Introduction: -

Brain tumour prediction is a crucial task in the field of medical imaging, as early and accurate detection can significantly improve patient outcomes. With the use of MRI scan images, this initiative seeks to categorize brain cancers into four groups: gliomas, meningiomas, pituitary tumours, and healthy instances. To do this, I have used Support Vector Machine (SVM) for classification and prediction and Principal Component Analysis (PCA) for extracting/capturing important features. SVM is a strong classifier that works well with high-dimensional data, whereas PCA works to reduce the dimensionality of the data while keeping important features.

Data Loading and Preprocessing: -

Firstly, I have loaded the MRI images from the dataset and pre-processed them for analysis. The dataset consisted of MRI scans stored in different folders, each representing a specific tumour category. In this step I have converted the scans to grayscale to simplify feature extraction, resized them to 64x64 pixels for consistency, and normalized to scale pixel values between 0 and 1. After that, I have flattened each image into a 1D array and had also encoded each category, so that machine learning models may be used more easily. After preprocessing the dataset was stored in NumPy arrays, with feature vectors representing the images and corresponding labels indicating the tumour category. The output displays 7023 photos in total, each having 4096 features (64x64 pixels flattened).

Applying PCA for Feature Extraction: -

Here I have created and selected 100 principal components, which captured the maximum variance of the dataset. PCA helps in improving computational efficiency and removing redundant features while retaining meaningful information from the original high-dimensional dataset. After applying PCA, I have used a cumulative explained variance plot to visualise the variance. This plot helped in determining how much variance was retained as the number of principal components keep ton increasing. The output displays that the first few components accounted for a significant percentage of the variance.

Visualizing Principal Components: -

To better understand the features captured by PCA, I had visualized the top five principal components as images. The images reveal the key structures in MRI scans that contribute most to variance in the dataset.

PC 1 captured the most dominant features, showing the overall structure of the brain.

PC 2 - PC 5 highlighted variations in contrast, shape, and texture that distinguish different tumour types from healthy brain images.

These visualizations display that PCA effectively captures meaningful structural patterns in MRI images.

OUTPUT-

Top Principal Components

PC 1   PC 2   PC 3   PC 4   PC 5
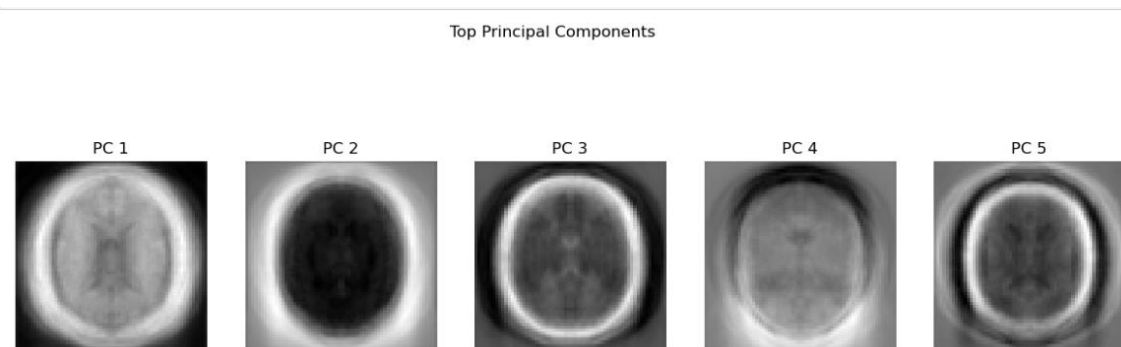
PCA Projection - Tumour Type Clustering: -

The scatter plot shows the dataset of MRI images projected onto the first two principal components. Each plot point represents a brain scan, and the colours represent the various types of tumours:

- Glioma (Red)
- Meningioma (Purple)
- Pituitary (Blue)
- Healthy (Green)

Since PCA reduces the original high-dimensional image data (4096 features) to just two principal components, this visualization allows us to see patterns and clusters that naturally form in the dataset.

Observations from the Graph:

The plot shows that the points were not randomly scattered but it showed some structure. The healthy cases (green) formed a distinct cluster, indicating that PCA has successfully captured important differences in brain scans. Tumour types (glioma, meningioma, and pituitary) overlapped slightly but still showed separability.
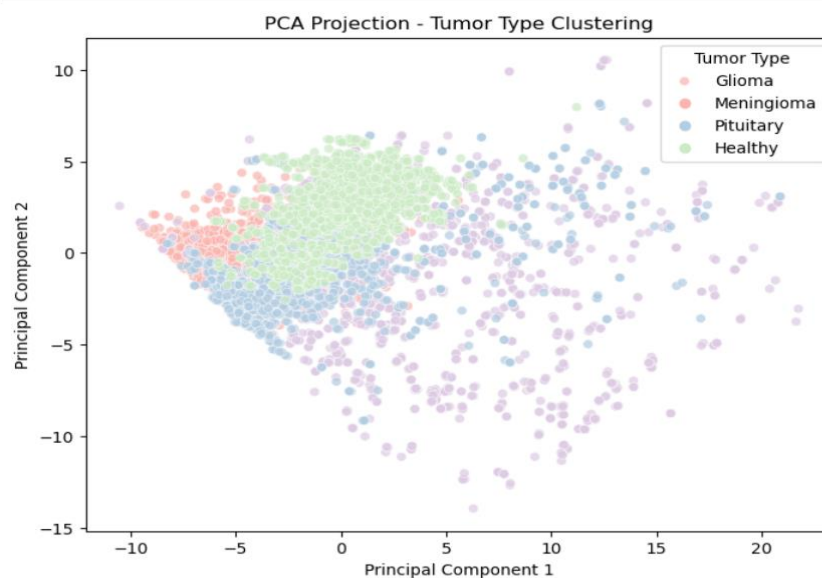
PCA Feature Visualization – Explanation: -
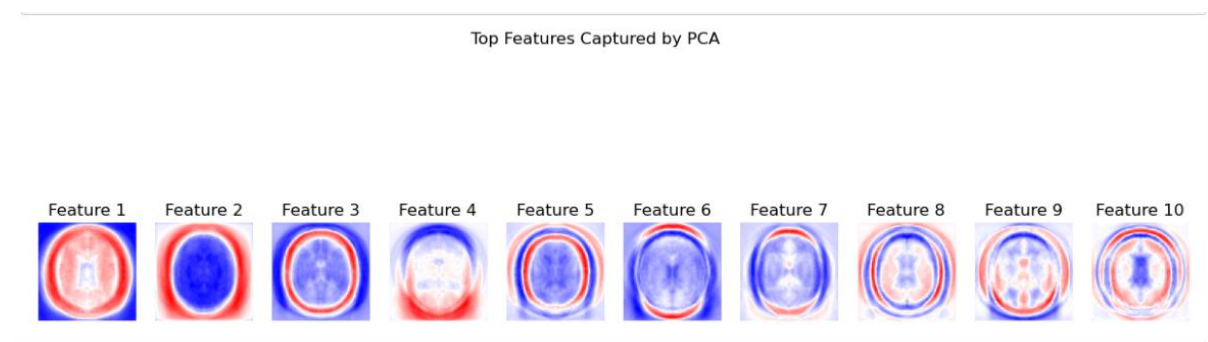
This visualization provides insights into the first 10 important features captured by PCA. Each principal component represents a weighted combination of original features, highlighting areas that contribute the most to variance in the dataset.

Understanding the Output:

- The images represent the first 10 principal components (features).

- Red regions indicate positive weights (strong positive correlation with variance).

- Blue regions indicate negative weights (strong negative correlation with variance).

- The contrast between red and blue regions shows which features are emphasized in each component.

- Principal components capture variations in shape, intensity, and texture patterns—useful for distinguishing tumour types.

Interpreting Key Features:

1. Feature 1: Strong focus on central brain structures, showing key differences in tumour vs. healthy brain images.

2. Feature 2: Highlights contrast around the edges, possibly distinguishing tumour boundaries.

3. Feature 3-5: Capture circular or elliptical patterns, potentially related to different tumour shapes.

4. Feature 6-10: Show finer details, including structural variations in different brain regions.

This visualization helps in understanding how PCA reduces dimensionality while preserving key structural patterns in the images.

SVM Classification Using PCA Features: -

In order to train an SVM classifier, I used PCA to extract significant features from MRI images. 80% of the dataset was used for training, and 20% was used for testing. For classification, an RBF kernel SVM (C=10, gamma=0.01) was employed. Patterns in PCA-reduced data were successfully learned by the model. The accuracy after testing was 94.95% . This illustrates how PCA can reduce dimensionality while still achieving good classification results.

SVM Classification Accuracy: 0.9494661921708185

Tumour Prediction on New MRI Scan: -

I had pre-processed a new MRI scan, applied PCA transformation, and predicted the image's class in order to evaluate the trained PCA + SVM model. The model accurately identified the scan as glioma. Prior to categorization, the grayscale MRI picture was scaled to 64 x 64, normalized, and PCA-transformed.

Prediction for New MRI Scan: Glioma

Predicted: Glioma