

ĐẠI HỌC QUỐC GIA TP. HCM  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



BÁO CÁO ĐỒ ÁN  
DỮ LIỆU LỚN

BIÊN THỂ MÔ HÌNH HỌC SÂU TIÊN TIẾN  
KẾT HỢP LOGISTIC REGRESSION TRONG XỬ LÝ ẢNH  
PHÁT HIỆN TẾ BÀO UNG THƯ CỔ TỬ CUNG

*ADVANCED DEEP LEARNING MODEL VARIANTS  
INTEGRATING LOGISTIC REGRESSION FOR IMAGE PROCESSING  
IN CERVICAL CANCER CELL DETECTION*

Khoa/ Bộ môn: Hệ Thống Thông Tin

Lớp: IS405.P23

Thời gian thực hiện: 4 tháng

Cán bộ hướng dẫn: Thầy Nguyễn Hồ Duy Trí

Tham gia thực hiện

TT	Họ và tên, MSSV	Chịu trách nhiệm	Điện thoại	Email
1.	Lê Châu Khánh Vy, 22521704	Chủ nhiệm	090912345 6	22521704@gm.uit.edu.vn
2.	Nguyễn Hạ Phương, 22521166	Tham gia	092743036 6	22521166@gm.uit.edu.vn



**ĐẠI HỌC QUỐC GIA TP. HCM**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

## **BÁO CÁO TỔNG KẾT**

**Tên đề tài tiếng Việt: BIẾN THỂ MÔ HÌNH HỌC SÂU TIỀN TIẾN KẾT HỢP  
LOGISTIC REGRESSION TRONG XỬ LÝ ẢNH PHÁT HIỆN TẾ BÀO UNG THƯ CỔ  
TỬ CUNG**

**Tên đề tài tiếng Anh: ADVANCED DEEP LEARNING MODEL VARIANTS  
INTEGRATING LOGISTIC REGRESSION FOR IMAGE PROCESSING IN  
CERVICAL CANCER CELL DETECTION**

*Ngày ... tháng ..... năm ....*

**Cán bộ hướng dẫn**

*(Họ tên và chữ ký)*

*Ngày 04 tháng ..... năm ....*

**Sinh viên chủ nhiệm đề tài**

*(Họ tên và chữ ký)*



**ĐẠI HỌC QUỐC GIA TP. HCM**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

## LỜI CẢM ƠN

Trước tiên, chúng em xin bày tỏ lòng biết ơn sâu sắc đến Thầy – giảng viên hướng dẫn môn Dữ liệu lớn – người đã luôn tận tâm đồng hành, chỉ dẫn và hỗ trợ nhóm trong suốt quá trình thực hiện đồ án với đề tài:

“Biến thể mô hình học sâu tiên tiến kết hợp Logistic Regression trong xử lý ảnh phát hiện tế bào ung thư cổ tử cung”.

Những định hướng quý báu cùng sự góp ý chi tiết từ Thầy đã giúp nhóm không ngừng hoàn thiện nội dung đề tài, khắc phục những thiếu sót và phát huy tốt thế mạnh sẵn có. Sự hướng dẫn tận tình ấy đã tạo nền tảng vững chắc để chúng em triển khai và hoàn thành đồ án một cách nghiêm túc và hiệu quả.

Thông qua những bài giảng chất lượng, tinh thần giảng dạy nhiệt huyết cùng những chia sẻ thiết thực của Thầy về chuyên môn và tư duy giải quyết vấn đề, chúng em đã hiểu sâu hơn về lĩnh vực Dữ liệu lớn – một hướng phát triển trọng tâm trong thời đại số hiện nay. Từ đó, chúng em không chỉ tích lũy kiến thức lý thuyết mà còn rèn luyện được kỹ năng ứng dụng vào thực tiễn qua quá trình nghiên cứu và triển khai đề tài.

Chúng em đặc biệt trân trọng sự đồng hành của Thầy trong suốt thời gian thực hiện đồ án, dù lịch trình công tác bận rộn. Sự kiên trì, tận tụy và những lời động viên chân thành từ Thầy đã tiếp thêm động lực để nhóm vượt qua những khó khăn, thử thách, và hoàn thành nhiệm vụ một cách tốt nhất.

Chúng em xin kính chúc Thầy luôn dồi dào sức khỏe, hạnh phúc và tiếp tục gặt hái nhiều thành công trong sự nghiệp giảng dạy và nghiên cứu.

Trân trọng,

# THÔNG TIN KẾT QUẢ NGHIÊN CỨU

## 1. Thông tin chung:

- Tên đề tài: Biến thể mô hình học sâu tiên tiến kết hợp Logistic Regression trong xử lý ảnh phát hiện tế bào ung thư cổ tử cung
- Mã số:
- Chủ nhiệm: Thầy Nguyễn Hồ Duy Trí
- Thành viên tham gia: Lê Châu Khánh Vy, Nguyễn Hạ Phương
- Cơ quan chủ trì: Trường Đại học Công nghệ Thông tin.
- Thời gian thực hiện: 4 tháng

## 2. Mục tiêu:

- Xây dựng và thử nghiệm các biến thể mô hình học sâu tiên tiến, kết hợp Logistic Regression để phân loại hình ảnh tế bào ung thư cổ tử cung với độ chính xác cao.
- Đề xuất một giải pháp hiệu quả để hỗ trợ chẩn đoán sớm ung thư cổ tử cung, giảm thiểu sai sót trong phân loại và tăng hiệu quả phân tích y tế.

## 3. Tính mới và sáng tạo:

- Tích hợp Logistic Regression: Kết hợp Logistic Regression với các biến thể mô hình học sâu tiên tiến để cải thiện hiệu quả phân loại, đặc biệt trên các bộ dữ liệu y tế nhỏ và phức tạp.
- Ứng dụng trên hình ảnh y tế đặc thù: Áp dụng các biến thể mô hình học sâu hiện đại như CvT, BoTNet, và CoAtNet trong lĩnh vực y tế, điều này chưa được nghiên cứu rộng rãi.
- Tăng cường hiệu quả học đặc trưng: Nghiên cứu tối ưu hóa kiến trúc mô hình và thử nghiệm các phương pháp cải thiện khả năng phân loại hình ảnh tế bào.

#### **4. Tóm tắt kết quả nghiên cứu:**

- Nghiên cứu cho thấy sự vượt trội của mô hình BoTNet kết hợp Logistic Regression, đạt độ chính xác cao nhất (96.5%) trên bộ dữ liệu hình ảnh tế bào ung thư cổ tử cung Sipakmed.
- Các biến thể của mô hình tiên tiến khác như CMT-Logistic Regression (96.21%) và CvT-Logistic Regression (96.05%) cũng đạt hiệu suất cao, cho thấy tiềm năng của phương pháp tích hợp này.
- Phương pháp Logistic Regression giúp tăng cường hiệu suất phân loại so với các mô hình gốc như CvT, BoTNet, và CoAtNet.

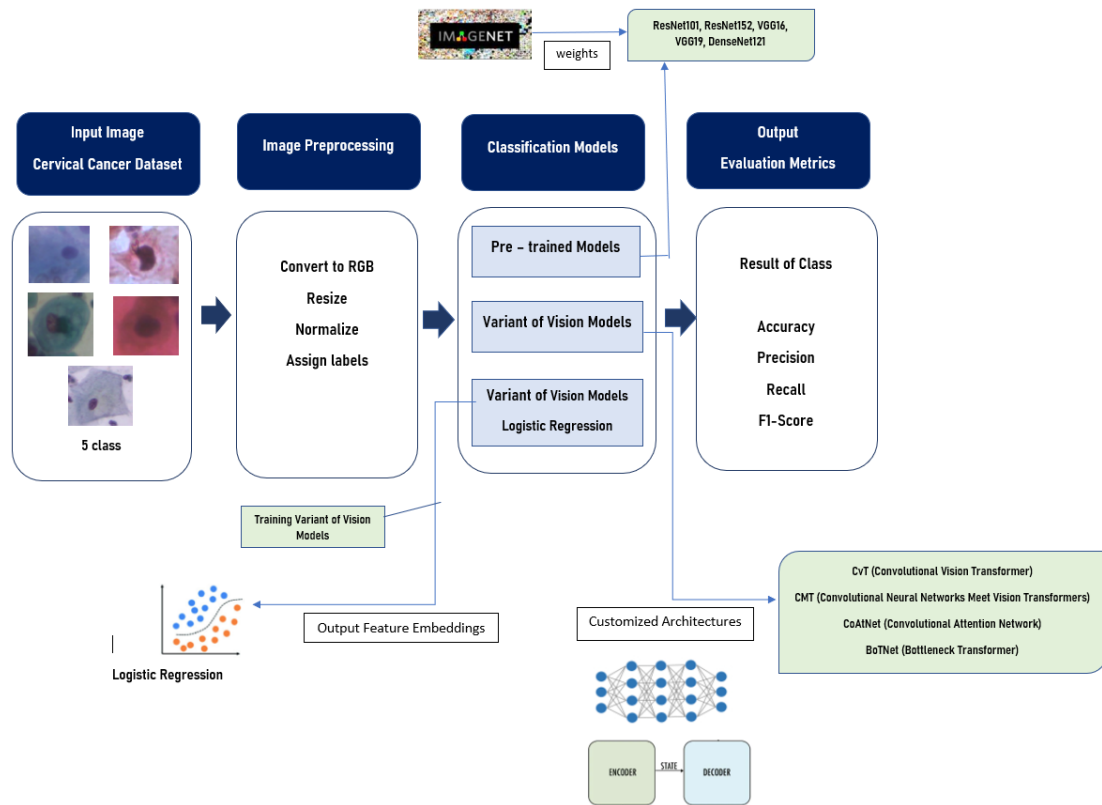
**5. Tên sản phẩm:** Mô hình hỗ trợ chẩn đoán sớm ung thư cổ tử cung dựa trên biến thể kiến trúc học sâu tiên tiến.

#### **6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng:**

- Hiệu quả:
  - Cải thiện đáng kể độ chính xác trong phát hiện tế bào ung thư, hỗ trợ bác sĩ trong việc chẩn đoán và điều trị.
  - Tăng tốc độ phân tích hình ảnh, giảm thiểu sai sót và tiết kiệm thời gian.
- Phương thức chuyển giao:
  - Triển khai phần mềm vào hệ thống bệnh viện và phòng khám chuyên khoa.
  - Chuyển giao mô hình và tài liệu hướng dẫn sử dụng cho các tổ chức y tế và trung tâm nghiên cứu.
- Khả năng áp dụng:
  - Áp dụng trong chẩn đoán sớm ung thư cổ tử cung và các loại ung thư khác liên quan đến tế bào.

- Mở rộng ứng dụng sang các bài toán khác như phân loại hình ảnh y tế (X-quang, MRI).

## 7. Hình ảnh, sơ đồ minh họa chính



Cơ quan Chủ trì  
(ký, họ và tên, đóng dấu)

Chủ nhiệm đề tài  
(ký, họ và tên)

## MỤC LỤC

<b>DANH MỤC HÌNH.....</b>	<b>9</b>
<b>DANH MỤC BẢNG.....</b>	<b>10</b>
<b>DANH MỤC TỪ VIẾT TẮT .....</b>	<b>11</b>
<b>TÓM TẮT ĐỀ TÀI .....</b>	<b>1</b>
<b>CHƯƠNG 1: MỞ ĐẦU .....</b>	<b>2</b>
1.1. Lý do chọn đề tài: .....	2
1.2. Mục đích nghiên cứu: .....	3
1.3. Đối tượng nghiên cứu: .....	3
1.4. Phạm vi nghiên cứu:.....	3
<b>CHƯƠNG 2: TỔNG QUAN .....</b>	<b>5</b>
2.1. Nghiên cứu trong và ngoài nước .....	5
2.2. Những vấn đề còn tồn tại:.....	6
2.3. Hướng tiếp cận của đề tài:.....	6
<b>CHƯƠNG 3: PHƯƠNG PHÁP NGHIÊN CỨU THỰC NGHIỆM VÀ LÝ THUYẾT.....</b>	<b>7</b>
3.1. Cơ sở lý thuyết .....	7
3.2. Giả thuyết khoa học .....	9
3.3. Phương pháp nghiên cứu.....	10
3.3.1. Thu thập và tiền xử lý dữ liệu .....	10
3.3.2. Thiết kế mô hình .....	10
3.3.3. Kết hợp với Logistic Regression .....	10
3.3.4. Huấn luyện và đánh giá.....	10
3.4. So sánh và phân tích.....	11
<b>CHƯƠNG 4: TRÌNH BÀY MÔ HÌNH THỰC NGHIỆM, ĐÁNH GIÁ BÀN LUẬN VỀ KẾT QUẢ .....</b>	<b>12</b>
4.1. Mục tiêu nghiên cứu.....	12
4.2. Phương pháp thực nghiệm đã tiến hành .....	12

4.2.1. Tiền xử lý dữ liệu:.....	12
4.2.2. Xây dựng kiến trúc biến thể của các mô hình .....	12
4.2.3. Đánh giá kết quả.....	20
4.3. Kết quả nghiên cứu.....	20
4.3.1. Kết quả thực nghiệm .....	20
<b>CHƯƠNG 5: KẾT LUẬN .....</b>	<b>22</b>
<b>CHƯƠNG 6: HƯỚNG PHÁT TRIỂN.....</b>	<b>23</b>



## **DANH MỤC HÌNH**

Hình 3.1: Sơ đồ cấu trúc chính của mô hình đồ án.....	7
Hình 4.1. Kiến trúc biến thể mô hình CoAtNet – Logistic Regression.....	14
Hình 4.2. Kiến trúc biến thể mô hình BoTNet – Logistic Regression.....	16
Hình 4.3. Kiến trúc biến thể mô hình CvT – Logistic Regression.....	18
Hình 4.4. Kiến trúc biến thể mô hình CMT – Logistic Regression.....	19

## **DANH MỤC BẢNG**

Bảng 1.1: Danh mục từ viết tắt.....	1
Bảng 4.1: Kết quả thực nghiệm trên các mô hình.....	21

## DANH MỤC TỪ VIẾT TẮT

STT	Từ viết tắt	Diễn giải
1	AI	Artificial Intelligence (Trí tuệ nhân tạo)
2	CNN	Convolutional Neural Network (Mạng nơ-ron tích chập)
4	CvT	Convolutional Vision Transformer
5	CMT	Convolutional Neural Networks Meet Vision Transformers
6	CoAtNet	Convolutional Attention Network
7	BoTNet	Bottleneck Transformer Network
8	Dyskeratotic	Dyskeratotic Dyskeratotic Cells (Tế bào loạn sừng)
9	Koilocytotic	Koilocytotic Cells (Tế bào không bào hóa)
10	Metaplastic	Metaplastic Cells (Tế bào chuyển sản)
11	Parabasal	Parabasal Cells (Tế bào nền)
12	Superficial-Intermediate	Superficial and Intermediate Cells (Tế bào bề mặt và trung gian)
13	DL	Deep Learning (Học sâu)
14	ML	Machine Learning (Học máy)

Bảng 1.1: Danh mục từ viết tắt

## TÓM TẮT ĐỀ TÀI

Ung thư cổ tử cung là một trong những nguyên nhân gây tử vong hàng đầu ở nữ giới, đặc biệt tại các quốc gia đang phát triển. Phát hiện sớm và chẩn đoán chính xác tế bào ung thư là yếu tố then chốt giúp nâng cao hiệu quả điều trị và tỷ lệ sống sót. Tuy nhiên, phương pháp truyền thống dựa vào phân tích hình ảnh tế bào dưới kính hiển vi còn phụ thuộc nhiều vào chuyên môn con người, dễ gây sai sót và mất thời gian.

Với sự phát triển mạnh mẽ của trí tuệ nhân tạo, đặc biệt là học sâu (Deep Learning), nhiều nghiên cứu đã ứng dụng các mô hình mạng nơ-ron tích chập (CNN) và Transformer vào xử lý ảnh y tế, đem lại kết quả khả quan. Đề tài này tập trung xây dựng mô hình phân loại tế bào ung thư cổ tử cung tự động dựa trên bộ dữ liệu SipakMed gồm 5 lớp tế bào: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, Superficial-Intermediate.

Đồ án thực hiện áp dụng các biến thể học sâu tiên tiến như CMT, CvT, CoAtNet, BoTNet và kết hợp đầu ra của các mô hình này với Logistic Regression – thuật toán học máy đơn giản nhưng hiệu quả – để nâng cao khả năng phân loại chính xác. Sự kết hợp này kỳ vọng sẽ tối ưu hiệu suất, giảm sai sót và hỗ trợ bác sĩ trong quá trình chẩn đoán.

Nghiên cứu hướng đến xây dựng hệ thống hỗ trợ chẩn đoán tự động, góp phần nâng cao chất lượng chăm sóc sức khỏe, đặc biệt trong phát hiện sớm ung thư cổ tử cung.

# Chương 1: MỞ ĐẦU

## 1.1. Lý do chọn đề tài:

Trong bối cảnh hiện nay, ung thư cổ tử cung vẫn là một trong những căn bệnh đe dọa nghiêm trọng đến sức khỏe phụ nữ, đặc biệt ở các khu vực có điều kiện y tế hạn chế. Việc chẩn đoán sớm và chính xác loại ung thư này đóng vai trò then chốt trong quá trình điều trị, nhưng lại gặp nhiều khó khăn do phụ thuộc lớn vào kỹ năng và kinh nghiệm của các chuyên gia y tế. Quá trình quan sát và phân loại tế bào qua kính hiển vi không chỉ đòi hỏi sự tỉ mỉ mà còn tiềm ẩn nguy cơ sai sót do yếu tố chủ quan.

Song song với những hạn chế trong y học truyền thống, sự phát triển nhanh chóng của lĩnh vực trí tuệ nhân tạo – đặc biệt là học sâu – đã mở ra hướng đi mới cho việc xử lý ảnh y tế. Những tiến bộ trong các mô hình mạng nơ-ron tích chập (CNN), Transformer và các biến thể kết hợp đã cho thấy khả năng vượt trội trong việc trích xuất đặc trưng hình ảnh và phân loại tự động. Những công nghệ này không chỉ giúp giảm gánh nặng cho đội ngũ y tế mà còn nâng cao độ chính xác, tính nhất quán trong chẩn đoán.

Xuất phát từ tiềm năng ứng dụng của các mô hình học sâu hiện đại trong lĩnh vực y học, nhóm nghiên cứu nhận thấy cần thiết phải xây dựng một hệ thống hỗ trợ chẩn đoán tự động tế bào ung thư cổ tử cung dựa trên hình ảnh. Bên cạnh đó, việc kết hợp đầu ra của các mô hình học sâu với Logistic Regression – một thuật toán có khả năng tổng quát hóa tốt – được kỳ vọng sẽ tạo ra một phương pháp phân loại hiệu quả, có tính ứng dụng thực tiễn cao. Do đó, đề tài được lựa chọn nhằm góp phần nghiên cứu và phát triển hướng tiếp cận này.

## 1.2. Mục đích nghiên cứu:

Mục đích của đề tài là xây dựng một mô hình phân loại tế bào cổ tử cung tự động, chính xác và hiệu quả, dựa trên hình ảnh tế bào trong bộ dữ liệu y tế. Đề tài hướng đến việc:

- Khảo sát và áp dụng các biến thể mô hình học sâu hiện đại như CMT, CvT, CoAtNet, BoTNet trong nhiệm vụ phân loại ảnh y tế.
- Kết hợp các đặc trưng học được từ các mô hình học sâu với thuật toán Logistic Regression nhằm tăng cường khả năng phân loại và tổng quát hóa.
- Đánh giá hiệu quả của mô hình thông qua các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu, F1-score,...

## 1.3. Đối tượng nghiên cứu:

Đối tượng nghiên cứu của đề tài là hình ảnh tế bào cổ tử cung được trích xuất từ bộ dữ liệu SipakMed, bao gồm 5 loại tế bào: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal và Superficial-Intermediate. Đây là tập dữ liệu đã được gán nhãn, thường dùng trong các nghiên cứu xử lý ảnh y tế.

## 1.4. Phạm vi nghiên cứu:

Phạm vi của đề tài tập trung vào việc xử lý ảnh tĩnh của tế bào cổ tử cung trong môi trường mô phỏng, không áp dụng trên dữ liệu thời gian thực hay kết hợp với dữ liệu văn bản y khoa. Cụ thể:

- Giới hạn trong việc sử dụng ảnh từ bộ dữ liệu SipakMed.
- Chỉ tập trung vào các mô hình học sâu về thị giác máy tính (computer vision), không mở rộng sang các kỹ thuật xử lý ngôn ngữ tự nhiên hay đa phương thức.
- Quá trình triển khai và đánh giá mô hình được thực hiện trong môi trường học thuật, chưa tích hợp vào hệ thống chẩn đoán thực tế.

- Môi trường thực nghiệm được triển khai trên nền tảng Kaggle Notebook, sử dụng thư viện TensorFlow để xây dựng và huấn luyện mô hình. Trong quá trình thử nghiệm, cấu hình GPU T4  $\times 2$  (Dual NVIDIA T4) được sử dụng nhằm tăng tốc độ huấn luyện và đảm bảo hiệu quả xử lý cho các mô hình học sâu tiên tiến.

## Chương 2: TỔNG QUAN

### 2.1. Nghiên cứu trong và ngoài nước

Trong những năm gần đây, ứng dụng của học sâu trong chẩn đoán hình ảnh y tế đã nhận được sự quan tâm rộng rãi từ cộng đồng nghiên cứu, đặc biệt là trong phân loại tế bào ung thư cổ tử cung. Các mô hình học sâu như Convolutional Neural Network (CNN) và các biến thể kết hợp attention hoặc Transformer đã cho thấy hiệu quả vượt trội so với các phương pháp truyền thống.

Một trong những công trình tiêu biểu là nghiên cứu “DeepPap” của Zhang et al. [1], sử dụng CNN để tự động phân loại tế bào cổ tử cung, cho kết quả đầy hứa hẹn trên tập dữ liệu Pap smear. Nghiên cứu này là một trong những bước đầu tiên ứng dụng học sâu trong chẩn đoán ung thư cổ tử cung.

Tiếp theo đó, Lu et al. [2] đã đề xuất mô hình sử dụng mạng CNN sâu với phương pháp học chuyển giao để cải thiện độ chính xác phân loại trên bộ dữ liệu Herlev và Pap smear. Trong khi đó, bài báo của Bora et al. [3] sử dụng mô hình kết hợp CNN và LSTM để tận dụng cả đặc trưng không gian và chuỗi, nhưng vẫn còn giới hạn khi xử lý hình ảnh độ phân giải cao.

Ngoài ra, CoAtNet [4] là một mô hình kết hợp giữa CNN và self-attention, giúp cải thiện khả năng khái quát hóa mô hình khi xử lý hình ảnh y tế, đã được áp dụng thành công trong nhiều bài toán phân loại ảnh y khoa. CvT (Convolutional vision Transformer) [5] và BoTNet (Bottleneck Transformer) [6] là các kiến trúc Transformer cải tiến có khả năng khai thác thông tin không gian tốt hơn khi tích hợp với CNN, và được kỳ vọng có thể áp dụng hiệu quả trong phân loại ảnh tế bào cổ tử cung.



Tại Việt Nam, một số nghiên cứu ban đầu cũng đã áp dụng mô hình học máy truyền thống như SVM, Random Forest, hoặc CNN cơ bản vào phân loại tế bào, tuy nhiên vẫn còn hạn chế về độ chính xác và khả năng tổng quát khi triển khai trên dữ liệu thực tế.

## **2.2. Những vấn đề còn tồn tại:**

Mặc dù các nghiên cứu trên đã đạt được những kết quả tích cực, vẫn tồn tại một số hạn chế như:

- Hầu hết chỉ sử dụng một kiến trúc đơn lẻ (CNN hoặc Transformer), chưa khai thác hiệu quả sức mạnh kết hợp giữa các mô hình học sâu.
- Việc áp dụng các mô hình phức tạp thường đi kèm chi phí tính toán cao, khó triển khai trên môi trường thực tế hoặc thiếu tài nguyên.
- Chưa có nhiều nghiên cứu thử nghiệm việc kết hợp đầu ra của mô hình học sâu với các thuật toán phân loại đơn giản để cải thiện khả năng phân biệt lớp.

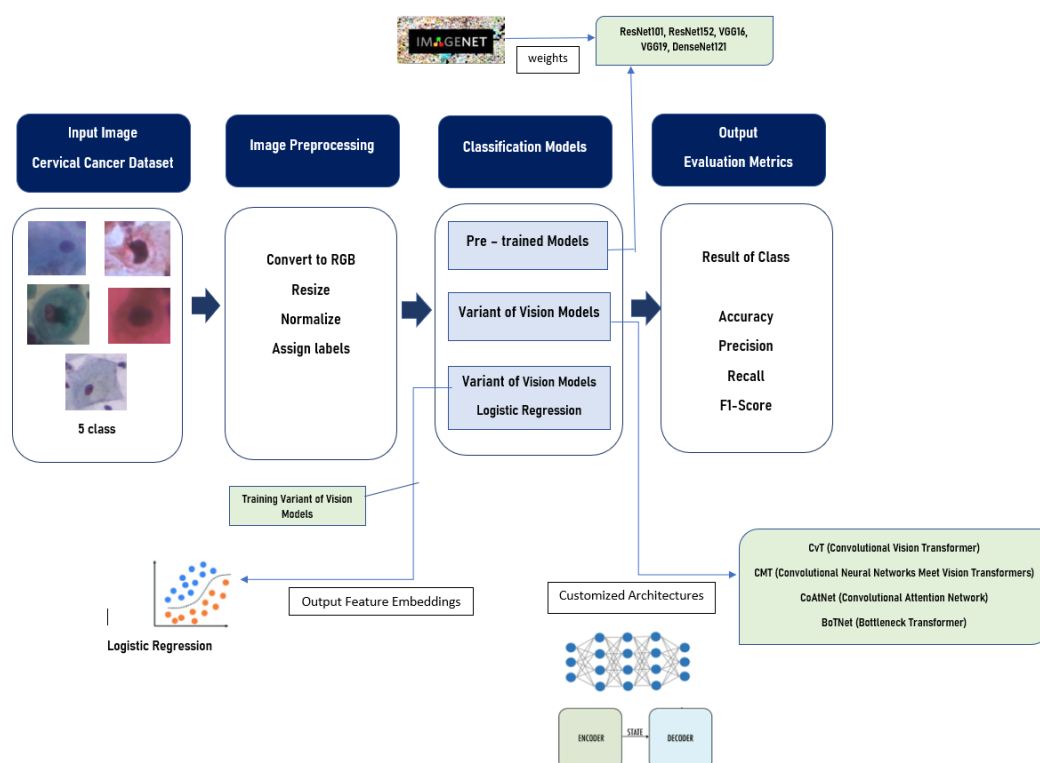
## **2.3. Hướng tiếp cận của đề tài:**

Đề tài này đề xuất một hướng tiếp cận mới: kết hợp các biến thể hiện đại của mô hình học sâu như CMT, CvT, CoAtNet và BoTNet để khai thác đa dạng đặc trưng từ ảnh tế bào, sau đó tích hợp đầu ra vào một mô hình Logistic Regression nhằm tăng cường khả năng phân loại và tổng quát hóa. Hướng tiếp cận này hứa hẹn sẽ:

- Tận dụng được sức mạnh biểu diễn của các mô hình học sâu hiện đại.
- Khai thác tính đơn giản và hiệu quả của Logistic Regression trong việc phân loại đầu ra cuối cùng.
- Giảm rủi ro overfitting và tăng khả năng triển khai thực tế nhờ tính tổng quát tốt.

Bên cạnh đó, toàn bộ mô hình sẽ được thực nghiệm trên bộ dữ liệu chuẩn SipakMed, và triển khai trên nền tảng Kaggle sử dụng GPU T4×2 với thư viện TensorFlow để đảm bảo khả năng tái lập và tối ưu tài nguyên.

# Chương 3: PHƯƠNG PHÁP NGHIÊN CỨU THỰC NGHIỆM VÀ LÝ THUYẾT



Hình 3.1: Sơ đồ cấu trúc chính của mô hình đề án

## 3.1. Cơ sở lý thuyết

Phân lớp tế bào cổ tử cung là một trong những ứng dụng quan trọng trong lĩnh vực y sinh học, đặc biệt nhằm hỗ trợ chẩn đoán và phát hiện sớm ung thư cổ tử cung. Đây là một vấn đề y học quan trọng vì ung thư cổ tử cung là một trong những nguyên nhân hàng đầu gây tử vong ở phụ nữ trên toàn thế giới. Việc phát hiện sớm thông qua phân tích hình ảnh tế bào có thể giúp tăng cơ hội điều trị thành công và giảm gánh nặng kinh tế lên hệ thống y tế.

Bộ dữ liệu SipakMed cung cấp một tập hợp các hình ảnh tế bào với các nhãn phân lớp rõ ràng, bao gồm năm loại tế bào: bình thường basal/parabasal, bình thường intermediate, bình thường superficial, và hai loại tế bào bất thường. Đây

là một bộ dữ liệu tiêu chuẩn thường được sử dụng trong nghiên cứu về phân loại tế bào cổ tử cung, đặc biệt trong việc thử nghiệm các mô hình học sâu và các thuật toán học máy khác. Bộ dữ liệu này đòi hỏi các mô hình không chỉ có khả năng phân biệt được các loại tế bào bình thường mà còn phải nhận diện chính xác các loại tế bào bất thường với đặc điểm phức tạp.

Việc áp dụng các biến thể tiên tiến của các mô hình học sâu như CoAtNet, CvT, CMT, và BotNet kết hợp với Logistic Regression được kỳ vọng sẽ cải thiện độ chính xác của phân lớp nhờ tận dụng sự kết hợp giữa khả năng trích xuất đặc trưng cục bộ của Convolutional Neural Networks (CNN) và khả năng học các mối quan hệ toàn cục của Transformer. Các mô hình này đại diện cho những bước tiến mới trong công nghệ học sâu, được thiết kế để vượt qua các hạn chế của các mô hình truyền thống.

CoAtNet kết hợp các cơ chế của convolution và attention để đạt được sự cân bằng giữa khả năng biểu diễn đặc trưng cục bộ và toàn cục. Điều này giúp nó có thể học tốt các đặc điểm từ hình ảnh y tế phức tạp như các cấu trúc tế bào cổ tử cung.

CvT sử dụng cơ chế attention kết hợp với các lớp convolution để giảm độ phức tạp tính toán trong khi vẫn duy trì khả năng học các mối quan hệ toàn cục trong dữ liệu. CvT phù hợp để xử lý hình ảnh tế bào có độ phân giải cao và các chi tiết nhỏ.

CMT kết hợp các ưu điểm của convolution và transformer trong một cấu trúc duy nhất, cho phép mô hình vừa học được các đặc trưng cục bộ thông qua convolution, vừa khai thác được các mối quan hệ toàn cục thông qua cơ chế attention.

BotNet sử dụng Bottleneck Attention để tăng cường khả năng học các đặc điểm không gian quan trọng, giúp cải thiện hiệu suất trong các bài toán phân loại hình ảnh y tế có độ phức tạp cao.

Logistic Regression, mặc dù là một phương pháp phân lớp tuyến tính đơn giản, lại có khả năng hoạt động hiệu quả khi được sử dụng làm lớp phân loại cuối cùng sau các mô hình học sâu. Điều này giúp giảm độ phức tạp của mô hình tổng thể, đồng thời tăng khả năng tổng quát hóa và giảm nguy cơ overfitting.

Khi kết hợp các mô hình học sâu tiên tiến này với Logistic Regression, hệ thống phân lớp tế bào cổ tử cung có thể đạt được sự cân bằng giữa tính hiệu quả và độ chính xác. Việc sử dụng các công cụ hiện đại này không chỉ giải quyết bài toán một cách chính xác mà còn mở ra khả năng ứng dụng trong thực tiễn, hỗ trợ các bác sĩ và nhà nghiên cứu trong việc chẩn đoán và điều trị bệnh một cách nhanh chóng và chính xác hơn.

### 3.2. Giả thuyết khoa học

Việc sử dụng các biến thể của các mô hình tiên tiến như CoAtNet, CvT, CMT, và BotNet sẽ tối ưu hóa khả năng trích xuất và biểu diễn đặc trưng từ hình ảnh tế bào cổ tử cung. Các mô hình này được thiết kế để tận dụng điểm mạnh của cả CNN và Transformer. CoAtNet kết hợp các lớp convolution và attention để học cả đặc trưng cục bộ và toàn cục. CvT sử dụng cơ chế attention với các lớp chuyển đổi convolution để giảm thiểu độ phức tạp tính toán. CMT tích hợp trực tiếp convolution và transformer trong một cấu trúc thống nhất, giúp mô hình có khả năng học các đặc trưng phức tạp hơn. BotNet, với Bottleneck Attention, tăng cường khả năng phát hiện các mối quan hệ không gian quan trọng. Việc sử dụng các mô hình này đảm bảo việc trích xuất các đặc trưng hữu ích từ hình ảnh tế bào cổ tử cung, góp phần tăng độ chính xác cho bài toán phân lớp.

Logistic Regression là một phương pháp phân lớp đơn giản nhưng hiệu quả, có khả năng hoạt động như một bộ phân lớp cuối cùng để tận dụng các đặc trưng học được từ các mô hình sâu. Logistic Regression, mặc dù không phức tạp như các phương pháp học sâu, lại có ưu điểm là dễ triển khai và tối ưu hóa. Khi được sử dụng làm bộ phân lớp cuối cùng, Logistic Regression có thể tận dụng tốt các đặc trưng đầu ra từ các mô hình học sâu để thực hiện phân lớp với độ chính xác cao. Hơn nữa, nó giúp giảm thiểu hiện tượng overfitting nhờ khả năng tổng quát hóa tốt trên các tập dữ liệu lớn.

Sự kết hợp này sẽ dẫn đến hiệu suất phân lớp cao hơn so với việc chỉ sử dụng riêng lẻ từng mô hình học sâu hoặc Logistic Regression. Các mô hình học sâu mạnh mẽ như CoAtNet, CvT, CMT, và BotNet có khả năng trích xuất các đặc

trung phức tạp từ dữ liệu hình ảnh, nhưng khi kết hợp với Logistic Regression, hệ thống phân lớp trở nên linh hoạt hơn, có thể khai thác tối đa tiềm năng của từng mô hình. Thay vì chỉ dựa vào một mô hình cụ thể, sự kết hợp này giúp giảm thiểu những hạn chế riêng lẻ và tạo ra một hệ thống phân lớp mạnh mẽ, có khả năng đối phó với các trường hợp khó phân loại. Kết quả là, hiệu suất phân lớp, đặc biệt trên các tập dữ liệu phức tạp như SipakMed, được cải thiện đáng kể, đồng thời đảm bảo tính ổn định và đáng tin cậy của mô hình.

### **3.3. Phương pháp nghiên cứu**

#### **3.3.1. Thu thập và tiền xử lý dữ liệu**

- Sử dụng bộ dữ liệu SipakMed, bao gồm các hình ảnh tế bào cổ tử cung với các nhãn phân loại.
- Thực hiện các bước tiền xử lý như cân bằng dữ liệu, chuẩn hóa hình ảnh.

#### **3.3.2. Thiết kế mô hình**

- CoAtNet: Áp dụng mô hình này để tận dụng khả năng kết hợp giữa CNN và Attention.
- CvT: Sử dụng các lớp chuyển đổi để học các đặc trưng toàn cục và cục bộ từ hình ảnh.
- CMT: Tích hợp cả Convolution và Transformer trong một cấu trúc duy nhất để trích xuất đặc trưng hiệu quả hơn.
- BotNet: Tăng cường khả năng học mối quan hệ không gian với các lớp Bottleneck Attention.

#### **3.3.3. Kết hợp với Logistic Regression**

- Sử dụng Logistic Regression làm bộ phân lớp cuối cùng, nhận các đặc trưng đã học từ các mô hình trên.
- Điều chỉnh tham số và tối ưu hóa để đạt độ chính xác cao nhất.

#### **3.3.4. Huấn luyện và đánh giá**

- Huấn luyện các mô hình trên tập huấn luyện và sử dụng tập kiểm tra để đánh giá hiệu suất.

- Các chỉ số đánh giá bao gồm Accuracy, Precision, Recall và F1-score.

### **3.4. So sánh và phân tích**

- So sánh hiệu suất của từng mô hình riêng lẻ với mô hình kết hợp để kiểm chứng giả thuyết khoa học.
- Phân tích các trường hợp sai phân lớp và tìm cách cải thiện hiệu suất.

## Chương 4: TRÌNH BÀY MÔ HÌNH THỰC NGHIỆM, ĐÁNH GIÁ BÀN LUẬN VỀ KẾT QUẢ

### 4.1. Mục tiêu nghiên cứu

Công việc nghiên cứu nhằm mục tiêu phân loại các tế bào cổ tử cung trong tập dữ liệu SipakMed, sử dụng các biến thể mô hình tiên tiến kết hợp với Logistic Regression. Cụ thể, nghiên cứu tập trung vào:

- Trích xuất đặc trưng từ hình ảnh tế bào cổ tử cung bằng biến thể các mô hình học sâu.
- Sử dụng Logistic Regression để phân lớp dựa trên các đặc trưng đã trích xuất nhằm cải thiện hiệu suất phân loại.

### 4.2. Phương pháp thực nghiệm đã tiến hành

#### 4.2.1. Tiền xử lý dữ liệu:

- Dữ liệu từ tập SipakMed được chia thành năm nhóm: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, và Superficial-Intermediate.
- Hình ảnh được resize về kích thước 64x64x3 và chuẩn hóa giá trị pixel về khoảng  $[0, 1]$ .
- Tập dữ liệu được chia thành ba phần: huấn luyện (70%), kiểm tra (15%), và kiểm định (15%).

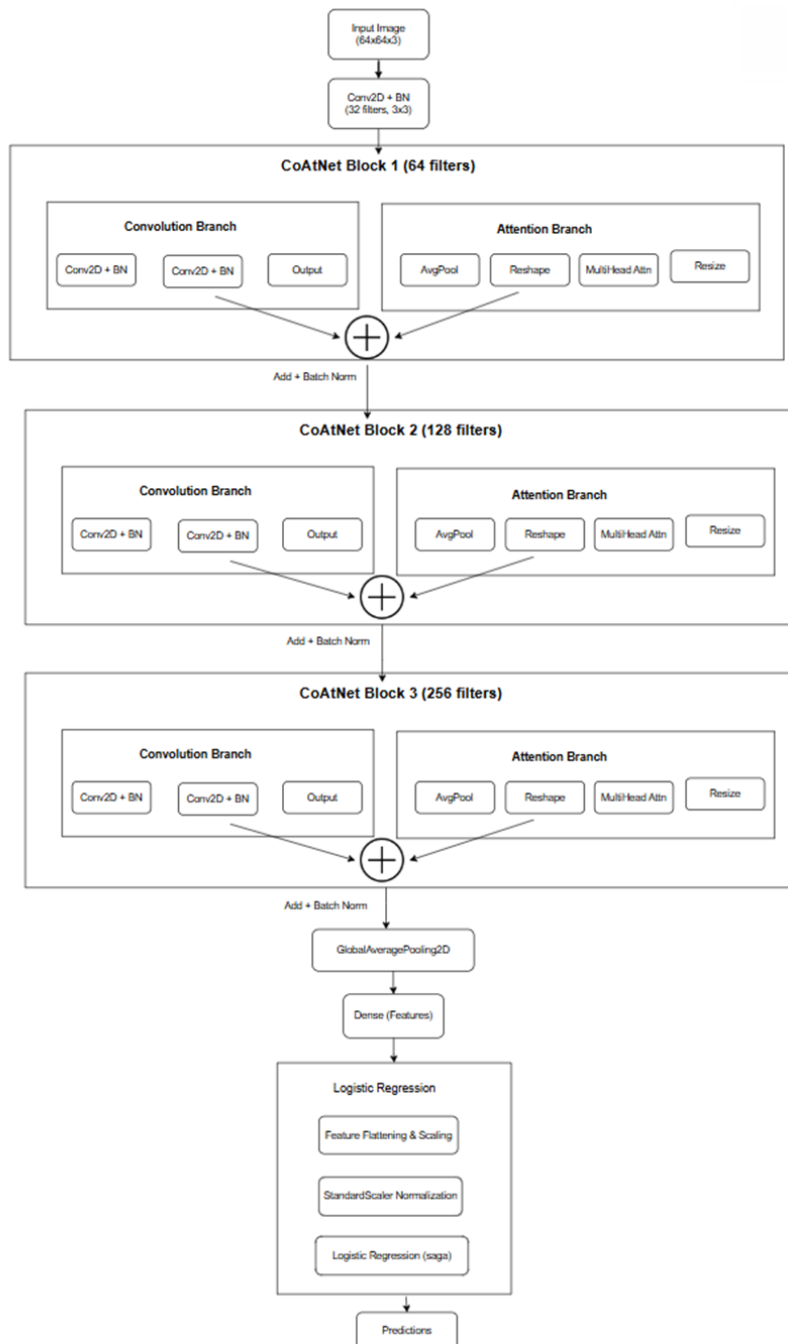
#### 4.2.2. Xây dựng kiến trúc biến thể của các mô hình

##### 4.2.2.1. Kiến trúc biến thể mô hình CoAtNet – Logistic Regression.

- Kiến trúc biến thể CoatNet:
  - Kiến trúc CoAtNet sử dụng kết hợp các khối convolution và attention để khai thác ưu điểm của cả hai kỹ thuật.
    - Nhánh Convolution: Sử dụng các lớp Conv2D và BatchNormalization để trích xuất đặc trưng cục bộ.

- Nhánh Attention: Sử dụng lớp MultiHeadAttention để khai thác các mối quan hệ không gian toàn cục sau khi giảm kích thước không gian bằng AveragePooling2D.
- Sử dụng chuẩn hóa BatchNormalization: chuẩn hóa các đặc trưng sau mỗi bước convolution hoặc attention để cải thiện khả năng hội tụ và ổn định trong huấn luyện.
- Khối coatnet\_block trong mã có cơ chế tương tự khối cơ bản của CoAtNet, với hai nhánh (convolution và attention) được gộp lại bằng phép cộng (Add) và sau đó được chuẩn hóa.
- Việc gộp không gian toàn cục (GlobalAveragePooling2D) được sử dụng để thu gọn đặc trưng trước tầng fully-connected, giúp giảm số lượng tham số.
- Training Logistic Regression:
  - Sử dụng mô hình biến thể CoAtNet như một bộ trích xuất đặc trưng.
  - Các đặc trưng cuối cùng được sử dụng để huấn luyện Logistic Regression cho nhiệm vụ phân loại.



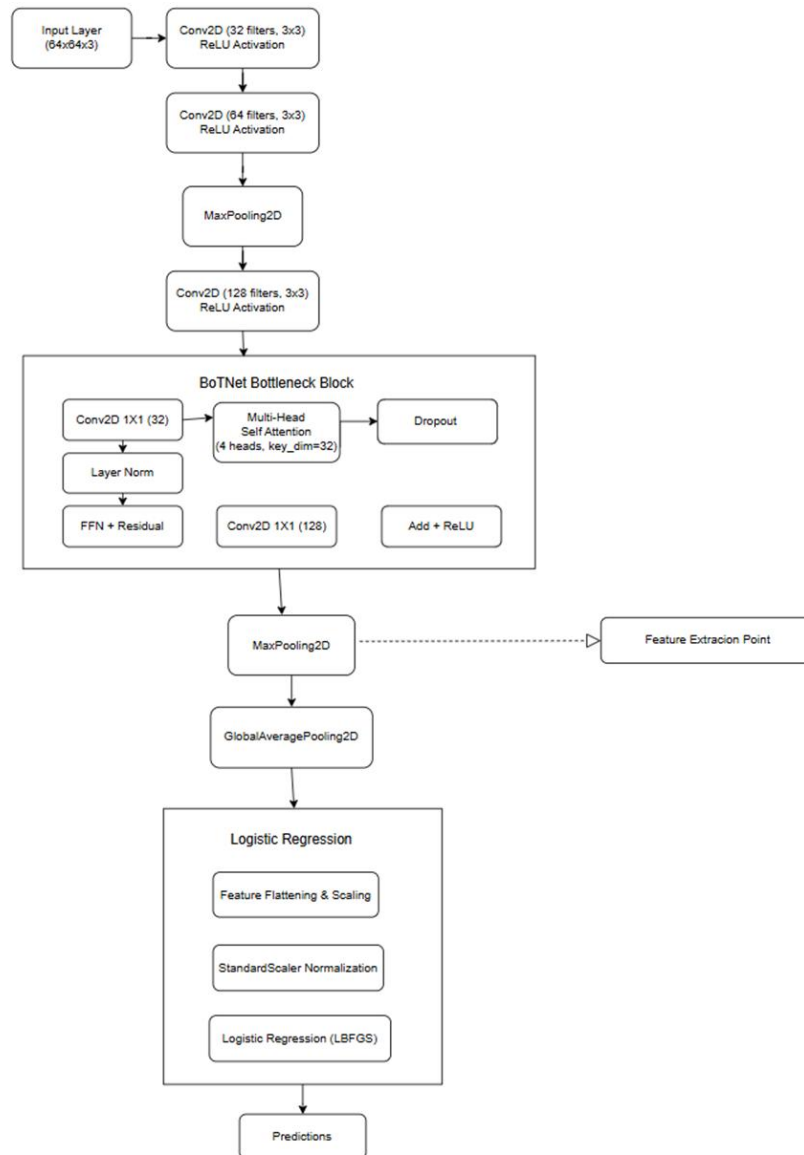


Hình 4.1. Kiến trúc biến thể mô hình CoAtNet – Logistic Regression

#### 4.2.2.2. Kiến trúc biến thể mô hình BoTNet – Logistic Regression.

- Kiến trúc biến thể BoTNet:
  - Phát triển kiến trúc BoTNet với các khối Bottleneck Transformer kết hợp Convolutional Layers.

- Các tầng CNN được sử dụng để trích xuất đặc trưng cục bộ, sau đó được đưa vào MHSA để học các đặc trưng phi cục bộ.
  - Sử dụng hàm kích hoạt ReLU, chuẩn hóa lớp (Layer Normalization), và giảm nhiễu qua Dropout.
  - Kết nối đầu ra cuối cùng với lớp Dense để phân loại thành 5 nhãn.
  - Áp dụng các callback (checkpoint, early stopping, reduce learning rate) để tối ưu hóa quá trình huấn luyện.
- Training Logistic Regression:
- Trích xuất đặc trưng từ tầng GlobalAveragePooling2D của mô hình biến thể BoTNet.
  - Sử dụng các đặc trưng này để huấn luyện Logistic Regression.

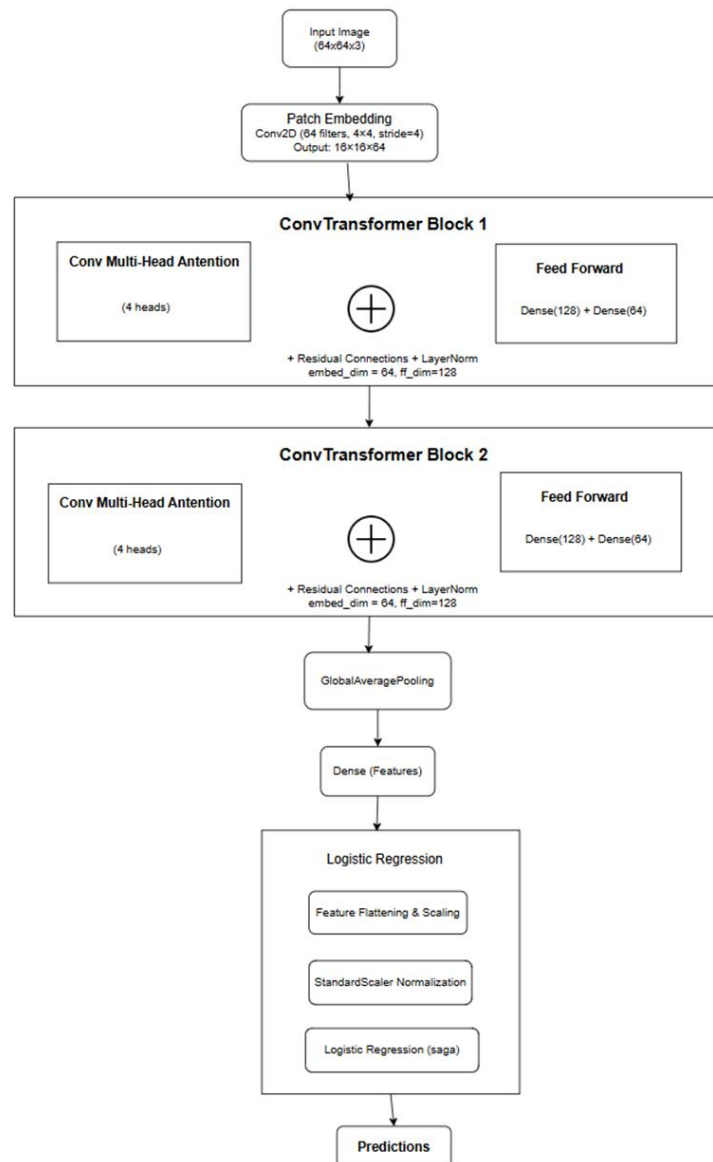


Hình 4.2. Kiến trúc biến thể mô hình BoTNet – Logistic Regression

#### 4.2.2.3. Kiến trúc biến thể mô hình CvT – Logistic Regression.

- Kiến trúc biến thể CvT:
  - Sự kết hợp convolution để trích xuất đặc trưng và transformer để học mối quan hệ phi tuyến tính từ các đặc trưng trích xuất.
  - Sử dụng phương pháp patch embedding dựa trên convolution để trích xuất các đặc trưng từ ảnh. Patch Embedding: Tách hình ảnh thành các đoạn nhỏ (patches) và mã hóa thành vector.

- ConvMultiHeadAttention: Sử dụng đầu đa luồng để học mối quan hệ không gian giữa các đặc trưng với các cơ chế như Q, K, V và tính scaled dot-product attention.
  - Áp dụng Layer Normalization và Residual Connections giúp ổn định gradient và tăng cường khả năng huấn luyện của mạng.
  - Sử dụng các khối ConvTransformerBlock với thành phần attention và feed-forward Layer: Kết hợp và tinh chỉnh thông tin từ các đặc trưng đã học.
- Training Logistic Regression:
- Trích xuất đặc trưng từ tầng GlobalAveragePooling2D của mô hình biến thể CvT.
  - Sử dụng các đặc trưng này để huấn luyện Logistic Regression.

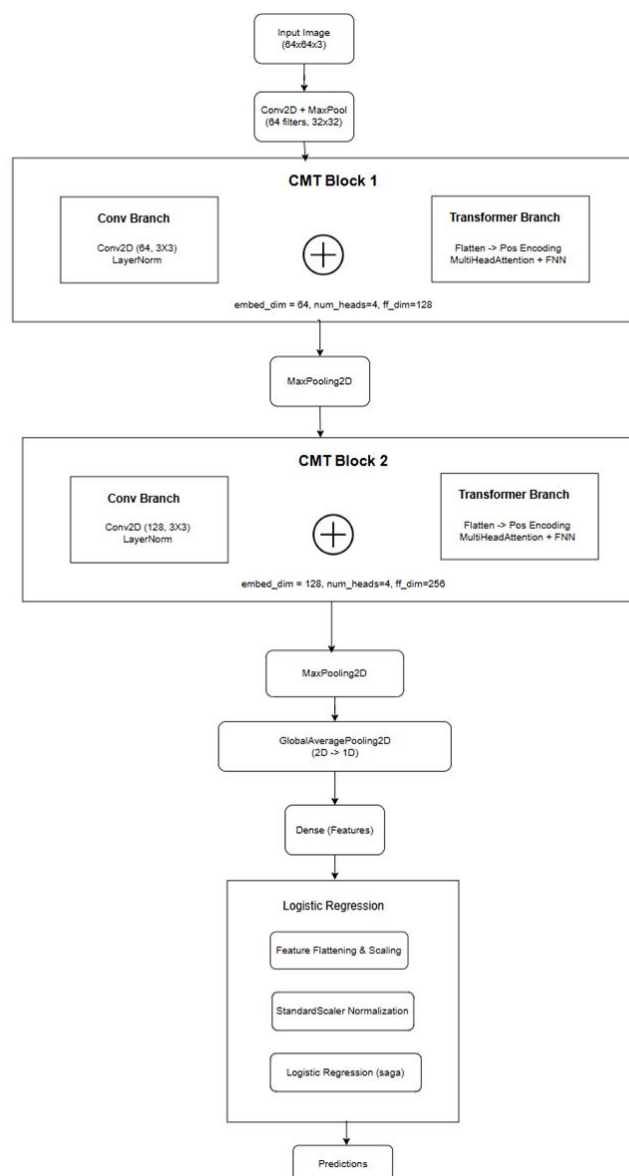


Hình 4.3. Kiến trúc biến thể mô hình CvT – Logistic Regression

#### 4.2.2.4. Kiến trúc biến thể mô hình CMT – Logistic Regression.

- Kiến trúc biến thể CMT:
  - Kết hợp các đặc trưng không gian cục bộ được trích xuất bởi các lớp CNN với các đặc trưng toàn cục từ các lớp Transformer.
  - Mô-đun CMTBlock trong mô hình có nhánh CNN (sử dụng Conv2D) và nhánh Transformer (sử dụng MultiHeadAttention). Cách kết hợp song song (convolution branch + transformer branch) trong một block rồi cộng kết quả.
  - Sử dụng mã hóa vị trí (Positional Encoding) để bổ sung thông tin vị trí vào các vector đặc trưng trước khi đưa vào Transformer.

- Cấu trúc module Transformer: Bao gồm attention và feed-forward network (FFN), cùng với residual connections giúp gradient flow tốt hơn và tránh vanishing gradient.
  - Xây dựng mạng với các tầng CMTBlock, pooling, và các tầng fully-connected cuối cùng.
- Training Logistic Regression:
- Trích xuất đặc trưng từ tầng GlobalAveragePooling2D của mô hình biến thể CMT.
  - Sử dụng các đặc trưng này để huấn luyện Logistic Regression.



#### Hình 4.4. Kiến trúc biến thể mô hình CMT – Logistic Regression

### 4.2.3. Đánh giá kết quả

- Huấn luyện các mô hình trên tập huấn luyện và sử dụng tập kiểm tra để đánh giá hiệu suất.
- Các chỉ số đánh giá bao gồm Accuracy, Precision, Recall và F1-score.

## 4.3. Kết quả nghiên cứu

### 4.3.1. Kết quả thực nghiệm

Mô hình biến thể	Accuracy (%)
VGG19	87.16
VGG16	88.52
ResNet152	48.03
ResNet101	55.26
DenseNet121	93.09
CvT	95.56
CvT + Logistic Regression	96.05
BoTNet	93.75
BoTNet + Logistic Regression	96.55
CMT	94.40
CMT + Logistic Regression	96.21
CoAtNet	94.73
CoAtNet + Logistic Regression	95.72

Bảng 4.1: Kết quả thực nghiệm trên các mô hình

### 4.3.2. Đánh giá

Bộ dữ liệu Sipakmed chứa các hình ảnh tế bào phục vụ chẩn đoán y tế, một bài toán yêu cầu độ chính xác cao trong việc phân loại hình ảnh. Vì dữ liệu y tế thường có các đặc điểm chi tiết, nhỏ và rất khó phân biệt, việc lựa chọn mô hình phù hợp trở nên rất quan trọng. Kết quả cho thấy các mô hình dựa trên

Transformer hoặc CNN cải tiến có hiệu suất cao hơn nhiều so với các mô hình CNN truyền thống.

Các mô hình CNN truyền thống (VGG, ResNet) hoạt động kém do weights không được fine-tune cho dữ liệu y tế. DenseNet là ngoại lệ nhờ kiến trúc liên kết tốt hơn. Mô hình quá sâu, không phù hợp với dữ liệu nhỏ. Các layers học đặc trưng không đủ khả năng tổng hợp thông tin từ ảnh y tế phức tạp. ResNet được thiết kế để giải quyết vấn đề gradient vanishing bằng các residual connections. Tuy nhiên, trên Sipakmed, ResNet hoạt động rất kém do hai nguyên nhân chính: Pretrained weights trên ImageNet không phù hợp cho hình ảnh y tế, kích thước và tính chất đặc thù của bộ dữ liệu gây khó khăn cho việc huấn luyện các mô hình sâu như ResNet152. Kết quả này khẳng định rằng mô hình truyền thống không thích hợp cho các bài toán y tế đặc thù. Lớp Dense trong các mô hình gốc thường có số lượng tham số lớn, dễ dẫn đến overfitting trên bộ dữ liệu nhỏ như Sipakmed. Điều này hạn chế khả năng phân loại chính xác. Logistic Regression không yêu cầu tham số phức tạp, giúp giảm overfitting và tập trung vào việc phân loại dựa trên các đặc trưng đã học. Điều này giải thích tại sao hầu hết các mô hình sử dụng Logistic Regression đều cho kết quả tốt hơn

Nhóm Transformer-based models chiếm ưu thế, với độ chính xác trung bình cao hơn so với các mô hình truyền thống. Biến thể BoTNet + Logistic Regression (96.5%) là mô hình tốt nhất nhờ sự kết hợp mạnh mẽ giữa self-attention và Logistic Regression và kiến trúc phù hợp với bộ dữ liệu. Đây là mô hình có hiệu suất tốt nhất. Sức mạnh của biến thể BoTNet trong việc học các đặc trưng dài hạn kết hợp với Logistic Regression giúp cải thiện khả năng phân loại chi tiết, đưa ra kết quả tốt. Logistic Regression cải thiện hiệu suất trong mọi trường hợp, nhờ giảm overfitting và tối ưu hóa phân loại đặc trưng.

Biến thể CvT sử dụng convolutional token embedding và projection, phù hợp với đặc thù của hình ảnh y tế. Kết hợp thêm Logistic Regression hoạt động như lớp phân loại cuối cùng, giảm overfitting và cải thiện khả năng phân biệt. Điều này lý giải độ chính xác khá cao (96.05%).



Biến thể CMT tích hợp convolution vào transformer blocks, tận dụng inductive bias của CNN cho medical images. Kết hợp Logistic Regression cải thiện thêm độ chính xác (96.21%), nhưng không vượt qua BoTNet + Logistic Regression do đặc thù kiến trúc có thể do kiến trúc chưa tối ưu cho dữ liệu đặc thù.

Biến thể CoAtNet kết hợp Convolution và Attention, học tốt cả đặc trưng cục bộ và tổng quát. Tuy nhiên, nó chưa đạt mức độ chi tiết như biến thể CMT hoặc BoTNet trên dữ liệu nhỏ. Logistic Regression tăng hiệu suất, nhưng không đạt mức tốt nhất. Điều này cho thấy CoAtNet có khả năng hạn chế khi học các đặc trưng nhỏ và phức tạp trong ảnh tế bào.

Kết hợp Logistic Regression giúp tăng độ hiệu quả cho các biến thể của mô hình phân loại: BoTNet: +2.75% (cải thiện nhiều nhất), CMT: +1.81%, CoAtNet: +0.97%, CvT: +0.49% (cải thiện ít nhất). Nguyên nhân do Logistic Regression classifier giúp tối ưu hóa decision boundary. Hiệu quả cao hơn với các mô hình có feature representation chưa tối ưu. Biến thể CvT đã có feature representation tốt nên cải thiện ít.

## Chương 5: KẾT LUẬN

Kết quả đạt được: Nghiên cứu đã thử nghiệm và so sánh hiệu quả của các mô hình biến thể trên bộ dữ liệu hình ảnh tế bào Sipakmed. Các mô hình như biến thể của BoTNet + Logistic Regression đạt độ chính xác cao nhất, vượt trội so với các mô hình truyền thống như VGG và ResNet. Kết quả cho thấy sức mạnh của các kiến trúc CNN, Transformer kết hợp Logistic Regression trong bài toán phân loại hình ảnh y tế.

Kết hợp Logistic Regression với biến thể các mô hình tiên tiến giúp cải thiện hiệu suất phân loại. Điều này mở ra hướng tối ưu hóa mô hình phân loại đặc thù cho dữ liệu y tế nhỏ và phức tạp.

Logistic Regression có tiềm năng như một công cụ cải thiện phân loại khi kết hợp với các kiến trúc hiện đại. Phương pháp này có thể áp dụng rộng rãi hơn trong các bài toán liên quan đến hình ảnh y tế.

## Chương 6: HƯỚNG PHÁT TRIỂN

Thử nghiệm các kiến trúc Transformer hiện đại như Swin Transformer hoặc ViT kết hợp cùng các kỹ thuật sử dụng cho bài toán phân lớp tế bào.

Xây dựng các mô hình phân lớp với các phương pháp sử dụng nhiều tài nguyên phần cứng hơn để đánh giá hiệu suất phân loại và kết quả.

Tiến hành thực nghiệm trên nhiều bộ dữ liệu hình ảnh y tế.

Sử dụng các kỹ thuật augmentation và sinh dữ liệu nhân tạo để cải thiện khả năng học của mô hình.

Triển khai hệ thống hỗ trợ chẩn đoán tự động dựa trên các mô hình đạt hiệu suất cao và rút ngắn thời gian chuẩn đoán phân loại.

Sử dụng phương pháp ensemble để kết hợp các mô hình tốt nhất, cải thiện độ chính xác tổng thể.

Nghiên cứu thêm các mô hình lightweight để triển khai trên thiết bị hạn chế tài nguyên.

## TÀI LIỆU THAM KHẢO

- [1] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, "DeepPap: Deep Convolutional Networks for Cervical Cell Classification," arXiv preprint arXiv:1801.08616, 2018. [Online]. Available: <https://arxiv.org/abs/1801.08616>
- [2] M. Lu, X. Wang, Y. Yu, "Cervical cell classification using transfer learning," *Computers in Biology and Medicine*, vol. 107, pp. 13-20, 2019.
- [3] A. Bora, A. Singh, P. Rani, "Cervical Cancer Diagnosis Using Hybrid Deep Learning Model," *Biomedical Signal Processing and Control*, vol. 68, 2021.
- [4] H. Dai, X. Liu, Y. Jiang, "CoAtNet: Marrying Convolution and Attention for All Data Sizes," arXiv preprint arXiv:2106.04803, 2021. [Online]. Available: <https://arxiv.org/abs/2106.04803>
- [5] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "CvT: Introducing Convolutions to Vision Transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 22–31. [Online]. Available: [https://openaccess.thecvf.com/content/ICCV2021/html/Wu\\_CvT\\_Introducing\\_Convolutions\\_to\\_Vision\\_Transformers\\_ICCV\\_2021\\_paper.html](https://openaccess.thecvf.com/content/ICCV2021/html/Wu_CvT_Introducing_Convolutions_to_Vision_Transformers_ICCV_2021_paper.html)
- [6] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck Transformers for Visual Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 16519–16529. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2021/html/Srinivas\\_Bottleneck\\_Transformers\\_for\\_Visual\\_Recognition\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content/CVPR2021/html/Srinivas_Bottleneck_Transformers_for_Visual_Recognition_CVPR_2021_paper.html)
- [7] J. Guo, K. Han, H. Wu, Y. Tang, X. Chen, Y. Wang, and C. Xu, "CMT: Convolutional Neural Networks Meet Vision Transformers," arXiv preprint arXiv:2107.06263, 2021. [Online]. Available: <https://arxiv.org/abs/2107.06263>
- [8] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck Transformers for Visual Recognition," arXiv preprint arXiv:2101.11605, 2021. [Online]. Available: <https://arxiv.org/abs/2101.11605>
- [9] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, S. Wang, and Y. Zhang, "DeepCervix: A Deep Learning-Based Framework for the Classification of Cervical Cells Using Hybrid Deep Feature Fusion Techniques," *Computers in Biology and Medicine*, vol. 136, p. 104649, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482521004431>