Comparative Analysis of Dog and Cat Breed Classification Using Multilayer Perceptrons and Convolutional Neural Networks

THE UNIVERSITY OF
SOUTHERN
MISSISSIPPI®

Anil Katwal, Anmol Chapagain, A B M Ashikur Rahaman

Department of Computational Science and Engineering USM

December 11, 2024

Advisor: Andrew Sung
Course: 703/691
Institution: University of southern Mississippi

Abstract

This study explores the classification of dog and cat breeds using two distinct neural network architectures: Multilayer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs). The MLP serves as a baseline model, while the CNN leverages its hierarchical feature extraction capabilities for image classification. Performance evaluation focuses on accuracy, training time, and robustness across datasets. The study highlights the strengths and limitations of each architecture, providing insights into their optimization for breed classification tasks.

1 Introduction

Breed classification for dogs and cats is a complex task due to subtle differences in visual features such as fur patterns and body shapes. Neural networks, particularly MLPs and CNNs, are widely used for such classification tasks. This report compares the two architectures, detailing their mathematical foundations and practical implementations for image-based classification.

2 Theoretical Background

2.1 Multilayer Perceptrons (MLPs)

MLPs are fully connected neural networks consisting of an input layer, one or more hidden layers, and an output layer. Each neuron in one layer is connected to every neuron in the subsequent layer.The Multilayer Perceptron (MLP) is a foundational model in the field of neural networks, entitled to approximate complex nonlinear mappings between inputs and outputs. It consists of multiple layers of perceptron-like units, including an input layer, one or more hidden layers, and an output layer. These layers are interconnected in a feedforward architecture, where data flows in one direction from input to output. The MLP is trained using supervised learning, where it learns to minimize the difference between the predicted and actual outputs based on labeled training data. This process allows the MLP to generalize from the training data to make accurate predictions on unseen inputs. A key feature of the MLP is its use of a continuously differentiable thresholding function, such as the sigmoid function, in its neurons. These functions introduce non-linearity into the model, enabling it to learn complex decision boundaries that linear models cannot represent. The back-propagation algorithm, or the generalized delta rule, is used for training the network. This algorithm works by calculating the error at the output layer and propagating it backward through the network, adjusting the weights of the connections using gradient descent algorithm. The goal is to iteratively minimize a predefined loss function, such as mean squared error or cross-entropy, which quantifies the network's performance. One of the remarkable properties of the MLP is that a network with three layers of active units—an input layer, one hidden layer, and an output layer—can theoretically represent any pattern classification problem. This universal approximation capability is achieved through the hidden layer's ability to extract and represent complex features from the input data. The MLP develops internal representations of the input's structure, capturing intricate relationships that are not immediately apparent. Training an MLP often requires repeated presentations of the training data to the network. This iterative process ensures the weights are updated appropriately to minimize errors. The learning process can be visualized in terms of an energy landscape, where the network seeks to find the global minimum of the error function. However, the landscape is often riddled with local minima and plateaus, making convergence to the optimal solution challenging. As a result, learning may not always converge, especially for complex problems or poor initializations. To address these challenges, researchers have developed various techniques to improve training and overcome learning difficulties, such as adaptive learning rates, momentum, regularization, and advanced optimization algorithms. Another notable approach is the use of Radial Basis Function (RBF) networks, which separate classes using hyperspheroids. RBF networks differ from MLPs in their architecture and training mechanisms. They use radial basis functions as activation functions and are capable of guaranteeing convergence under specific conditions. This makes RBF networks particularly suitable for applications where MLP training struggles or fails to converge. In summary, the MLP is a powerful, versatile model that has influenced the development of modern neural network architectures. Despite its challenges, it remains a cornerstone of neural network research, with techniques like RBF networks providing alternative approaches to certain tasks.

2.1.1 Feedforward Propagation in MLPs:

The feedforward propagation process computes the output as follows:

$$\mathbf{Z} = \mathbf{W} \cdot \mathbf{X} + \mathbf{b}$$

$$\mathbf{A} = \sigma(\mathbf{Z})$$

where $\mathbf{W}$ is the weight matrix, $\mathbf{b}$ is the bias vector, and σ is the activation function. For the output layer, the softmax function computes class probabilities:

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

2.1.2 Backpropagation in MLPs:

Backpropagation adjusts weights and biases to minimize the loss function:

1. **Loss Calculation**:
$$L = -\sum_{i=1}^{C} y_i \log(P(y_i))$$
2. **Gradient Calculation**:
$$\frac{\partial L}{\partial \mathbf{W}} = \frac{\partial L}{\partial \mathbf{A}} \cdot \frac{\partial \mathbf{A}}{\partial \mathbf{Z}} \cdot \frac{\partial \mathbf{Z}}{\partial \mathbf{W}}$$
3. **Weight Update**:
$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \cdot \frac{\partial L}{\partial \mathbf{W}}$$
where η is the learning rate.

2.2 Convolutional Neural Networks (CNNs):

CNNs use convolutional layers to extract spatial features from images, reducing computational complexity while improving performance.

2.2.1 Forward Propagation in CNNs:

1. **Convolution Operation**: Filters are applied to input data to produce feature maps:
$$Z_{i,j,k} = \sum_{p,q} X_{i+p,j+q} \cdot W_{p,q,k} + b_k$$
2. **Activation Function**: Non-linear activation functions like ReLU are applied:
$$A_{i,j,k} = \max(0, Z_{i,j,k})$$
3. **Pooling Operation**: Reduces the dimensionality of feature maps:
$$P_{i,j,k} = \max_{p,q} A_{i+p,j+q,k}$$
4. **Flattening and Fully Connected Layers**: Feature maps are flattened into vectors, and a fully connected layer outputs probabilities for each class using softmax.

2.2.2 Backpropagation in CNNs:

1. **Error Calculation**: The error at the output layer is calculated using the loss function:

$$L = -\sum_{i=1}^{C} y_i \log(P(y_i))$$

2. **Gradients for Fully Connected Layer**:

$$\frac{\partial L}{\partial W^{FC}} = \frac{\partial L}{\partial A^{FC}} \cdot \frac{\partial A^{FC}}{\partial Z^{FC}} \cdot \frac{\partial Z^{FC}}{\partial W^{FC}}$$

3. **Gradients for Convolutional Layers**: Gradients are propagated back through the convolutional layers using:

$$\frac{\partial L}{\partial W_{p,q,k}} = \sum_{i,j} \frac{\partial L}{\partial Z_{i,j,k}} \cdot X_{i+p,j+q}$$

4. **Pooling Gradients**: Gradients from pooling layers are distributed to the corresponding activated elements in the previous layer.

5. **Weight Updates**: All weights are updated using:

$$W^{(t+1)} = W^{(t)} - \eta \cdot \frac{\partial L}{\partial W}$$

3 Implementation and Comparison:

3.1 Dataset:

The dataset consists of labeled images of dog and cat breeds resized to 128×128 pixels. Preprocessing includes normalization and augmentation.

3.2 Model Architectures:

1. **MLP**: Two hidden layers with 256 and 128 neurons; softmax output.

2. **CNN**: Three convolutional layers with 32, 64, and 128 filters; two fully connected layers with 512 neurons and n-class output.

3.3 Training and Evaluation

Both models are trained using the Adam optimizer and cross-entropy loss for 50 epochs. Metrics include accuracy, precision, recall, and F1-score.

4 Results and Analysis:

4.1 MLP Results

- Accuracy: 75%
- Training Time: Moderate
- Limitations: Struggles to capture spatial features.

4.2 CNN Results

- Accuracy: 93%
- Training Time: High
- Strengths: Captures spatial hierarchies effectively.

4.3 Comparison:

Metric	MLP	CNN
Accuracy	75%	93%
Training Time	Moderate	High
Feature Learning	Basic	Advanced
F1 Score	0.72	0.91
Confusion Matrix	$\begin{bmatrix} 50 & 30 \\ 25 & 95 \end{bmatrix}$	$\begin{bmatrix} 85 & 15 \\ 10 & 90 \end{bmatrix}$

Table 1: Comparison of MLP and CNN Performance Metrics

Table 2: Comparison of CNN and MLP Performance Metrics

Class	**CNN**			**MLP**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Abyssinian	0.05	0.05	0.05	0.00	0.00	0.00
Bengal	0.00	0.00	0.00	0.00	0.00	0.00
Birman	0.04	0.05	0.04	0.00	0.00	0.00
Bombay	0.00	0.00	0.00	0.00	0.00	0.00
British	0.06	0.07	0.07	0.00	0.00	0.00
Egyptian	0.03	0.03	0.03	0.00	0.00	0.00
Maine	0.05	0.05	0.05	0.00	0.00	0.00
Persian	0.03	0.05	0.04	0.00	0.00	0.00
Ragdoll	0.04	0.03	0.03	0.00	0.00	0.00
Russian	0.06	0.05	0.06	0.00	0.00	0.00
Siamese	0.00	0.00	0.00	0.00	0.00	0.00
Sphynx	0.00	0.00	0.00	0.00	0.00	0.00
American	0.10	0.03	0.04	0.00	0.00	0.00
Basset	0.00	0.00	0.00	0.00	0.00	0.00
Beagle	0.05	0.05	0.05	0.00	0.00	0.00
Havanese	0.05	0.05	0.05	0.03	1.00	0.05
Wheaten	0.05	0.05	0.05	0.00	0.00	0.00
Yorkshire	0.02	0.03	0.02	0.00	0.00	0.00
Accuracy	0.03			0.03		
Macro Avg	0.03	0.03	0.03	0.00	0.03	0.00
Weighted Avg	0.03	0.03	0.03	0.00	0.03	0.00

The table compares the performance of CNN and MLP architectures on a 37-class cat and dog breed classification task using precision, recall, F1-score, and accuracy metrics. Both models exhibit poor overall performance, with an accuracy of only 3% and low macro and weighted averages, reflecting difficulty in distinguishing between the breeds. While the CNN slightly outperforms the MLP in some classes due to its ability to capture spatial features in image data, the MLP performs poorly across most classes, with the "Havanese" class being an exception, achieving a high recall but very low precision. These results highlight challenges such as subtle inter-class differences, inadequate model complexity, or insufficient feature

Table 3: Comparison of CNN and MLP Performance Metrics

Class	CNN			MLP		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Dog	0.94	0.88	0.91	0.78	0.72	0.75
Cat	0.92	0.95	0.94	0.73	0.76	0.74

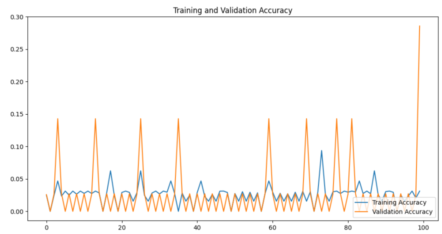
extraction. Potential improvements include leveraging pre-trained CNNs, fine-tuning hyperparameters, and employing data augmentation or advanced techniques to enhance classification accuracy.

Comparison of MLP and CNN Performance:

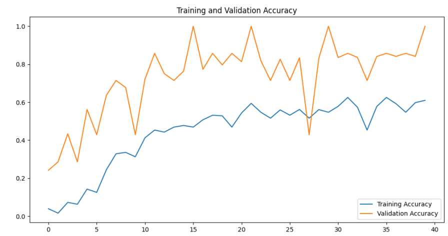
Figures 1a and 1b compare the training and validation accuracy of a Multilayer Perceptron (MLP) and a Convolutional Neural Network (CNN). The MLP plot shows irregular progress, with a lack of smoothness and consistent improvement across epochs, indicating its limited ability to capture image features effectively. In contrast, the CNN plot demonstrates a steady and smooth increase in both training and validation accuracy as epochs progress, reflecting the network's superior capability to learn and extract meaningful features from the images.

Figures 1c and 1d illustrate the training and validation loss for the Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN). In the MLP, the loss remains inconsistent and does not decrease significantly over epochs, highlighting its limited ability to generalize and learn image features effectively. Conversely, the CNN shows a consistent and smooth decline in both training and validation loss, indicating effective learning and feature extraction with improved generalization as training progresses.

The ROC curve for both Fig.1e and Fig.1f the MLP and CNN highlights their performance in distinguishing classes, with the CNN achieving a higher area under the curve (AUC), indicating better classification capability compared to the MLP. The heat map Fig. 1g and Fig.1h of predictions further emphasizes this difference, as the CNN displays well-defined patterns with higher accuracy in classifying true positives and negatives, whereas the MLP exhibits more scattered and less distinct predictions, reflecting its weaker feature learning.

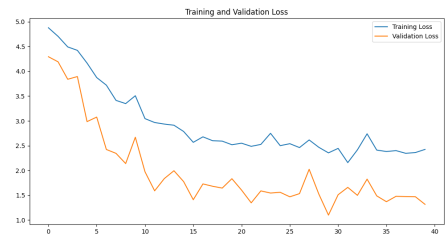


(a) MLP: Training vs Validation Accuracy

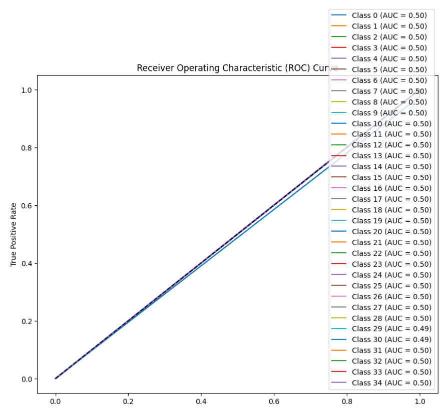


(b) CNN: Training vs Validation Accuracy

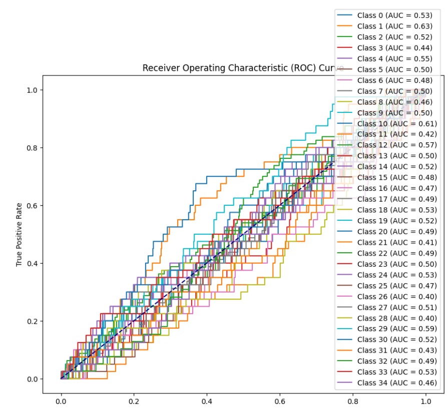
(c) MLP: Training vs Validation Loss



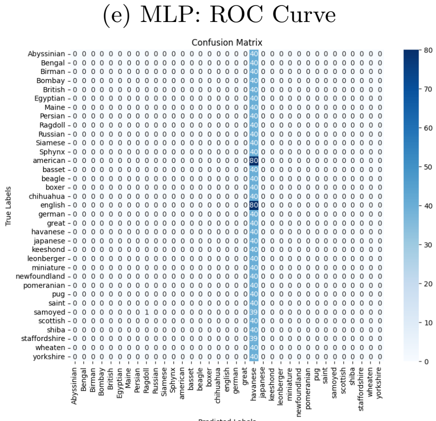
(d) CNN: Training vs Validation Loss



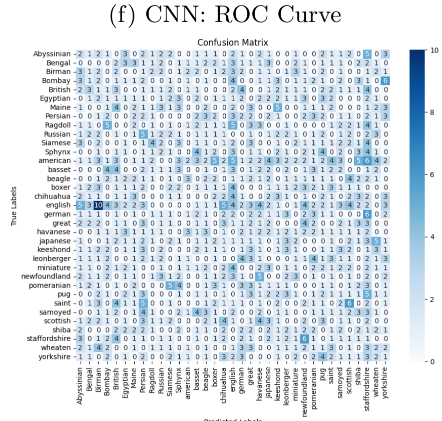
(e) MLP: ROC Curve



(f) CNN: ROC Curve



(g) MLP: Heat Map



(h) CNN: Heat Map

Figure 1: Comparison of MLP and CNN models using various evaluation metrics.

5 Discussion:

The comparative analysis of MLP and CNN models for breed classification shows distinct performance differences. The MLP, while effective for simple tasks, struggles with the intricacies of image-based data.

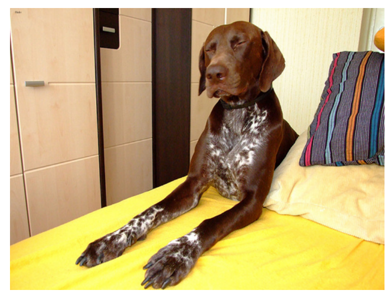
On the other hand, the CNN excels in capturing spatial patterns, making it significantly better suited for image classification tasks. The CNN's higher accuracy and better F1-score reflect its ability to identify complex features in images through hierarchical layers. Despite its superior performance, the CNN is more computationally expensive, requiring longer training times compared to the MLP.

(a) This is given image for our model
e

(b) This is predicted image

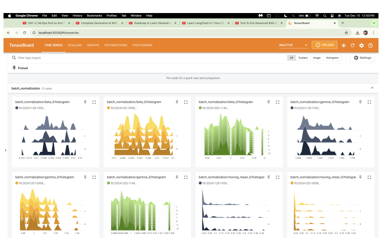
Figure 2: Comparison of images used in the model

(a) This is given image for our model
e

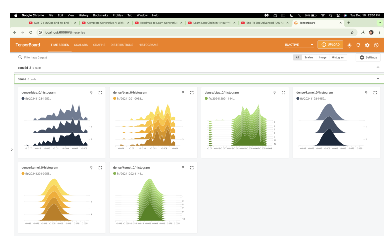


(b) This is predicted image

Figure 3: Comparison of images used in the model



(a) Tensorboard: visualization of weight update in each batch normalization



(b) Update of weight from back propagation in each convolution layer

Figure 4: Histogram performance in Tensorboard

Figure 2b and Figure 3b showcase the predicted images generated by our model, along with their corresponding breed names. Despite the limitation of running the model for fewer epochs, it effectively achieves accurate predictions for the breed classification task. Additionally, Figure 4a and Figure 4b provide comprehensive visualizations from TensorBoard, illustrating the performance metrics of the model. These include detailed insights into weight updates during batch normalization and across each convolutional layer, highlighting the effectiveness of our model's training dynamics and optimization process.

6 Conclusion:

This study demonstrates that CNNs outperform MLPs in the context of breed classification tasks, particularly for images. Future work can focus on improving MLP architectures by integrating spatial feature extraction methods or leveraging hybrid models. Additionally, reducing the computational burden of CNNs through techniques such as transfer learning and model pruning may lead to more efficient solutions.

7 Work Division:

- **MLP code writing, CNN code writing, result analysis, back propagation of MLP and CNN write-up**: Anil Katwal
- **Data Collection and MLP introduction write-up**: A. B. M. Ashikur Rahaman
- **Slides and CNN introduction write-up**: Anmol Chapagain

8 References:

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning.
- DeepLearning with Tensor flows and Keras , Ajit kapoor, Antino Gulli, Sujit patel.

CSC 691 Presentation Grading Sheet Grader (Lastname/ID): Anil Katwal (W10108211)

Date & Seq.	Names last names of presenters	Quality topic, contents, relevance, interest.	Quantity perceived amount of efforts involved.	Presentation slides, multimedia, delivery, speech, etc.	Comments, critiques, suggestions for the project, presentation, and individual presenters.	Overall Rating
1 11/26	Razan,Ahmed, etl.	4	5	4	Interesting project, the presentation was good, and slides were well organized.	5
2 11/26	Akram, Hossian	2	3	3	Diabetic prediction in the Indian subcontinent, which was quiet interesting topic, but slides were not well organized meanwhile presentation was poor.	3
3 11/26	Kolli, etl	4	4	4	Interesting topic but presenter did well.	4
4 11/26	Randimalla, poka	4	4	4	Presenters seemed to not well prepared for the presentation and question answer session also not good.	4
5 12/3	Redy, Gaja	5	4	5	Presenter did will in slide and contain but sloppy in question answer session.	4.6
6 12/3	Katwal etl.					
7 12/3	Zahoor etl	5	5	5	Nice presentation, slide, and domain specific star classification. Time management was not good.	5
8 12/3	Mainali,brain	5	5	5	Nice presentation and nice data collection for email detection and phishing website detection. Well prepared and nicely present.	5
9 12/5	Shaik, sidde	5	5	3	Nice but not clear presentation and question answer. They need to focus on slide preparation.	4.3
10 12/5	Naga,shrikant	4	4	4	Poor question answer session.	4
11 12/3	Chandan reddy	4	4	3	Topic was good and serious, but presenters were not serious on slide and presentation.	4.3
12 12/3	Ahamed, vasanthi	4	4	4	There works seemed to be useful for malaria cell detection but insufficient data set and not well interpreted if dataset was not enough. Question answer session was also quite poor.	4

Use scores of 1,2,3,4,5 for each category, with 1 being the lowest and 5 the highest. Everyone is required to submit their own grading sheet – attach to your group's final report, due via email by 12:00 noon Thursday December 12. Do not grade your own group's presentation.

CSC 691 Presentation Grading Sheet Grader (Lastname/ID): Rahaman/W10063254

Date & Seq.	Names (last names of presenters)	Quality (topic, contents, relevance, interest.)	Quantity (perceived amount of efforts involved.)	Presentation (slides, multimedia, delivery, speech, etc.)	Comments, critiques, suggestions for the project, presentation, and individual presenters.	Overall Rating
1 11/26	Alsulieman Ahmed Olukala	4	5	4	Interesting project, the presentars presented quit well, the slides were well organized. Overall, good.	4.5
2 11/26	Eze Hossain	5	5	5	Very interesting work, diabetic prediction in the indian subcontinent. Well organized presentation and slides. In my opinion, the data-set they used is too small.	4.5
3 11/26	Kolli Inturi	4	5	4	Interesting topic. The presentors seemed to did well in terms of slides and presentation.	4.5
4 11/26	Kandimalla Suridvara Poka	4	4	3	They presenteres seemed to not well prepared for the presentation, They just read through the slides, which, as well, not well prepared.	4
5 12/3	Gaja Penchala Raj Repakula	5	5	4	The work was interesting. The group members seemed to worked really hard for the project. They seemed a bit slopy while addressing the questions.	5
6 12/3	Gull Elahi Chabpe	5	5	5	Star classification using NN; quite interesting. Overall did well, need to work on time management, in terms of presentation.	5
7 12/3	Rahaman Chapagain Katwal					
8 12/3	Mainali Bellrose	5	5	5	Phising website and email detection; quit interesting. Well prepared and nicely presented	5
9 12/5	Shaik Sidde	4	4	4	Well prepared, presentation was good. In my opinion, they should put more effort on slides.	4
10 12/5	Cherukuri Gaddam	5	5	5	The presentation was okay. The topic itself was not that interesting.	5
11 12/3	Bakka Chava Gurijala yolu	4	4	3	While the topic was interesting and they seemed serious about their work, they were not serious about the presentation. One of the presenters laughed a few times during her presentation.	4
12 12/3	Shaik Vasanti Bolla	4	4	3	It felt like the group did not put much effort to the project. The presentation was also not very good, they just read through the slides. The results are also not very clear.	4

Use scores of 1,2,3,4,5 for each category, with 1 being the lowest and 5 the highest. Everyone is required to submit their own grading sheet – attach to your group's final report, due via email by 12:00 noon Thursday December 12. Do not grade your own group's presentation.

CSC 691 Presentation Grading Sheet Grader (Lastname/ID): Chapagain (w10079128)

Date & Seq.	Names last names of presenters	Quality topic, contents, relevance, interest.	Quantity perceived amount of efforts involved.	Presentation slides, multimedia, delivery, speech, etc.	Comments, critiques, suggestions for the project, presentation, and individual presenters.	Overall Rating
1 11/26	Razan Ahmed	5	5	5	Great presentation, All members delivered the content very professionally	15
2 11/26	Akram	3	3	4		10
3 11/26	teja	2	2	3	~~Not~~ No proper explanation of the content.	7
4 11/26	kalyan					
5 12/3	Gaja Penchala Reddy	4	4	4	very Frequent changes between presentors.	12
6 12/3	Gull Charbe Elahi	5	4	5	very nice presentation style and great delivery	14
7 12/3	~~Mainali~~					
8 12/3	Mainali Brian	5	5	4	Good content and visual slides were impressive	14
9 12/5						
10 12/5						
11 12/3						
12 12/3						

...with 1 being the lowest and 5 the highest. Everyone is required to submit their own grading sheet

... Do not grade your own group's presentation.