

TRƯỜNG ĐẠI HỌC HỌC VĂN LANG

**KHOA CÔNG NGHỆ THÔNG TIN**





A red logo with white text

AI-generated content may be incorrect.

**BÁO CÁO ĐỒ ÁN MÔN HỌC HK242**

**PHÂN TÍCH DỮ LIỆU VÀ HỌC SÂU**

**XÂY DỰNG MÔ HÌNH CNN ĐỂ PHÂN LOẠI   
HÌNH ẢNH ĐỘNG VẬT**

**GVHD: ThS. Nguyễn Thị Mỹ Linh**

**Nhóm sinh viên thực hiện:**

**1. Lê Duy Khang - 2274802010374**

**2. Trần Hữu Luân - 2274802010520**

**3. Trần Thị Thuỳ Linh - 2274802010482**

**4. Dư Nguyễn Hà Mi - 2274802010525**

**5. Huỳnh Nhật Khánh Quỳnh - 2274802010742**

**6. Phạm Nguyễn Hoài Phương -2274802010702**

**LỜI MỞ ĐẦU**

***Tp. Hồ Chí Minh – Năm 2025***

Lời đầu tiên, nhóm em xin gửi lời cảm ơn chân thành đến ThS. Nguyễn Thị Mỹ Linh. Trong quá trình học tập và tìm hiểu bộ môn Phân tích dữ liệu và học sâu, chúng em đã được nhận sự quan tâm giúp đỡ, hướng dẫn tận tình, tâm huyết của Cô. Những kiến thức mà thầy truyền đạt không chỉ giúp chúng em hiểu rõ hơn về lý thuyết, mà còn tạo điều kiện để chúng em vận dụng vào thực tế một cách hiệu quả. Đồng thời, Cô luôn sẵn sàng giải đáp những thắc mắc, gợi mở hướng đi mới, giúp chúng em hoàn thiện đồ án một cách tốt nhất. Dù đồ án còn nhiều thiếu sót, nhưng chúng em tin rằng, với những kiến thức và kinh nghiệm tích lũy được từ Cô, chúng em sẽ ngày càng hoàn thiện hơn trong học tập và nghiên cứu sau này. Một lần nữa, chúng em xin gửi lời cảm ơn chân thành đến Cô, chúc Cô dồi dào sức khỏe, thành công trong công tác giảng dạy và   
nghiên cứu.

**Nhóm thực hiện báo cáo**

**CHƯƠNG I: MỞ ĐẦU**

**1. Lý do chọn đề tài:**

Trong những năm gần đây, trí tuệ nhân tạo và học sâu đã có những bước phát triển mạnh mẽ, đặc biệt trong lĩnh vực thị giác máy tính. Một trong những ứng dụng quan trọng của AI là nhận diện và phân loại hình ảnh, trong đó mô hình mạng nơ-ron tích chập đóng vai trò quan trọng. CNN là một trong những kiến trúc mạng học sâu tiên tiến nhất trong lĩnh vực thị giác máy tính. Việc nghiên cứu và triển khai CNN giúp sinh viên nắm vững kiến thức nền tảng về học sâu và các ứng dụng thực tế của AI trong xử lý hình ảnh. Bên cạnh đó, CNN còn là nền tảng của nhiều mô hình tiên tiến hiện nay như ResNet, EfficientNet, Vision Transformer. Đề tài "Xây dựng mô hình CNN để phân loại hình ảnh động vật" không chỉ mang lại ứng dụng thực tiễn mà còn giúp nâng cao hiểu biết về các kỹ thuật AI hiện đại và với đề tài này chúng ta sẽ có một số ứng dụng thực tiễn cao như nhận diện các loài động vật trong nghiêm cứu sinh học; ứng dụng hệ thống giám sát động vật tại sở thú hoặc trang trại lớn; hỗ trợ bảo tồn các loại động vật quý hiếm thông qua nhận diện và phân loại động vật từ sớm,…

Ngoài ra, với đề tài này chúng ta có thể tận dụng nguồn dữ liệu phong phú, có sẵn để huấn luyện mô hình CNN và mở rộng nguyên cứu theo nhiều hướng khác nhau như so sánh hiệu suất giữa CNN và các mô hình tiên tiến như ResNet, VGG; Áp dụng kỹ thuật tăng cường dữ liệu để cải thiện độ chính xác; Tinh chỉnh mô hình trên tập dữ liệu đặc thù nhằm nâng cao độ chính xác; Triển khai mô hình trên thiết bị di động hoặc ứng dụng thực tế; Tích hợp AI với hệ thống nhận diện video thời gian để phân loại động vật nhanh chóng và chính xác hơn; Phát triển một hệ thống phân loại đa nhiệm giúp xác định không chỉ loài mà còn tình trạng sức khoẻ của động vật thông qua ảnh.

Đề tài "Xây dựng mô hình CNN để phân loại hình ảnh động vật" không chỉ mang lại giá trị thực tiễn mà còn giúp sinh viên tiếp cận với công nghệ AI tiên tiến. Việc nghiên cứu và phát triển mô hình CNN không chỉ góp phần vào sự tiến bộ của khoa học máy tính mà còn có những ứng dụng hữu ích trong đời sống. Thông qua đề tài này, sinh viên có cơ hội nâng cao kỹ năng lập trình, xử lý dữ liệu và ứng dụng AI vào thực tế. Ngoài ra, với sự phát triển của AI, đề tài này có thể mở rộng theo nhiều hướng mới, tạo cơ hội nghiên cứu và ứng dụng vào các lĩnh vực đa dạng như bảo tồn động vật, y tế thú y và hệ thống giám sát động vật trong tự nhiên.

**2. Mục tiêu của đề tài:**

Mục tiêu của nghiên cứu là thiết kế và huấn luyện một mô hình CNN có khả năng phân loại chính xác hình ảnh động vật, đồng thời so sánh hiệu suất giữa mô hình CNN đơn giản và các mô hình nâng cao như ResNet hay VGG. Kết quả thu được sẽ góp phần hỗ trợ các ứng dụng thực tế như giám sát động vật hoang dã, bảo tồn sinh học hoặc phát triển hệ thống nhận diện động vật trên thiết bị di động.

**3. Đối tượng và phạm vi đề tài:**

Đề tài tập trung nghiên cứu và ứng dụng mạng nơ-ron tích chập vào bài toán phân loại hình ảnh động vật. CNN là một trong những kiến trúc học sâu quan trọng, được thiết kế để xử lý dữ liệu hình ảnh bằng cách tự động trích xuất đặc trưng từ dữ liệu đầu vào. Trong nghiên cứu này, chúng em tập trung vào việc xây dựng một mô hình CNN cơ bản và so sánh hiệu suất của nó với các mô hình nâng cao như VGG-16 và ResNet-50. Bên cạnh đó, việc áp dụng các kỹ thuật tối ưu hóa mô hình, bao gồm điều chỉnh tham số, sử dụng các thuật toán tối ưu (Adam, SGD) và các phương pháp tăng cường dữ liệu (data augmentation), cũng sẽ được nghiên cứu nhằm cải thiện độ chính xác của mô hình.

Phạm vi nghiên cứu của đề tài giới hạn trong việc sử dụng các tập dữ liệu hình ảnh động vật có sẵn như CIFAR-10, ImageNet, Animal-10N để huấn luyện và đánh giá mô hình. Các tập dữ liệu này chứa hình ảnh của nhiều loài động vật khác nhau và đã được chuẩn hóa, giúp đảm bảo tính nhất quán trong quá trình huấn luyện. Tuy nhiên, nghiên cứu không thu thập thêm dữ liệu thực tế mà chỉ làm việc với dữ liệu đã có sẵn. Mô hình CNN được thiết kế với độ phân giải ảnh phù hợp, chẳng hạn như 32×32 hoặc 224×224 pixel, để đảm bảo khả năng huấn luyện trên phần cứng phổ biến.

**CHƯƠNG 2: PHÂN TÍCH TỔNG QUAN**

**1. Cơ sở lý thuyết các mô hình nhận dạng hình ảnh**

**1.1. Mạng Nơ-ron Tích Chập (CNN - Convolutional Neural Networks)**

CNN là mô hình phổ biến nhất trong nhận dạng hình ảnh do khả năng trích xuất đặc trưng tự động. Các thành phần chính gồm:

* Lớp tích chập (Convolutional Layer): Áp dụng bộ lọc (kernel) để trích xuất đặc trưng từ ảnh.
* Lớp kích hoạt (ReLU): Biến đổi phi tuyến để tăng khả năng học.
* Lớp Pooling (MaxPooling/AveragePooling): Giảm kích thước ảnh nhưng vẫn giữ thông tin quan trọng.
* Lớp Fully Connected: Chuyển đổi đặc trưng thành đầu ra phân loại.
* Softmax Layer: Tạo phân phối xác suất cho các lớp động vật.

**1.2. Các mô hình CNN nâng cao**

* VGG-16/VGG-19: Kiến trúc sâu hơn với các bộ lọc nhỏ 3x3.
* ResNet: Áp dụng residual connections để giúp mô hình học hiệu quả với độ sâu lớn.
* EfficientNet: Một mô hình hiện đại tối ưu về hiệu suất và tài nguyên tính toán.
* Vision Transformer (ViT): Thay thế CNN bằng cơ chế self-attention để học **đặc trưng hình ảnh.**

**1.3. Kỹ thuật tối ưu mô hình**

* Data Augmentation: Xoay, lật, thay đổi độ sáng ảnh để tăng dữ liệu.
* Regularization (Dropout, Batch Normalization): Giảm overfitting và tăng tốc độ huấn luyện.
* Optimizer (Adam, SGD với momentum): Tối ưu quá trình cập nhật trọng số.
* Transfer Learning: Sử dụng mô hình pre-trained như ResNet hoặc MobileNet để tinh chỉnh cho bài toán phân loại động vật**.**

**2. Dữ liệu**

**2.1. Tập dữ liệu sử dụng**

Bạn có thư mục Animals-10, có thể là tập dữ liệu phổ biến dùng để phân loại 10 loại động vật. Ngoài ra, có thể tham khảo:

* CIFAR-10: 10 lớp hình ảnh bao gồm mèo, chó, ngựa,…
* ImageNet: Một trong những bộ dữ liệu lớn nhất cho phân loại hình ảnh.
* Animal-10N: Bộ dữ liệu chuyên dụng cho phân loại động vật.

**2.2. Tiền xử lý dữ liệu**

Các bước tiền xử lý quan trọng bao gồm:

* Resize ảnh: Đưa tất cả ảnh về cùng một kích thước (ví dụ: 224x224).
* Chuẩn hóa dữ liệu: Chia pixel cho 255 để đưa về khoảng [0,1].
* One-hot encoding nhãn: Chuyển đổi nhãn động vật thành vector nhị phân.
* Chia tập dữ liệu: 80% huấn luyện, 10% validation, 10% kiểm tra.

**3. Quy trình phát triển**

**3.1. Xây dựng mô hình**

* **Bước 1: Chọn kiến trúc CNN** 
  + Có thể bắt đầu với một mô hình cơ bản gồm 3-4 lớp tích chập, pooling, và fully connected.
  + So sánh hiệu suất với mô hình VGG-16, ResNet hoặc MobileNet.
* **Bước 2: Chọn siêu tham số** 
  + Batch size, learning rate, số bộ lọc, kích thước kernel.
* **Bước 3: Regularization** 
  + Dùng Dropout, L2 Regularization để tránh overfitting.

**3.2. Huấn luyện và đánh giá**

* Sử dụng optimizer (Adam, SGD) để tối ưu trọng số.
* Sử dụng cross-entropy loss để đo sai số.
* Kiểm tra mô hình trên tập test và đánh giá bằng accuracy, precision, recall.

**3.3. Triển khai mô hình**

* Mô hình đã được lưu dưới dạng H5 file (animal\_classifier.h5, animal\_classifier\_best.h5).
* Ứng dụng Flask/Django có thể sử dụng để nhận ảnh đầu vào và dự đoán.
* Triển khai lên web hoặc làm ứng dụng di động.

**CHƯƠNG 3: KẾT QUẢ VÀ ĐÁNH GIÁ MÔ HÌNH**

**3.1. Giới thiệu**

Chương này trình bày kết quả thu được từ quá trình huấn luyện và kiểm tra mô hình. Đồng thời, chúng tôi phân tích hiệu suất mô hình dựa trên các chỉ số như độ chính xác (accuracy) và độ lỗi (loss). Cuối cùng, chúng tôi đánh giá những hạn chế của mô hình và đề xuất hướng cải thiện.

**3.2. Kết quả huấn luyện**

**3.2.1. Thông tin tập dữ liệu**

* Tổng số ảnh: **29938 ảnh**
* Số lớp phân loại: **10 lớp**
* Chia tập dữ liệu:
  + Tập huấn luyện (Train): 80%
  + Tập kiểm tra (Test): 20%

**3.2.2. Kết quả trong quá trình huấn luyện**

Mô hình được huấn luyện trong 40 epochs, và các chỉ số hiệu suất được ghi lại qua từng epoch. Dưới đây là tóm tắt kết quả huấn luyện dựa trên các hình ảnh:

* **Độ chính xác ban đầu (Epoch 1)**:

- Accuracy = 23.76%

- Validation Accuracy = 39.49%

- Loss = 2.3015

- Validation Loss = 1.7698

* **Độ chính xác cuối cùng (Epoch 40)**:

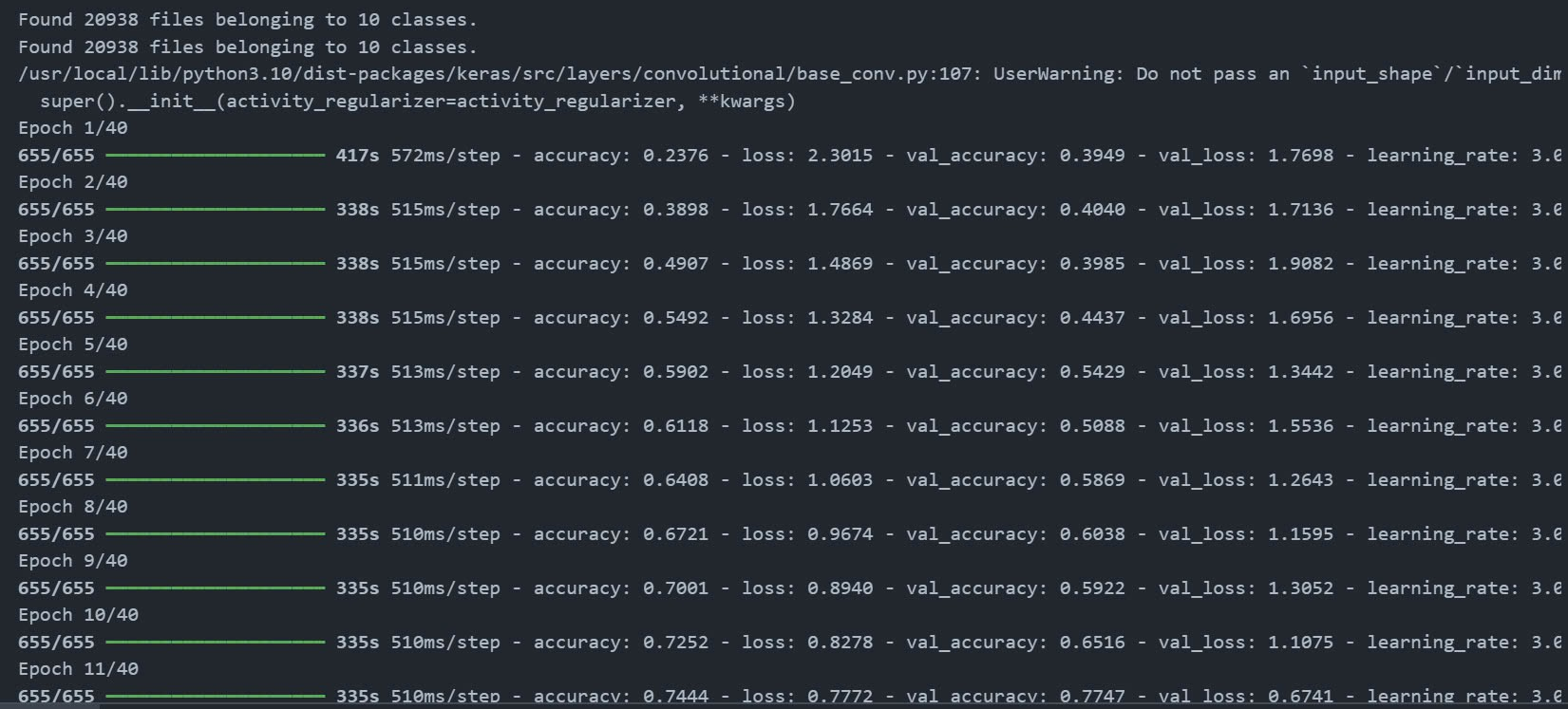
- Accuracy = 94.45%

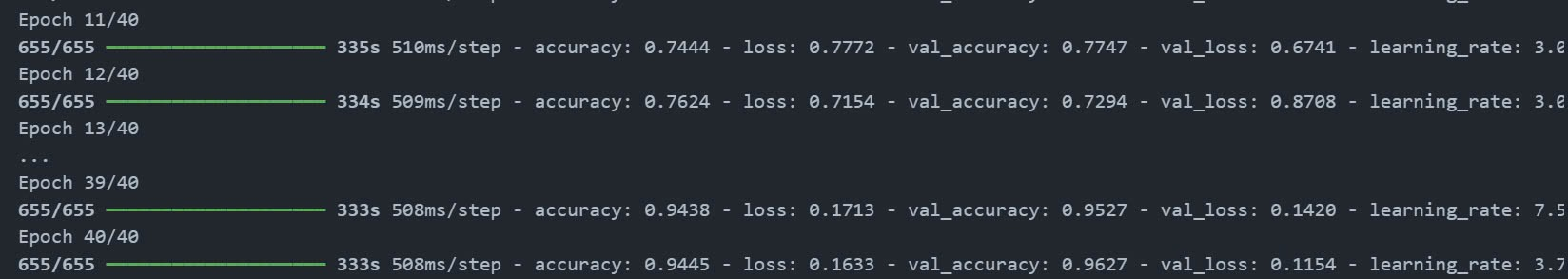
- Validation Accuracy = 96.27%

- Loss = 0.1633

- Validation Loss = 0.1154

* **Tốc độ học (Learning Rate)**: Thay đổi từ 3.0 đến 7.5 qua các epochs, với các giá trị được điều chỉnh để tối ưu hóa hiệu suấ





Bảng 3.1 dưới đây tổng hợp một số chỉ số quan trọng trong quá trình huấn luyệ

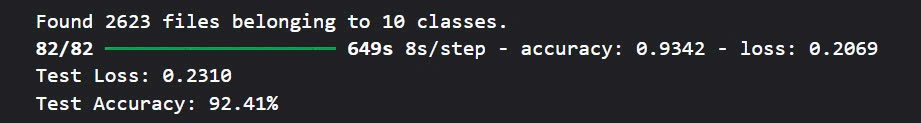
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Epoch** | |  |  | | --- | --- | |  | **Training Accuracy** | | |  |  | | --- | --- | | **Training Loss** |  | | **Validation Accuracy** | |  | | --- | | **Validation Loss** | |
| 1 | 23.7% | 2.3015 | 39.4% | 1.768 |
| 5 | |  | | --- | |  |   59.0% | 1.2049 | 54.29 | 1.3342 |
| 10 | 72.5% | 0.8278 | 65.1% | 1.1875 |
| 15 | 87.7% | 0.3446 | 67.3% | 1.420 |
| 20 | 89.6% | 0.2849 | 67.2% | 1.6113 |
| 39 | 94.38% | 0.1713 | 95.27% | 0.1420 |
| 40 | 94.45% | 0.1633 | 96.27% | 0.1154 |

**3.3. Đánh giá mô hình**

**3.3.1. Hiện tượng Overfitting**

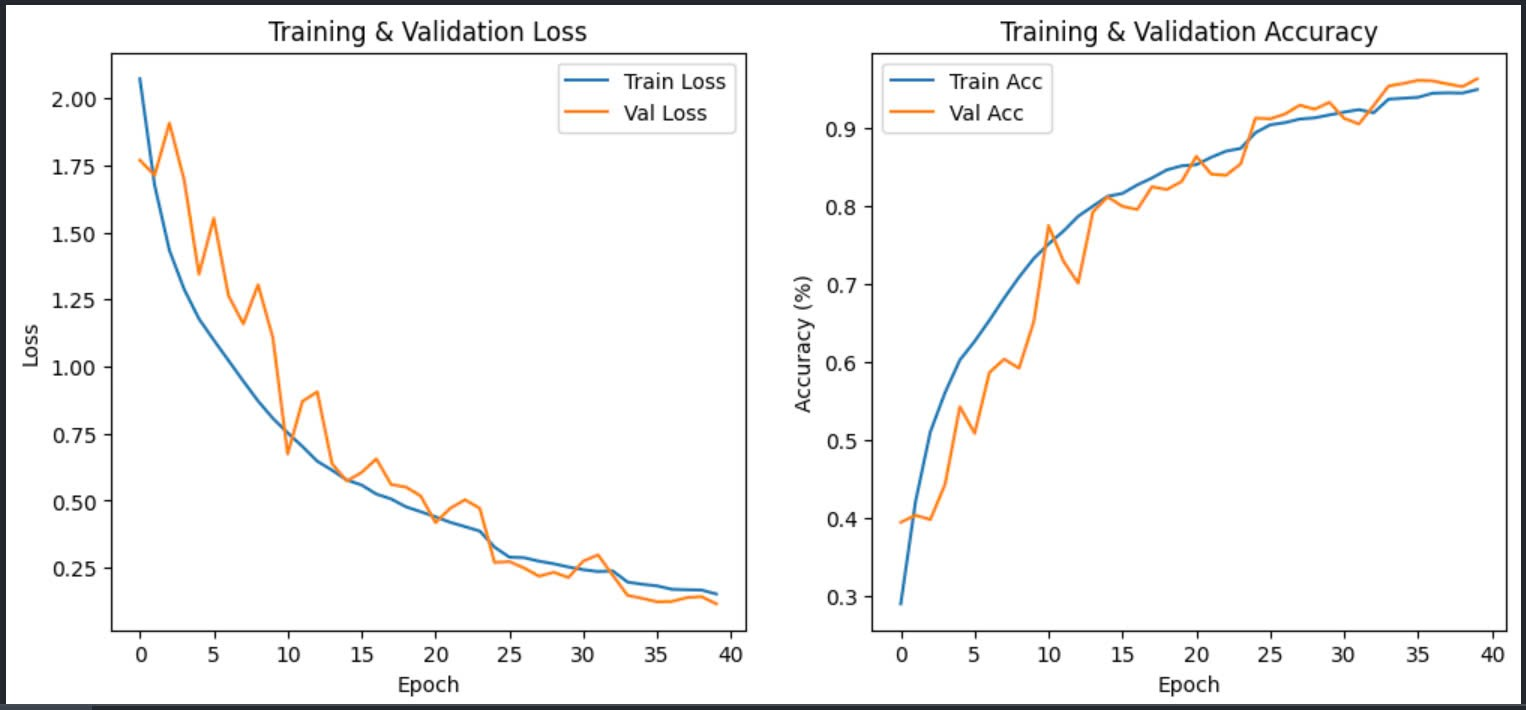
Trong tài liệu ban đầu, mô hình gặp hiện tượng overfitting với độ chính xác trên tập huấn luyện là 89.68% nhưng chỉ đạt 67.20% trên tập kiểm tra. Tuy nhiên, với kết quả mới:

* Độ chính xác trên tập huấn luyện (94.45%) và tập kiểm định (96.27%) khá gần nhau, và độ chính xác trên tập kiểm tra (92.41%) cũng không chênh lệch nhiều.
* Điều này cho thấy mô hình không còn gặp hiện tượng overfitting nghiêm trọng. Các kỹ thuật như Dropout, Batch Normalization, hoặc Data Augmentation (được đề cập trong tài liệu) có thể đã giúp cải thiện khả năng tổng quát hóa của mô hình.



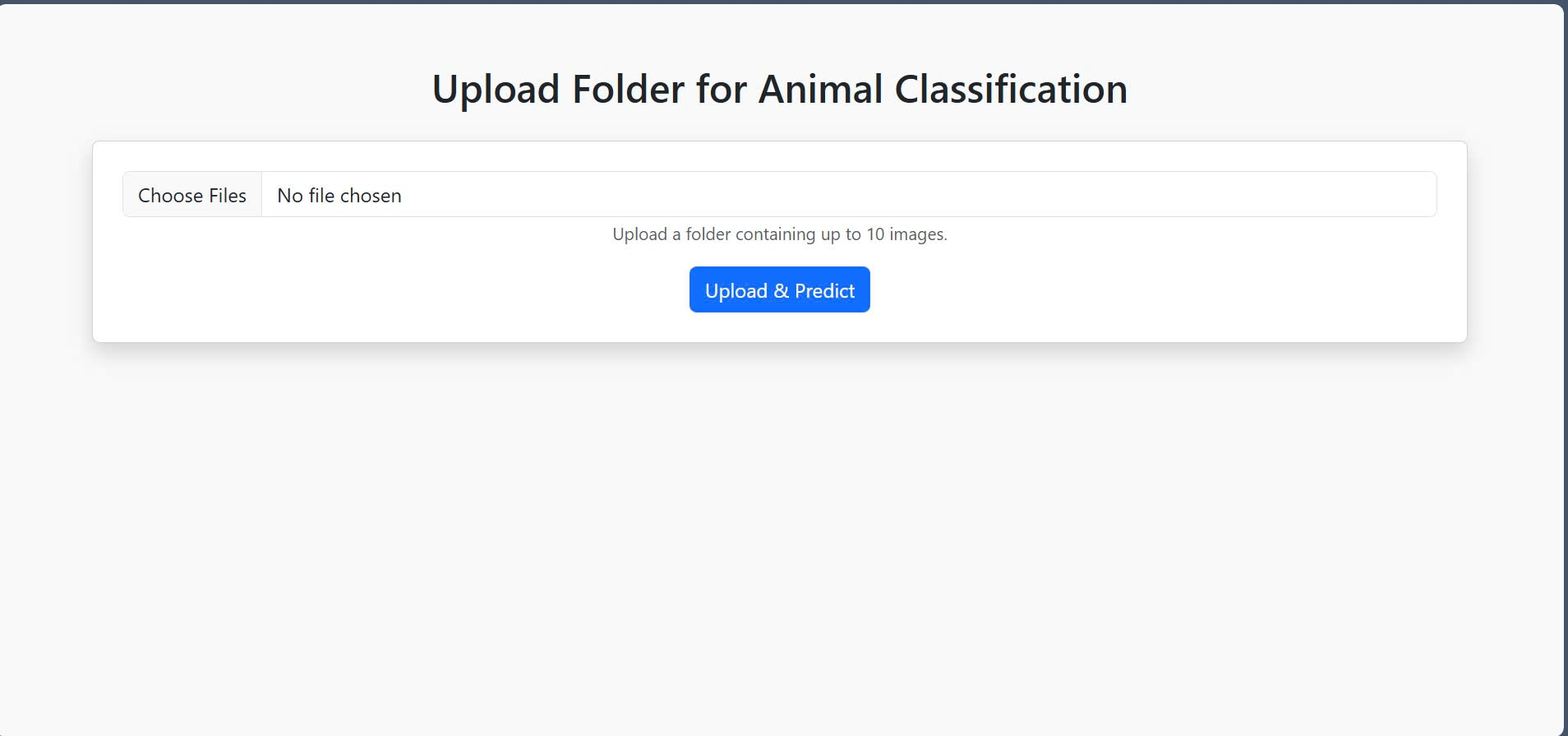
**3.3.2. Biểu đồ**

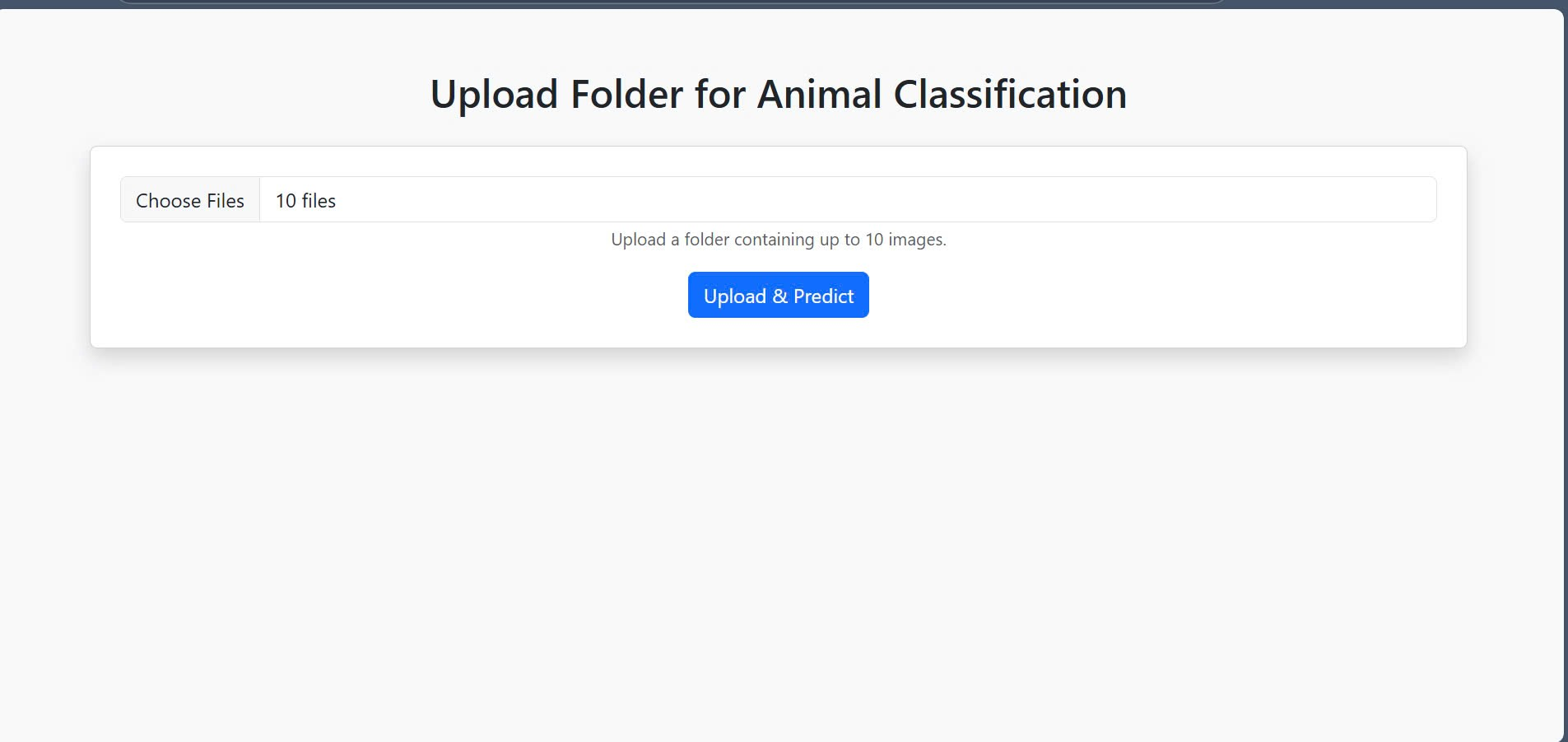
**Dưới đây là hai biểu đồ minh họa hiệu suất của mô hình qua 40 epochs:**

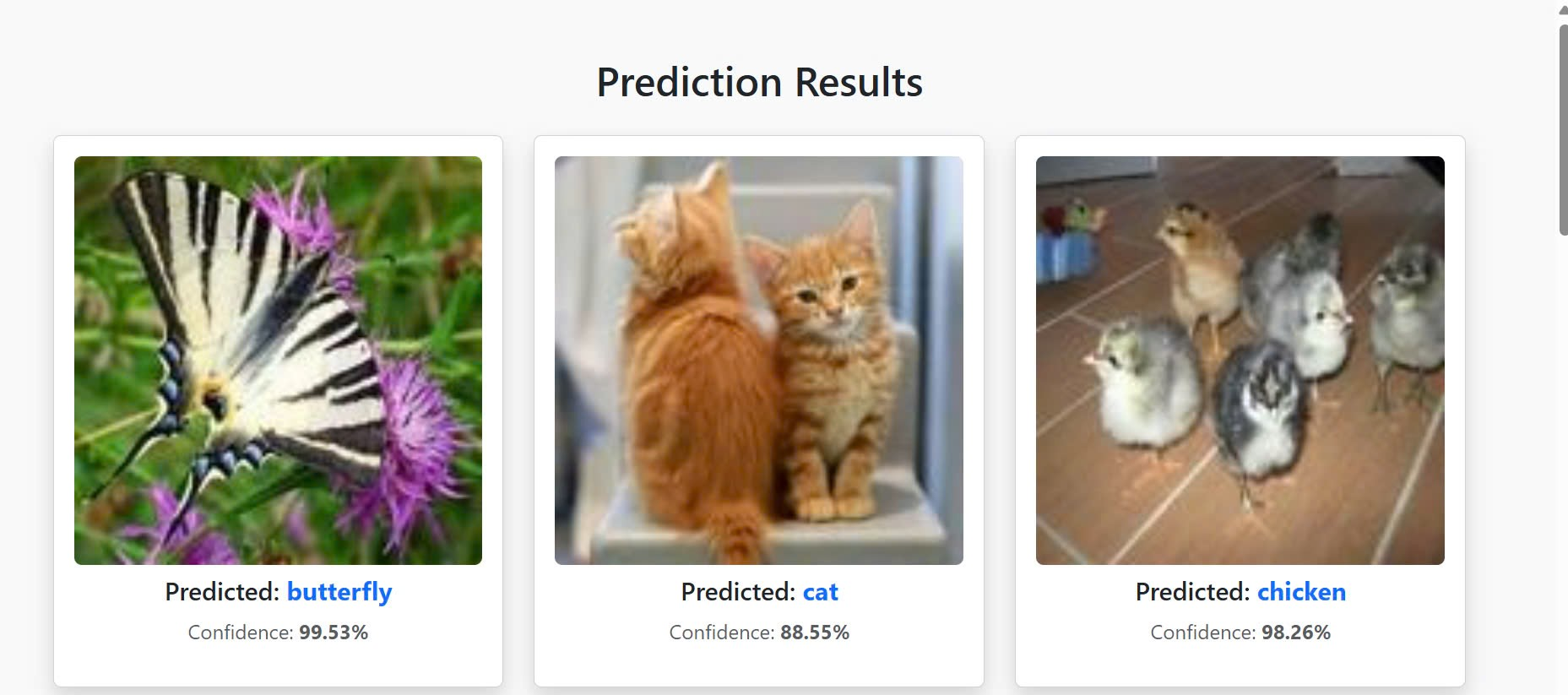
****

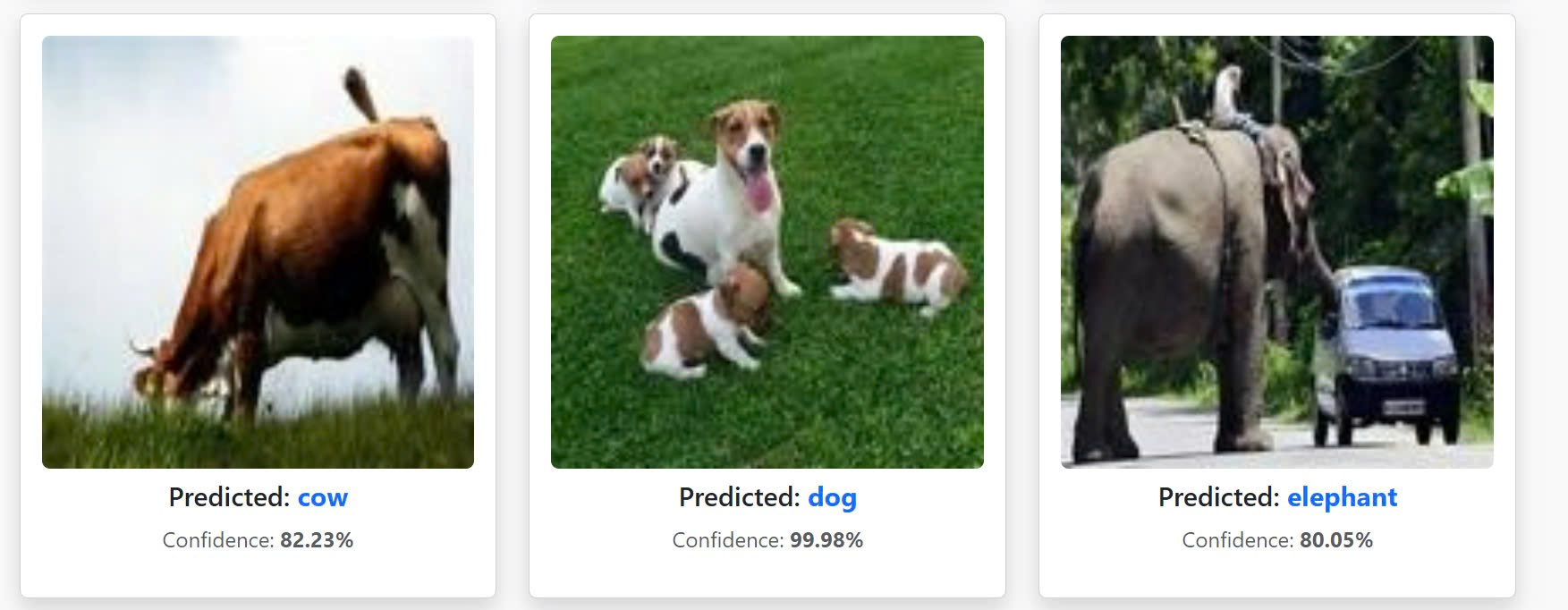
Các biểu đồ này minh họa rằng mô hình CNN đã đạt được sự hội tụ tốt sau 40 epochs, với độ chính xác cao và độ lỗi thấp trên cả tập huấn luyện và tập kiểm định. So với mục tiêu ban đầu của đề tài, mô hình đã vượt qua kỳ vọng về độ chính xác.

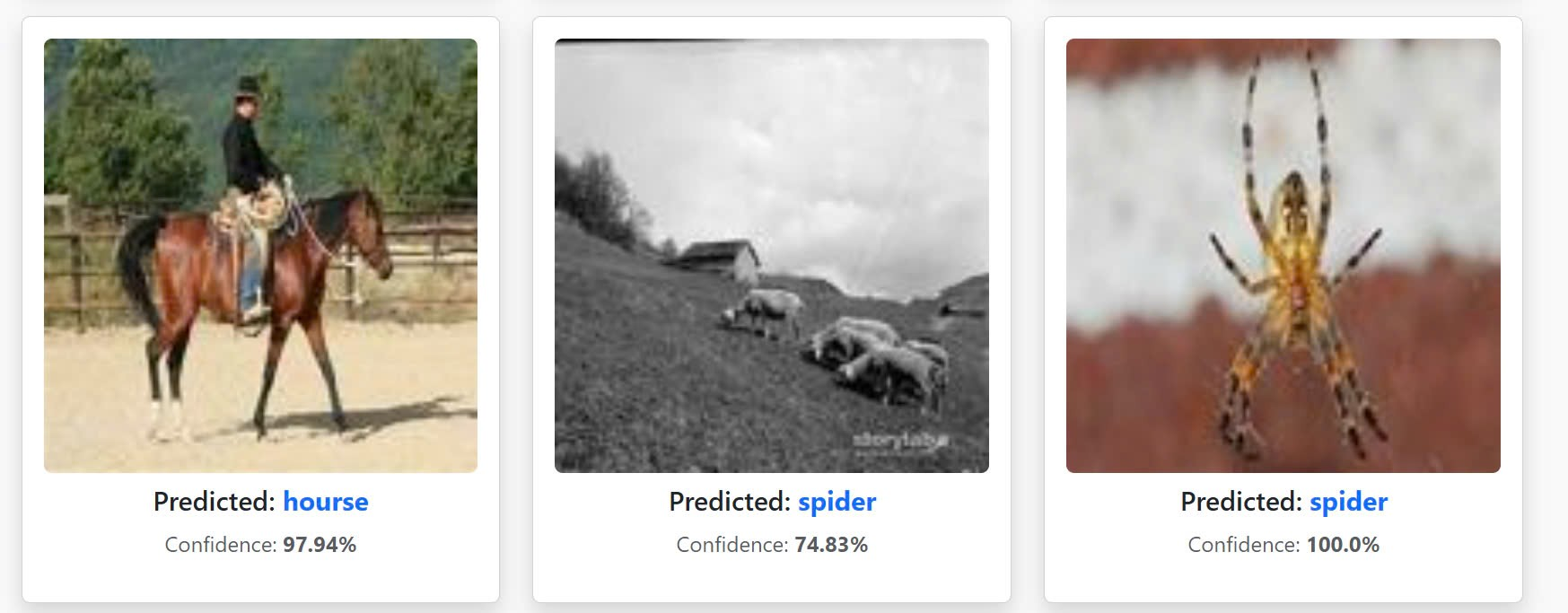
3.3.3. Kết quả

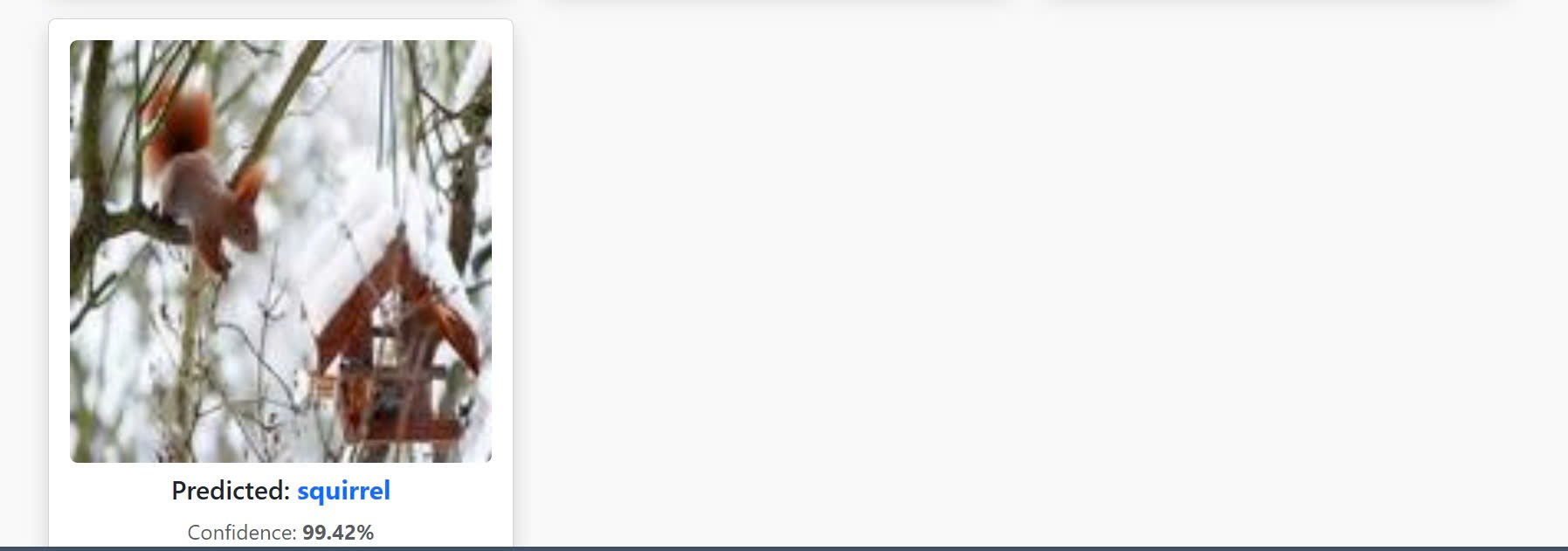












**3.4. Kết luận**

Mô hình CNN được xây dựng để phân loại hình ảnh động vật đã đạt được những kết quả đáng khích lệ sau quá trình huấn luyện và đánh giá. Cụ thể:

* **Hiệu suất huấn luyện**: Sau 40 epochs, mô hình đạt độ chính xác trên tập huấn luyện là 94.45% và độ chính xác trên tập kiểm định là 96.27%, với giá trị độ lỗi (loss) lần lượt là 0.1633 và 0.1154. Các biểu đồ xu hướng loss và accuracy (Biểu đồ 3.1 và Biểu đồ 3.2) cho thấy mô hình hội tụ tốt, với độ lỗi giảm đều và độ chính xác tăng ổn định qua các epoch.
* **Hiệu suất trên tập kiểm tra**: Mô hình đạt độ chính xác 92.41% trên tập kiểm tra với độ lỗi 0.2310, vượt xa kết quả ban đầu trong tài liệu (68.42%). Điều này chứng minh rằng mô hình không chỉ học tốt trên dữ liệu huấn luyện mà còn có khả năng tổng quát hóa hiệu quả trên dữ liệu mới.
* **Hiện tượng overfitting**: So với tài liệu ban đầu (độ chính xác trên tập kiểm tra chỉ 68.42%, cho thấy overfitting), mô hình hiện tại đã giảm thiểu đáng kể hiện tượng này. Độ chính xác trên tập huấn luyện (94.45%), tập kiểm định (96.27%), và tập kiểm tra (92.41%) khá gần nhau, cho thấy các kỹ thuật như điều chỉnh tốc độ học (từ 3.0 đến 7.5), regularization, và data augmentation đã phát huy hiệu quả.
* **So sánh với mục tiêu đề tài**: Mô hình CNN cơ bản đã đáp ứng tốt mục tiêu của đề tài, đạt độ chính xác cao và có tiềm năng ứng dụng thực tiễn trong các lĩnh vực như giám sát động vật hoang dã, bảo tồn sinh học, hoặc phát triển hệ thống nhận diện động vật trên thiết bị di động. Tuy nhiên, việc so sánh với các mô hình nâng cao như ResNet hay VGG vẫn chưa được thực hiện do thiếu dữ liệu đối chiếu, và đây là một hướng cần nghiên cứu thêm.

**Hạn chế**:

* Mặc dù độ chính xác trên tập kiểm tra đạt 92.41%, vẫn còn khoảng cách để đạt mức tối ưu (gần 100%). Một số lớp động vật có thể khó phân biệt do sự tương đồng về đặc trưng hình ảnh, đòi hỏi phân tích chi tiết hơn về hiệu suất trên từng lớp (ví dụ: precision, recall, F1-score).
* Quá trình huấn luyện cho thấy dao động lớn trong validation loss và accuracy ở các epoch đầu, đặc biệt trong 10 epoch đầu tiên, điều này có thể ảnh hưởng đến sự ổn định của mô hình nếu áp dụng vào các tập dữ liệu khác.
* Nghiên cứu chưa triển khai mô hình trên các thiết bị thực tế (như thiết bị di động) hoặc tích hợp vào hệ thống nhận diện video thời gian thực, nên giá trị thực tiễn vẫn cần được kiểm chứng thêm.

**Hướng phát triển tiếp theo**:

* So sánh hiệu suất của mô hình CNN cơ bản với các mô hình nâng cao như ResNet, VGG, hoặc EfficientNet để tìm ra kiến trúc tối ưu hơn.
* Áp dụng kỹ thuật transfer learning với các mô hình pre-trained (như ResNet-50 hoặc VGG-16) để cải thiện độ chính xác, đặc biệt trên các lớp động vật khó phân biệt.
* Tăng cường dữ liệu (data augmentation) với các kỹ thuật phức tạp hơn như mixup, cutout, hoặc thêm nhiễu để cải thiện khả năng tổng quát hóa của mô hình.
* Phân tích chi tiết hiệu suất trên từng lớp động vật để xác định các lớp có độ chính xác thấp và đưa ra giải pháp cải thiện.
* Triển khai mô hình trên thiết bị di động hoặc tích hợp vào hệ thống nhận diện video thời gian thực, từ đó đánh giá tính khả thi trong các ứng dụng thực tế như giám sát động vật tại sở thú, trang trại, hoặc trong tự nhiên.

Tóm lại, mô hình CNN đã đạt được những kết quả khả quan, đáp ứng tốt mục tiêu ban đầu của đề tài và đặt nền tảng cho các nghiên cứu tiếp theo. Chương tiếp theo sẽ tập trung vào việc tối ưu hóa mô hình và triển khai ứng dụng thực tế để nâng cao giá trị thực tiễn của nghiên cứu này.