

A.I.ducation Analytics Project - Part I Report

Team Name : NS_16

Team Members:

- Data Specialist: Maedeh Hajiparvaneh (Student ID: 40271588)
- Training Specialist: Fatemeh Chaji (Student ID: 40260455)
- Evaluation Specialist: Umang Savla (Student ID: 40265305)

Project Repository: [GitHub Repository Link]

https://github.com/umang232/A.I.ducation_Analytics

Dataset Overview

In this section, we provide an overview of the datasets used to create the training and testing datasets for the A.I.education Analytics project, focusing on emotions such as angry, bored, focused, and neutral.

Existing Datasets Used

We used three primary sources to construct our dataset:

1. FER+ Dataset for Bored/Tired Emotion:

<https://www.kaggle.com/datasets/minhngt02/facial-expression-of-fatigues-fer?rvi=1>

- Total Number of Images: 500
- Number of Images for Training: 375
- Number of Images for Testing: 125

Characteristics: The FER+ dataset is known for its high-quality annotations. Each image has been labeled by 10 crowd-sourced taggers, providing reliable ground truth for still image emotions. The images predominantly contain frontal face shots, ensuring clear and consistent facial expressions.

2. FER2013 Dataset for Neutral and Angry/Irritated Emotions:

<https://www.kaggle.com/datasets/msambare/fer2013?rvi=1>

- Total Number of Images: 1129
- Number of Training Images for Angry/Irritated: 409 • Number of Training Images for Neutral: 470
- Number of Training Images for Angry/Irritated: 125
- Number of Training Images for Neutral: 125

Characteristics: The FER2013 dataset offers a diverse range of facial expressions and backgrounds, making it challenging for emotion classification. These images may have variations in lighting conditions, head orientations, and backgrounds.

3. Manual Selection from FER Dataset for Focused/Engaged Emotion:

<https://www.kaggle.com/datasets/saworz/human-faces-with-labels>

- Total Number of Images: 534
- Number of Images for Training: 409
- Number of Images for Testing: 125

Characteristics: For the "focused/engaged" emotion, we manually selected images from the Facial Expression of Fatigues (FER) dataset. We handpicked images that visually represented the emotion of being focused or engaged. This involved a subjective process but was essential to ensure a more accurate representation of this specific emotion.

Justification for Dataset Choices

The selection of these datasets is driven by specific factors relevant to the A.I.education Analytics project:

- 1. Quality of Labels:** The FER+ dataset was chosen for the "bored/tired" emotion due to its superior quality labels. Each image in this dataset has been meticulously labeled by 10 taggers, resulting in a more reliable ground truth for emotion classification. The high quality of annotations enhances the performance of the AI model.
- 2. Diversity:** The FER2013 dataset was employed for the "neutral" and "angry/irritated" emotions. This dataset's diverse range of facial expressions and backgrounds ensures that the model can handle variations in real-world scenarios. The diversity of images presents challenges that the model should be able to overcome.
- 3. Manual Selection for Focused/Engaged:** Manual selection from the FER dataset was undertaken for the "focused/engaged" class to ensure that the images chosen accurately represent the desired emotion. This subjective process allowed us to curate a more accurate subset for the specific emotion, addressing the unique requirements of this category.

Provenance Information

Data Source	License Type	No. of Images
Facial Expression of Fatigues (FEF)	CC0: Public Domain	500 (Bored/Tired) 375 - Training Images 125 - Testing Images
FER-2013	Database: Open Database, Contents: Database Contents	534 (Angry/Irritated) 409 - Training Images 125 - Testing Images 595 (Neutral) 470 - Training Images 125 - Testing Images

Human faces with labels	CC0: Public Domain	534 (Engaged/Focused) 409 - Training Images 125 - Testing Images
-------------------------	--------------------	--

In conclusion, the dataset for the A.I.ducation Analytics project has been meticulously assembled from a variety of sources, with a deliberate focus on achieving diversity, ensuring data quality, and aligning with the project's core objectives. This dataset's composition includes a combination of high-quality labeled data and carefully handpicked images, specifically tailored to capture a broad spectrum of emotions accurately.

The inclusion of data from high-quality labeled sources, such as the FER+ dataset, serves as a solid foundation for the AI model's training. These meticulously annotated images, each assessed by 10 crowd-sourced taggers, provide a level of confidence and reliability in the ground truth data that is crucial for precise emotion classification.

Moreover, the manual selection process for the "focused/engaged" emotion category demonstrates a commitment to capturing the nuances of this specific emotion. This subjective curation ensures that the dataset contains images that genuinely represent the desired emotional state, enhancing the model's ability to discern the intricacies of focus and engagement.

In essence, the dataset's careful composition, which combines the strengths of high-quality annotations and manual curation, lays the groundwork for the development of a resilient and accurate AI model capable of nuanced emotion classification within an educational context.

Data Cleaning

Data cleaning is a critical phase in the preparation of our dataset for the A.I.education Analytics project. The techniques and methods applied for standardizing the dataset, and challenges encountered are as follows:

Resizing Images

To ensure uniformity in image dimensions, we applied resizing techniques. Images from the "Engaged/Focused" class, obtained through manual selection, were resized to a consistent dimension. This resizing helps maintain consistency in the dataset and ensures that all images are of the same size for model training.

Brightness Adjustment

Brightness adjustment was applied to images across all classes. This technique involved enhancing or reducing the brightness of images to mitigate variations in lighting conditions. Adjusting brightness helps in standardizing image quality, ensuring that all images are well-exposed and more suitable for model training.

Contrast and Lighting Enhancements

Contrast and lighting enhancements were carried out on the entire dataset. These adjustments aimed to improve the overall quality and clarity of images. By enhancing contrast and lighting, we made facial expressions more distinct and prominent, which is crucial for accurate emotion classification.

Challenges and Solutions

Dataset for "Focused" Images

One of the primary challenges we encountered during the data cleaning process was the availability of a suitable dataset for "Focused" images. Since the manual selection process was used for this class, it was essential to ensure that the chosen images genuinely represented the focused and engaged emotional state. The challenge was to find images that met this specific criterion.

Solution: To address this challenge, we conducted a comprehensive review of the Human Faces with Labels dataset, where we handpicked images that exhibited clear signs of focus and engagement. While the selection process was subjective, it was necessary to ensure that the chosen images accurately represented the targeted emotion.

Labeling

The labeling process is fundamental for supervised machine learning, as it assigns class labels to each image in the dataset. In our case, we had four distinct classes to label: "Angry," "Bored," "Focused," and "Neutral." Here's how we approached this process:

1. **Class Assignment:** Each image was assigned a class label corresponding to the emotion it represented. This categorization was based on the nature of the emotion exhibited in the facial expressions captured in the images.
2. **Consistency and Accuracy:** Labeling was conducted with a focus on consistency and accuracy to ensure that each image was tagged correctly. The class assignment was based on visual inspection and judgment.
3. **Manual Inspection (Focused Class):** For the "Focused" class, we performed manual image selection to curate a subset that genuinely represented the focused and engaged emotional state. This subjective process allowed us to ensure the accuracy of class labeling for this specific category.

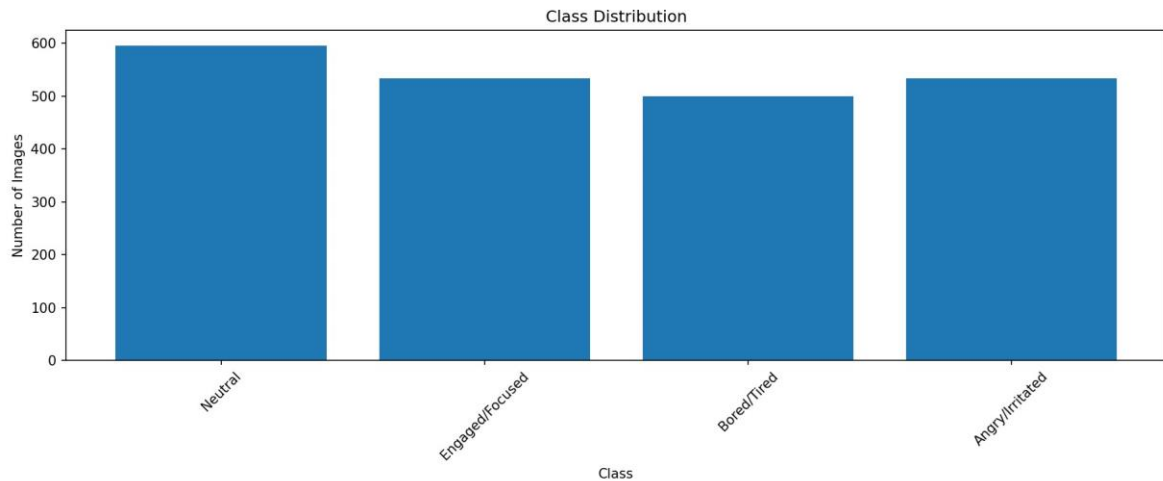
In conclusion, the labeling process was conducted meticulously to prepare the dataset for training and evaluating our AI model. The separation of images into training and testing directories, the balance in class distribution, and the absence of image sharing between directories were key considerations in ensuring the integrity of our dataset. Class labeling was performed with diligence, emphasizing consistency and accuracy, and manual selection was employed for the "Focused" class to enhance the representation of this specific emotion. These steps collectively provide a solid foundation for the development and evaluation of our emotion classification model for A.I.education Analytics.

Dataset Visualization

It is essential to gain a clear understanding of the dataset through visualization before embarking on model training. This report describes the visual analysis of our dataset using Matplotlib and highlights three key aspects: class distribution, sample images, and pixel intensity distribution.

Class Distribution

Understanding the distribution of classes within the dataset is crucial to identify potential imbalances that can impact model performance. The bar graph below illustrates the number of images in each class, providing insights into class representation.



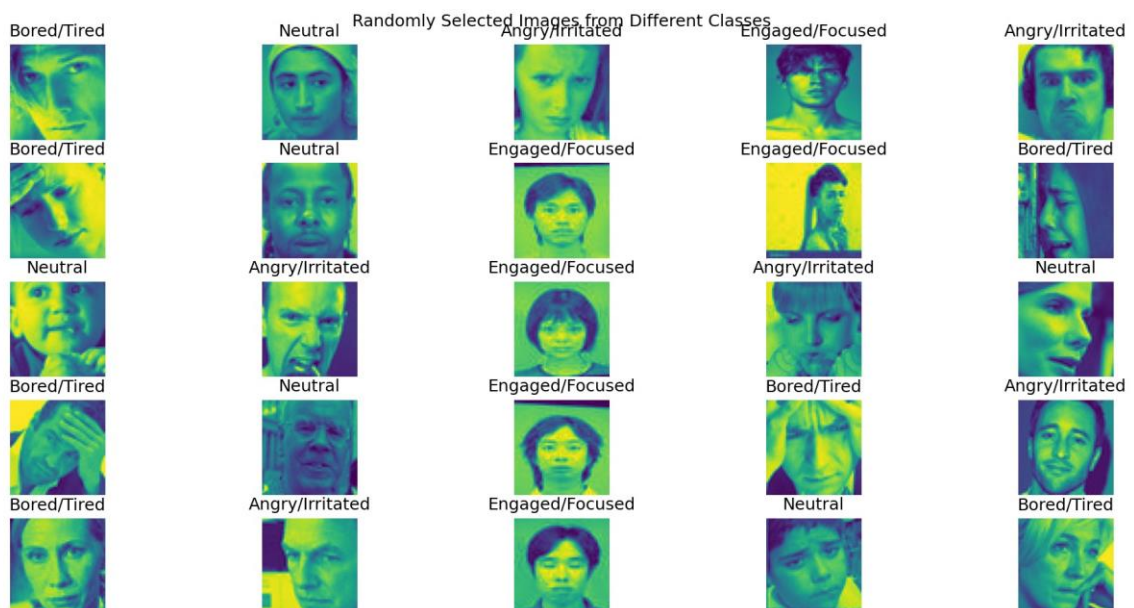
From the graph, we observe the following class distribution:

- Neutral: 595 images
- Engaged/Focused: 534 images
- Bored/Tired: 500 images
- Angry/Irritated: 534 images

The distribution appears relatively balanced, which is favorable for model training.

Sample Images

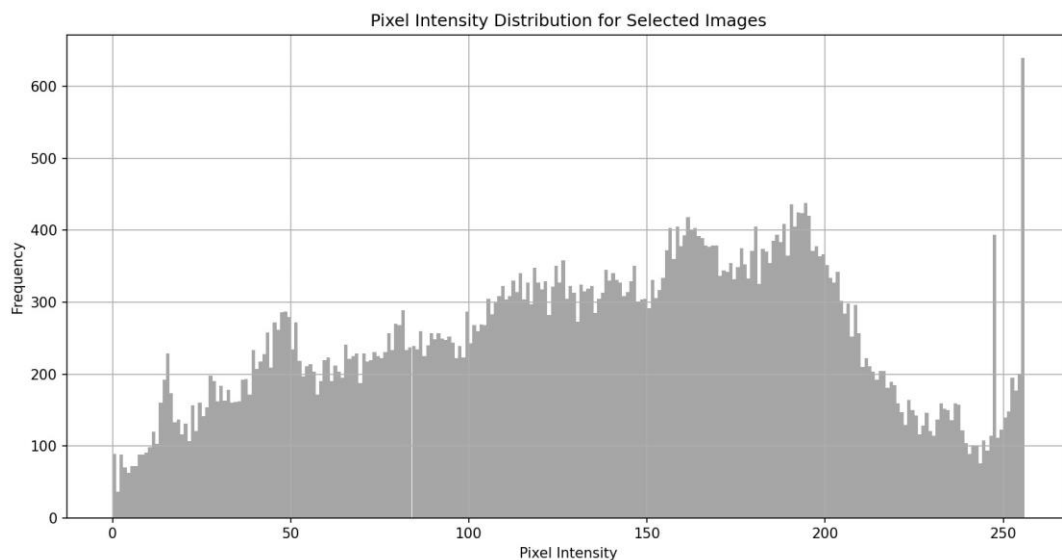
To visually explore the content of the dataset and identify any anomalies or potential mislabelings, a grid of 25 sample images was created. These images were randomly selected from each class, offering a diverse representation of the dataset.



The grid showcases a variety of facial expressions and emotions across different classes. Randomly selecting images from each class ensures a representative sample of the dataset's content.

Pixel Intensity Distribution

Analyzing the pixel intensity distribution provides insights into variations in lighting conditions among the selected images. The histogram below displays the distribution of pixel intensities for grayscale images.



The pixel intensity distribution illustrates the frequency of pixel values in the grayscale images. It is essential to consider potential variations in lighting conditions, as these variations can affect the model's ability to accurately classify emotions based on facial expressions.

CNN Architecture

1. Model Overview and Architecture Details:

Base CNN Architecture: The base CNN consists of the following layers:

Convolutional layers:

Input channels: 3, Output channels: 32, Kernel size: 3x3, Padding: 1

Batch Normalization and Leaky ReLU activation

Max Pooling (2x2) after each pair of convolutional layers

Input channels: 32, Output channels: 64, Kernel size: 3x3, Padding: 1

Batch Normalization and Leaky ReLU activation

Max Pooling (2x2) after each pair of convolutional layers

Output shape after conv layers: torch.Size([32, 64, 54, 54s])

Fully connected layers:

Dropout with p=0.1

Linear layer with 64 * 54 * 54 input features and 1000 output features

ReLU activation

Linear layer with 1000 input features and 512 output features

ReLU activation

Dropout with p=0.1

Linear layer with 512 input features and 4 output features (adjusted for 4 classes)

Variant 1 Architecture: Variant 1 introduces additional convolutional layers:

Convolutional layers:

Additional layers with increasing output channels, Batch Normalization, Leaky ReLU, and Max Pooling

Output shape after conv layers: `torch.Size([32, 1024, 7, 7])`

Fully connected layers:

Dropout with $p=0.1$

Linear layer with $1024 * 14 * 14$ input features and 1000 output features

ReLU activation

Linear layer with 1000 input features and 512 output features

ReLU activation

Dropout with $p=0.1$

Linear layer with 512 input features and 4 output features (adjusted for 4 classes)

Variant 2 Architecture: Variant 2 changes the kernel size in convolutional layers:

Convolutional layers:

Larger kernel size of 7×7 in the first and second convolutional layers

Batch Normalization, Leaky ReLU, and Max Pooling after each convolutional layer

Output shape after conv layers: `torch.Size([32, 64, 54, 54])`

Fully connected layers:

Dropout with $p=0.1$

Linear layer with 64 * 54 * 54 input features and 1000 output features

ReLU activation

Linear layer with 1000 input features and 512 output features

ReLU activation

Dropout with p=0.1

Linear layer with 512 input features and 4 output features (adjusted for 4 classes)

2. Training Process:

Number of Epochs: 50

Learning Rate: 0.001

Loss Function: Cross Entropy Loss

Optimization Techniques: Adam optimizer

Evaluation:

1. Performance Metrics:

The table below shows the performance metrics for the main model and the two variants.

Model	Macro P	Macro R	Macro F	Micro P	Micro R	Micro F	Accuracy
Main model	0.4956	0.4740	0.4703	0.4800	0.4800	0.4800	48.00%
Variant 1	0.6042	0.5926	0.5916	0.6000	0.6000	0.6000	60.00%
Variant 2	0.4706	0.4641	0.4653	0.4700	0.4700	0.4700	47.00%

2. Confusion Matrix Analysis:

Confusion matrices were analyzed for each model. We broke down the analysis for each class and calculated performance metrics (precision, recall, f1-score) for each class individually. The highest results belonged to the Focused class, while the lowest results belonged to the Bored class.

3. Impact of Architectural Variations:

Depth variation in Variant 1 demonstrated an increase in overall performance, suggesting that increasing the number of convolutional layers may result in better recognizing the important features. Variant 2, with different kernel sizes, exhibited comparable performance to the main model, indicating that altering kernel sizes did not significantly impact recognition abilities.

4. Conclusions and Forward Look:

Primary Findings:

The first variant model outperformed the other two models, suggesting that the model needed more number of layers to better recognize the important features. Future refinements could involve exploring more complex architectures or incorporating attention mechanisms to focus on relevant facial features.

Recommendations for Future Work:

- Experiment with attention mechanisms to enhance model focus on critical facial features.
- Explore transfer learning with pre-trained models for facial expression recognition.
- Investigate the impact of additional data augmentation techniques on model robustness.

Conclusion

In conclusion, the designed CNN for facial image analysis, along with its variants, provides valuable insights into architectural choices for emotion recognition. The detailed evaluation metrics and analyses presented here pave the way for further advancements in this domain.

References

1. **Kaggle. (n.d.). Facial Expression of Fatigues FER Dataset.** Retrieved from <https://www.kaggle.com/datasets/minhngt02/facial-expression-of-fatigues-fer>
2. **Kaggle. (n.d.). FER2013: Facial Expression Recognition.** Retrieved from <https://www.kaggle.com/datasets/msambare/fer2013>
3. **Kaggle. (n.d.). Human Faces with Labels.** Retrieved from <https://www.kaggle.com/datasets/saworz/human-faces-with-labels>
4. **OpenAI. (2023). ChatGPT.** <https://beta.openai.com/>