Distance Analysis and Dimensionality Reduction using PCA on Brain Tumour MRI Scans

Aniket Jhariya¹, Dhvani Parekh¹, Joshua Lobo¹, Anupkumar Bongale¹,*, Ruchi Jayaswal¹, Prachi Kadam¹, Shruti Patil¹, and Tanupriya Choudhury²

¹Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India
²Graphic Era Deemed to be University, Dehradun, 248002, Uttarakhand, India

Abstract

In this paper, Distance Analysis and Principal Component Analysis (PCA) are used to reduce the dimension of Brain Tumor MRI scans. The presented work aims to facilitate the differentiation between various tumor types while evaluating the impact of PCA on image representation modification. To accomplish this, PCA was applied to the MRI scans to determine the number of principal components. The distance analysis between several cancer classes, as well as between the original images and those that underwent PCA within a class is carried out. Visualization of both the most distant and the least distant images among and within the classes are plotted. The outcomes of the presented analysis revealed compelling results such as variance retention of 88.98% through PCA, signifying a substantial reduction in dimensions within the brain tumor scans. Furthermore, our evaluation showcased a lower Mean Squared Error (MSE) of 0.004, indicating improved performance, while the Peak Signal-to-Noise Ratio (PSNR) scored an impressive 28.24, denoting high fidelity in the reconstructed images. The achieved Structural Similarity Index (SSI) stood at a commendable 0.84, signifying strong similarity between the original and compressed images. Moreover, it is observed a disparity of 0.132, indicating superior alignment and emphasizing the efficacy of the PCA approach in retaining essential structural information. The significant emphasis on structural changes in distinct brain tumor types was evidenced by substantial alterations observed in the distance matrices. In conclusion, our utilization of PCA successfully reduced the dimensions of brain tumor scans while retaining critical information, enabling a clearer distinction between tumor types. The notable SSIM score of 0.84, the high PSNR score of 28.24, the low MSE of 0.004, the disparity of 0.132, and the substantial variance retention of 88.98% collectively validate the efficacy of PCA in enhancing image analysis and interpretation within the context of brain tumor classification.

Received on 28 December 2023; accepted on 27 March 2024; published on 04 April 2024

Keywords: Dimensionality Reduction, Brain MRI Scans, Principal Component Analysis, Distance Analysis, Correlation Heatmap, Imaging Variance

Copyright © 2024 A. Jhariya et al., licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetpht.10.5632

1. Introduction

Brain tumors can be diagnosed and monitored via medical imaging. Magnetic Resonance Imaging (MRI) is one of the most regularly used medical imaging procedures that deliver high-resolution pictures of the brain. Due to the quantity of data produced by the imaging process, MRI scan interpretation can be difficult. In addition, precise segmentation and identification of the tumor's position and size are necessary for the interpretation of brain tumor MRI images. Distance analysis and dimensionality reduction employing Principal Component Analysis (PCA) are frequently used methods for the analysis of medical images, in particular brain tumor MRI data. In MRI images, distinct regions of interest (ROIs) are compared or contrasted using distance analysis. The complexity of the data is reduced through dimensionality reduction.

*Corresponding author. Email: anupkumar.bongale@sitpune.edu.in
using PCA by projecting the high-dimensional MRI scans onto a lower-dimensional space while retaining the majority of the data. This allows for the detection of pertinent characteristics that are crucial for the appropriate diagnosis and monitoring of brain tumors. For the analysis of MRI scans for brain tumors, we will examine the application of distance analysis and dimensionality reduction utilizing PCA in this research issue. We will look into how these methods may be applied to find pertinent characteristics and trends in the data that help radiologists arrive at precisediagnoses and choose the best course of action. In general, this study issue intends to contribute to the development of advanced tools and methodologies for the interpretation of brain tumor MRI scans, which may ultimately result in improved patient outcomes. (see, for example, [1–3]).

1.1. Dataset Description

The dataset [1] is a collection of 3064 brain tumor images, taken from a sample of 233 people, with three types of cancers. The data is in the format of .mat files.

And the structure of each file goes as follows:

- Label - 1 for meningioma ; 2 for glioma ; 3 for pituitary tumor ; PID- patient ID ; image- the actual image data ; tumorBorder- a vector of the coordinates aligning the tumor area ; tumorMask- a binary image of the tumor.

2. Literature Survey

The field of MRI-based brain tumor detection and classification has undergone significant advancements in recent years, with researchers exploring various methodologies to enhance accuracy and efficiency. This literature survey provides a chronological overview of key contributions in this domain, presenting the evolution of techniques and insights.

MRI Segmentation with Rough-Fuzzy C-Means Segmentation is a crucial step in the analysis of brain tumor MRI scans, aiding in the identification and isolation of tumor cells [2]. Bal and colleagues introduced a segmentation method utilizing rough-fuzzy C-means and shape-based properties [3]. Published in the Journal of King Saud University - Computer and Information Sciences, their work focuses on precise tumor delineation, demonstrating promising results in accurate segmentation and analysis. The technique employs the spatial Fuzzy C-Means algorithm, a semi-automated and interactive approach [4]. This algorithm successfully segments the tumor region in MRI brain images, providing valuable information for diagnosis and treatment planning. Additionally, researchers have suggested the use of dimensionality reduction techniques, such as Principal Component Analysis (PCA), to further analyze and interpret brain tumor MRI scans [5]. Applying PCA to wavelet coefficients obtained from image decomposition reduces the dimensionality of the data [6], enabling a more concise representation of image features for improved analysis and classification of brain tumors. Overall, the combination of segmentation techniques like Rough-Fuzzy C-Means and dimensionality reduction methods like PCA plays a pivotal role in accurately analyzing and interpreting brain tumor MRI scans, facilitating enhanced diagnosis and treatment planning for patients.

Feature Selection and Extraction: In the context of brain tumor MRI scans, feature selection and extraction are vital steps in the classification process [7]. These steps involve identifying and extracting relevant features from MRI scans to distinguish between normal brain tissue and tumor tissue. Various techniques, including transform-based dimensionality reduction techniques like Principal Component Analysis and Independent Component Analysis [8], are employed for feature selection and extraction. These techniques analyze data to identify the most informative features contributing to the classification task. One commonly used technique for feature extraction in brain tumor MRI scans is the Discrete Wavelet Transform [9]. This technique allows for the extraction of important image features usable for classification purposes. Additionally, principal component analysis is widely adopted for feature reduction and dimensionality reduction, simplifying data representation while retaining distinguishing image characteristics. Rathi [10] conducted research on brain tumor MRI image classification, emphasizing the significance of feature selection and extraction using linear discriminant analysis. Published in the International Journal of Information Sciences and Techniques, the study underscores the importance of robust feature engineering for improved classification accuracy. Furthermore, K-means clustering is employed for image segmentation, providing a method to identify and delineate tumor objects in MRI brain images[11]. The utilization of K-means clustering, coupled with feature extraction techniques like Discrete Wavelet Transform (DWT) and dimensionality reduction methods such as Principal Component Analysis (PCA), facilitates accurate image segmentation. This process simplifies the isolation of tumor cells from the rest of the image, playing a crucial role in precisely identifying the tumor's location and boundaries for more accurate diagnosis and treatment planning.

Tumor region augmentation and partitioning are additional steps that can enhance the analysis of brain tumor MRI scans [12][13]. Tumor region augmentation involves improving the visibility and clarity of the tumor region in the MRI scan. Techniques such as contrast enhancement [14] and intensity adjustment [15] are employed to make the tumor region more distinguishable from the surrounding healthy brain.
tissue. Additionally, partitioning the tumor region is vital in the analysis of brain tumor MRI scans. In 2015, Cheng and collaborators [16] enhanced brain tumor classification by incorporating tumor region augmentation and partition techniques. Published in PLOS ONE, their research demonstrated notable improvements in classification accuracy, highlighting the effectiveness of data augmentation strategies. Techniques like K-means clustering for segmentation and cohesion-based merging aid in determining the exact location and boundaries of the tumor [17]. The process of partitioning the tumor region enables the identification of specific areas of abnormality within the MRI scan, contributing to accurate diagnosis and assessment of the tumor. A brain tumor segmentation approach employing a modified fuzzy metric-based strategy with adaptive techniques is presented in [18]. Featured in the International Journal of Advanced Trends in Computer Science and Engineering, this work emphasizes the importance of adaptability in achieving effective segmentation.

Another approach for brain tumor segmentation and analysis in MRI scans involves the use of biologically inspired techniques, such as the Biologically Inspired Binary Wavelet Transform and Support Vector Machine classification [19]. These techniques leverage the principles and algorithms inspired by biological systems to enhance the accuracy and efficiency of brain tumor segmentation. The Biologically Inspired Binary Wavelet Transform analyzes MRI scans by decomposing images into frequency subbands using wavelet transform. These subbands capture different levels of details and variations in MRI images, allowing for a comprehensive analysis of brain tumors. Inspired by the visual system in humans and animals, which decomposes visual information into different frequency channels, this approach assigns each subband a binary representation based on a thresholding technique [20]. The binary representation highlights relevant structures and features in MRI images, including tumor boundaries and textures. Support Vector Machine classification is then applied to the binary representations of wavelet subbands. This classification algorithm utilizes a trained model to categorize different regions in MRI scans as either tumor or non-tumor. Bahadure and team employed the Biologically Inspired BWT and SVM for image analysis in MRI-based brain tumor detection [21]. Published in the International Journal of Biomedical Imaging, their study explores innovative methodologies inspired by biological systems. The combination of the Biologically Inspired Binary Wavelet Transform and Support Vector Machine classification has shown promising results in accurate tumor segmentation [22].

This approach, based on the Biologically Inspired Binary Wavelet Transform and Support Vector Machine classification, not only accurately segments brain tumors but also allows precise identification of the tumor's location within brain tissue. The proposed methodology achieved good results with an accuracy of 96.51.

Furthermore, the application of advanced algorithms such as Support Vector Machine for classification plays a pivotal role in the accurate categorization of brain tumor MRI images [24][25]. These algorithms effectively classify images as normal or abnormal and further distinguish between benign and malignant tumors. Overall, the combination of techniques, including K-means clustering for segmentation, DWT for feature extraction, and PCA for dimensionality reduction, allows accurate analysis of brain tumor MRI scans [26]. This integrated approach aids in the identification, segmentation, and classification of brain tumors, facilitating more precise diagnosis and treatment planning for patients.

Unsupervised Anomaly Detection with AI has emerged as a powerful tool in various fields, including healthcare [27][28][29]. In the context of brain tumor MRI scans, unsupervised anomaly detection with AI can be used to identify and classify abnormal regions in images without the need for pre-labeled data. This approach involves training an AI model on a large dataset of normal brain MRI scans to learn the patterns and characteristics of healthy brain tissue. Once trained, the model can analyze new MRI scans and identify deviations or anomalies from the learned patterns, aiding in detecting and localizing brain tumors. The use of unsupervised anomaly detection with AI in brain tumor MRI scans has shown promising results in accurately identifying and classifying abnormal regions. In addition to traditional machine learning algorithms like support vector machines and wavelet transforms, researchers have explored unsupervised anomaly detection with AI methods such as autoencoders and generative adversarial networks for brain tumor segmentation in MRI scans [20]. Baur et al. [30] explored modeling healthy anatomy with artificial intelligence for unsupervised anomaly detection in brain MRI, presenting novel approaches for anomaly detection in Radiology: Artificial Intelligence.

The combination of unsupervised anomaly detection with AI methods, such as autoencoders [31] and generative adversarial networks, has shown great potential in accurately identifying and segmenting brain tumors in MRI scans.

Transfer Learning in Deep Learning for Brain Tumor Detection has also shown promise in improving the accuracy and efficiency of brain tumor detection in MRI scans [32]. By leveraging pre-trained models on large datasets, transfer learning allows the model to learn generic features applicable to specific tasks, reducing the need for extensive training on limited
data. Fine-tuning pre-trained models on brain tumor MRI scans enables the model to detect specific features and patterns relevant to brain tumors, resulting in improved accuracy and efficiency. Additionally, transfer learning in deep learning for brain tumor detection can overcome challenges of limited labeled data and address class imbalance in brain tumor detection.

PCA, or Principal Component Analysis, is a powerful technique for both distance analysis [33] and dimensionality reduction [34] in brain tumor MRI scans. It transforms high-dimensional data into a lower-dimensional space, capturing the most important information while reducing noise and redundancy. PCA aids in better understanding and interpretation of the data by identifying the most important features or components of brain tumor MRI scans. Alla and Athota [35] investigated brain tumor detection using transfer learning in deep learning, highlighting the efficacy of transfer learning in improving classification accuracy. Gokila Brindha and co-authors [36] focused on brain tumor detection from MRI images using deep learning techniques, contributing to the exploration of deep learning methodologies for enhanced detection.

Adaptive spatial pooling [37] is another technique proposed for the retrieval of brain tumors in MRI scans. This technique dynamically adjusts pooling regions based on the characteristics of brain tumor images, allowing more accurate and targeted retrieval of tumors. The use of unsupervised PCA and adaptive spatial pooling techniques enhances the detection, localization, and retrieval of brain tumors in MRI scans, ultimately improving accuracy and efficiency in diagnosis and treatment planning for patients with brain tumors. Cheng et al. [37] proposed a method for the retrieval of brain tumors using adaptive spatial pooling and Fisher Vector Representation, emphasizing the importance of adaptive spatial pooling in improving tumor retrieval. Cheng and collaborators [16] revisited their work on enhanced brain tumor classification through tumor region augmentation and partition, reinforcing the significance of their proposed techniques in the broader landscape of brain tumor research. The summarized survey is represented in Table 1.

3. Methodology

The flowchart figure 1 shows the methodology, as first the dataset was taken into the system as an input. The images are in the format of MAT files, so the image data was extracted into a directory, the information about the images was conflated into a separate CSV file and the tumor masks were extracted into another directory. The visualization of the dataset on the factors of ratios between tumor and brain, the types of tumors, and the size of tumor give insight into the structure and distribution of the dataset. The images were read into an array in the converted grayscale form, first into three subsets based on the type of cancer that the MRI shows, that is, Meningioma, Glioma, and Pituitary Tumor, taking 100 samples of each. The subset arrays are then flattened on the dimensions of images, 512 x 512, which gives a linear array of each image in an array of 100 points. The linear array of the Meningioma and the Glioma subsets are calculated for their corresponding distances and a heatmap was plotted for the distance matrix Figure 2 between them. A manual search for the approximate most distant and the least distant images between the classes and in the classes was visualized to give insights on the drawn inferences Figure 3.

The images can be analyzed for their visual distinction and similarity as visually seen. A similar was also achieved for the Glioma class taken as the original sample in the second row of images. Principal component analysis (PCA) dimensionality reduction was applied on each of the subsets, and the process was iterated over nine different values of components ranging from fifty to ten to visually analyze until which value of the component would an agreeable resolution would still hold. The images after each implementation of PCA was displayed in a gridded image Figure 4. And as can be seen, the images were well perceivable until the thirty component value mark,
Table 1. Literature Survey on MRI-Based Brain Tumor Detection and Classification

<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Methodology Used</th>
<th>Key Findings</th>
<th>Identified Research Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bal et al. (2022)</td>
<td>Rough-fuzzy C-means, Shape-based properties</td>
<td>Demonstrated promising results in accurate MRI brain tumor segmentation and classification</td>
<td>Further exploration needed to assess the generalizability of the proposed method to diverse MRI datasets</td>
</tr>
<tr>
<td>Rathi (2012)</td>
<td>Linear discriminant analysis, Feature selection and extraction</td>
<td>Emphasized robust feature engineering for improved brain tumor MRI image classification</td>
<td>Limited exploration of the impact of different feature selection/extraction methods on classification</td>
</tr>
<tr>
<td>Cheng et al. (2015)</td>
<td>Tumor region augmentation, Partition techniques</td>
<td>Showcased notable improvements in classification accuracy with data augmentation</td>
<td>Need for validation on larger datasets and exploration of real-world applicability of augmentation</td>
</tr>
<tr>
<td>S (2019)</td>
<td>Modified fuzzy metric, Adaptive techniques</td>
<td>Introduced an adaptive segmentation approach with improved fuzzy metrics</td>
<td>Evaluation of adaptability on varied tumor types and sizes, and comparison with other segmentation methods</td>
</tr>
<tr>
<td>Bahadure et al. (2017)</td>
<td>Biologically inspired BWT, SVM</td>
<td>Employed effective image analysis using biologically inspired methodologies</td>
<td>Exploration of the generalizability of biologically inspired techniques across different datasets</td>
</tr>
<tr>
<td>Soomro et al. (2023)</td>
<td>Review of image segmentation for MR brain tumor detection, Machine learning</td>
<td>Provided a comprehensive review offering insights into evolving segmentation methodologies</td>
<td>Need for more recent reviews incorporating the latest advancements in machine learning for segmentation</td>
</tr>
<tr>
<td>Baur et al. (2021)</td>
<td>Unsupervised anomaly detection with AI</td>
<td>Proposed novel approaches for unsupervised anomaly detection in brain MRI</td>
<td>Exploration of the interpretability of AI-driven anomaly detection and its impact on clinical workflows</td>
</tr>
<tr>
<td>Alla and Athota (2022)</td>
<td>Transfer learning in deep learning</td>
<td>Reported improved classification accuracy using transfer learning</td>
<td>Evaluation of the transferability of the approach to different datasets and acquisition conditions</td>
</tr>
<tr>
<td>Gokila Brindha et al. (2021)</td>
<td>Deep learning techniques</td>
<td>Contributed to the exploration of deep learning methodologies for enhanced brain tumor detection</td>
<td>Investigation of the scalability and resource requirements of deep learning models for large-scale implementation</td>
</tr>
<tr>
<td>Cheng et al. (2016)</td>
<td>Adaptive spatial pooling, Fisher Vector Representation</td>
<td>Demonstrated enhanced retrieval of brain tumors with adaptive spatial pooling</td>
<td>Investigation into the impact of adaptive spatial pooling on different tumor characteristics and imaging conditions</td>
</tr>
<tr>
<td>Cheng et al. (2015)</td>
<td>Tumor region augmentation, Partition techniques (Revisited)</td>
<td>Reinforced the importance of tumor region augmentation and partition for brain tumor classification</td>
<td>Further exploration needed on the long-term impact of these techniques on classification stability</td>
</tr>
</tbody>
</table>

after which a gradual degradation and noise could be seen in the images. PCA was implemented on each subset of the cancer classes, with component thirty. The PCA transform was then inverted back into, and reshaped into the dimension of the images to be able to visualize. A distance matrix was calculated between the classes of Meningioma and Glioma after the PCA had been implemented on each of those subsets. A heatmap was then plotted between those classes and another heatmap for the distance matrix between the original image subset and the PCA subset, particularly for the Meningioma class. Procrustes analysis was then performed between each of the separate distance matrices, that was the distance matrix between the original Meningioma class images and the original Glioma class images, and the distance matrix between the PCA implemented Meningioma class images and the PCA implemented Glioma class images. The sum squared difference or Disparity were then evaluated between the two distance matrices. The images from the complete dataset were then read into an array and the complete process of PCA implementation was repeated on the array. The resultant images Figure 7 were agreeable of a clear resolution with reduced dimensions.

4. Results

4.1. Procrustes Analysis

Procrustes analysis can be used to compare two distance matrices by minimising the difference between them. The goal is to find a transformation that minimises the sum of squared differences between the two distance matrices. Given by:

$$\min ||B - AR||_F : R^T R = I_n, \det(R) = 1$$  (1)
Figure 2. Heatmap depicting the distance matrix between the Meningioma and Glioma original images, highlighting their similarities and differences.

Figure 3. Classwise comparison of the reduced pictures following dimensionality reduction, illuminating the distinctive characteristics of Meningioma and Glioma.

Where,
‘B’ refers to the target matrix
‘A’ refers to the original matrix
‘R’ refers to the transformation matrix

\[ \|B - AR\|_F \] refers to the Frobenius norm of the matrix B - AR.

The Frobenius norm considers each individual element of a matrix, squares it, adds up all these squared values, and then takes the square root of the sum. In this context, it quantifies the error between the target matrix and the transformed original matrix, capturing its overall spread or extent.

The Procrustes transformation involves rotating, scaling, and translating one of the matrices to match the other matrix as closely as possible. This transformation is calculated using singular value decomposition (SVD) of the two matrices. The first constraint ensures that R is an orthogonal matrix which has the property that its
Distance Analysis and Dimensionality Reduction using PCA on Brain Tumour MRI Scans

4.2. Variance retained

Variance retained is an evaluation metric of PCA that measures the amount of variance in the original data that is preserved by the reduced set of principal components. It is calculated by dividing the total variance of the principal components retained by the total variance of the original data. Mathematically, the variance retained can be expressed as:

\[
\text{Variance Retained} = \frac{\text{Total Variance of Retained Principal Components}}{\text{Total Variance of Original Data}}
\]  

where,
- ‘Total variance of the retained principal components’ refers to the sum of the variances of each principal component that is retained.
- ‘Total variance of the original data’ refers to the sum of the variances of each variable in the original dataset.

The percentage of variance retained in the final PCA in the complete image dataset was 88.98.

4.3. Mean Squared Error

A popular evaluation metric in principal component analysis (PCA) is mean squared error (MSE), which calculates the difference between the original data and the rebuilt data using a smaller number of principal components. The equation is as follows:

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)
\]

where,
- ‘\(y_i\)’ refers to the original data
- ‘\(\hat{y}_i\)’ refers to the reconstructed data
- ‘\(n\)’ refers to the number of data points

A lower MSE indicates better performance of PCA in capturing the most important features of the data. The Mean Squared Error here was 0.004.

4.4. Peak Signal to Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) is a commonly used evaluation metric in image compression, restoration, and other related tasks, including PCA. It calculates the ratio of a signal’s maximal potential power to the noise level that compromises the accuracy of its representation. The equation is as follows:

\[
PSNR = 10 \cdot \log_{10} \left( \frac{r^2}{MSE} \right)
\]

where,
- ‘\(r\)’ refers to the largest variation in the type of input image data.
- ‘MSE’ is Mean Square Error

In PCA, PSNR is used to evaluate the quality of the reconstructed images after reducing their...
4.5. Structural similarity index

A metric called the structural similarity index (SSIM) is used to compare two photographs. It takes into account three components of an image: luminance, contrast, and structure. A score of one on the SSIM scale denotes complete resemblance between the two images. The value has a range from -1 to 1. In PCA, SSIM is used as an evaluation metric to compare the reconstructed value has a range from -1 to 1. In PCA, SSIM is used to compare two photographs. It takes into account three components of an image: luminance, contrast, and structure. A score of one on the SSIM scale denotes complete resemblance between the two images. The value has a range from -1 to 1.

\[
SSIM(a, b) = [l(a, b)]^\alpha \cdot [c(a, b)]^\beta \cdot [s(a, b)]^\gamma
\]

where,
nl(a,b) refers to the luminance function c(a,b) refers to the contrast function s(a,b) refers to the structure function α, β, and γ are the weights

A higher SSIM score indicates better reconstruction quality. The ideal SSIM value depends on the specific application and the level of similarity required between the two images. Generally, a score of 0.9 or higher is considered good, while a score below 0.6 indicates poor similarity. The SSIM achieved was 0.84.

5. Conclusion

The method mainly includes the steps of Data Preprocessing and Visualization, division of dataset on basis of types of cancer and creating a distance matrix and a heatmap, PCA implementation on complete dataset, Comparison of distance matrices (between Cancer class and between original and PCA applied images), Evaluation of the dimensionality reduction. The variance retained was 88.98 percent, Disparity was 0.132, MSE: 0.004, PSNR equalled 28.24, Smaller disparity obtained indicates better alignment. The low MSE error indicates a good performance of PCA in capturing important features. Hence our proposed method was successful in reducing the dimensions of the MRI scans with the help of PCA.

References


