Brain Tumour Feature Extraction Using Improved Principal Component Analysis

R. Sasikala[1]*, Dr. S. P. Swornambiga[2]

[1]*Assistant professor, Department of computer science, Sankara college of science and commerce, Tamilnadu, India.
[2]Associate professor, Department of computer application, sCMS college of science and commerce, Tamilnadu, India.

***Corresponding Author: -** R. Sasikala
*Assistant professor, Department of computer science, Sankara college of science and commerce, Tamilnadu, India.

Abstract – MRI (Magnetic Resonance Imaging) is a crucial medical imaging modality for diagnosing and characterizing brain tumors. Extracting informative features from MRI brain images is essential for accurate tumor analysis and decision-making. Traditional Principal Component Analysis (PCA) has been widely used for feature extraction, but its linear assumption limits its ability to capture the complex nonlinear structures present in MRI images. To overcome this limitation, an improved version of PCA, known as Kernel PCA or KPCA, has been introduced. This research investigates the application of KPCA for MRI brain image feature extraction, aiming to enhance the representation of intricate tumor patterns and structures.

Keywords: Magnetic Resonance Imaging, Brain imaging, Tumor detection, Feature extraction, Image analysis.

1. Introduction

MRI (Magnetic Resonance Imaging) is a widely used medical imaging modality that provides detailed structural and functional information of the human body. The analysis of MRI images plays a crucial role in medical diagnosis, treatment planning, and monitoring of various diseases. One important aspect of MRI analysis is feature extraction, which involves extracting meaningful and representative information from the images. MRI image feature extraction aims to capture relevant characteristics or patterns in the images, enabling subsequent analysis and interpretation by machine learning algorithms or domain experts.

The extraction of features from MRI images is a challenging task due to the complex and high-dimensional nature of the data. Traditional feature extraction methods, such as manual delineation or handcrafted feature design, are often time-consuming, subjective, and limited in capturing subtle variations and intricate structures present in the images. Therefore, there is a growing interest in developing automated and data-driven approaches that can effectively extract relevant features from MRI images. In recent years, advanced techniques, including machine learning and deep learning, have shown promising results in MRI image feature extraction. Machine learning algorithms, such as Support Vector Machines (SVM), Random Forests, or Convolutional Neural Networks (CNNs), can learn discriminative patterns and structures directly from the raw MRI image data. These algorithms can automatically extract relevant features by leveraging large datasets and powerful learning capabilities.

MRI image feature extraction is a critical step in the analysis of MRI data. Advanced techniques such as machine learning and deep learning, particularly CNNs, have shown great potential in automatically extracting relevant and discriminative features from MRI images. These features can enhance diagnostic accuracy, aid in treatment planning, and provide valuable insights into the underlying biological processes. With continued advancements in imaging technology and machine learning algorithms, MRI image feature extraction is expected to play an increasingly vital role in personalized medicine and precision healthcare.

Extracting meaningful features from MRI images plays a crucial role in automated analysis and decision-making processes. Traditional Principal Component Analysis (PCA) is a popular technique for feature extraction, but it assumes linearity in the data distribution. However, many real-world datasets, including MRI images, exhibit complex nonlinear structures. To address this limitation, an improved version of PCA, known as Kernel PCA or KPCA, has been introduced. KPCA leverages kernel functions to implicitly map the data into a higher-dimensional feature space, where linear PCA can be performed. This paper proposed Improved PCA for MRI image feature extraction, aiming to enhance the representation of complex patterns and structures in brain tumor images.

2. Literature Survey

2.1 Convolutional Neural Network (CNN)

Weiguang Wang (2020) et.al proposed convolutional neural network Combined with Image Feature Extraction in Brain Tumor Detection. CNNs combined with MRI detection technology to construct a model adapted to brain tumor feature detection. The main function of this research model is to segment and recognize MRI brain tumors and use convolutional layer to perform convolution operation to improve recognition efficiency and rate and combine artificially selected features with machine learning features. In addition, this article uses feature fusion to further improve the diagnostic results.

4.2 Principle Component Analysis (PCA)

Pronab Kumar Mondal (2020) et.al proposed Principle Component Analysis (PCA) for feature extraction. The statistical features are extracted from the MRI images by wavelet decomposition followed by PCA algorithm for dimensionality reduction. The extracted features of the training set of images constitute the training feature database of MRI image. This database is used when a test input image is given to classify the brain tumor into either benign or malignant. Brain MRI Images are used to diagnose such as Benign and Malignant based on the proposed supervised learning SVM classification algorithm.

4.3 Discrete Wavelet Transform (DWT)

Rinky B.P (2012) et.al proposed DWT Based Feature Extraction using Edge Tracked Scale Normalization for Enhanced face recognition. This paper presented a new approach for improved rate in face recognition systems. They proposed ETSN process which performed edge detection (as a preprocessing) along with the use of scale normalization to remove the background details. In their paper DWT is used for wavelet feature extraction followed by BPSO (Binary Particle Swarm Optimization) as a feature selection technique. As a classification technique they used Euclidean classifier. They found the efficient results by their proposed technique. In the end they recommended the SVM classifier and Gabor wavelet for better performance considerably.

4.4 Mammogram Image Features Extraction

Mane A.S, Kulhalli K.V. (2015) et.al proposed Mammogram Image Features Extraction and Classification for Breast Cancer Detection presented some techniques for feature extraction and classification of medical image using digital mammography which is most reliable way to detect the breast cancer. Initially they proposed some advanced preprocessing techniques that provide accuracy in mammographic images of breast cancer. Secondly to extract the optimum feature set, they suggested about the Gabor Wavelet function which is considered as a well suited method for feature extraction. According to them the best selection process of feature sets is crucial for having higher accuracy and reliable classification.

3. Research Methodology

MRI (Magnetic Resonance Imaging) is a broadly involved medical imaging technique for diagnosing different diseases, including brain tumors. Traditional Principal Component Analysis (PCA) is a famous method for feature extraction; however it expects linearity in the information dispersion. In any case, some genuine world datasets, including MRI images, display complex nonlinear structures. To address this impediment, an improved variant of PCA, known as Kernel PCA or KPCA, has been presented. Proposed Improved Standard Component Analysis (IPCA) for MRI picture feature extraction upgrades the portrayal of perplexing examples and structures in brain tumor images.

3.1 Improved PCA with Kernel Functions

Improved PCA extends the traditional PCA framework by integrating kernel functions, for example, the radial basis function (RBF) or polynomial kernel. These kernel functions measure the likeness or distance between sets of data points in the original feature space, considering catching nonlinear connections. The top-k eigenvectors, which catch the main varieties in the data, are chosen as the principal components for dimensionality reduction.

3.1.1 Advantages of KPCA for MRI Feature Extraction

Using KPCA for MRI image feature extraction offers a few benefits. It, first and foremost, takes into consideration catching complex nonlinear examples and structures that might be missed by linear techniques like traditional PCA. This is especially advantageous for brain tumor analysis, where the presence and location of tumors can shift fundamentally across patients. The extracted features acquired from KPCA can be utilized for different applications, including tumor classification, segmentation, visualization, or anomaly detection, adding to the improvement of automated diagnostic devices and personalized treatment systems.

Kernel Function Selection is a basic move toward Kernel Principal Component Analysis (KPCA). The kernel function computes the closeness or distance between sets of data points in the original feature space, empowering KPCA to catch nonlinear connections that PCA can't deal with. Pick a kernel function, for example,

Radial basis function (RBF) kernel, denoted as K(x, x'), or

Polynomial kernel, denoted as $K(x, x') = (x^T x' + c)^d$

Where x and x' are data points, and c and d are kernel parameters.

Subsequent to choosing the kernel function, the Kernel Matrix Calculation step follows. Every section in the kernel matrix addresses the likeness between two data points in light of the picked kernel function. This matrix shapes the basis for ensuing calculations in KPCA. Compute the kernel matrix, signified as K, where every section K_{ij} represents the similarity (or distance) between data points x_i and x_j based on the chosen kernel function:

$$K_{ij} = K(x_i, x_j)$$

To ensure compelling feature extraction, the Focusing the Kernel Matrix step is performed. The mean of every column (or row) in the kernel matrix is deducted, focusing the matrix.

Center the kernel matrix by subtracting the mean of each column (or row):

$$K_c = K - 1/n * 1 * K - 1/n * K * 1 + 1/n * 1 * K * 1$$

Where 1 is an n × n matrix of ones, and n is the number of data points.
Eigenvalue Decomposition is then performed on the centered kernel matrix. This step includes computing the eigenvectors and eigenvalues of the matrix.
Perform eigenvalue decomposition on the centered kernel matrix

$$K_c = U * \Lambda * U^T$$

Where U contains the eigenvectors as columns, and Λ is a diagonal matrix of eigenvalues.
Then, in Choosing Principal Components, the top-k eigenvectors with the biggest eigenvalues are picked as the principal components. These principal components catch the most significant data in the data and act as the basis for subsequent analysis.
Choose the top-k eigenvectors corresponding to the largest eigenvalues as the principal components:

$$U_k = [u_1, u_2, \dots, u_k]$$

Where u_i represents $i - th$ eigenvector
In the Projection and Reconstruction step, the original data is projected onto the chose principal components, bringing about a lower-dimensional representation. This considers reconstruction of the original data points from the decreased feature space representation.
Project the original data onto the selected principal components to obtain the lower-dimensional representation:

$$z_i = [K_c(x_i, u_1), K_c(x_i, u_2), \dots, K_c(x_i, u_k)]$$

Where z_i represents the reduced-dimensional representation of data point x_i
To reconstruct the data, project the lower-dimensional representation back into the original feature space using the eigenvectors and eigenvalues:

$$x_i' = \Sigma_j (z_i)_j * u_j$$

Where $(z_i)_j$ represents the j-th element of the reduced-dimensional representation z_i.
At last, the extracted features got from KPCA have various applications. They can be used for errands like classification, clustering, visualization, or anomaly detection. The nonlinear planning given by the kernel function empowers KPCA to catch complex examples and structures in the data, making it an important device for extracting meaningful features from high-dimensional datasets like MRI images.
The extracted features obtained from KPCA, represented by z_i

Input: *MRI brain tumor segmented image dataset*
Parameters: *Kernel function, Number of principal components (k)*

Step 1: *Initialize an empty kernel matrix K.*
Step 2: For each pair of data points (x_i, x_j) *in the dataset:*
Step 3: Compute the kernel value $K(x_i, x_j)$ *based on the chosen kernel function.*
Step 4: Set $K[i, j] = K(x_i, x_j)$.

Compute the Kernel Matrix

Step 5: Compute the mean vector mu of each column (or row) in the kernel matrix K.
Step 6: Subtract mu from each column (or row) of the kernel matrix K.

Center the Kernel Matrix

Step 7: Compute the eigenvectors U and eigenvalues lambda of the centered kernel matrix K.
Step 8: Sort the eigenvectors in descending order based on their corresponding eigenvalues.

Eigen value Decomposition

Step 9: Select the first k eigenvectors from the eigenvector matrix U, corresponding to the largest eigenvalues.

Select Principal Component

Step 10: Create a projection matrix P using the selected eigenvectors.
Step 11: Initialize an empty feature matrix X.
Step 12: For each segmented image x in the dataset:
Step 13: Compute the kernel values between x and each sample based on the chosen kernel function.
Step 14: Multiply the kernel values by the corresponding eigenvectors from P and sum them up to obtain the projected vector z.
Step 15: Append z to the feature matrix X.
Step 16: Evaluate the performance of the trained model using appropriate evaluation metrics for the specific task.
Output: *Extracted MRI brain tumor image features using KPCA as Improved PCA.*

4. Experimental Results

4.1 Precision

Images	PCA	DWT	Proposed IPCA
Image 1	0.85	0.93	0.93
Image 2	0.92	0.88	0.90
Image 3	0.78	0.85	0.87
Image 4	0.89	0.86	0.91
Image 5	0.91	0.87	0.94

Table 1. Comparison table of Precision

The Comparison table 1 of Precision values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). The existing algorithm values start from 0.78 to 0.92, 0.85 to 0.93 and proposed IPCA values starts from 0.87 to 0.94. The proposed method provides the great results.

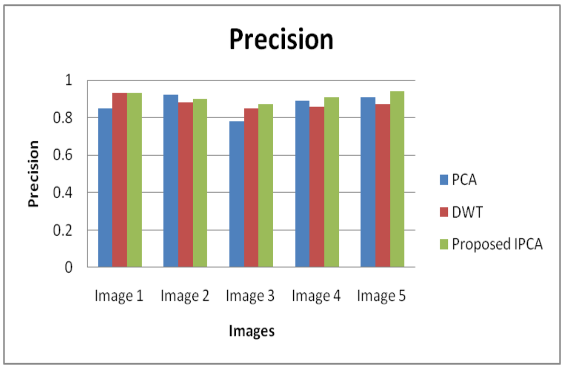

Figure 1. Comparison chart of Precision

The Figure 1 of Precision values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). X axis denote the Dataset and y axis denotes the Precision ratio. The existing algorithm values start from 0.78 to 0.92, 0.85 to 0.93 and proposed IPCA values starts from 0.87 to 0.94. The proposed method provides the great results.

4.2 Recall

Images	PCA	DWT	Proposed IPCA
Image 1	0.78	0.86	0.89
Image 2	0.86	0.82	0.89
Image 3	0.75	0.82	0.84
Image 4	0.86	0.84	0.89
Image 5	0.87	0.85	0.92

Table 2. Comparison table of Recall

The Comparison table 2 of Recall values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). The existing algorithm values start from 0.78 to 0.87, 0.82 to 0.86 and proposed IPCA values starts from 0.84 to 0.92. The proposed method provides the great results.

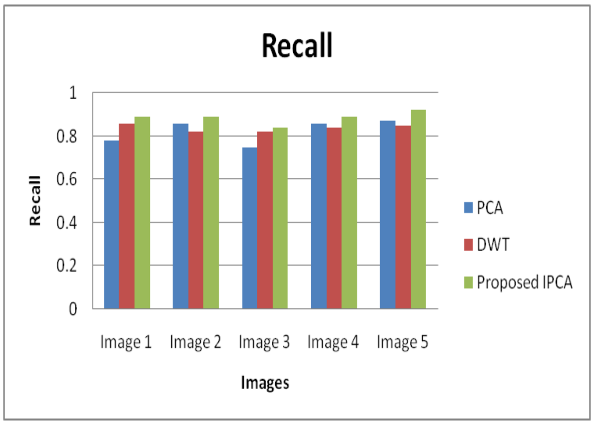

Figure 2. Comparison chart of Recall

The Figure 2 of Recall values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). X axis denote the Dataset and y axis denotes the Recall ratio. The existing algorithm values start from 0.78 to 0.87, 0.82 to 0.86 and proposed IPCA values starts from 0.84 to 0.92. The proposed method provides the great results.

4.3 F - Measure

Images	PCA	DWT	Proposed IPCA
Image 1	0.88	0.92	0.93
Image 2	0.90	0.88	0.91
Image 3	0.76	0.83	0.87
Image 4	0.88	0.85	0.91
Image 5	0.90	0.86	0.93

Table 3. Comparison table of F-Measure

The Comparison table 3 of F-Measure values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). The existing algorithm values start from 0.76 to 0.90, 0.83 to 0.92 and proposed IPCA values starts from 0.87 to 0.93. The proposed method provides the great results.

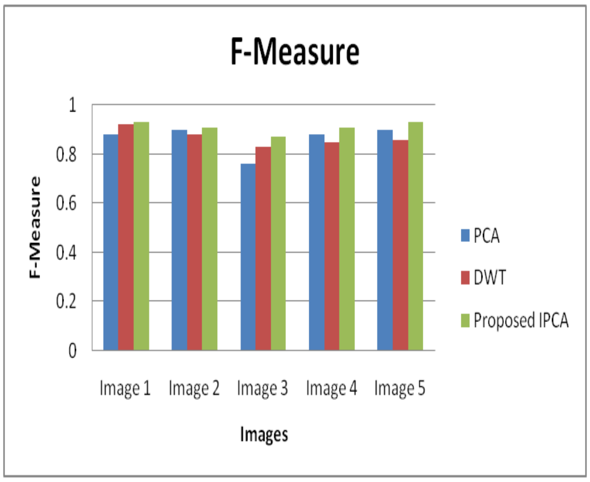

Figure 3.Comparison chart of F-Measure

The Figure 3 of F-Measure values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). X axis denote the Dataset and y axis denotes the F-Measure ratio. The existing algorithm values start from 0.78 to 0.92, 0.85 to 0.93 and proposed IPCA values starts from 0.87 to 0.94. The proposed method provides the great results.

4.4 Accuracy

Images	PCA	DWT	Proposed IPCA
Image 1	68	73	89
Image 2	70	70	90
Image 3	75	66	91
Image 4	80	69	94
Image 5	87	64	98

Table 4. Comparison table of Accuracy

The Comparison table 4 of Accuracy values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). The existing algorithm values start from 68 to 87, 64 to 73 and proposed IPCA values starts from 89 to 98. The proposed method provides the great results.

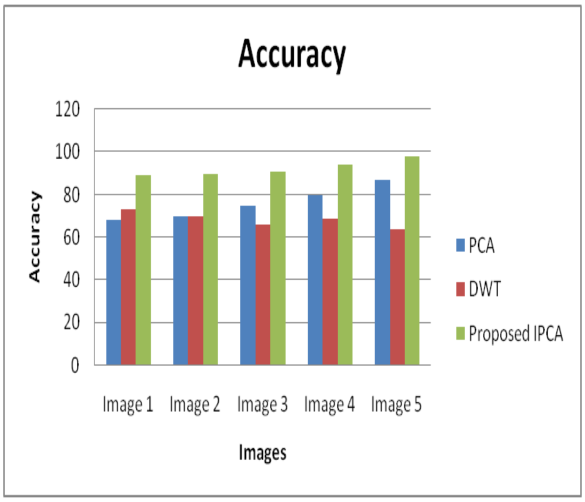

Figure 4.Comparison chart of Accuracy

The Figure 4 of Accuracy values for three different feature extraction methods: PCA, DWT (Discrete Wavelet Transform), and Proposed IPCA (Improved Principal Component Analysis). X axis denote the Dataset and y axis denotes the Accuracy ratio. The existing algorithm values start from 68 to 87, 64 to 73 and proposed IPCA values starts from 89 to 98. The proposed method provides the great results.

5. Conclusion

In this paper we proposed Improved Principal Component Analysis (IPCA), specifically Kernel PCA (KPCA), for MRI brain image feature extraction holds extraordinary commitment in propelling the field of medical imaging analysis, especially with regards to brain tumor analysis. The use of improved PCA, specifically KPCA, for MRI brain image feature extraction exhibits the potential for advancements in medical imaging analysis. By successfully catching the complex examples and structures intrinsic in MRI brain images, KPCA adds to the improvement of automated diagnostic devices, personalized treatment procedures, and a more profound understanding of brain tumor qualities. Preceded with examination and exploration of KPCA in this space without doubt prompt further headways in the field of medical imaging and ultimately work on patient care.

References

1. Weiguang Wang , Fanlong Bu , Ziyi Lin , And Shuangqing Zhai (2020), “Convolutional neural network Combined With Image Feature Extraction in Brain Tumor Detection”, https://creativecommons.org/licenses/by/4.0/, IEEE.
2. Pronab Kumar Mondal , Syed Zenith Rhyhan (2020), “Brain Tumor Detection, Classification and Feature Extraction from MRI Brain Image”, International Journal Of Engineering Research And Development e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 16, Issue 2 (Feb. 2020), PP. 09-18.
3. Siva Sankari. S, Sindhu. M, Sangeetha. R, Shenbaga Rajan. A, “Feature Extraction of Brain Tumor Using MRI”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 3, March 2014, ISSN: 2319-8753.
4. Nag, Sanjay & Maitra, Dr. Indra & Sudipta, Maitra& Roy, Sudipta & Bandyopadhyay, Kumar (2014),“A Review of Image Segmentation Methods On Brain MRI For Detection Of Tumor And Related Abnormalities” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 5, May 2014, ISSN: 2277 128X.
5. Rajeshwari, S. &Sharmila, T (2013). Efficient quality analysis of MRI image using preprocessing techniques”, IEEE Conference on Information and Communication Technologies, ICT 2013. 391-396. 10.1109/CICT.2013.6558127.
6. ChengLiu, WeibinLiu, WeiweiXing, “A weighted edge-based level set method based on multi-local statistical information for noisy image segmentation”, Journal of Visual Communication and Image Representation, Volume 59, February 2019, Pages 89-107.
7. Intedhar Shakir Nasir, “The proposed image segmentation method based on adaptive k-means algorithm”, Journal of Theoretical and Applied Information Technology,15th November 2018. Vol.96. No 21.
8. Rinky B.P, Mondal P, K Manikantan S Ramachandran. “DWT Based Feature Extraction using Edge Tracked Scale Normalization for Enhanced face recognition”. International Conference on Communication, Computing & Security. Procedia Technology 6(2012), 344-353. Doi: 10.1016| j.protcy.2012.10.041.
9. Mane A.S, Kulhalli K.V. “Mammogram Image Features Extraction and Classification for Breast Cancer Detection”. International Research Journal of Engineering and Technology (IRJET). 2015. Vol2, Issue-7, ISSN: 2395-0056.
10. Eqbal S, Ansari A.M. “Medical Image Feature Extraction for Computer Aided Diagnosis of Lung Cancer”. International Journal of Advanced Research in Computer Science and Software Engineering. 2015. Vol-5, Issue 6, ISSN: 2277- 128X.