Group E

Comparative Analysis of Emotion Detection using Deep-Neural Networks

1. Problem Statement and Application

Emotion recognition is critical in applications like human-computer interaction, healthcare, and marketing, by personalizing interactions, identifying emotions through facial expressions can significantly enhance user engagement and experience. The variability of facial expressions across individuals, lighting conditions, and occlusions, are crucial challenges all of which affect the performance of emotion detection algorithms [7]. Our goal is to evaluate the performance of AlexNet, ResNet18 and VGG16 on three distinct emotion detection datasets varying in size and complexity, in order to study the strengths and weaknesses of the different architectures to recognize emotions from images accurately. The goal is to achieve high performance in recognizing expressions such as anger, happiness, and fear from image data, and performing and ablative study to improve the model's ability to generalize across emotions. [2]

2. Image Dataset Selection

For emotion detection, it is essential to gather sufficient, high-quality data to enable the model to learn important patterns and variations across different emotions. The datasets are chosen for their size, diversity, and format variations, crucial for building a robust emotion recognition system.

	FER Emotion	Emotion Recognition	Sentiment Images
Classes	7	7	6
#Images	35,887	15,503	1,200
Image Size	48x48	48x48	224x224
Image Format	.png (35,887)	.png (7,003), .jpg (8,229), .jpeg (216), Other (6)	.jpg (1,109), .jpeg (36), .webp (50), Other (5)

Table 1. Dataset Statistics

6 Human Emotions Dataset: The Sentiment Images or 6 Human Emotions dataset [6] includes 1,200 high-resolution images of varying dimensions spread across six emotion categories: Anger (214), Disgust (201), Fear (163), Happy (230), Pain (168) and Sad (224), in JPG, JPEG, WEBP, and other formats, this dataset offers greater visual detail, which is beneficial for capturing subtle emotional expressions.

Facial Emotion Recognition Dataset: The Facial Emotion Recognition dataset [4] contains 15,503 images, each representing one of seven emotions Ahegao (1,205), Angry (1,313), Happy (3,740), Neutral (4,027), Sad (3,934), and Surprise (1,234) mainly in PNG, JPG, and JPEG formats, making the dataset diverse in structure, with varying pixel dimensions, contributing to a versatile dataset that can train models to handle various image formats and sources.

Emotion Detection Dataset: The Emotion Detection dataset [1] contains 35,887 grayscale images distributed across 7 emotion classes: Angry (4,953), Disgusted (547), Fearful (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral (6,198). With uniform 48x48 pixel images in PNG format. It has an imbalance of classes which can be addressed by implementing appropriate sampling.

3. Possible Methodology

The dataset will undergo preprocessing, which will involve removing corrupted images, ensuring consistent RGB format across all images, handling any missing labels, and eliminating duplicate images to prevent bias. Following this, all images will be resized to 224x224 pixels, adhering to the standard input size required by the selected models while maintaining image quality. The pixel values will be normalized with mean [0.485, 0.456, 0.406] and standard deviations of [0.229, 0.224, 0.225] from ImageNet statistics, for faster convergence during training and effective transfer learning. The images are then converted into tensors for PyTorch compatibility with the following models:

AlexNet: An 8-layer network known for pioneering innovations like ReLU activations and GPU-based training. While it's simpler and faster, it performs well on smaller datasets but lacks the depth of modern architectures. [5]

ResNet18: A lightweight network with 18 layers, using residual connections to prevent vanishing gradients. It offers a balance between performance and computational efficiency, making it suitable for tasks requiring quick processing and lower memory usage. [3]

VGG16: It consists of 16 layers with small receptive fields, allowing it to capture fine details in images. It's highly accurate but computationally expensive, ideal for tasks prioritizing precision over speed. [8]

The models will be evaluated using accuracy, precision, recall, F1 score, training time and confusion matrix to ensure a comprehensive comparison. Additionally, TSNE will be used to visualize and understand how each architecture captures and represents the data.

The ResNet18 is expected to offer a balance between speed and classification accuracy, AlexNet to be the fastest performing well on simpler tasks, and VGG16 to excel in accuracy at the cost of higher resource consumption. By comparing the results across models, we aim to provide insights into the trade-offs between model performance, and computational complexity. This analysis will be valuable for selecting appropriate models in resource-limited research environments.

Group E

4. Gantt Chart

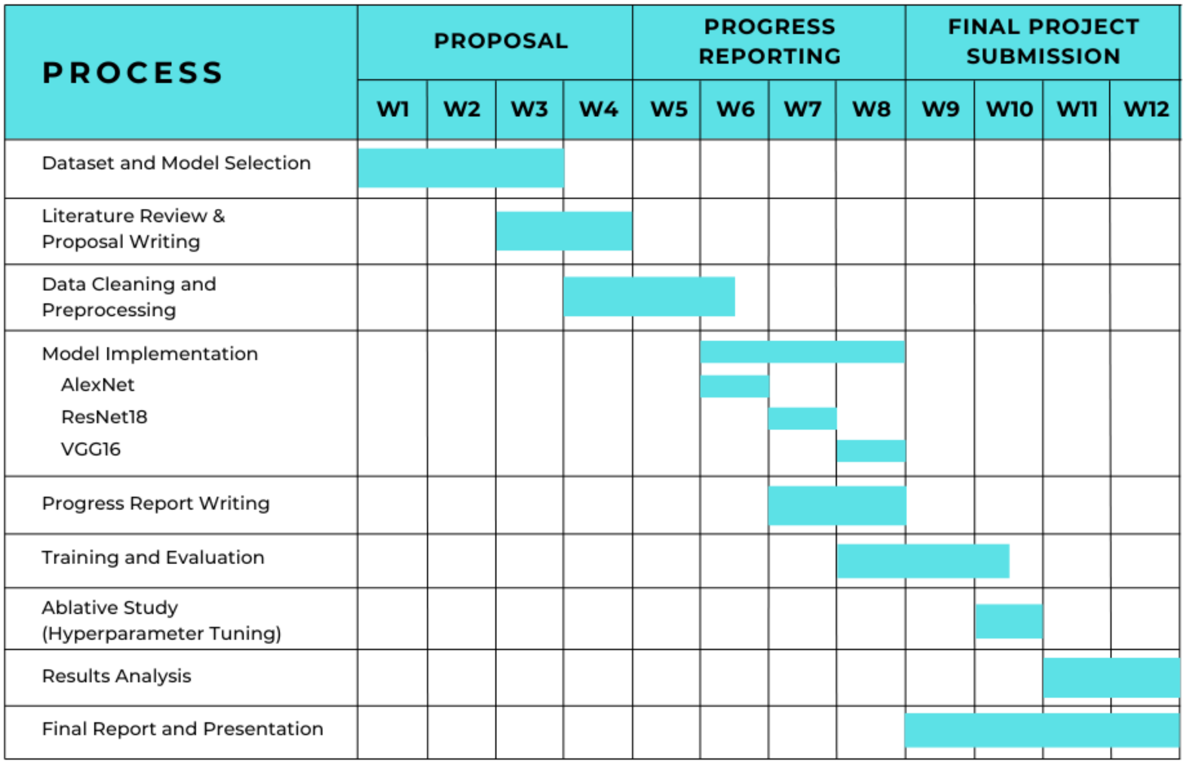

Figure 1. Gantt Chart Representing the various stages of the project.

The initial phase focuses on Dataset and Model Selection, alongside Literature Review and Proposal Writing, which culminates in Week 4. Data Cleaning and Preprocessing follow in Weeks 4-6. Model Implementation is done across Weeks 6-8, with AlexNet, ResNet18, and VGG16 being implemented sequentially. A Progress Report is scheduled for writing during Weeks 7 and 8, parallel to the model implementation. Training and Evaluation of the models occur in Weeks 8-10, followed by an Ablative Study for hyperparameter tuning in Week 10. The final stages involve Results Analysis in Weeks 11 and 12 and the preparation of the Final Report and Presentation done all through Weeks 9-12. This structure allows for a systematic approach to the project, with clear milestones and deadlines for each major phase of the research.

Group E

References

[1] Ares. Emotion detection — kaggle.com. https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer/data. Accessed 22-Sep-2024. 1

[2] Runfang Guo et al. Development and application of emotion recognition technology. *PMC*, 2024. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[4] Sujay Kapadnis. Facial emotion recognition — kaggle.com. https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer/data. Accessed 22-Sep-2024. 1

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 1

[6] Yousef Mohamed. 6 human emotions — kaggle.com. https://www.kaggle.com/datasets/yousefmohamed20/sentiment-images-classifier. Accessed 22-Sep-2024. 1

[7] Noldus. Emotion analysis - facial expression recognition software. *Noldus.com*, 2023. 1

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

5. Group Information

Bhavya Bugude - 40270772
Sweta Sarat - 40263287
Mahimur Rahman Khan - 40271113
Shivansh Gupta - 40320977