc trưng không gian từ ảnh. Một số mô hình CNN kinh điển bao gồm:
  + LeNet-5: Một trong những kiến trúc CNN đầu tiên, được dùng để nhận diện chữ viết tay.
  + AlexNet: Góp phần làm Deep Learning bùng nổ với kiến trúc sâu và hiệu quả cao.
  + VGGNet: Mạng sâu, sử dụng các kernel đồng nhất kích thước 3x3 giúp trích xuất đặc trưng chi tiết.
  + ResNet: Sử dụng các khối Residual giúp vượt qua vấn đề vanishing gradient, cho phép xây dựng mạng cực sâu.
* RNN (Recurrent Neural Network): Được thiết kế để xử lý dữ liệu chuỗi có thứ tự như âm thanh, văn bản. Các biến thể như LSTM và GRU giúp cải thiện khả năng ghi nhớ thông tin dài hạn.

1. **Deep Learning trong xử lý ảnh y tế**

Xử lý ảnh y tế là một trong những ứng dụng quan trọng và nhạy cảm của Deep Learning. Các loại ảnh như X-quang, MRI, CT, sinh thiết (biopsy) đều chứa thông tin quan trọng hỗ trợ bác sĩ trong việc chẩn đoán và điều trị bệnh lý. Việc ứng dụng học sâu trong phân đoạn hình ảnh y tế đã giúp giảm thiểu sự can thiệp thủ công và tăng tốc độ xử lý dữ liệu. Điều này không chỉ tiết kiệm thời gian mà còn nâng cao hiệu quả công việc.

Bên cạnh đó, độ chính xác cao của các mô hình học sâu giúp nhận diện và phân đoạn các vùng bất thường trong hình ảnh y tế với độ chính xác vượt trội, từ đó hỗ trợ bác sĩ đưa ra chẩn đoán chính xác hơn. Đáng kể đến là tính linh hoạt của các mô hình học sâu, cho phép áp dụng cho nhiều loại hình ảnh y tế khác nhau, từ hình ảnh 2D đến 3D, góp phần cải thiện hiệu quả chẩn đoán và điều trị trong các lĩnh vực y tế đa dạng.

* 1. **U-net trong phân đoạn hình ảnh y tế**
     1. **Kiến trúc và nguyên lý hoạt động của U-net**

U-Net là một mô hình mạng nơ-ron tích chập được thiết kế đặc biệt cho phân đoạn ảnh y tế, đặc biệt là các bài toán phân đoạn các cấu trúc trong ảnh sinh học, chẳng hạn như tế bào, mô hoặc các khối u trong ảnh chụp y tế. U-Net được giới thiệu bởi Ronneberger và các cộng sự vào năm 2015 với mục tiêu cung cấp một giải pháp hiệu quả cho việc phân đoạn trong các bức ảnh y tế có độ phân giải thấp và thường xuyên gặp vấn đề với sự thiếu hụt dữ liệu [3].

Các thành phần chính trong kiến trúc U-net bao gồm: phần mã hóa (Encoder), phần giải mã (Decoder) và Skip connections.

* Encoder (bộ mã hóa):
  + Phần này bao gồm các lớp convolution kép kèm theo max-pooling.
  + Nhiệm vụ chính là mã hóa thông tin từ hình ảnh đầu vào thành các đặc trưng quan trọng thông qua việc giảm dần kích thước không gian và tăng số lượng kênh đặc trưng.
  + Max-pooling giúp giảm kích thước ảnh để trích xuất các đặc trưng tổng quát hơn, nhưng vẫn giữ lại các thông tin cần thiết.
* Decoder (bộ giải mã):
  + Phần này thực hiện tái tạo lại kích thước không gian của hình ảnh từ các đặc trưng đã mã hóa, sử dụng các lớp up-convolution (hoặc transposed convolution).
  + Các đặc trưng từ encoder được liên kết (skip connections) với các đặc trưng trong decoder để giữ lại thông tin chi tiết từ các bước mã hóa trước đó, giúp cải thiện độ chính xác của phân đoạn.
* Skip connections:
  + Các kết nối này truyền trực tiếp thông tin từ encoder sang decoder, đảm bảo rằng các thông tin chi tiết không bị mất đi trong quá trình mã hóa.
  + Skip connections là điểm nổi bật của U-net, giúp mô hình vượt trội trong các bài toán phân đoạn hình ảnh có biên phức tạp.

Phần mã hóa giúp giảm kích thước ảnh qua các lớp tích chập và pooling, trong khi phần giải mã giúp khôi phục thông tin không gian bằng upsampling và sử dụng skip connections để giữ lại thông tin chi tiết từ các tầng trước. Cơ chế này giúp U-Net đạt độ chính xác cao, đặc biệt với các đối tượng nhỏ trong ảnh y khoa.

* + 1. **Lợi thế của U-net trong phân đoạn 2D và 3D**

U-net đã chứng minh hiệu quả vượt trội trong nhiều bài toán phân đoạn hình ảnh y tế nhờ vào các ưu điểm nổi bật sau:

* Khả năng xử lý dữ liệu nhỏ:
  + Một trong những lợi thế lớn của U-net là khả năng hoạt động tốt với các bộ dữ liệu nhỏ, điều này rất quan trọng trong y tế khi dữ liệu thường bị hạn chế do các quy định bảo mật và quyền riêng tư.
  + Các kỹ thuật tăng cường dữ liệu (Data Augmentation) có thể được áp dụng để mở rộng tập dữ liệu đầu vào mà không làm mất đi tính chính xác của mô hình.
* Độ chính xác cao:
  + Nhờ vào skip connections, U-net có khả năng xác định các đặc trưng quan trọng ngay cả ở các vùng biên phức tạp hoặc các khu vực nhỏ, giúp cải thiện đáng kể độ chính xác phân đoạn.
  + Điều này đặc biệt hữu ích trong phân đoạn các cấu trúc nhỏ như khối u hoặc tổn thương xương.
* Khả năng xử lý hình ảnh 3D:
  + U-net có thể mở rộng để xử lý các hình ảnh 3D (3D U-net), điều này rất quan trọng trong lĩnh vực y tế, đặc biệt là khi làm việc với dữ liệu CT hoặc MRI.
  + Trong các ứng dụng này, U-net sử dụng các lớp convolution 3D và pooling 3D để học các đặc trưng không gian ba chiều.
  1. **Cơ chế Attention**
     1. **Khái niệm về Attention trong học sâu**

Cơ chế Attention là một khái niệm quan trọng trong học sâu, giúp mô hình học cách tập trung vào các vùng quan trọng trong dữ liệu đầu vào thay vì xử lý toàn bộ dữ liệu một cách đồng nhất. Trong các bài toán về ảnh, Attention hoạt động như một trọng số, ưu tiên các đặc điểm quan trọng hơn trong không gian hoặc kênh của dữ liệu hình ảnh [3, 4].

* + 1. **Attention kết hợp với U-net trong phân đoạn hình ảnh y tế**

Khi được tích hợp vào U-net, Attention giúp mô hình tập trung hơn vào các vùng quan trọng trong hình ảnh, đặc biệt là các vùng tổn thương hoặc khối u nhỏ và bỏ qua các vùng không liên quan. Điều này giúp:

* Tăng độ chính xác: Mô hình dễ dàng nhận diện các khu vực quan trọng trong ảnh y tế.
* Giảm sai số: Attention giúp giảm thiểu các lỗi phân đoạn không cần thiết trong các vùng ít quan trọng.
* Cải thiện khả năng tổng quát hóa: Mô hình trở nên linh hoạt hơn khi xử lý dữ liệu mới.
  1. **Ứng dụng thực tiễn của U-net và Attention trong y tế**

Các ứng dụng phổ biến của U-net và Attention trong y tế bao gồm:

* Phân đoạn khối u và tổn thương từ ảnh CT hoặc MRI.
* Hỗ trợ lập kế hoạch phẫu thuật và xạ trị.
* Theo dõi tiến triển bệnh và đánh giá hiệu quả điều trị.

1. **Công nghệ PyTorch, MONAI và PyTorch Lightning**

Trong nghiên cứu này, một số các công cụ và thư viện chính được sử dụng nhằm hỗ trợ xây dựng và huấn luyện mô hình học sâu hiệu quả.

Đầu tiên là PyTorch. PyTorch được lựa chọn nhờ tính linh hoạt và khả năng hỗ trợ toàn diện trong thiết kế và triển khai các thuật toán học sâu. Framework này cung cấp các công cụ tối ưu cho việc xây dựng kiến trúc mạng và quản lý quá trình huấn luyện.

Tiếp đến là MONAI, được sử dụng để đáp ứng các yêu cầu đặc thù trong xử lý hình ảnh y tế. Thư viện này cung cấp các công cụ chuyên biệt, hỗ trợ nhiều định dạng ảnh y khoa và tích hợp các phương pháp tăng cường dữ liệu, tối ưu hóa cho các bài toán y tế phức tạp.

Cuối cùng, PyTorch Lightning hỗ trợ quản lý và chuẩn hóa quy trình huấn luyện mô hình học sâu. Framework này giúp tự động hóa các bước trong quá trình phát triển, giảm bớt công việc lặp lại và tối ưu hóa hiệu quả nghiên cứu.

Sự kết hợp giữa các công cụ trên mang lại nền tảng vững chắc, giúp cải thiện đáng kể chất lượng và hiệu quả của mô hình học sâu trong y tế.

**CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU**

1. **Thu thập và tiền xử lý dữ liệu**

.

1. **Thiết kế mô hình**
2. **Huấn luyện mô hình**
3. **Đánh giá và tối ưu hóa**

.

**CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM**

1. **Kết quả thực nghiệm**

.

1. **Phân tích kết quả**
2. **Ứng dụng thực tiễn**

.

# **KẾT LUẬN VÀ KHUYẾN NGHỊ**

1. **Kết luận**

.

* 1. **Ưu điểm**

.

* 1. **Khó khăn, thách thức**
* .

1. **Khuyến nghị / Hướng phát triển**

.

# **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] 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.

[4] O. Oktay, J. Schlemper, and e. al, "Attention U-Net: Learning Where to Look for the Pancreas.," Medical Imaging with Deep Learning 2018, 2018.