Brain Tumor Classification: Model Evaluation Report

Dataset Overview:

- **Dataset structure**: Split into train, validation, and test folders.
- **Classes**:
    - glioma
    - meningioma
    - no_tumor
    - pituitary
- **Input Image Size**: 224x224 (resized for model compatibility)
- **Image Normalization**: Pixel values scaled to range [0, 1].

1. Model Comparison

Metric	Custom CNN	Fine-Tuned MobileNetV2
Accuracy	25%	**80%**
Macro F1	0.24	**0.78**
Precision	0.25 (avg)	**0.84** (macro avg)
Recall	0.24 (avg)	**0.79** (macro avg)

1.1 Class-wise Performance

Custom CNN (502 test images):

Class	Precision	Recall	F1-score
glioma	0.32	0.30	0.31
meningioma	0.29	0.15	0.19
no_tumor	0.20	0.20	0.20
pituitary	0.20	0.31	0.24

Fine-Tuned MobileNetV2 (246 test images):

Class	Precision	Recall	F1-score
glioma	0.72	0.99	0.84
meningioma	0.96	0.37	0.53
no_tumor	0.91	0.82	0.86

pituitary	0.78	1.00	0.88

1.2 Key Observations

- **Custom CNN severely underperforms**, likely due to:
    - Limited depth and capacity
    - Insufficient training data for generalization
- **Fine-Tuned MobileNetV2 significantly improves** across all metrics:
    - Especially effective for glioma, no_tumor, and pituitary
    - Struggles a bit on meningioma recall, but still outperforms the custom model

2. Project Summary

Objective:

To develop an automated brain tumor classification system using MRI scans categorized into four types: glioma, meningioma, pituitary, and no tumor.

Key Components:

2.1 Data Handling:

- Dataset loaded from local directory with structured folders.
- Preprocessing included:
    - Resizing to 224x224
    - Normalizing to [0,1]
- Augmentation applied on training data:
    - Rotation, flipping, shifting, brightness variation

2.2 Model Architectures:

- **Custom CNN**:
    - 3 Conv2D layers + MaxPooling + Dense
    - Moderate performance on training but overfits
- **MobileNetV2**:
    - Pretrained on ImageNet
    - Fine-tuned last 30 layers
    - Showed significant improvement in generalization

2.3 Evaluation Metrics:

- Accuracy, Precision, Recall, F1-score
- Confusion matrix for visualization
- Classification report from sklearn

2.4 Inference Pipeline:

- Model tested on unseen test images.
- Streamlit app built for interactive inference.
    - Allows user to upload image and see predicted tumor type
    - Also shows prediction probabilities as a bar chart

Deliverables:

- Trained and fine-tuned .h5 model
- Classification reports
- Confusion matrix visualizations
- Streamlit web app for deployment

3. Conclusion

The fine-tuned **MobileNetV2** model clearly outperformed the custom CNN in both accuracy and class-wise performance. It can be reliably used in a practical application like a diagnostic assistant for radiologists or healthcare professionals.

For production, further improvements may include:

- Adding Grad-CAM to visualize model attention
- Using a larger and more diverse dataset
- Ensemble of multiple pretrained models