Machine learning approach for brain tumor detection

Conference Paper · April 2012

DOI: 10.1145/2222444.2222467

4 authors:

Mahmoud Al-Ayyoub
Jordan University of Science and Technology

Ghaith Husari
University of North Carolina at Charlotte

Omar Darwish
Western Michigan University

Ahmad Alabed
Universiti Sultan Zainal Abidin | UniSZA

Machine Learning Approach for Brain Tumor Detection

Mahmoud Al-Ayyoub
Jordan University Of Science and Technology
Irbid, Jordan
maalshbool@just.edu.jo

Ghaith Husari
Jordan University Of Science and Technology
Irbid, Jordan
ghusari@hotmail.com

Ahmad Alabed-alaziz
Jordan University Of Science and Technology
Irbid, Jordan
ahmad.mma87@gmail.com

Omar Darwish
Jordan University Of Science and Technology
Irbid, Jordan
omardrweesh@yahoo.com

ABSTRACT
Medical diagnosis via image processing and machine learning is considered one of the most important issues of artificial intelligence systems. In this paper, we present a machine learning approach to detect whether an MRI image of a brain contains a tumor or not. The results show that such an approach is very promising.

Categories and Subject Descriptors
I.2.1 [**Applications and Expert Systems**]: Medicine and science

General Terms
Design, Experimentation

Keywords
Brain MRI images; NN Classifier; Decision Tree Classifier; Feature Extraction; Image Processing

1. INTRODUCTION
Medical images are one of the most important resources used by doctors to diagnose brain tumors. A tool with high accuracy to automate this process can be extremely valuable. However, because of issues related to legal liabilities, such a tool cannot replace the expert opinions of trained physicians.

In this paper, we design a system to correctly classify new brain MRI images into images with tumor and images without tumor. This has to be done with no human intervention. In order to apply several types of classifiers, we need to preprocess several aspects of the images such as the color, area of interest, image file extension, and contrast level. We used two popular tools to achieve this, viz. ImageJ and MATLAB. Afterwards, we extracted the most important and discriminating features of the preprocessed images. In this phase, we extract ten different features (discussed in Section 5). Finally, we use a tool called WEKA 3.6 to apply four different classification algorithms on these features and calculate the precision/recall, the F-measure, the percentage of correctly classified images and the time taken to build each model.

2. RELATED WORKS
In recent years, interest in designing tools for diagnosing brain tumors has been increasing. The work of Gopal and Karnan [1] uses image processing clustering algorithms to classify images into a group that has a brain tumor and another group which does not. The dataset used in this work is composed of 42 MRI images obtained from the KG hospital database. In the preprocessing phase, the authors remove the film artifacts (labels and X-ray marks). They also use the filter Median to remove high frequency components in the MRI image. The authors then use an algorithm called Fuzzy C Means (FCM) as an image clustering algorithm, in addition to using a Genetic Algorithm (GA) as an intelligent optimization tool. The results of the experiments showed that, the classification algorithm FCM achieved a classification accuracy of 74.6% with less than 0.4% error rate. To enhance the accuracy, the authors used an optimization technique called Particle Swarm Optimization (PSO). They managed to reach an accuracy level of 92%.

In [3], Othman and Ariffanan propose a new system for brain tumor automatic diagnosis (shown in Figure 1). The Probabilistic Neural Network (PNN) provides a solution to pattern classification problems [4]. The paper uses a dataset from University Teknologi Malaysia (UTK) and the dataset goes through a preprocessing phase as follows. The MRI images are first converted to matrices by using MATLAB. Then, the classification algorithm PNN is used to classify the MRI images. The results show that the proposed system achieves a diagnosis accuracy of more than 73%. The accuracy level can even be higher than that depending on what the authors call “a smoothing factor” [3].

Finally, Najadat at al. [2] design a classifier to detect abnormalities in CT brain images caused by the following diseases/cases: Atrophic, Hemorrhage, Hematoma, Infract and Craniotomy.

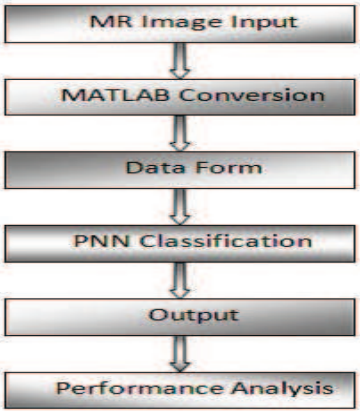

Figure 1: The proposed system by [3].

3. THE DATASET

The dataset we use in this work consists of 27 MRI images of human brain. Seven of those images are normal human brain images, and the remaining twenty show brains suffering from tumors. The images were mainly acquired by searching the Internet as it appears that acquiring such images from local hospitals is harder task that initially anticipated. To verify the correctness of the images' classification, two experts were consulted, viz Samir Abu-nameh, M.D., and Omar Borini, final year medical student.

4. DATASET PREPROCESSING

As with any other approach using medical images, a preprocessing phase is vital. Below, we explain the steps taken in this phase.

1. The dataset is converted from RGB images into gray-scale images using MATLAB image processing tools imread and rgb2gray.
2. The contrast of each image is increased to enhance the features appearance in each image using the tool imcontrast.
3. The area of interest is defined using the tool ImageJ, which allows us to draw a circle around the area of interest (the brain itself without the background of the MRI image). Figure 2 shows an example of how to specify the area of interest.
4. MRI images are converted into a single compression type. We use "TIFF" as a compression format for all images using the tool ImageJ.

After preprocessing the dataset, the next step is to extract the discriminating features.

5. THE FEATURES

In this work, ten features were used as explained below.

1. Mean Gray Value; for the area of interest, the sum of gray values of all pixels is divided on the number of pixels.

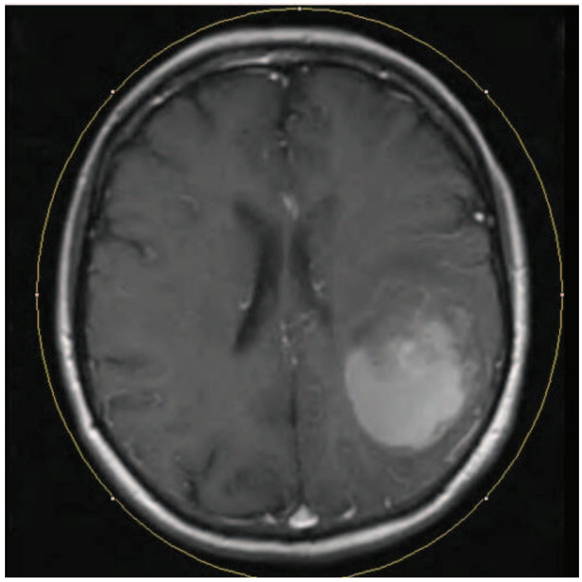

Figure 2: Area of interest as computed by the tool ImageJ. Note the yellow circle around the skull.

2. Standard deviation of the gray values of the area of interest.
3. Modal gray value of the area of interest.
4. Circularity; we use an algorithm that estimates how perfect a circle is. The output of this algorithm ranges from 0 to 1, with 1 representing a perfect circle.
5. Area Fraction; the percentage of pixels in the image with non-zero values.
6. Integrated Density; we use an algorithm that calculates and displays two values:
   - IntDen, which is simply the Area multiplied by the Mean Gray Value.
   - RawIntDen, which represents the values of the pixels in the selected area of interest.
7. Aspect Ratio of the image.
8. Round (roundness), which is the inverse of the aspect ratio.
9. Solidity, which is the area/convex area. The longest distance between any two points along the selected area.

The previously explained features were extracted using the tool ImageJ 1.45s.

6. CLASSIFICATION ALGORITHMS

The success of a system like the proposed in this work depends on whether an efficient classification algorithm can be designed that can accurately classify new images into a group of images with tumors and a group of images with no tumor. This is done by identifying some patterns in the

Table 1: Results of classifiers used in this work

Algorithm	Correct	Recall	Precision	F-Measure
NN	66.6%	0.667	0.63	0.623
J48	59.2%	0.593	0.628	0.603
Naïve Bayes	59.2%	0.593	0.691	0.6
Lazy-IBk	62.9%	0.63	0.652	0.637

selected feature values to distinguish between the images of the two types.

Four different types of classification algorithms were used as follows.

1. Artificial Neural Network.
2. Tree J48.
3. Naïve Bayes.
4. Lazy-IBk.

We use the popular classification tool "WEKA 3.6" to run our tests.

7. RESULTS

In this section, the results are shown for each one of the four classifiers mentioned above. Specifically, for each classifier, we show the following.

1. The percentage of correctly classified images.
2. The recall value.
3. The precision value.
4. The F-Measure value.

The widely-used ten-fold cross-validation technique was used to test all classification algorithms. The following table shows the results of these tests.

In order to have a better look at these values, we put them in a 2D clustered column (see Figure 3). This way, it will be easier to compare the different algorithms.

As can be seen from the result, the neural network (NN) algorithm exceeds all other algorithms in correctness and recall. Lazy-IBk surprisingly came in second in terms of correctness and recall. Moreover, it exceeds all other algorithms in the F-Measure value and achieved higher precision than NN. J48 and Naïve Bayes achieved the lowest correctness and recall.

The J48 decision tree (depicted in Figure 4) shows that the most important features used in classification were gray mean value and Int. Density.

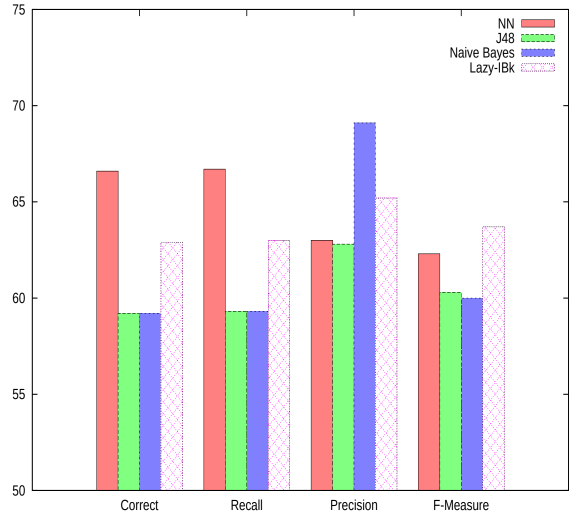

Figure 3: The results of the four classifiers used in this project.

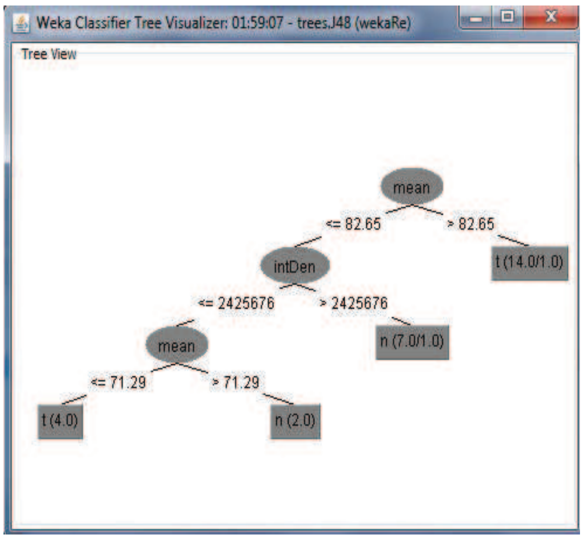

Figure 4: The decision tree of classifier J48.

8. CONCLUSIONS AND FUTURE WORK

Algorithms for analyzing and classifying medical images have gained a great level of attention recently. The experiments we present in this work show that after preprocessing MRI images, neural network classification algorithm was the best. Lazy-IBk did very well and came in second. Naïve Bayes and J48 decision tree came in last.

A much higher accuracy can be achieved by gaining a better dataset with high-resolution images taken directly from the MRI scanner. Moreover, classifier boosting techniques can be used to raise the accuracy even higher and reach a level that will allow this tool to be a significant asset to any medical facility dealing with brain tumors.

9. REFERENCES

[1] N. N. Gopal and M. Karnan. Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques. In *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–4, Dec 2010.

[2] H. Najadat, Y. Jaffal, O. Darwish, and N. Yasser. A classifier to detect abnormality in CT brain images. In *The 2011 IAENG International Conference on Data Mining and Applications*, pages 374–377, Mar 2011.

[3] M. F. Othman and M. A. M. Basri. Probabilistic neural network for brain tumor classification. In *Second International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, pages 136–138, Jan 2011.

[4] D. F. Specht. Probabilistic neural networks. *Neural Networks*, 3(1):109–118, 1990.