Department of

Electrical & Electronics Engineering

Abdullah Gül University

BIOMEDICAL SYSTEM DESIGN CAPSULE

Lecture Free Week Activity Report

Exploring, Critiquing, and Critical Thinking with Project Datasets Using GenAI

Submitted on: 09.04.2025

Submitted by: Mehmet Fatih Göğüş

Group Mamber: Elif Nur Baysar – Davut Uysal

Grade: / 100

1. Objective

The aim of this study is to encourage us to use Generative AI tools to analyze project datasets in depth. During this process, we are expected to recognize the strengths and weaknesses of the datasets, identify potential research questions, and improve their critical thinking skills by detecting possible biases and hallucinations in the outputs of GenAI.

2. Background

Generative Artificial Intelligence is a term used to describe advanced artificial intelligence systems that create original content or data such as images, text, or audio by learning from existing examples. However, the outputs of GenAI systems may include biases and incorrect information depending on the datasets they are trained on and the structure of the models. Especially in tasks like classification, interpretation, and making suggestions, such errors can negatively affect decision-making processes. Therefore, it is important for people to develop a mindset that questions the accuracy and fairness of GenAI outputs. In this study, we will actively use GenAI tools, ChatGPT, while working with datasets and will analyze the strengths, weaknesses, and potential biases that emerge during the process, helping them build a more conscious approach to data science.

3. Activity Procedure

Strengths, Weaknesses, and Possible Research Questions of the Dataset Obtained from GenAI (The answer below given by the AI was taken exactly as it was for better comparison)

Summary

This dataset contains 7,023 brain MRI images categorized into four classes: Glioma Tumor, Meningioma Tumor, Pituitary Tumor, and No Tumor. The data is labeled by medical experts and is considered reliable for classification tasks. Each class has a balanced number of images, which helps in fair training of machine learning models. However, the images are in JPG format, which is not ideal for medical use due to compression. Also, there is no metadata (like patient age or image angle), which limits clinical interpretation. Despite these, the dataset is useful for building and testing models in medical image analysis, especially with proper preprocessing and augmentation.

1. Strengths and Weaknesses of the Dataset

<u>Strengths:</u>

1. *Balanced Classes:*
   Each of the four categories (Glioma, Meningioma, Pituitary Tumor, No Tumor) contains approximately the same number of images, which helps the model learn each class fairly. This reduces the need for additional preprocessing to handle class imbalance.

2. *Large and Rich Dataset:*
   With 7,023 MRI images, the dataset is sufficiently large—especially when supported by techniques like transfer learning or data augmentation.
3. *Clinical Validity:*
   The dataset is labeled by medical experts and considered clinically reliable, making it highly suitable for building models that can be used in real healthcare applications.
4. *Pre-divided into Training and Testing Sets:*
   Having the dataset already split into training and test sets enables more realistic model evaluation and supports reproducibility and comparative studies.
5. *Open License (CC BY 4.0):*
   The dataset can be freely used and published in academic research, encouraging open and collaborative work.

<u>Weaknesses</u>:

1. *Image Format – JPG:*
   Since JPG is a lossy compression format, the image quality may be reduced. Lossless formats such as DICOM or PNG are generally preferred for medical imaging.
2. *Lack of Metadata:*
   There is no additional patient information such as age, gender, tumor size, or location. This kind of metadata could improve the clinical interpretability of the models.

2. Possible Research Questions

<u>Basic Classification Question:</u>

- How does the performance of pre-trained models (e.g., ResNet, VGG, EfficientNet) vary when applied to this dataset through transfer learning?

<u>Data Processing and Modeling Questions:</u>

- How do data augmentation techniques affect model accuracy and generalization?
- Do image enhancement or segmentation preprocessing steps improve classification performance?

<u>Explainability and Clinical Integration:</u>

- Can methods like Grad-CAM be used to make the model's decision-making process interpretable? Which brain regions are most influential for each class?
- How can the developed models assist doctors in diagnosis? Can they be validated for real-world clinical applications?

Advanced Research Topics:

- Can models trained on this dataset perform well on other brain MRI datasets? (Generalization ability)
- Do ensemble methods (combining multiple models) yield a significant improvement in accuracy?

Our Dataset Evaluation

Strengths:

- There are a total of 7,023 brain MRI images in this dataset. This means that there are enough examples for the model to learn.
- Thanks to this diversity, the model can learn and distinguish different situations.
- Data is labeled by medical people, so it is categorized correctly. This shows that we are working with reliable data.

Weaknesses:

- It would be better if we had formats (such as PNG, DICOM), since they are preferred in medical images.
- MRI images are usually taken from three angles: axial (from above), sagittal (from the side), and coronal (from the front). This information is not provided.

Possible Research Questions for This Dataset

- Does reducing the size of the images or converting them to grayscale before giving them to the model affect the results?
- How does varying the train/test split ratio influence model success?
- How much does the model's success affect if we reduce the number of training data?

Ideas That Could Be Project Subjects:

- Can a lightweight model that can work in mobile applications be designed?
- Can the dataset be combined with another tumor dataset and a larger system be established?

3. Comparative Analysis of Student and GenAI Responses

We noticed that both our answer and the GenAI answer talk about the large size of the dataset (7,023 MRI images) and the fact that the images are labeled by medical experts. We both mention that JPG images may not be the best option, and that formats like PNG or DICOM are usually preferred in medical imaging.

However, the GenAI answer also highlights that the dataset has an open license (CC BY 4.0) and that it is already split into training and testing sets. It points out missing patient details (like age or gender), which could help with better clinical interpretation. We do not mention these points in our own response.

Instead, we focus on the absence of information about MRI angles (axial, sagittal, coronal). We also suggest different ways to modify the data, such as resizing or converting it to grayscale, and we consider changing the train/test split. The GenAI answer goes into more technical details, like using Grad-CAM for explainability or using ensemble methods.

In summary, we and the GenAI, ChatGPT, agree on the main strengths and weaknesses. GenAI provides more extensive technical points, while we focus on simpler suggestions and observations.

4. Finding examples of bias, inconsistency, or hallucination in GenAI Responses

We checked the GenAI, ChatGPT, response for three main issues: bias, inconsistency, and hallucination. First, we found that the AI response seems to favor deep learning approaches, especially pre-trained models like ResNet or EfficientNet. It doesn't mention any traditional machine learning methods or simpler models. This creates a kind of bias, as it assumes deep learning is always the best choice, which may not be true depending on the size and nature of the dataset. Second, there was a point about inconsistency. Although the text calls the dataset sufficient, 7,023 images might not always be enough for deep learning. Lastly, there was a possible case of hallucination. The text said the image planes (axial, coronal, sagittal) were not labeled, but without checking file names or folders, we cannot be sure if that is really true. This kind of statement might be incorrect if the writer did not fully examine the dataset.

5. Results and Discussion

In this study, we looked at the dataset with help from GenAI and also shared our own thoughts. Both we and the AI agreed that the dataset has some strong parts. It has 7,023 images, the classes are balanced, and the images are labeled by medical experts. These are helpful for training a machine learning model.

We also noticed some weak points. The images are in JPG format, which is not the best for medical use. We also don’t know which angle (axial, sagittal, or coronal) each image was taken from. This makes it harder to understand the full picture.

Our ideas were more focused on simple things, like changing image size or using grayscale. GenAI gave more advanced suggestions like using Grad-CAM or combining models.

When we checked the AI's answer, we found a few issues. It seemed to prefer deep learning too much and didn't mention simpler models. It also called the dataset "sufficient," but that might not always be true. Finally, it said the image angles weren't given, but we're not sure if that's 100% correct without checking the dataset.

6. Conclusion

In this activity, we learned how to use, correct, and understand GenAI to analyze a dataset and also made our own evaluation. We found both strengths and weaknesses in the dataset and suggested possible research questions. By comparing our ideas with GenAI's answers, we understood how AI can help, but also how important it is to think critically and check the information. This work helped us improve our analysis skills and see the value of combining human and AI perspectives in research.