SỞ GIÁO DỤC VÀ ĐÀO TẠO THÀNH PHỐ ĐÀ NẴNG

**TRƯỜNG THPT CHUYÊN LÊ QUÝ ĐÔN**

**BÁO CÁO KHOA HỌC KĨ THUẬT**

**Đề tài:**

*ỨNG DỤNG CÔNG NGHỆ HỌC MÁY ĐỂ PHÁT HIỆN TRIỆU CHỨNG BỆNH ĐẬU GÀ*



**1. Họ và tên tác giả 1:** Phan Trí Hải **2. Lớp:** 11A1

**3. Họ và tên tác giả 2:** Đỗ Quảng Giác **4. Lớp:** 11A1

**5. Người hướng dẫn:** Trần Thị Anh Đào

**6. Nhóm nghiên cứu:** Kĩ thuật

**7. Lĩnh vực nghiên cứu:** Phần mềm hệ thống

**MỤC LỤC**

[MỞ ĐẦU 3](#_Toc150148442)

[CHƯƠNG 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU 5](#_Toc150148443)

[1.1. Phương pháp Học máy (Machine Learning) và Học sâu (Deep Learning) 5](#_Toc150148444)

[1.2. Cơ sở dữ liệu 6](#_Toc150148445)

[1.3. Mô hình Học máy (Machine Learning model) 6](#_Toc150148446)

[1.4. Bệnh đậu gà (Fowlpox) 7](#_Toc150148447)

[CHƯƠNG 2. PHƯƠNG PHÁP NGHIÊN CỨU 8](#_Toc150148448)

[2.1. Thu thập tập dữ liệu ảnh (datasets) về các cá thể gà nhiễm bệnh 8](#_Toc150148449)

[2.2. Huấn luyện mô hình nhận diện 8](#_Toc150148450)

[2.3. Mô hình Học máy được sử dụng trong nghiên cứu 9](#_Toc150148451)

[CHƯƠNG 3. KẾT QUẢ NGHIÊN CỨU 11](#_Toc150148452)

[3.1. Kết quả nhận dạng 11](#_Toc150148453)

[3.2. Độ chính xác của mô hình 11](#_Toc150148454)

[CHƯƠNG 4. KẾT LUẬN 13](#_Toc150148455)

[4.1. Tóm tắt kết quả đạt được 13](#_Toc150148456)

[4.2. Hoàn thiện và đề xuất hướng phát triển 13](#_Toc150148457)

[TÀI LIỆU THAM KHẢO 14](#_Toc150148458)

# MỞ ĐẦU

1. **Lý do chọn** đề **tài**

Hiện nay, gia cầm là một trong những loại thực phẩm đóng vai trò quan trọng trong nền công nghiệp chế biến thực phẩm, trong đó thịt gà được tiêu thụ nhiều nhất [1, 2]. Vì thế, gà là động vật được chăn nuôi với số lượng nhiều nhất thế giới [3]. Tuy vậy, dịch bệnh là một mối đe doạ lớn đến việc chăn nuôi gia cầm nói chung và gà nói riêng, gây ra nhiều thiệt hại to lớn về mặt kinh tế [4, 5]. Đáng chú ý, bệnh đậu gà là một trong những bệnh phổ biến nhất [6, 7], đặc biệt nguy hiểm đối với các cá thể còn non, gây giảm sản lượng trứng và chất lượng thịt. [8] Do đó, việc phát triển một mô hình nhận dạng các triệu chứng bệnh đậu gà nhằm tối ưu hoá năng suất chăn nuôi là cần thiết. Việc xét nghiệm nhằm phát hiện bệnh đậu gà, tuy hiệu quả nhưng lại tốn thời gian và công sức. Với sự phát triển mạnh mẽ của Khoa học Máy tính, đặc biệt là công nghệ Học máy (Machine Learning) và Học sâu (Deep Learning), cần có một phương pháp hiện đại, nhanh chóng hơn nhằm phát hiện kịp thời các triệu chứng của bệnh đậu gà.

Từ những lí do trên, nhóm tác giả quyết định nghiên cứu tạo ra mô hình có khả năng nhận diện được triệu chứng bệnh đậu gà thông qua các hình ảnh chụp. Nếu nghiên cứu thành công, thì dựa vào hình ảnh dữ liệu, người chăn nuôi có thể phát hiện hoặc phát hiện sớm các cá thể nhiễm bệnh, từ đó có kế hoạch tách khu chăn nuôi, điều trị và phòng chống cho các cá thể khoẻ mạnh.

1. **Mục tiêu nghiên cứu**

Nhóm nghiên cứu đã đề ra các mục tiêu:

- Xây dựng thành công mô hình nhận diện triệu chứng bệnh ở gà.

- Thực hiện cài đặt và tinh chỉnh mô hình huấn luyện Machine Learning nhằm đạt được độ chính xác cao nhất.

1. **Đối tượng nghiên cứu**

Bài toán nhận diện triệu chứng bệnh qua hình ảnh, loại bệnh được nhận dạng là Bệnh đậu gà (Fowl pox).

1. **Giả thuyết khoa học**

- Có thể nhận diện được triệu chứng bệnh đậu gà qua các hình ảnh.

- Sử dụng được mô hình để nhận diện các triệu chứng bệnh đậu gà.

1. **Phương pháp nghiên cứu**

- Thu thập dữ liệu ảnh của các cá thể gà cần nhận dạng.

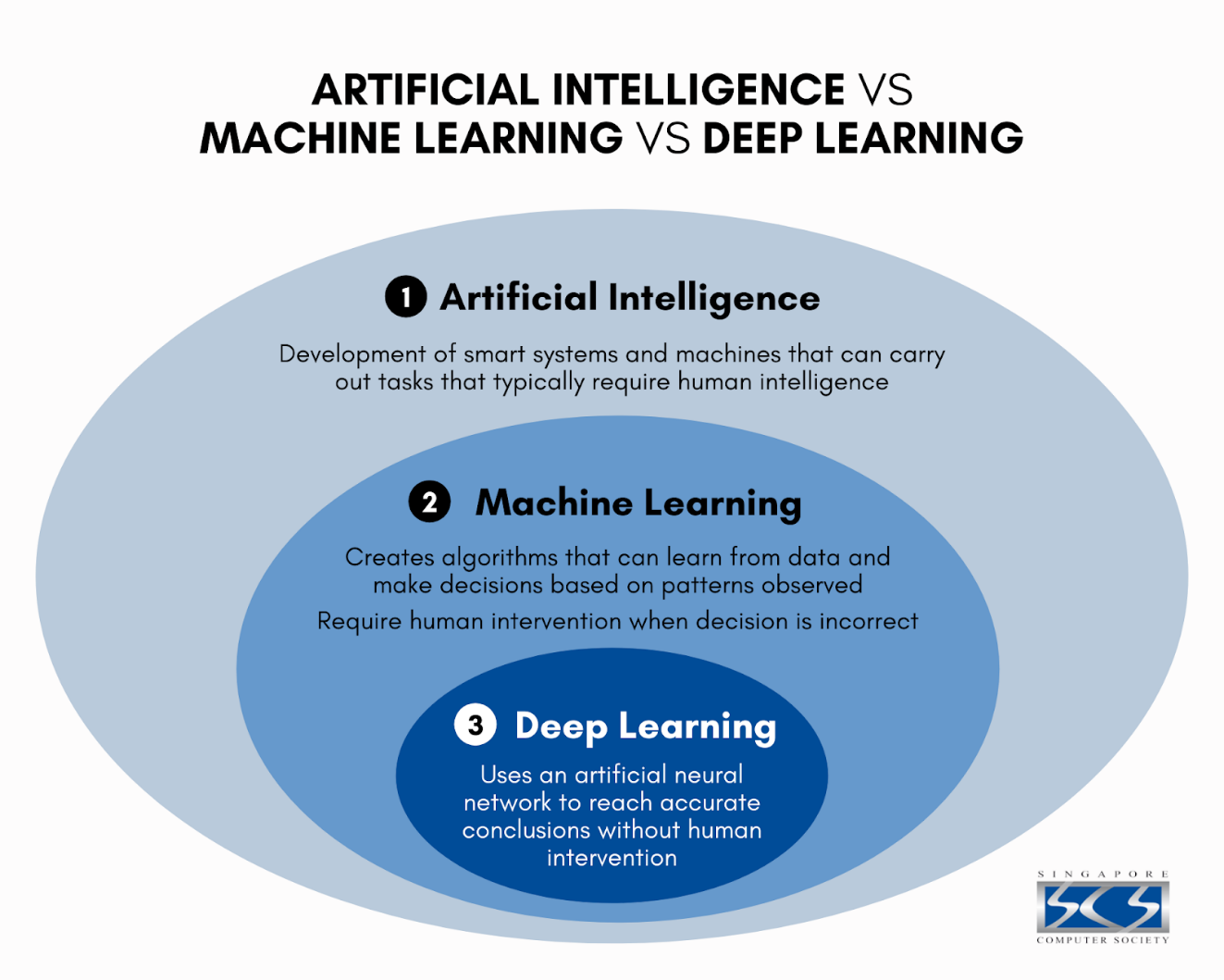
- Phân tích ảnh được để nhận dạng triệu chứng bệnh và đưa ra kết quả.

# CHƯƠNG 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU

* 1. Phương pháp Học máy (Machine Learning) và Học sâu (Deep Learning) **[9, 10]**

Học máy, hay Machine Learning, là một lĩnh vực nhỏ của Trí tuệ nhân tạo, có khả năng tự học hỏi dựa trên dữ liệu (data) đưa vào mà không cần được lập trình một cách cụ thể, bằng việc sử dụng các phương pháp thống kê, các thuật toán để phân loại hoặc dự đoán.

Học sâu là một “tập hợp con” của ngành “Học máy”, mô phỏng dựa theo mạng lưới nơ-rơn (neuron) trong não bộ con người. Học sâu dần trở nên phổ biến vì sự đột phá trong khả năng tính toán của các máy tính và nguồn dữ liệu khổng lồ được thu thập bởi các công ty công nghệ lớn.



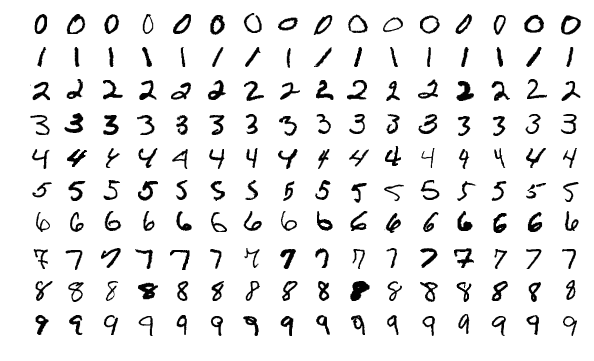
*Hình 1.1. Mối quan hệ của Học sâu, Học máy và Trí tuệ nhân tạo*

*(Nguồn: Singapore Computer Society)*

##### Cơ sở dữ liệu

Cơ sở dữ liệu là tập hợp các dữ liệu (data) liên quan đến nhau, sắp xếp theo một hệ thống nhất định và được lưu trữ và truy cập điện tử từ hệ thống máy tính. [11]

Trong các phương pháp Học máy nói chung, bộ cơ sở dữ liệu (dataset) có vai trò đặc biệt quan trọng. Các bộ cơ sở dữ liệu thường được sử dụng trong việc huấn luyện (train) và tinh chỉnh các mô hình Học máy. Thông thường, bộ cơ sở dữ liệu càng lớn thì độ chính xác của mô hình càng cao, nhưng trong một số trường hợp không yêu cầu bộ cơ sở dữ liệu quá lớn để đạt được độ chính xác cao.

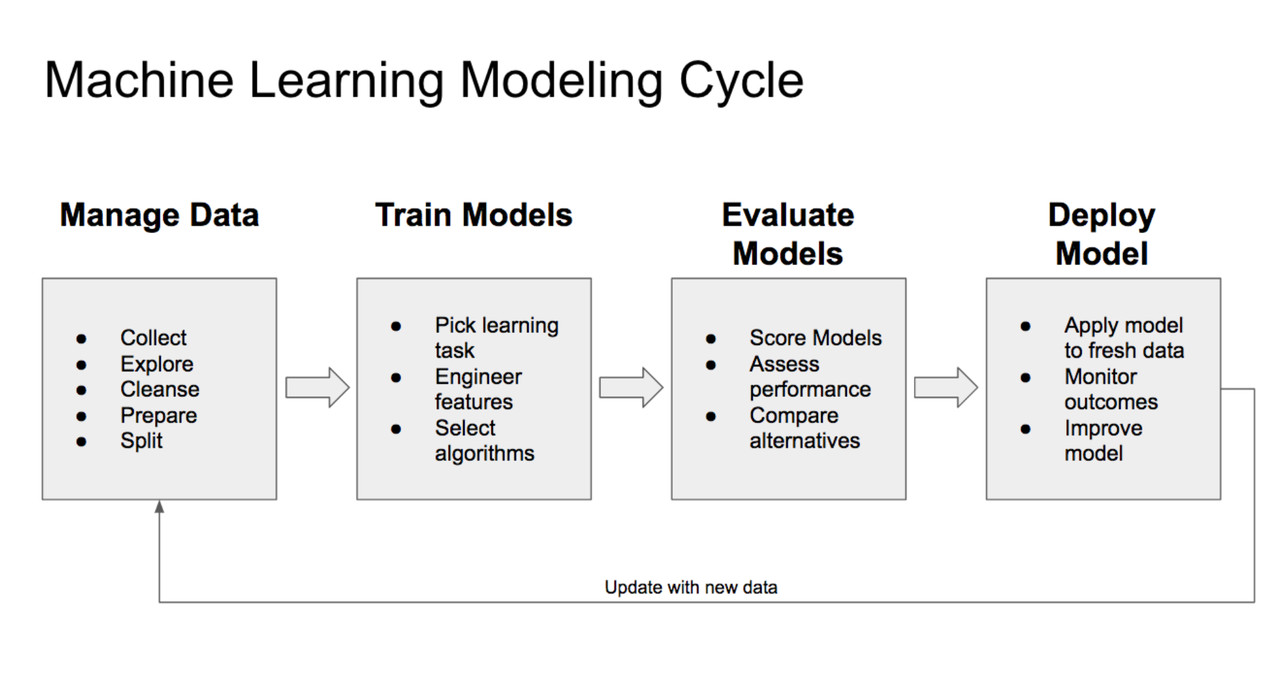


*Một phần của Bộ cơ sở dữ liệu MNIST (Modified National Institute of Standards and Technology database)*

*(Nguồn: ttested.com)*

##### Mô hình Học máy (Machine Learning model)

Mô hình Học máy là một chương trình có khả năng tìm các mẫu hình hoặc tự đưa ra quyết định từ một tệp dữ liệu bất kì. Để có thể làm được những điều này, mô hình phải được “huấn luyện” bằng một cơ sở dữ liệu đủ lớn. Trong quá trình huấn luyện, mô hình sẽ học các thuật toán tối ưu nhằm tìm ra một vài mẫu hình hoặc kết quả nhất định từ tệp dữ liệu, tùy thuộc vào tác vụ cần thực hiện. Kết quả của quá trình này được gọi là Mô hình Học máy. [12, 13]



*Quá trình huấn luyện một Mô hình Học máy*

*(Nguồn: Linkedin.com)*

* 1. Bệnh đậu gà (Fowlpox) **[14]**

Bệnh đậu gà (*Fowlpox* hay *Avian* *pox*) là một trong những loại bệnh phổ biến ở gà [6, 7]. Bệnh này do các chủng vi-rút (virus) thuộc họ *Poxviridae,* chi *Avipoxvirus* gây ra không chỉ ở gà mà còn ở các loại động vật họ chim khác như vịt, gà tây, bồ câu, etc.

Tuy bệnh không quá nguy hiểm, nhưng với những con non, và gà chưa trưởng thành thì bệnh hoàn toàn có thể lây lan phát triển thành dịch, hay gây tử vong, ảnh hưởng đến đến chất lượng của lứa gà [8]. Hiện nay, dù đã có vắc-xin (vaccine) phòng bệnh đậu gà, tuy nhiên vẫn chưa có thuốc chữa cụ thể dành cho bệnh này.

# CHƯƠNG 2. PHƯƠNG PHÁP NGHIÊN CỨU

## Thu thập tập dữ liệu ảnh (datasets) về các cá thể gà nhiễm bệnh

Nghiên cứu này sử dụng bộ cơ sở dữ liệu gồm 70 hình ảnh các cá thể gà, cả bị nhiễm bệnh và khoẻ mạnh, được tự tổng hợp, lựa chọn từ nhiều nguồn trên Internet, và không có nguồn dữ liệu tự thu thập thực tế. Đây là điểm hạn chế của nghiên cứu vì những điều kiện về thời gian, phạm vi nghiên cứu. Tuy nhiên, bộ cơ sở dữ liệu này vẫn đảm bảo được tính chính xác của nghiên cứu, do tính chọn lọc kĩ càng về chất lượng hình ảnh và nguồn tài liệu tham khảo.

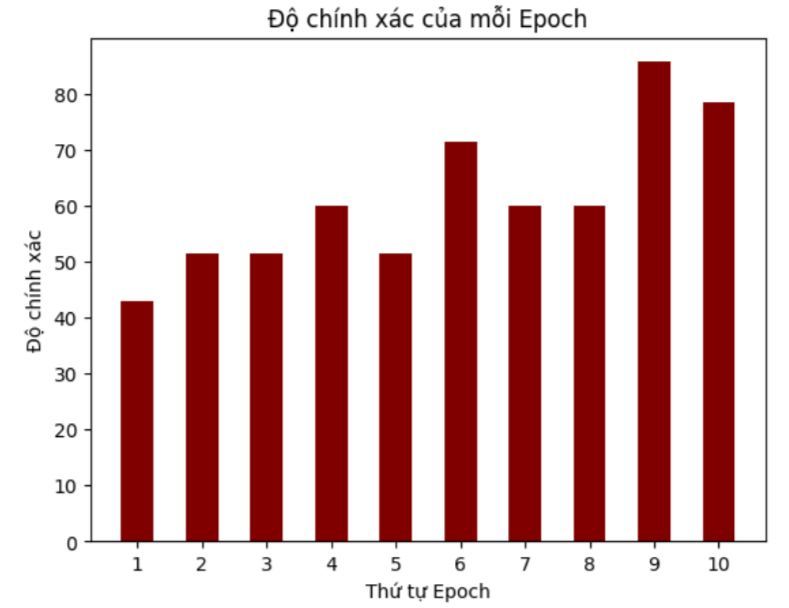


*Hình ảnh cá thể gà con mắc bệnh Đậu gà*

*(Nguồn: backyardchickens.com)*

## Huấn luyện mô hình nhận diện

Để huấn luyện mô hình nhận diện được cá thể gà bị nhiễm đậu gà, mô hình sẽ cần phải nhận dữ liệu được phân định sẵn. Trong trường hợp này, mô hình được cung cấp 35 hình ảnh đầu của cá thể gà bị nhiễm bệnh và 35 hình ảnh đầu của cá thể gà khỏe mạnh. Khi đó, máy sẽ tiếp nhận và chạy qua 10 epoch (một epoch được tính khi đưa tất cả dữ liệu trong tập “train” vào trong mô hình) theo từng cụm (batch) (một cụm là một lần phân tích đã được chia nhỏ để tối ưu hóa bộ nhớ của máy) dựa vào kích cỡ của cụm (batch size), với số lượng mẫu lấy ngẫu nhiên từ hình ảnh mẫu vào mỗi cụm. [15]



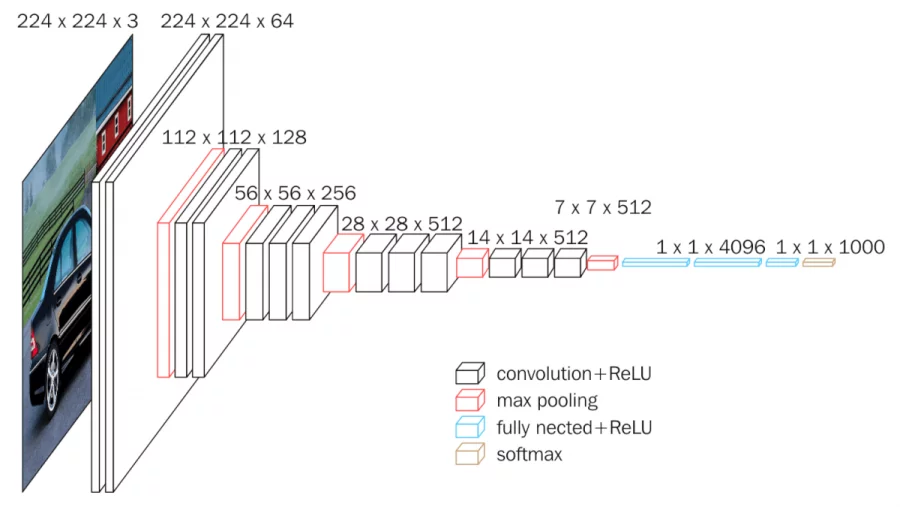
*Độ chính xác qua từng Epoch*

## Mô hình Học máy được sử dụng trong nghiên cứu

TensorFlow là một thư viện mã nguồn mở miễn phí dành cho Học máy và Trí tuệ nhân tạo, được phát triển bởi đội ngũ Google Brain. TensorFlow có thể sử dụng trong nhiều ngôn ngữ lập trình khác nhau, bao gồm Python, JavaScript, C++ và Java, với nhiều mục đích khác nhau, nhưng chủ yếu là cho quá trình huấn luyện, tinh chỉnh các mô hình Học máy. Chính nhờ sự linh hoạt, đa dụng và hỗ trợ mạnh mẽ các phép toán học để tính toán trong Machine Learning mà TensorFlow giúp việc tiếp cận các bài toán đơn giản, nhanh chóng và hiệu quả hơn. [16, 17]

Keras là mã nguồn mở cho Neural Network được viết bằng ngôn ngữ lập trình Python, được phát triển vào năm 2005 bởi Francois Chollet, một kỹ sư Deep Learning. Keras có thể sử dụng chung với các thư viện nổi tiếng như TensorFlow, CNTK, Theano,... phù hợp với việc phân tích dữ liệu số, dữ liệu văn bản và dữ liệu hình ảnh. [17]

Nghiên cứu này sử dụng mô hình VGG-16, một mô hình Học sâu được lấy từ Keras trên TensorFlow [18] cùng các module có sẵn trên Keras và chạy chẩn đoán nhờ hàm fit của module keras.model.

**

*Sơ đồ tổng quan cách hoạt động của Mô hình VGG-16*

*(Nguồn: neurohive.io)*

# CHƯƠNG 3. KẾT QUẢ NGHIÊN CỨU

### Kết quả nhận dạng

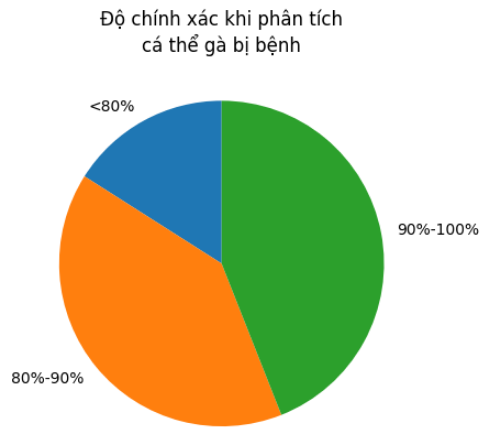
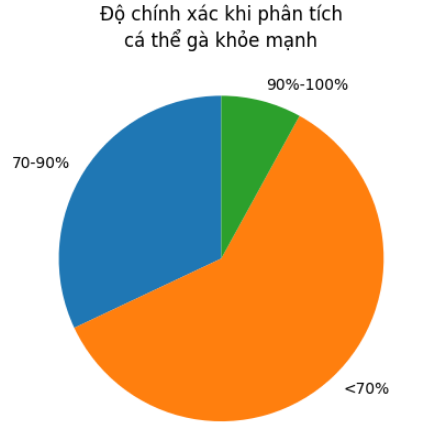
Nhìn chung, mô hình có thể nhận dạng được những cá thể có triệu chứng của bệnh đậu gà. Tuy nhiên, vẫn còn tồn tại một vài trường hợp việc nhận diện sai, một phần do bộ cơ sở dữ liệu chưa đủ lớn dẫn đến việc tinh chỉnh mô hình gặp khó khăn, nhưng cũng một phần vì chất lượng hình ảnh đầu vào kém.



*Kết quả nhận diện (bên trái: khoẻ mạnh; bên phải: nhiễm bệnh)*

### Độ chính xác của mô hình

Sau khi huấn luyện, mô hình có thể nhận diện được cá thể bị nhiễm bệnh với độ chính xác tương đối cao, lên đến 90%. Tuy nhiên, đối việc nhận dạng các cá thể khoẻ mạnh, độ chính xác cho ra khá thấp (phần lớn dưới 70%). Điều này là hệ quả từ việc thiếu cơ sở dữ liệu hình ảnh để phục vụ cho việc tinh chỉnh mô hình nhận dạng.

**

*Biểu đồ biểu thị độ chính xác qua các lần thử*

# CHƯƠNG 4. KẾT LUẬN

#### Tóm tắt kết quả đạt được

- Dữ liệu ảnh về các cá thể gà bị nhiễm bệnh còn hạn chế do điều kiện về thời gian và phạm vi nghiên cứu, tuy vậy dữ liệu ảnh vẫn đảm bảo được tính chính xác và đáng tin cậy.

- Huấn luyện, tinh chỉnh thành công mô hình Học máy có thể nhận diện các triệu chứng bệnh đậu gà qua hình ảnh bằng mô hình VGG-16.

#### Hoàn thiện và đề xuất hướng phát triển

- Mở rộng, tiếp tục phát triển mô hình để có thể nhận diện thêm các loại bệnh ở gà khác, mở rộng thêm phạm vi nhận diện ra các khía cạnh khác của cá thể nhiễm bệnh.

- Tìm cách phát triển mô hình để có thể tiếp cận đến nhiều người dùng hơn.

- Phát triển mô hình không chỉ dừng lại ở việc nhận dạng mà còn có thể đưa ra các thông tin, biện pháp xử lí đối với từng loại bệnh.

- Phát triển, xây dựng một hệ thống quản lí, giám sát sức khoẻ của gà tự động sử dụng trong các trại chăn nuôi.

# 

# TÀI LIỆU THAM KHẢO

1. [*https://en.wikipedia.org/wiki/Poultry*](https://en.wikipedia.org/wiki/Poultry)
2. [*https://worldanimalfoundation.org/advocate/most-consumed-meat-in-the-world/*](https://worldanimalfoundation.org/advocate/most-consumed-meat-in-the-world/)
3. [*https://www.ufaw.org.uk/why-ufaws-work-is-important/farm-animals*](https://www.ufaw.org.uk/why-ufaws-work-is-important/farm-animals)
4. [*https://www.ajol.info/index.php/sokjvs/article/view/157721*](https://www.ajol.info/index.php/sokjvs/article/view/157721)
5. *A.K. Bera, Evaluation of Economic Losses due to Coccidiosis in Poultry Industry in India, (2010)*
6. [*https://www.bentoli.com/chicken-problems-common/*](https://www.bentoli.com/chicken-problems-common/)
7. *Dr. Abdikarim Osman Mooge, A survey on the prevalence of fowl pox and associated risk factors in Hargesa, Somaliland, (2017)*
8. [*https://www.woah.org/fileadmin/Home/fr/Health\_standards/tahm/3.03.10\_FOWLPOX.pdf*](https://www.woah.org/fileadmin/Home/fr/Health_standards/tahm/3.03.10_FOWLPOX.pdf)
9. [*https://vietnix.vn/machine-learning-la-gi/*](https://vietnix.vn/machine-learning-la-gi/)
10. *Vũ Hữu Tiệp, Machine Learning cơ bản, (2017).*
11. [*https://vi.wikipedia.org/wiki/Cơ\_sở\_dữ\_liệu*](https://vi.wikipedia.org/wiki/Cơ_sở_dữ_liệu)
12. [*https://www.databricks.com/glossary/machine-learning-models*](https://www.databricks.com/glossary/machine-learning-models?fbclid=IwAR1hRCG_7CgkubooB6YDDyXXTI2mi8HpqDfZbM1mE6yIJmGvqkfQtiE7OMA)
13. [*https://cole.edu.vn/mo-hinh-machine-learning-la-gi-cac-buoc-xay-dung-va-phuong-phap-danh-gia-mo-hinh-hoc-may/*](https://cole.edu.vn/mo-hinh-machine-learning-la-gi-cac-buoc-xay-dung-va-phuong-phap-danh-gia-mo-hinh-hoc-may/)
14. [*https://en.wikipedia.org/wiki/Fowlpox*](https://en.wikipedia.org/wiki/Fowlpox)
15. [*https://www.phamduytung.com/blog/2018-10-02-understanding-epoch-batchsize-iterations*](https://www.phamduytung.com/blog/2018-10-02-understanding-epoch-batchsize-iterations)
16. [*https://en.wikipedia.org/wiki/TensorFlow*](https://en.wikipedia.org/wiki/TensorFlow)
17. [*https://topdev.vn/blog/tensorflow-la-gi/*](https://topdev.vn/blog/tensorflow-la-gi/)
18. [*https://keras.io/api/applications*](https://keras.io/api/applications)