Addressing Class Imbalance in Brain Tumor Detection Using GAN-Based Data Augmentation

Ihmaidan Waseem Alhaj 2220516

Department of Artificial Intelligence, The University of Jordan

Special Topics in Artificial Intelligence

Dr. yousef sanjalawe yousef

May,2025

Problem Statement

Medical image datasets often suffer from class imbalance, which adversely affects the performance of machine learning models. In the context of brain tumor classification, the number of images depicting tumors is significantly less than those without tumors. This imbalance leads to biased models that favor the majority class (no tumor), resulting in low recall and poor generalization. This project investigates the use of Generative Adversarial Networks (GANs) to generate synthetic tumor images, thereby mitigating imbalance and improving classification performance.

Description of Dataset & Imbalance Analysis

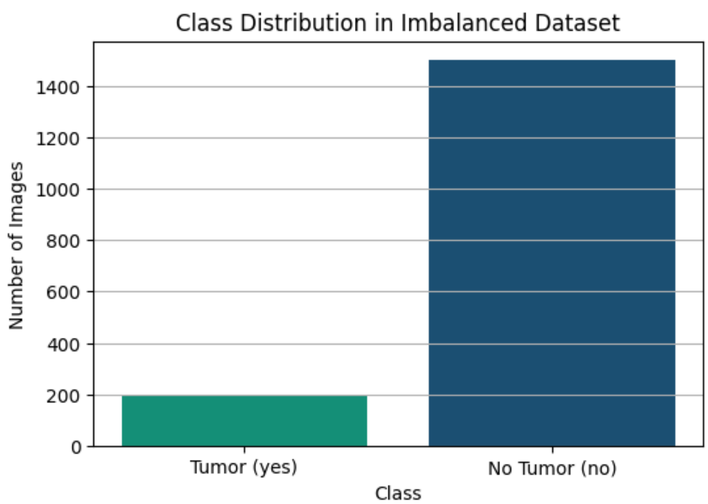
The dataset used contains two classes:

- **Tumor (yes)**: 200 images
- **No Tumor (no)**: 1500 images

This severe imbalance (1:7.5 ratio) skews the training process. A simple CNN trained on this dataset achieved high accuracy but poor recall for the tumor class, indicating that it often failed to detect tumors.

Imbalance Visualization:

A bar chart illustrates the class distribution:



GAN Architectures & Training Details

To balance the dataset, three GAN variants were trained solely on the minority class (tumor images) to generate synthetic tumor samples. The synthetic data was then used to augment the training dataset.

a. Vanilla GAN

- **Generator**: Fully connected layers (Linear → ReLU)
- **Discriminator**: Fully connected classifier with Sigmoid output

b. Deep Convolutional GAN (DCGAN)

- **Generator**: ConvTranspose2D layers with BatchNorm and ReLU
- **Discriminator**: Convolutional layers with LeakyReLU and Sigmoid

c. Wasserstein GAN (WGAN)

- **Generator**: Similar to DCGAN, optimized with Wasserstein loss
- **Critic (Discriminator)**: Uses no Sigmoid, with weight clipping and multiple training iterations

All networks were trained using the following parameters:

- **Noise vector (z_dim)**: 100
- **Image size**: 64x64 (grayscale)
- **Epochs**: 500
- **Optimizer**: Adam (or RMSprop for WGAN)
- **Batch size**: 64

Classifier Setup and Evaluation

After training the GANs, the classifiers were trained using:

- **CNN architecture** (Conv → ReLU → Pooling → FC layers)
- **Binary cross-entropy loss**
- **Evaluation metrics**: Accuracy, Precision, Recall, F1-Score, AUC-ROC

Each classifier was tested on a held-out validation set (20%).

Results & Comparisons

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
CNN (Imbalanced Dataset)	0.9471	0.8696	0.5714	0.6897	0.9642
CNN + Vanilla GAN	0.9617	0.9894	0.9331	0.9604	0.9941
CNN + DCGAN	0.9750	0.9895	0.9592	0.9741	0.9968
CNN + WGAN	0.9867	0.9899	0.9833	0.9866	0.9976

Observations and Conclusions

The experiment aimed to assess the impact of GAN-based data augmentation techniques on addressing severe class imbalance in a brain tumor classification dataset. The original dataset consisted of only 200 tumor images compared to 1500 no-tumor images, creating a significant imbalance that negatively affected the performance of traditional classifiers.

Baseline Model (CNN without GAN-generated data)

- The CNN trained directly on the imbalanced dataset achieved high accuracy (**94.71%**) and AUC-ROC (**0.9642**), but the **recall was alarmingly low (57.14%)**, indicating that the model failed to identify a large portion of actual tumor cases.
- This imbalance caused the model to be biased toward predicting the majority class (no tumor), which is unacceptable in medical diagnostics where missing a tumor detection can have severe consequences.

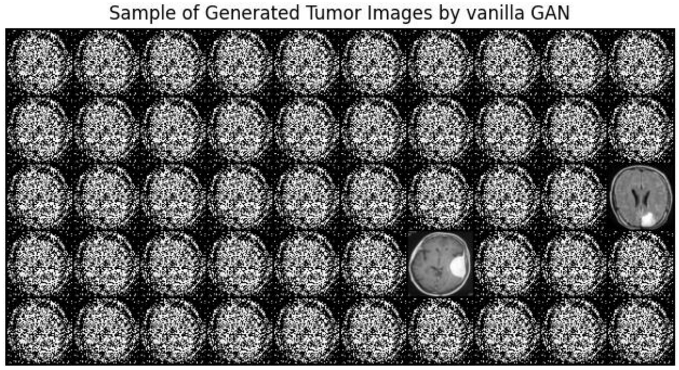
Impact of GAN-Generated Synthetic Data

To mitigate the class imbalance, synthetic tumor images were generated using three different GAN architectures: **Vanilla GAN**, **DCGAN**, and **WGAN**. Each GAN was trained exclusively on the minority class (tumor images) for **500 epochs** to ensure sufficient learning and diversity in the generated samples.

Vanilla GAN

- This fully connected GAN significantly improved the performance of the CNN classifier when synthetic tumor images were added.
- **Recall increased to 93.31%**, with an overall F1-Score of **96.04%**, indicating a much better balance in detecting both classes.
- However, due to its simple architecture, some of the generated images lacked fine-grained features and clarity.

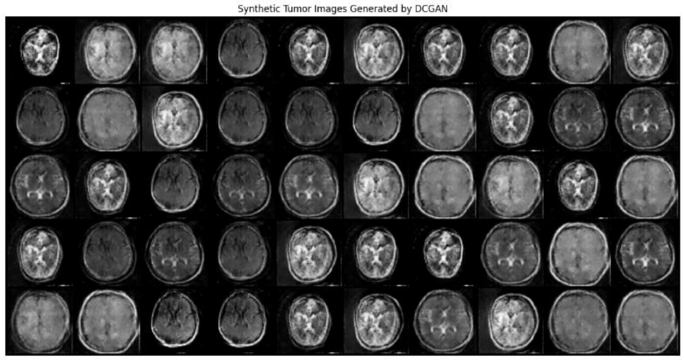
Sample of Generated Tumor Images by vanilla GAN



DCGAN

- The use of convolutional layers allowed DCGAN to generate more realistic and sharper tumor images.

Synthetic Tumor Images Generated by DCGAN

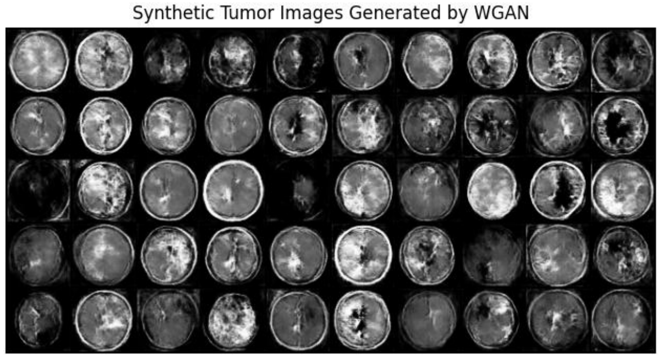


- The classifier trained with DCGAN-augmented data achieved **97.5% accuracy** and a **recall of 95.92%**, outperforming Vanilla GAN.
- The F1-Score reached **97.41%**, demonstrating a well-balanced performance with both high precision and recall.

Wasserstein GAN (WGAN)

- WGAN provided the most stable training due to its improved loss formulation and weight clipping in the critic.
- It produced the **highest quality synthetic tumor images**, characterized by better texture, edge definition, and diversity.

Synthetic Tumor Images Generated by WGAN



- The resulting classifier reached an **accuracy of 98.67%**, with an outstanding **recall of 98.33%** and F1-Score of **98.66%**. AUC-ROC was also the highest at **0.9976**, indicating excellent discriminative ability.

Conclusion

The results strongly demonstrate that **GAN-based data augmentation is a powerful approach** to address class imbalance in medical imaging tasks. While the baseline classifier struggled to detect tumor cases effectively, the introduction of synthetic samples generated from GANs—particularly **WGAN**—substantially improved all evaluation metrics, especially **recall**, which is critical in medical diagnosis.

Among the three GANs used:

- **WGAN** proved to be the most effective due to its ability to generate high-quality and diverse samples, contributing to a well-balanced and generalizable classifier.
- **DCGAN** also performed well and can be considered a strong alternative when training stability or computational constraints are a concern.
- **Vanilla GAN**, while still beneficial, had limitations due to its simplistic architecture.

Overall, this study confirms that **integrating advanced generative models like WGAN into the data preprocessing pipeline can significantly enhance classification performance in imbalanced medical datasets**, ultimately leading to safer and more reliable diagnostic tools.