Week 2 Narrative Report: Synesthetic Learning Project

Chris Jallaine Mugot

Department of Data Science - College of Information Technology and Computing
University of Science and Technology of Southern Philippines
Lapasan, Cagayan de Oro City 9000, Philippines

This week marked a transition from conceptual development to the concrete groundwork needed to operationalize our study on synesthetic learning between audio and abstract art. After establishing the intellectual foundation and research direction in Week 1, Week 2 focused on assembling, validating, and preparing the multimodal datasets that will serve as the basis for model training and cross-modal alignment. Our goal for this phase was to ensure that both the audio and artwork datasets were complete, well-organized, and harmonized under a consistent emotional label space.

As a team, we worked on gathering the required datasets, reviewing their structure, and constructing scripts that would allow us to process them systematically. My contribution this week centered on designing and implementing the data exploration and preprocessing workflows. I took the lead in writing the routines that validated the directory structures, checked for missing files, and generated summaries of available classes and samples. This ensured that we had full confidence in the integrity of the Audio Emotion Dataset and the D-ViSA abstract artwork collection before moving into deeper processing. I also handled the parsing of metadata for both modalities. For the audio dataset, I developed a function that automated the extraction of emotion labels and sample identifiers from the filenames, transforming them into a clean and usable DataFrame. For the artwork dataset, I wrote scripts that loaded and cataloged all image files according to their associated emotion categories. This organization was essential for establishing a consistent interface through which the model can later access paired visual and auditory inputs.

One of the more critical tasks I contributed to was the emotion label harmonization. Because both datasets came from different sources with varying emotion taxonomies, I worked on constructing a mapping that aligned the artwork categories with their closest audio emotion equivalents. This required analyzing the emotional characteristics expressed by each category and designing a sensible cross-modal mapping scheme. This harmonization step lays the groundwork for embedding the two modalities within a shared emotional space, which is central to the objectives of the project. To help the group understand dataset balance and potential preprocessing challenges, I also generated multiple visualizations. These included distribution bar charts, comparative pie charts, and a cross-tabulation heatmap that illustrates how artwork emotions map onto audio labels. These visual tools made it easier for the team to see where imbalances or mismatches might occur, informing next week's strategies for sampling, augmentation, or normalization. Overall, my contribution this week focused heavily on building the technical pipelines needed for data preparation. I ensured that the datasets were properly validated, parsed, standardized, and analyzed for alignment potential. By structuring the data this way and sharing all outputs with the team, I helped establish a smooth transition into the preprocessing and feature extraction phase that follows. This week's work provides the operational foundation required for training a model capable of learning emotional resonance across modalities, bringing us closer to realizing the synesthetic framework envisioned in the early stages of the project.