**Marwa Madni**

**BSCIS 18-22**

**IML**

**Project Report**

**First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images Using Support Vector Machines**

**First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images Using Support Vector Machines**

**ABSTRACT**

The project presents an approach for automated diagnosis, based on classification of T1-weighted Magnetic Resonance (MR) human brain images. I investigated performance of texture-based features with support vector machine and K-nearest Neighbour Algorithms for the classification of MRI disease based on T1-weighted MRI brain image. MRI images of size 256x256 have been used in the project. The number of features extracted using first and second order statistics are 14. The performance is evaluated in terms of confusion matrix and testing accuracies. An intelligent classification rate of 82% could be achieved using the support vector machine.

**Keywords:** Brain Tumour; Magnetic Resonance Imaging; Feature Extraction; First and Second Order

Statistical Features, GLCM

1. **Introduction**

A brain tumor is a collection, or mass, of abnormal cells in your brain. Your skull, which encloses your brain, is very rigid. Any growth inside such a restricted space can cause problems. Brain tumors can be cancerous (malignant) or noncancerous (benign). Brain tumors are categorized as primary or secondary. Primary brain tumors originate in your brain. They can develop from your:

* brain cells
* the membranes that surround your brain, which are called meninges
* nerve cells
* glands

Primary tumors can be benign or cancerous. In adults, the most common types of brain tumors are gliomas and meningiomas.

Among various imaging modalities, Magnetic Resonance Imaging (MRI) is most preferred as it is non-invasive technique with no side effects of rays and suitable for the internal study of human brain which provide better information about soft tissue anatomy. However, there is a huge MRI repository, which makes the task of manual interpretation difficult. Hence, computer aided analysis and diagnosis of MRI brain images have become an important area of research in recent years.

For proper analysis of MRI images, it is essential to extract a set of discriminative features which provide better classification of MRI images. In literature, various feature extraction methods have been proposed such as Independent Component Analysis, Fourier Transform, Wavelet Transform, and Texture based features.

Features based on statistics of texture gives less number of relevant, non-redundant, interpretable and distinguishable features. Motivated by this, in our proposed method, we use First and Second Order Statistics for feature extraction. In this report, I have investigated performance of First and Second order-based features. Since, the classification accuracy of a decision system also depends on the choice of a classifier. I have used most commonly and widely used classifiers for the classification of MRI brain images. The performance is evaluated in terms of confusion matrix and accuracy.

The paper is organized as: Section 2 explains Dataset used for classification 2. An overview of related SVM implementation is presented in Section 3 and the brain tumor classification and its evaluation are presented in Section 4. Finally, Section 5 concludes the paper, with some possible future directions.

1. **Dataset**

This brain tumor dataset used for project contains 3064 T1-weighted contrast-inhanced images from 233 patients with three kinds of brain tumor: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices).

This data was organized in matlab data format (.mat file). Each file stores a struct containing the following fields for an image:

**cjdata.label:** 1 for meningioma, 2 for glioma, 3 for pituitary tumor

**cjdata.PID:** patient ID

**cjdata.image:** image data

**cjdata.tumorBorder**: a vector storing the coordinates of discrete points on tumor border.

**cjdata.tumorMask:** a binary image with 1s indicating tumor region

1. **Preprocessing:**

Pre-processing is the name used for operations on images at the lowest level of abstraction. In this paper the pre-processing includes:

* Image filtering through median filter
* Image resizing from (512x512) to (256x256)
* Image conversion to uint8 type

1. **Feature detection and extraction**

The texture of an image region is determined by the way the gray levels are distributed over the pixels in the region. Although there is no clear definition of “texture” in literature, often it describes an image looks by fine or coarse, smooth or irregular, homogeneous or inhomogeneous etc. The features are described to quantify properties of an image region by exploiting space relations underlying the gray-level distribution of a given image.

* 1. **First-Order Statistics**

The most frequently used first-order statistics are Variance, Skewness and Kurtosis. The Variance is a measure of the histogram width that measures the deviation of gray levels from the Mean. Skewness is a measure of the degree of histogram asymmetry around the Mean and Kurtosis is a measure of the histogram sharpness.

* 1. **Second-Order Statistics**

The features generated from the first-order statistics provide information related to the gray-level distribution of the image. An occurrence of some gray-level configuration can be described by a matrix of relative frequencies Pθ,d(I1, I2). It describes how frequently two pixels with gray-levels I1, I2 appear in the window separated by a distance d in direction θ. The information can be extracted from the co-occurrence matrix that measures second-order image statistics , where the pixels are considered in pairs.

The co-occurrence matrix is a function of two parameters: relative distance measured in pixel numbers (d) and their relative orientation θ. The orientation θ is quantized in four directions that represent horizontal, diagonal, vertical and anti-diagonal by 0˚, 45˚, 90˚ and 135˚ respectively.

Using Co-occurrence matrix, features can be defined which quantifies coarseness, smoothness and texture— related information that have high discriminatory power. Among them, Angular Second Moment (ASM), Contrast, Correlation, Homogeneity and Entropy are few such measures.

All features are functions of the distance d and the orientation θ. Thus, if an image is rotated, the values of the features will be different. In practice, for each d the resulting values for the four directions are averaged out. This will generate features that will be rotations invariant.

4. **Experimental Setup and Results**

In this section, we investigate different combination of feature extraction methods and classifiers for the classification of two different types of MRI images i.e. Meningiom tumor MRI and Glioma Tumore MRI. The feature extraction methods under investigation is “Features based on First and second order statistics (FSStat)” .

We will explore the classifiers

* (SVM with linear (SVM-L),
* Polynomial kernel (SVM-P)
* Radial kernel (SVM-R))
* (K-nearest neighbor (KNN)

The polynomial kernel of SVM is used with degrees 2 & 3.

Textural features of an image are represented in terms of four first order statistics (Mean, Variance, Skewness, Kurtosis) and five-second order statistics (Angular second moment, Contrast, Correlation, Homogeneity, Entropy). Since, second order statistics are functions of the distance d and the orientation θ , hence, for each second order measure, the mean and range of the resulting values from the four directions are calculated. Thus, the number of features extracted using first and second order statistics are 14

To evaluate the performance, we have considered medical images from figshare. All MRI images are T2-weighted of 256 × 256 size. For our study, we have considered a total of 2134 image slices (1426 belonging to glioma tumor and 708 belonging to meningioma tumor).

In literature, various performance measures have been suggested to evaluate the learning models. Among them, I have used 1). confusion matrix and 2). accuracy.

The dataset was arbitrarily divided into a training set with 80% samples and a test set of 20% samples. The feature matrix was standardized using min-max algorithm and then the models were trained and tested.

1. **Results**

|  |  |  |
| --- | --- | --- |
| **Classifier** | **Accuracy** | **Confusion Matrix** |
| **Linear SVM** | **76.58** | **[[119 13]**  **[87 208]]** |
| **RBF SVM** | **76.81** | **[[117 15]**  **[84 211]]** |
| **2degrees Poly SVM** | **78.68** | **[[117 15]**  **[76 219]]** |
| **3degrees Poly SVM** | **80.32** | **[[118 14]**  **[70 225]]** |
| **KNN(1)** | **84.1** | **[[119 13]**  **[87 208]]** |
| **KNN(3)** | **81.03** | **[[119 13]**  **87 208]** |
| **KNN(5)** | **81.03** | **[[119 13]**  **[87 208]]** |

1. **Conclusion**

In this paper, we investigated features based on First and Second Order Statistics (FSStat) for classification of MRI images.

Since, the classification accuracy of a pattern recognition system not only depends on features extraction method but also on the choice of classifier. Hence, I investigated performance of FSStat based features with commonly used classifiers for the classification of MRI brain images. The performance is evaluated in terms of accuracy and confusion matrix.

It is found that FSStat features’ takes very less time for training and testing. This is because First and Second Order Statistics gives less number of relevant and distinguishable features and does not involve in computational intensive transformation in comparison to method proposed in literature.

In future, the performance of our proposed approach can be evaluated on other disease MRI images to evaluate its efficacy. We can also explore some feature extraction/construction techniques which provide invariant and minimal number of relevant features to distinguish two or more different kinds of MRI.

1. **Questionnaire**

1**. Presentation was accomplished**

Yes

2. **Used GLCM : how did you find which five features are top ranking?**

I found GLCM 5 top ranking features from the paper “ Namita Aggarwal, R. K. Agrawal, First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images <http://dx.doi.org/10.4236/jsip.2012.32019>

3**. Use images instead of ground truth: Major fault in the work found**

I had used tumor masks for feature extraction, that was a fault. I have now made this report while using images of brain tumors from datset.

**4. Share the link of dataset**

<https://figshare.com/articles/dataset/brain_tumor_dataset/1512427?file=7953679>

5**. K NN always uses k as 1 3 5 7 etc. to avoid a tie in voting during classification**

As increasing the number of neighbours decreased accuracy, I have now used odd number of neighbours in KNN where n<10.

6**. How class imbalance was handled?**

The class imbalance was handled through SMOTE algorithm. I used 80% data for training and 20% for testing. After applying the SMOTE, the number of instances of both classes for training were 1135 each

7.  **Use data augmentation techniques.**

According to your suggestion, I have skipped applying augmentation technique as I have used SMOTE already and the data is sufficient.

8**. The use of images is mandatory**

Yes, I have used the images

9. **Share the complete code of your project**

Attached under the file name “IML Project(1).ipynb”.

10.**Send the copy of presentation slides as ppt file**

Attached under the file name “Project Presentation”.