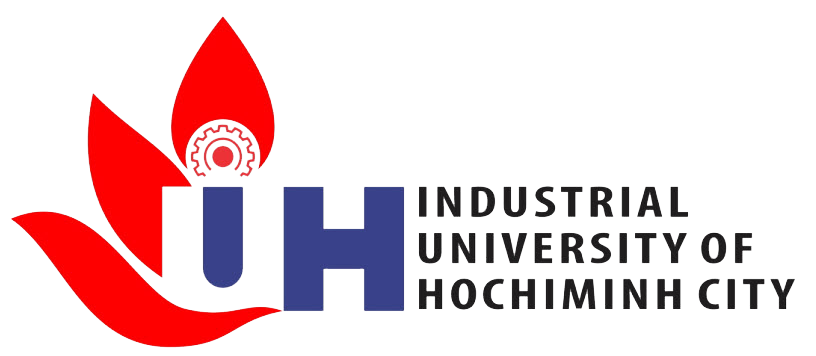
**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC**

**Ứng Dụng Nhận Biết Các Loài Rắn Độc ở Việt Nam**

**GIẢNG VIÊN HƯỚNG DẪN: Võ Quang Hoàng Khang**

**SINH VIÊN THỰC HIỆN 1: Nguyễn Võ Đức**

**MÃ SỐ SINH VIÊN: 21061511**

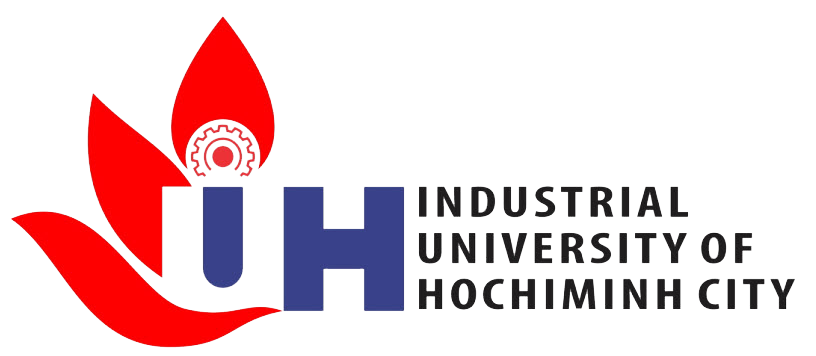
**SINH VIÊN THỰC HIỆN 2: Lê Hoàng Nam**

**MÃ SỐ SINH VIÊN: 21017311**

*TP HỒ CHÍ MINH, Tháng 04 năm 2025*

**INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**GRADUATION THESIS**

**APPLICATION FOR IDENTIFYING VENOMOUS SNAKE SPECIES IN VIETNAM**

**SUPERVISOR: MSc. Vo Quang Hoang Khang**

**STUDENT NAME: Nguyen Vo Duc**

**STUDENT CODE: 21061511**

**STUDENT NAME: Le Hoang Nam**

**STUDENT CODE: 21017311**

*HO CHI MINH CITY, Month 04 year 2025*

**TÓM TẮT**

Nghiên cứu của chúng tôi hướng đến việc giải quyết bài toán nhận diện rắn độc, một yếu tố then chốt trong cả lĩnh vực y tế cộng đồng và nỗ lực bảo tồn sự phong phú của các loài sinh vật. Các phương pháp nhận diện rắn độc hiện tại thường gặp nhiều thách thức do sự khác biệt lớn về hình dáng giữa các loài và sự đa dạng của môi trường sống. Để vượt qua những hạn chế này, chúng tôi giới thiệu một phương pháp mới, sử dụng nhiều hình ảnh của cùng một con rắn và một mạng nơ-ron tích chập sâu (CNN) để phân loại chúng. Mô hình mà chúng tôi đề xuất khai thác thông tin từ nhiều góc độ khác nhau của rắn (ví dụ: ảnh toàn thân và ảnh chụp gần) bằng cách kết hợp các đặc trưng một cách hiệu quả. Cách tiếp cận này giúp mô hình phân biệt các loài rắn độc tốt hơn. Chúng tôi đã thử nghiệm phương pháp này trên một tập dữ liệu ảnh rắn độc, đồng thời áp dụng các kỹ thuật tiền xử lý và tăng cường dữ liệu để làm cho mô hình hoạt động ổn định hơn. Kết quả cho thấy mô hình của chúng tôi đạt được độ chính xác cao hơn trên tập dữ liệu kiểm thử so với các mô hình cơ sở và các phương pháp chỉ dùng một ảnh duy nhất. Nghiên cứu này cho thấy tiềm năng lớn của việc kết hợp phân loại đa ảnh và học sâu để nhận diện rắn độc một cách chính xác và hiệu quả, mở ra nhiều ứng dụng thực tế trong việc hỗ trợ chẩn đoán y tế và quản lý bảo tồn.

**LỜI CẢM ƠN**

Trong quá trình thực hiện báo cáo môn học, nhóm chúng em đã nhận được sự quan tâm sâu sắc, hướng dẫn tận tâm và sự hỗ trợ nhiệt tình từ quý thầy cô Khoa Công nghệ Thông tin, cùng với sự động viên, giúp đỡ từ bạn bè và gia đình.

Lời đầu tiên, chúng em xin bày tỏ lòng biết ơn sâu sắc đến Ban Giám hiệu Trường Đại học Công nghiệp Thành phố Hồ Chí Minh, Ban Chủ nhiệm Khoa Công nghệ Thông tin đã luôn tạo điều kiện và hỗ trợ chúng em trong suốt thời gian học tập tại trường.

Đặc biệt, chúng em xin gửi lời tri ân chân thành và sâu sắc nhất đến thầy hướng dẫn, Thạc sĩ Võ Quang Hoàng Khang, người đã trực tiếp chỉ bảo, tận tình giúp đỡ chúng em hoàn thành bài báo cáo này.

Chúng em xin trân trọng cảm ơn!

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**TP. Hồ Chí Minh, ngày…. tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**LỜI MỞ ĐẦU**

Trong bối cảnh kỷ nguyên số hiện nay, các mô hình máy học và học sâu đã và đang chứng minh vai trò then chốt trong nhiều lĩnh vực của đời sống. Để thỏa mãn nhu cầu lớn hơn về ứng dụng công nghệ thông minh, ngành Khoa học Máy tính không ngừng đòi hỏi những chuyên gia có kiến thức chuyên sâu và kỹ năng thực hành để phát triển các mô hình máy học và học sâu tiên tiến, từ đó giải quyết những thách thức và đáp ứng nhu cầu đa dạng của thế giới hiện đại.

Trong bối cảnh đó, chúng em chọn chủ đề **“Nhận diện rắn độc ở Việt Nam”** bằng cách sử dụng phân loại đa ảnh cho báo cáo này. Đây là một chủ đề có ý nghĩa thiết thực, đặc biệt trong các khu vực có sự đa dạng sinh học cao và tỷ lệ tai nạn do rắn cắn đáng kể. Việc phát triển một hệ thống nhận diện rắn độc nhanh chóng và chính xác có thể hỗ trợ công tác y tế, giảm thiểu các trường hợp chẩn đoán sai sót và đưa ra phương pháp điều trị kịp thời. Hơn nữa, nó còn có tiềm năng ứng dụng trong lĩnh vực bảo tồn đa dạng sinh học, giúp các nhà nghiên cứu và quản lý môi trường giám sát và phân loại các loài rắn hiệu quả hơn.

Báo cáo này tập trung vào việc nghiên cứu và xây dựng một mô hình nhận diện rắn độc cải tiến, sử dụng phương pháp phân loại đa ảnh kết hợp với các kiến trúc học sâu hiện đại. Mô hình được đánh giá trên một tập dữ liệu ảnh đa dạng về các loài rắn độc, sử dụng các chỉ số đánh giá hiệu suất phổ biến như: Độ chính xác (accuracy), F1-score, precision, recall, specificity và AUC. Kết quả thực nghiệm cho thấy mô hình đề xuất có khả năng nhận diện chính xác các loài rắn độc trên ảnh đa dạng, làm nổi bật tiềm năng ứng dụng thực tế của nó trong việc hỗ trợ nhận diện và phân loại rắn độc.

Cuối cùng, Chúng em hy vọng rằng những kiến thức và kinh nghiệm thu được từ công trình này sẽ đóng góp vào sự phát triển của các ứng dụng học sâu trong lĩnh vực nhận diện sinh vật, đặc biệt là trong bài toán nhận diện rắn độc.

**MỤC LỤC**

[**CHƯƠNG 1.**  **TỔNG QUAN** 1](#_Toc198588013)

[**1.1 Lý do chọn đề tài** 1](#_Toc198588014)

[**1.2 Mục tiêu nghiên cứu** 2](#_Toc198588015)

[**1.3 Phạm vi nghiên cứu** 2](#_Toc198588016)

[**1.4 Phương pháp nghiên cứu** 3](#_Toc198588017)

[**1.5 Kết cấu đồ án.** 4](#_Toc198588018)

[**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT** 5](#_Toc198588019)

[**2.1. Học sâu (Deep learning)** 5](#_Toc198588020)

[**2.2. Chuyển giao học tập (Transfer Learning)** 5](#_Toc198588021)

[**2.3. Tăng cường dữ liệu (Data Augmentation)** 5](#_Toc198588022)

[**2.4. Các thành phần kiến trúc mạng và tối ưu hóa** 6](#_Toc198588023)

[**2.5. Các phương pháp đánh giá mô hình** 6](#_Toc198588024)

[**2.5.1. Chỉ số độ chính xác (Accuracy) và độ mất mát (Loss)** 6](#_Toc198588025)

[**2.5.2**. **Ma trận nhầm lẫn (Confusion Matrix)** 7](#_Toc198588026)

[**2.5.3. Báo cáo phân loại (Classification Report)** 7](#_Toc198588027)

[**CHƯƠNG 3. Phương Pháp và dữ liệu** 8](#_Toc198588028)

[**3.1 Mô hình tổng quát** 8](#_Toc198588029)

[**3.2 Mô tả dataset** 8](#_Toc198588030)

[**3.3 Tổng quan về phương pháp.** 9](#_Toc198588031)

[3.3.1 Kiến trúc VGG 9](#_Toc198588032)

[3.3.2 Kiến trúc EfficientNet 10](#_Toc198588033)

[3.3.3 Học chuyển giao 12](#_Toc198588034)

[**CHƯƠNG 4. THỰC NGHIỆM** 13](#_Toc198588035)

[**4.1 Môi trường thực nghiệm** 13](#_Toc198588036)

[**4.1.1. Cấu hình phần mềm** 13](#_Toc198588037)

[**4.1.2. Thiết lập môi trường** 13](#_Toc198588038)

[Các thiết lập khác bao gồm: 13](#_Toc198588039)

[**4.2 Tập dữ liệu** 14](#_Toc198588040)

[4.2.1. Nguồn dữ liệu 14](#_Toc198588041)

[4.2.2. Số lượng dữ liệu 14](#_Toc198588042)

[4.2.3. Tiền xử lý dữ liệu 14](#_Toc198588043)

[4.2.4. Tăng cường dữ liệu 14](#_Toc198588044)

[4.2.5. Chia dữ liệu 15](#_Toc198588045)

[**4.3 Ứng dụng thực nghiệm** 15](#_Toc198588046)

[4.3.1. Quy trình huấn luyện mô hình 15](#_Toc198588047)

[4.3.2. Cấu hình huấn luyện 16](#_Toc198588048)

[4.3.3. Kết quả thực nghiệm 16](#_Toc198588049)

[**CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** 19](#_Toc198588050)

[**5.1 Kết luận** 19](#_Toc198588051)

[**5.2 Hướng phát triển** 19](#_Toc198588052)

**DANH MỤC CÁC THUẬT NGỮ VIẾT TẮT**

|  |  |  |
| --- | --- | --- |
| Từ viết tắt | Từ đầy đủ | Nghĩa |
| **GAP** | Global Average Pooling | Kỹ thuật gôp trung bình toàn cầu dùng để giảm kích thước của các đặc trưng dữ liệu và chuẩn bị dữ liệu cho các lớp tích chập sau bằng cách tính trung bình các giá trị pixel trong feature map tạo ra một vector có kích thước bằng số lượng feature maps, mỗi phần tử trong vector đại diện cho giá trị trung bình của một feature map. |
| **BN** | Batch Normalization | Chuẩn hoá hàng loạt là một kỹ thuật trong học sâu. Nó tính toán trung bình và phương sai của dữ liệu trong một batch (một nhóm dữ liệu nhỏ), sau đó chuẩn hóa dữ liệu dựa trên các giá trị này. Từ đó làm cho dữ liệu phân phối ổn định giúp các mạng nơ-ron học nhanh, hiệu quả hơn. |
| **CNN** | Convolutional Neural Network | Mạng nơ-ron tích chập, một mô hình học sâu, đặc biệt hiệu quả trong các tác vụ xử lý ảnh và thị giác máy tính. |
| **LSTM** | Long Short-Term Memory | Mạng nơ-ron hồi tiếp đặc biệt dùng cho dữ liệu tuần tự hoặc chuỗi thời gian. |

# **CHƯƠNG 1.** **TỔNG QUAN**

## **1.1 Lý do chọn đề tài**

Tình trạng tai nạn do rắn độc cắn hiện đang đặt ra một gánh nặng đáng kể cho hệ thống y tế công cộng ở nhiều quốc gia trên toàn cầu, đặc biệt là tại các vùng nông thôn và những khu vực có khí hậu nhiệt đới. Hậu quả nghiêm trọng của nó thể hiện qua tỷ lệ thương tật và tử vong vẫn còn ở mức cao. Do đó, việc nhận diện một cách nhanh chóng và chính xác loài rắn gây ra vết cắn trở thành yếu tố then chốt, ảnh hưởng trực tiếp đến việc lựa chọn liệu pháp điều trị bằng huyết thanh kháng nọc độc một cách thích hợp và kịp thời, góp phần quan trọng trong việc giảm thiểu các biến chứng nguy hiểm. Tuy nhiên, các phương pháp nhận dạng rắn truyền thống thường dựa vào kinh nghiệm cá nhân hoặc kiến thức chuyên môn của các nhà khoa học, điều này 3 .đòi hỏi thời gian và sự am hiểu sâu sắc về đặc điểm hình thái, đồng thời tiềm ẩn nguy cơ sai sót chủ quan.

Ứng dụng học sâu trong nhận diện ảnh y tế: Trong bối cảnh đó, việc khai thác các kỹ thuật của ngành khoa học máy tính, đặc biệt là học sâu, để hỗ trợ công tác nhận diện rắn độc trở nên hết sức cần thiết. Các mô hình phân loại ảnh trong lĩnh vực y tế như CNN, VGG16, ResNet, InceptionV3 và DenseNet121 đã chứng minh được hiệu quả vượt trội trong việc xử lý hình ảnh và được ứng dụng rộng rãi trong nhiều bài toán thị giác máy tính. Dựa trên các công trình nghiên cứu gần đây [1] [2] [3], sự kết hợp giữa mạng nơ-ron tích chập sâu và các cơ chế tập trung sự chú ý đã cho thấy những triển vọng đầy hứa hẹn trong việc phân loại ảnh một cách tinh tế, giúp mô hình tập trung vào những đặc điểm quan trọng nhất để phân biệt giữa các loài.

Bài nghiên cứu này đề xuất một phương pháp cải tiến mô hình học sâu nhằm mục đích nhận diện các loài rắn độc phổ biến tại Việt Nam. Để đánh giá hiệu quả hoạt động của mô hình được đề xuất, chúng tôi sử dụng các phương pháp đánh giá định lượng như Độ chính xác, Độ chuẩn xác, Độ nhạy, Điểm số F1 và đường cong AUC-ROC. 5 .Các chỉ số này giúp xác định mức độ chính xác và khả năng khái quát hóa của mô hình. Cuối cùng, mô hình của chúng tôi sẽ được so sánh và đánh giá với các mô hình học sâu khác để lựa chọn ra mô hình phù hợp nhất cho bài toán nhận diện rắn độc. Quá trình so sánh này dựa trên các chỉ số đánh giá tiêu chuẩn như Độ chính xác, Độ nhạy, Độ chuẩn xác và Điểm số F1, nhằm xác định mô hình có hiệu suất và khả năng tổng quát hóa tốt nhất trên tập dữ liệu thực tế.

## **1.2 Mục tiêu nghiên cứu**

1. Phát triển mô hình phân loại đa ảnh**:** xây dựng một kiến trúc học sâu độc đáo, có khả năng tổng hợp và xử lý hiệu quả thông tin từ nhiều góc độ hình ảnh khác nhau của cùng một cá thể rắn.
2. Tích hợp các kỹ thuật tiên tiến**:** Nghiên cứu và tích hợp các kỹ thuật học sâu hiện đại như cơ chế tập trung (attention mechanisms), các phương pháp tối ưu hóa hiệu năng (Efficient Net), và các chiến lược kết hợp đặc trưng đa dạng để nâng cao khả năng phân biệt giữa các loài rắn độc.
3. So sánh hiệu suất với các mô hình khác**:** Đánh giá và so sánh hiệu suất của mô hình đề xuất với các kiến trúc mạng nơ-ron tích chập phổ biến và các phương pháp tiếp cận khác trong bài toán phân loại ảnh.
4. Đánh giá tính ổn định và tổng quát hóa: Áp dụng các kỹ thuật kiểm tra chéo để đảm bảo tính ổn định của mô hình và khả năng dự đoán tốt trên dữ liệu mới.
5. Phân tích và đề xuất hướng phát triển**:** Xác định những hạn chế của mô hình hiện tại, phân tích các yếu tố ảnh hưởng đến hiệu suất, và đề xuất các hướng nghiên cứu tiềm năng trong tương lai để cải thiện khả năng nhận diện và ứng dụng của mô hình trong thực tế

## **1.3 Phạm vi nghiên cứu**

* **Đối tượng nghiên cứu**: Hình ảnh các loài rắn, tập trung vào các đặc điểm hình thái có giá trị trong việc phân loại và nhận diện loài, bao gồm ảnh toàn thân, ảnh cận cảnh đầu, ảnh vảy và các góc chụp khác nhau.
* **Phạm vi kỹ thuật**: Sử dụng các mô hình học sâu, đặc biệt là kiến trúc mạng nơ-ron tích chập (CNN) như Efficient B0, ResNet,v.v**.** kết hợp với các kỹ thuật tối ưu hóa hiệu năng như Adam Optimizer, fine-tune Efficient Net. Bên cạnh đó, nghiên cứu cũng tập trung vào việc khai thác hiệu quả thông tin từ đa ảnh thông qua các chiến lược hợp nhất đặc trưng (feature fusion) và có thể bao gồm việc áp dụng các cơ chế chú ý (attention mechanisms) để tập trung vào các vùng quan trọng trên ảnh
* **Phạm vi kiểm tra**: Đề tài tập trung vào việc đánh giá hiệu suất của mô hình học sâu trên tập dữ liệu ảnh rắn được tự thu thập hoặc tổng hợp từ nguồn công khai. Việc đánh giá được thực hiện bằng các phương pháp và chỉ số phù hợp cho bài toán phân loại đa lớp, bao gồm: độ chính xác (Accuracy), độ chính xác dự đoán (Precision), độ nhạy (Recall), F1-score và AUC-ROC. Các chỉ số này được sử dụng nhằm phản ánh khả năng tổng quát hóa và hiệu năng thực tế của mô hình khi áp dụng vào bài toán nhận diện loài rắn.

## **1.4 Phương pháp nghiên cứu**

1. **Nghiên cứu tài liệu**:

* Tìm hiểu về đặc điểm hình thái học của các loài rắn độc và các phương pháp nhận diện truyền thống.
* Nghiên cứu các mô hình học sâu hiện đại, đặc biệt là các kiến trúc mạng nơ-ron tích chập (CNN) tiên tiến, đã được chứng minh hiệu quả trong bài toán phân loại ảnh.
* Tham khảo các công trình nghiên cứu trước đây liên quan đến nhận diện động vật hoặc rắn bằng học sâu (nếu có) và các kỹ thuật xử lý đa ảnh.

1. **Xây dựng mô hình**:

* Lựa chọn và xây dựng kiến trúc mô hình học sâu phù hợp cho bài toán phân loại đa ảnh rắn độc.
* Tích hợp các cơ chế attention để tăng cường khả năng trích xuất đặc trưng từ ảnh.
* Sử dụng kỹ thuật nhóm tích chập (Grouped Convolution) để tăng tính đa dạng của đặc trưng của ảnh và tối ưu hoá hiệu năng của mô hình.

1. **Kiểm thử và đánh giá**:

* Sử dụng phương pháp đánh giá chéo trên tập dữ liệu ảnh rắn để đánh giá khách quan hiệu suất và độ ổn định của mô hình
* So sánh kết quả với các mô hình khác như VGG16, ResNet50, MobileNet, InceptionV3 và DenseNet121.

1. **Phân tích và cải tiến**:

* Sử dụng chiến lược trộn và hợp nhất đặc trưng (feature fusion) dựa trên attention để tối ưu hiệu quả mô hình.
* Áp dụng các chiến lược tăng cường dữ liệu và tối ưu hóa thông qua các khối dư (residual block) và phân bổ động (dynamic allocation).

## **1.5 Kết cấu đồ án.**

Chương 1: Tổng quan

Chương 2: Cơ sở lý thuyết

Chương 3: Phân tích yêu cầu và thiết kế mô hình

Chương 4: Thực nghiệm

Chương 5: Kết luận và hướng phát triển

Tài liệu tham khảo.

# **CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

## **2.1. Học sâu (Deep learning)**

Học sâu (Deep Learning) là nền tảng của mô hình phân loại ảnh được xây dựng trong code. Đây là một nhánh của học máy sử dụng các mạng nơ-ron nhân tạo sâu (nhiều tầng) để học các biểu diễn phức tạp từ dữ liệu thô, cụ thể ở đây là hình ảnh các loài rắn. Các mô hình học sâu có khả năng tự động trích xuất các đặc trưng cấp cao từ ảnh, giúp phân biệt các loại rắn dựa trên các đặc điểm ngoại hình như màu sắc, họa tiết, hình dạng cơ thể. Việc sử dụng kiến trúc EfficientNetB0, một mô hình CNN hiệu quả, cho phép mô hình học các đặc trưng mạnh mẽ để phân loại chính xác.

## **2.2. Chuyển giao học tập (Transfer Learning)**

Kỹ thuật chuyển giao học tập (Transfer Learning) là một phương pháp quan trọng được áp dụng trong code. Thay vì xây dựng và huấn luyện mô hình từ đầu với một tập dữ liệu nhỏ, code sử dụng mô hình EfficientNetB0 đã được huấn luyện trước trên tập dữ liệu lớn ImageNet. Các lớp ban đầu của mô hình được "đóng băng", chỉ các lớp cuối được thêm vào để huấn luyện trên tập dữ liệu ảnh rắn. Điều này cho phép tận dụng kiến thức (các đặc trưng ảnh chung) mà mô hình đã học từ ImageNet và áp dụng chúng vào bài toán phân loại rắn, giúp đạt hiệu quả cao hơn với lượng dữ liệu huấn luyện hạn chế và giảm thời gian huấn luyện.

## **2.3. Tăng cường dữ liệu (Data Augmentation)**

Tăng cường dữ liệu (Data Augmentation) được thực hiện thông qua ImageDataGenerator. Kỹ thuật này tạo ra các phiên bản biến đổi của ảnh huấn luyện gốc (ví dụ: xoay, lật, thay đổi kích thước) để làm phong phú tập dữ liệu và tăng tính đa dạng. Mục tiêu chính của tăng cường dữ liệu là giúp mô hình ít nhạy cảm hơn với các biến thể nhỏ trong ảnh và giảm hiện tượng quá khớp (overfitting), đặc biệt khi tập dữ liệu không lớn. Điều này giúp mô hình có khả năng tổng quát hóa tốt hơn khi gặp các ảnh rắn mới.

## **2.4. Các thành phần kiến trúc mạng và tối ưu hóa**

Các thành phần phổ biến trong kiến trúc mạng nơ ron cho bài toán phân loại:

* **Global Average Pooling 2D:** Giúp giảm chiều dữ liệu trước khi đưa vào lớp cuối cùng.
* **Dropout:** Một kỹ thuật điều chuẩn giúp giảm quá khớp bằng cách ngẫu nhiên loại bỏ các nơ-ron trong quá trình huấn luyện.
* **Dense (Fully Connected) Layer:** Lớp cuối cùng để thực hiện phân loại, với số lượng nơ-ron bằng số lượng lớp rắn và sử dụng hàm kích hoạt softmax để đưa ra xác suất của từng lớp.
* **Optimizer (Adam):** Thuật toán được sử dụng để cập nhật trọng số của mô hình trong quá trình huấn luyện nhằm giảm thiểu hàm mất mát.
* **Loss function (Categorical Crossentropy):** Hàm mất mát được sử dụng để đo lường sự khác biệt phân phối xác suất dự đoán và phân phối xác suất thực tế của nhãn, phù hợp với bài toán phân loại đa lớp.

## **2.5. Các phương pháp đánh giá mô hình**

Sau khi mô hình học sâu được huấn luyện, việc đánh giá hiệu suất của nó là bước cực kỳ quan trọng để hiểu rõ mô hình hoạt động tốt như thế nào trên dữ liệu chưa từng thấy.

#### **2.5.1. Chỉ số độ chính xác (Accuracy) và độ mất mát (Loss)**

* Mô hình được đánh giá trực tiếp trên tập dữ liệu kiểm tra để tính toán độ chính xác và độ mất mát.
* **Độ chính xác:** Đo lường tỉ lệ các mẫu (ảnh) mà mô hình dự đoán đúng lớp của chúng. Đây là một chỉ số đơn giản và trực quan về hiệu suất tổng thể của mô hình.
* **Độ mất mát:** Là giá trị của hàm mất mát trên tập kiểm tra. Độ mất mát thấp cho thấy mô hình dự đoán gần với nhãn thực tế hơn.

#### **2.5.2**. **Ma trận nhầm lẫn (Confusion Matrix)**

Ma trận nhầm lẫn [5] là một công cụ mạnh mẽ để đánh giá hiệu suất của mô hình phân loại trên từng lớp riêng lẻ. Nó hiển thị số lượng các mẫu được phân loại đúng và sai cho mỗi lớp

Trong đó:

* TP: Số lượng mẫu của một lớp được dự đoán đúng là lớp đó (True Positives).
* FP: Số lượng mẫu không thuộc một lớp nhưng bị dự đoán sai là thuộc lớp đó (False Positives).
* TN: Số lượng mẫu không thuộc một lớp được dự đoán đúng là không thuộc lớp đó (True Negatives).
* FN: Số lượng mẫu thuộc một lớp nhưng bị dự đoán sai là không thuộc lớp đó (False Negatives).

#### **2.5.3. Báo cáo phân loại (Classification Report)**

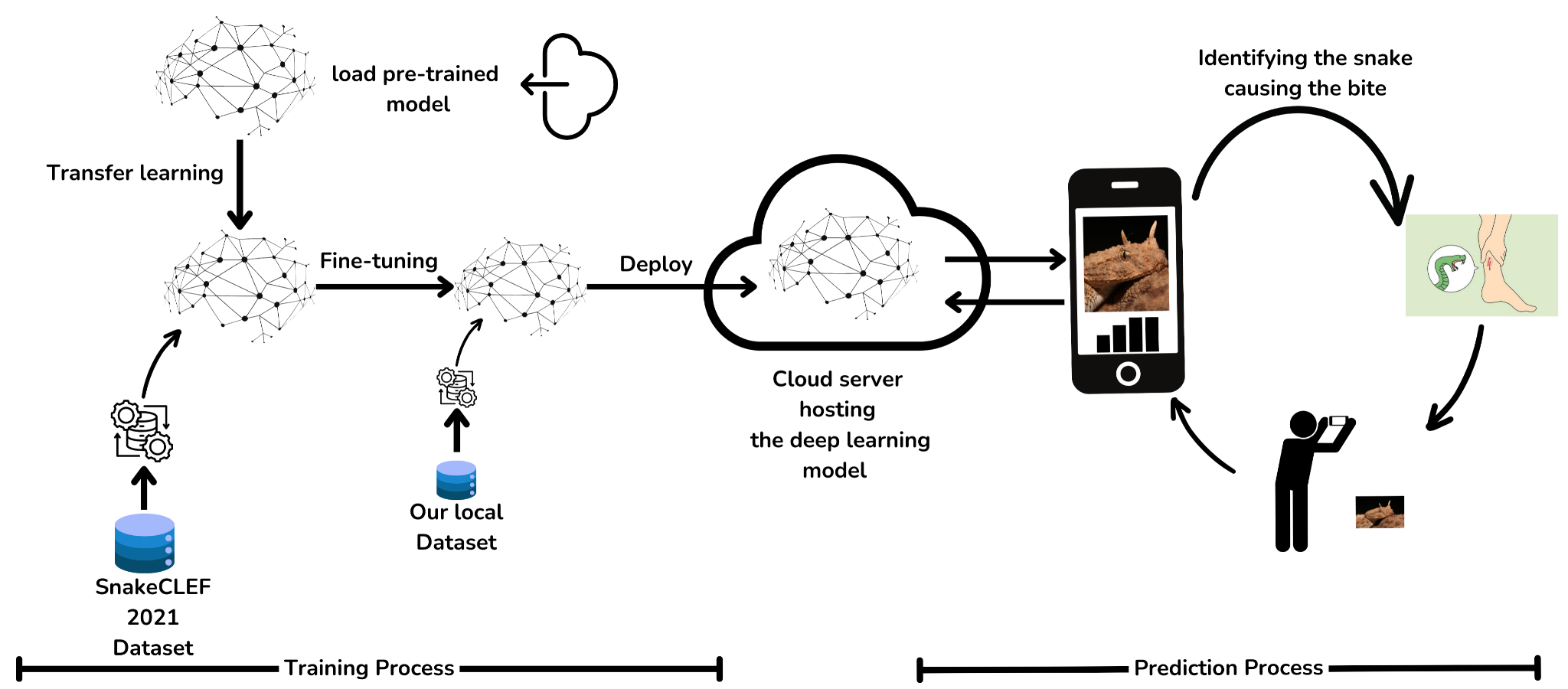
Báo cáo phân loại[6] cung cấp các chỉ số đánh giá chi tiết cho từng lớp bao gồm Precision, Recall và F1-score

* **Precision (Độ chính xác):** Đối với một lớp, Precision là tỷ lệ các mẫu được dự đoán là lớp đó mà thực sự thuộc về lớp đó (TP/ (TP + FP)). Nó cho biết khi mô hình dự đoán một mẫu thuộc lớp X, khả năng nó thực sự là lớp X là bao nhiêu.
* **Recall (Độ thu hồi hay Độ nhạy):** Đối với một lớp, Recall là tỷ lệ các mẫu thực sự thuộc lớp đó mà mô hình dự đoán đúng là lớp đó (TP / (TP + FN)). Nó cho biết khả năng mô hình tìm thấy tất cả các mẫu thuộc lớp đó là bao nhiêu.
* **F1-Score:** Là trung bình điều hóa của Precision và Recall (2 \* (Precision \* Recall) / (Precision + Recall)). F1-Score cân bằng giữa Precision và Recall và thường được sử dụng khi có sự mất cân bằng giữa các lớp.

# **CHƯƠNG 3. Phương Pháp và dữ liệu**

## **3.1 Mô hình tổng quát**

Phần này trình bày chi tiết phương pháp được áp dụng trong nghiên cứu, bao gồm các kỹ thuật nhận dạng loài rắn, đặc điểm của bộ dữ liệu, các bước tiền xử lý, và các kiến trúc học sâu được sử dụng. Những hiểu biết thu được từ tổng quan tài liệu đã định hướng việc tích hợp cả dữ liệu địa phương và dữ liệu toàn cầu nhằm đảm bảo hiệu suất mô hình mạnh mẽ và phù hợp về mặt sinh thái, như minh họa trong Hình 1.



**Hình 1**: Tổng quan hệ thống nhận diện loài rắn dựa trên học sâu.

## **3.2 Mô tả dataset**

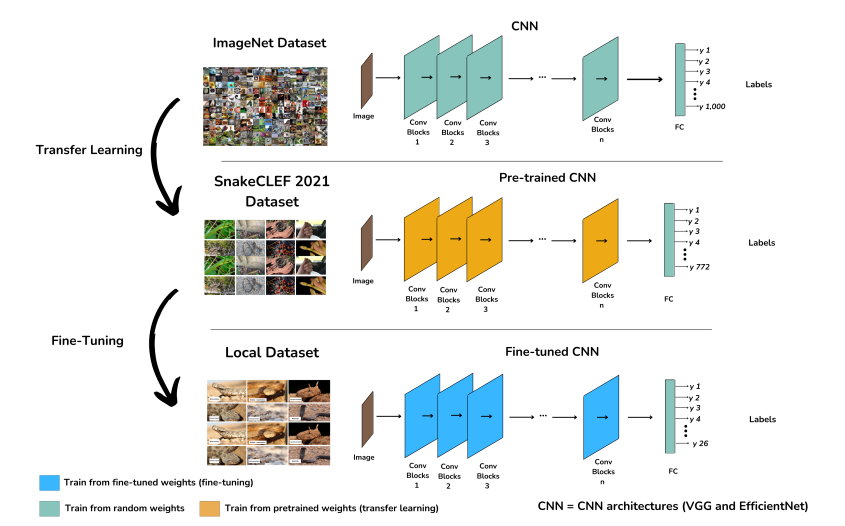
Chúng em đã thu thập bộ dữ liệu sử dụng trong nghiên cứu này được thu thập từ nền tảng **iNaturalist**, một cộng đồng trực tuyến lớn về khoa học công dân, nơi người dùng chia sẻ các quan sát về động thực vật trên khắp thế giới. Chúng tôi đã tận dụng nguồn tài nguyên phong phú này để thu thập hình ảnh của các loài rắn độc phổ biến ở Việt Nam.

Đối với mỗi loài rắn độc phổ biến tại Việt Nam mà chúng tôi tập trung nghiên cứu, số lượng ảnh thu thập được dao động từ 500 đến 1000 hình ảnh. Điều này cung cấp một lượng dữ liệu đáng kể cho việc huấn luyện các mô hình học sâu.Bộ dữ liệu được gán nhãn dựa trên tên khoa học của từng loài rắn độc. Chúng tôi tập trung vào những loài rắn độc phổ biến ở Việt Nam như cạp nong, cạp nia, hổ mang chúa, rắn lục đuôi đỏ, v.v. Các loài này được chọn dựa trên mức độ độc hại và tần suất xuất hiện trong các trường hợp tai nạn do rắn cắn ở Việt Nam.

## **3.3 Tổng quan về phương pháp.**

Phương pháp được sử dụng xây dựng mô hình nhận diện loài rắn chính xác với ba nội dung chính:

1. Lựa chọn kiến trúc mạng nơ-ron tích chập (CNN) – bao gồm VGG và EfficientNet
2. Áp dụng học chuyển tiếp (Transfer Learning)
3. Thực hiện tinh chỉnh mô hình (Fine-tuning)



**Hình 2**  Minh họa quy trình học chuyển tiếp và tinh chỉnh, sử dụng các mô hình CNN được huấn luyện trước (VGG và EfficientNet) cho bài toán phân loại loài rắn.

### 3.3.1 Kiến trúc VGG

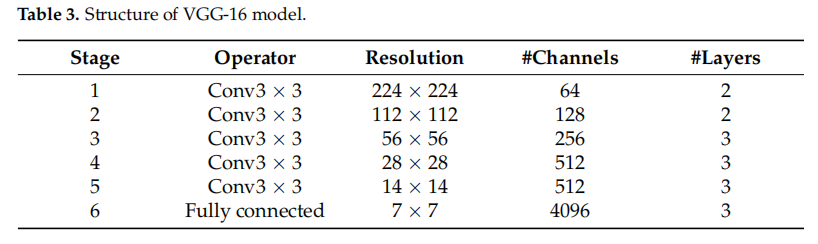
Kiến trúc **VGG**, do Simonyan và cộng sự đề xuất, ấn tượng nhờ **thiết kế đơn giản nhưng sâu**.

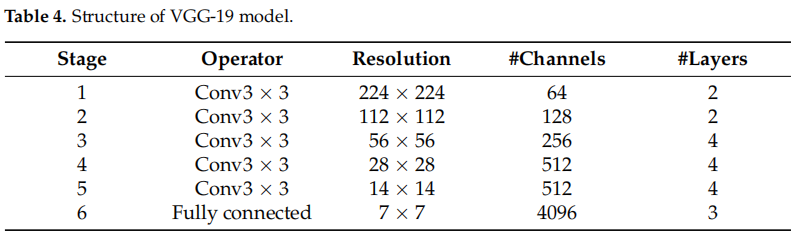
* **VGG-16** gồm 13 lớp tích chập (convolutional layers) và 3 lớp fully connected.
* **VGG-19** có cấu hình sâu hơn, bổ sung thêm nhiều lớp.

Tất cả đều sử dụng bộ lọc tích chập **3 × 3**, bước nhảy bằng 1, giúp trích xuất đặc trưng hình ảnh từ đường viền đơn giản đến kết cấu phức tạp.

Các lớp được tổ chức thành **các khối (blocks)**, mỗi khối gồm các lớp Conv và MaxPooling, giúp giảm dần kích thước ảnh đầu vào nhưng vẫn giữ lại thông tin quan trọng.

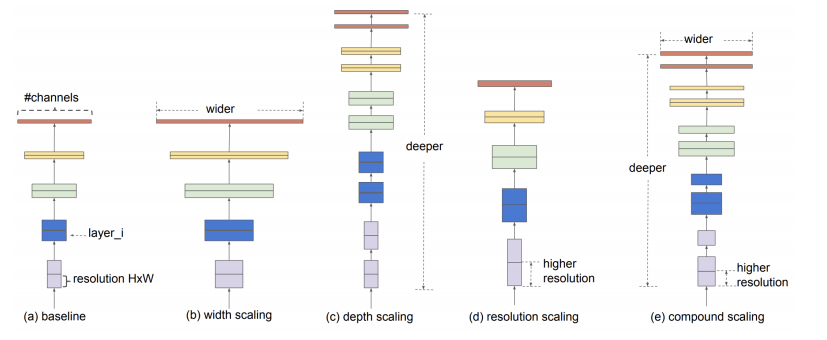
Các lớp fully connected cuối cùng thực hiện nhiệm vụ phân loại.  
 Cấu trúc sâu của VGG, đặc biệt là VGG-16 và VGG-19, cho phép nhận diện chính xác các đặc trưng hình ảnh tinh vi, đã được chứng minh hiệu quả trên tập ImageNet.





### 3.3.2 Kiến trúc EfficientNet

EfficientNet là một kiến trúc mạng nơ-ron tích chập (CNN) hiện đại được thiết kế dựa trên nguyên lý compound scaling – cho phép mở rộng đồng thời ba yếu tố chính của mô hình là chiều rộng (width), chiều sâu (depth) và độ phân giải đầu vào (resolution) một cách tỷ lệ. Cơ chế này được điều chỉnh bởi một hệ số tổng hợp (ϕ), từ đó giúp mô hình đạt được hiệu quả tính toán tối ưu mà vẫn duy trì độ chính xác cao.

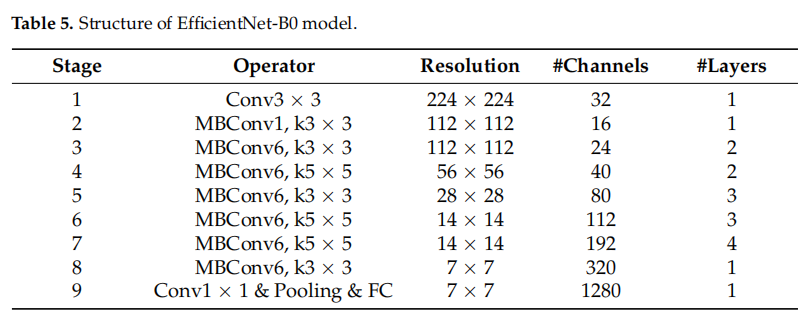


**Hình 3** Minh họa so sánh giữa các phương pháp mở rộng truyền thống và compound scaling.

Kiến trúc EfficientNet sử dụng các khối tích chập MBConv (Mobile Inverted Bottleneck) cùng với khối SE (Squeeze-and-Excitation), cho phép mô hình tập trung mạnh hơn vào các đặc trưng quan trọng trong từng kênh của ảnh. Nhờ vào sự kết hợp này, EfficientNet vừa giữ được khả năng học sâu, vừa giảm thiểu số lượng tham số và thời gian huấn luyện.

Một ưu điểm nổi bật của EfficientNet là khả năng đạt được hiệu suất cao với số lượng tham số thấp hơn so với các mô hình CNN truyền thống. Chẳng hạn, trong khi mô hình GPipe yêu cầu tới 556 triệu tham số để đạt độ chính xác cao, thì EfficientNet có thể đạt độ chính xác tương đương với chỉ 66 triệu tham số, từ đó rất phù hợp với các hệ thống có hạn chế về tài nguyên tính toán.

EfficientNet bao gồm các biến thể từ B0 đến B7, mỗi phiên bản có kích thước mô hình tăng dần. Trong khuôn khổ nghiên cứu này, nhóm tác giả đã lựa chọn EfficientNet B0 – phiên bản nhẹ nhất, có độ cân bằng tốt giữa hiệu suất và độ chính xác, phù hợp với điều kiện dữ liệu rắn và mục tiêu ứng dụng thực tế.



### 3.3.3 Học chuyển giao

Học chuyển giao (Transfer Learning) là một chiến lược hiệu quả trong học sâu, đặc biệt khi dữ liệu huấn luyện cho tác vụ cụ thể có giới hạn. Trong nghiên cứu này, chúng tôi đã áp dụng học chuyển giao để tận dụng kiến thức đã được học từ các bộ dữ liệu lớn hơn, giúp mô hình khởi đầu tốt hơn và giảm nhu cầu về dữ liệu huấn luyện cho bài toán nhận diện các loài rắn. Chúng em đã sử dụng hai phương pháp chính :

* **Trích xuất đặc trưng (Feature Extraction):** Chúng em sử dụng các trọng số đã được huấn luyện trước (pre-trained weights) từ các mô hình CNN (VGG và EfficientNet) đã được huấn luyện trên bộ dữ liệu ImageNet.
* **Tinh chỉnh mô hình (Fine-Tuning):** Chúng em cũng tiến hành điều chỉnh một phần hoặc toàn bộ các trọng số trong các mô hình CNN đã được huấn luyện trước (VGG và EfficientNet). Mục tiêu là điều chỉnh các tham số của mô hình để phù hợp hơn với các đặc điểm hình thái và màu sắc cụ thể của các loài rắn. Cách tiếp cận này giúp tận dụng các đặc trưng tổng quát đã học được từ ImageNet đồng thời cho phép mô hình học các đặc điểm riêng biệt của bộ dữ liệu địa phương, mang lại hiệu quả tính toán và khả năng tái sử dụng mô hình cao hơn.

# **CHƯƠNG 4. THỰC NGHIỆM**

## **4.1 Môi trường thực nghiệm**

#### **4.1.1. Cấu hình phần mềm**

* **Hệ điều hành**: Windows hoặc Linux
* **Ngôn ngữ lập trình**: Python 3.12
* **Thư viện chính sử dụng**:
  + Pytorch: Sử dụng để huấn luyện mô hình học sâu.
  + Numpy: Sử dụng cho các tính toán số học.
  + Albumentations: Sử dụng để tăng cường dữ liệu.
  + Matplotlib/Seaborn: Sử dụng để trực quan hóa dữ liệu

#### **4.1.2. Thiết lập môi trường**

Các thiết lập khác bao gồm:

* Bộ nhớ RAM: 24 GB.
* Dung lượng lưu trữ tạm thời: 1T.
* Thời gian thực nghiệm:Vòng lặp huấn luyện được thiết kế cho 10 epoch với batch size là 32. Thời gian huấn luyện phụ thuộc vào phần cứng.

## **4.2 Tập dữ liệu**

### 4.2.1. Nguồn dữ liệu

Dữ liệu sử dụng trong dự án được lấy từ **iNaturalist**, gồm các ảnh chụp rắn để có thể giúp nhận diện loại rắn, tập dữ liệu gồm 2 nhóm chính là:

* Ảnh rắn độc (VIPER).
* Ảnh rắn thường (SNAKE).

### 4.2.2. Số lượng dữ liệu

Tập dữ liệu ban đầu có tổng cộng 802 ảnh được phân bổ như sau:

* 392 ảnh rắn thường.
* 410 ảnh rắn độc.

### 4.2.3. Tiền xử lý dữ liệu

Để chuẩn bị cho các bước tiếp theo thì ta sẽ tiến hành tiền sử lý dữ liệu trước, ở đây ta sẽ áp dụng kỹ thuật RandomResizedCrop, Resize để đảm bảo kích thước đầu vào cố định và Normalize, ToTensorV2 để chuẩn hóa và định dữ liệu.

### 4.2.4. Tăng cường dữ liệu

Tập dữ liệu có tình trạng mất cân bằng giữa 2 nhóm rắn thường và rắn độc nên ta áp dụng các kỹ thuật tăng cường ảnh để giải quyết vấn đề mất cân bằng dữ liệu đó Các kỹ thuật tăng cường ảnh trong tập train được sử dụng để giúp cho mô hình học được các đặt trưng khác và tránh overfit được liệt kê trong **Bảng 1**:

**Bảng 1.** Các phương pháp và tham số tăng cường dữ liệu

|  |  |
| --- | --- |
| **Phương pháp** | **Tham số** |
| Resized | 0.8 - 1.0 |
| Rotation | ±15 |
| Shift | 0.05 |
| Scale | 0.05 |
| Vertical Flip | True |
| Horizontal flip | True |

### 4.2.5. Chia dữ liệu

Tập dữ liệu sau khi được tăng cường sẽ được chia thành 3 phần chính:

* Tập huấn luyện( training set): 80% tập dữ liệu.
* Tập kiểm định(validation set): 20% tập dữ liệu.

Quá trình chia dữ liệu được thực hiện ngẫu nhiên nhưng đảm bảo duy trì tỷ lệ cân đối giữa hai nhóm ảnh trong mỗi tập.

### 

## **4.3 Ứng dụng thực nghiệm**

### 4.3.1. Quy trình huấn luyện mô hình

Mô hình được triển khai trên nền tảng Kaggle, sử dụng GPU NVIDIA Tesla P100 để tăng tốc quá trình tính toán. Quy trình huấn luyện bao gồm các bước:

* Chuẩn bị dữ liệu.
* Thiết lập mô hình.
* Thiết lập hàm mất mát và bộ tối ưu.
* Huấn luyện mô hình.
* Đánh giá mô hình.
* Lưu mô hình.

### 4.3.2. Cấu hình huấn luyện

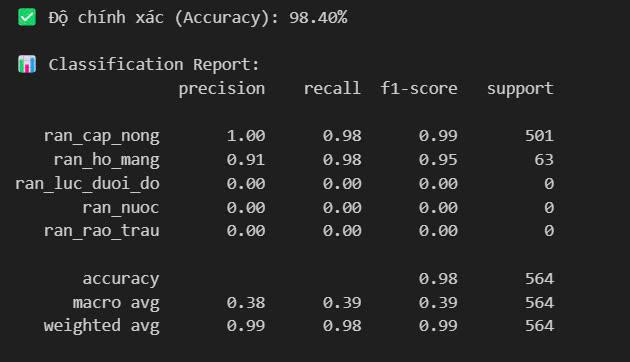
Mô hình được huấn luyện với các tham số sau:

* Số epoch: 10 epoch.
* Batch size: 32.
* Num\_classes: 5
* Device: “cuda” nếu GPU khả dụng, ngược lại là “cpu”
* Data\_Dir: ‘./dataset/dataset\_raw/’
* Img\_Size: 224
* Mô hình: EfficientNet-B0
* Hàm mất mát: Cross Entropy Loss
* Bộ tối ưu: Adam
* Learning rate: 1e-4

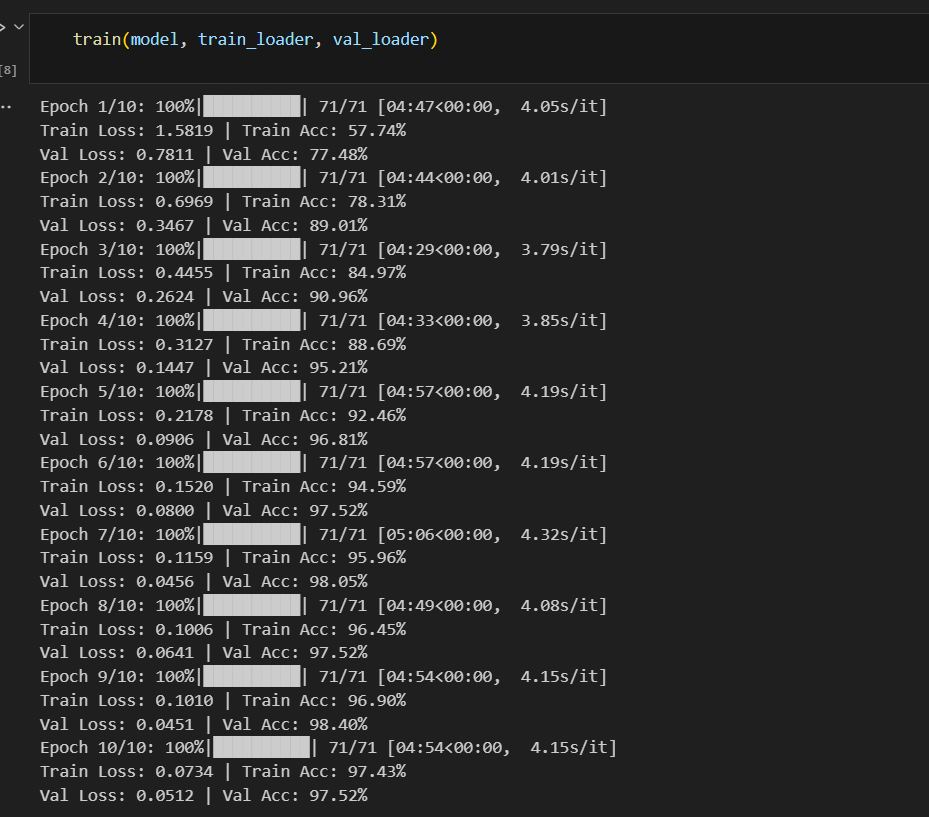
### 4.3.3. Kết quả thực nghiệm

Sau khi training với 10 epoch, mỗi epoch mất khoảng 169-261 giây với tổng thời gian đào tạo là 942 giây ta và chạy Classification Report ta có được kết quả sau:

**Hình 1.** Kết quả thử nghiệm và xác thực chéo của mô hình đề xuất



**Hình 2.** Độ chính xác và loss trong quá trình huấn luyện



**Hình 3.** Ma trận nhầm lẫn của mô hình đề xuất

# **CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

## **5.1 Kết luận**

Trong nghiên cứu này, chúng tôi đã đề xuất một mô hình học sâu cải tiến nhằm nhận diện loài rắn từ hình ảnh, với mục tiêu hỗ trợ phân loại rắn độc và không độc một cách chính xác và nhanh chóng. Mô hình được thiết kế theo kiến trúc mạng nơ-ron tích chập sâu (CNN) kết hợp với cơ chế Attention Ensemble. Bằng cách tích hợp Grouped Convolution và khối SE (Squeeze-and-Excitation), mô hình có khả năng học được các đặc trưng đa dạng giữa các loài rắn – vốn có hình dạng, màu sắc và họa tiết khác biệt – từ đó nâng cao chất lượng biểu diễn đặc trưng cho các lớp chú ý.

Mô hình đề xuất đã đạt hiệu suất ấn tượng trên tập dữ liệu kiểm tra, với độ chính xác 97,75% . Những kết quả này cho thấy mô hình có khả năng phân biệt chính xác giữa các loài rắn khác nhau – đặc biệt là giữa rắn độc và không độc – và có tiềm năng ứng dụng thực tế cao trong việc hỗ trợ bác sĩ, nhân viên cứu hộ hoặc người dân nhận dạng rắn để xử lý kịp thời các tình huống nguy hiểm, nhất là tại các khu vực miền núi hoặc vùng sâu vùng xa.

## **5.2 Hướng phát triển**

Dựa trên các kết quả đạt được, chúng tôi đề xuất một số hướng nghiên cứu và cải tiến trong tương lai như sau:

1. **Tăng cường dữ liệu**:

* Mở rộng tập dữ liệu bằng cách thu thập ảnh từ nhiều khu vực địa lý khác nhau, nhằm tăng tính đa dạng về môi trường sống, góc chụp và điều kiện ánh sáng.
* Sử dụng các kỹ thuật tăng cường dữ liệu tiên tiến như **GAN** để sinh thêm ảnh rắn trong điều kiện thực tế, đặc biệt là với các loài rắn hiếm gặp hoặc dữ liệu ít.

1. **Mở rộng phạm vi nhận diện:**

* Phát triển mô hình nhận diện nhiều loài rắn hơn, không chỉ phân biệt rắn độc/không độc, mà còn xác định tên loài cụ thể, nguồn gốc phân bố hoặc nhóm sinh học (ví dụ: họ rắn hổ, họ rắn cạp nia...).
* Kết hợp thêm dữ liệu về hành vi hoặc môi trường sống để mô hình dự đoán chính xác hơn trong bối cảnh thực địa.

1. **Tích hợp vào ứng dụng thực tế:**

* Xây dựng ứng dụng trên thiết bị di động giúp người dân và nhân viên y tế có thể chụp ảnh rắn và nhận kết quả ngay lập tức.
* Kết hợp mô hình vào hệ thống hỗ trợ cấp cứu hoặc thông tin y tế để đưa ra hướng xử lý ban đầu khi bị rắn cắn.

1. **Cải tiến hiệu suất mô hình**:

* Áp dụng các kỹ thuật fine-tuning và tối ưu hóa siêu tham sốđể tăng độ chính xác và khả năng khái quát hóa.
* Khám phá các kiến trúc mới như Vision Transformers (ViTs) hoặc mô hình tự giám sát (self-supervised learning) để học đặc trưng mà không cần quá nhiều ảnh gán nhãn.

1. **Giảm chi phí tính toán**:

* Tối ưu hóa mô hình để có thể triển khai hiệu quả trên thiết bị thông minh, hoặc các thiết bị không có GPU – đặc biệt quan trọng tại các khu vực vùng sâu, vùng xa có điều kiện hạ tầng hạn chế.

**TÀI LIỆU THAM KHẢO**

[1] Snake Species Identification with Deep Convolutional Neural Networks. <https://link.springer.com/article/10.1007/s11042-023-16773-0>

[2] Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. <https://www.pnas.org/doi/10.1073/pnas.1719367115>

[3] Deep Learning-Based Snake Species Identification for Enhanced Snakebite Management. <https://www.mdpi.com/2673-2688/6/2/21>

[4] Học sâu. <https://vi.wikipedia.org/wiki?curid=3339820>.

[5]“Ma trận nhầm lẫn.” *Ma Trận Nhầm Lẫn: Công Cụ Đánh Giá Hiệu Suất Mô Hình Học Máy*,https://rdsic.edu.vn/blog/toan/van-de-va-thach-thuc-cua-ma-tran-nham-lan-vi-cb.html.

[6] Các phương pháp đánh giá độ chính xác của các mô hình học sâu.<https://www.smartvisiontech.vn/forum/threads/cac-phuong-phap-danh-gia-do-chinh-xac-cua-cac-mo-hinh-hoc-sau.154/>