Classification of CT Brain Images of Head Trauma

Tianxia Gong[1], Ruizhe Liu[1], Chew Lim Tan[1], Neda Farzad[2], Cheng Kiang Lee[3], Boon Chuan Pang[3], Qi Tian[4], Suisheng Tang[4], and Zhuo Zhang[4]

[1] Department of Computer Science, School of Computing, National University of Singapore, 3 Science Drive 2, Singapore 117543
{gong_tianxia,liurz,tancl}@comp.nus.edu.sg

[2] Department of Learning, Management, Informatics & Ethics (LIME), Karolinska Institute Berzelius v. 3, Stockholm, Sweden 17177
neda.farzad.182@student.ki.se

[3] National Neuroscience Institute, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433
{cheng_kiang_lee,boon_chuan_pang}@nni.com.sg

[4] Insitute of Infocomm Research, 21 Heng Mui Keng Terrace, Singapore, 119613
{tian,suisheng,zzhang}@i2r.a-star.edu.sg

Abstract. A method for automatic classification of computed tomography (CT) brain images of different head trauma types is presented in this paper. The method has three major steps: 1. The images are first segmented to find potential hemorrhage regions using ellipse fitting, background removal and wavelet decomposition technique; 2. For each region, features (such as area, major axis length, etc.) are extracted; 3. Each extracted feature is classified using machine learning algorithm; the images are then classified based on its component regions' classification. The automatic medical image classification will be useful in building a content-based medical image retrieval system.

1 Introduction

Due to the advances of multi-slice Computed Tomography (CT) Scan with up to 64 slices per scan, a huge amount of CT images are produced in modern hospitals. Today, CT scan images are in the standard DICOM (Digital Imaging and Communications in Medicine) format which incorporates textual information together with the images. Display and retrieval of CT scan images are via PACS (Picture Archives and Communication System) hardware [1]. However with such standards and hardware, the CT scan images currently can only be retrieved using patient names or identity card numbers. To retrieve an image pertaining to a particular anomaly without the patient name is literally like looking for a needle in a haystack. In the domain of CT brain images, very often doctors already overloaded with day-to-day medical consultation simply could not remember patients' names when they need to refer to cases of certain type of brain trauma seen before and as such valuable information are lost in the sea of raw image pixels.

However, if the CT brain images are automatically classified according to trauma types and incorporated to the medical image search system, then the system with search

J.C. Rajapakse, B. Schmidt, and G. Volkert (Eds.): PRIB 2007, LNBI 4774, pp. 401–408, 2007.

functions not just by patients' names but by trauma types provides solution to the problem. In this paper, we propose a method to classify CT brain images of head trauma automatically and quickly, so that it facilitates the building of such search systems.

Head trauma has the following major types [2]: epidural hemorrhage[1] (EDH), acute subdural hemorrhage (SDH_Acute), chronic subdural hemorrhage (SDH_Chronic), intracerebral hemorrhage (ICH), intraventricular hemorrhage (IVH) and subarachnoid hemorrhage (SAH). In this paper, we focus on classification of EDH, SDH_Acute and ICH, for they are the dominant types in most head trauma cases. Our images are from CT brain scans performed in the two-year period of 2003 and 2005 as a result of hospital admission for mild head injured patients in National Neuroscience Institute, Tan Tock Seng Hospital [3]. Some of the mild head injuries were later found to be insignificant with no hemorrhage detected. Such cases are treated as belonging to the "normal" class in our training data.

The rest of the paper is organized as follows. In section 2, we will present our method of automatic classification which basically consists of three phases: namely, pre-processing, feature extraction and classification. In section 3, we will discuss our experimental results involving machine learning and validation. Finally, section 4 concludes the paper with our future works.

2 A Method for Automatic Classification of CT Brain Images

Our proposed method to automatically classify CT brain images consists of three phases: preprocessing, feature extraction and classification. In the preprocessing phase, we segment the hemorrhage regions from the CT brain image using ellipse fitting [4], background removal and wavelet decomposition technique [5, 6]. The segmented result is a binary image with potential hemorrhage regions in white and the others in black. Then for each of the potential hemorrhage regions, we extract information about size, shape and position, and create a feature vector accordingly. Lastly, we use a machine learning algorithm to classify the potential hemorrhage regions into different hemorrhage types or normal regions according to the extracted features. The CT brain images are then classified according to the classification of its potential hemorrhage regions.

2.1 Preprocessing

Preprocessing algorithm consists of 4 steps. Step one removes the skull and fits an ellipse to the skull to construct an "interior region", which is the brain inside the skull. Step two removes the gray matter. Step three uses a wavelet decomposition to reduce noise and set a threshold automatically to identify the hemorrhage regions. The last step generates a binary image containing the hemorrhage regions in white and the others in black.

Step 0: Input CT brain image in JPEG format of dimension 512×512. (Figure 1)

[1] The terms "hemorrhage" and "hematoma" are often used interchangeably. In this paper, we use "hemorrhage" for consistency.

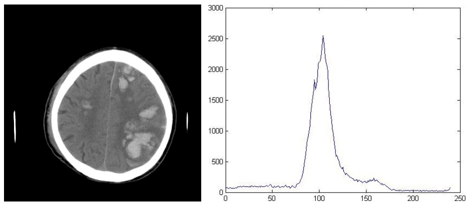

Fig. 1. Image ***I***, Original raw input image; left: image; right: intensity distribution[2]

Step 1: Remove the skull and segment the "interior region"

The skull is in white color, whose intensity is above 250 in a gray scale map. Hence, we simply treat those pixels with intensity 250 and above as the skull. The interior region refers to the brain content inside the skull. Since most traumas are diagnosed according to blood clots or edema inside the skull, it is important to segment the interior region. Firstly, we do a boundary detection based on the skull removed. The boundary contains points with intensity above 250, which belong to the skull. Note that there are two other regions that are also in white color. These two regions belong to the CT scan device. However, since they are much smaller than the skull, they can be removed by doing a simple area comparison. Next, we do an ellipse fitting on the boundary points, and compute the center (X_c, Y_c), the major axis, the minor axis, and the parameters of the ellipse. There are 6 parameters, a,b,c,d,e and f, and thus the ellipse has an equation of the following form:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

Hence, a point [x, y] given to the equation that has a result less than zero is inside the ellipse. Finally, we segment the interior points based on the following rules on the original image ***I***.

1. The point should be inside the ellipse;
2. The point should be set apart from the center of the ellipse with a distance less than 80% of the average of the major and minor axes of the ellipse.
3. Its intensity is between 10 and 250.

We denote the interior image to be $\boldsymbol{T}_0$. (Figure 2)

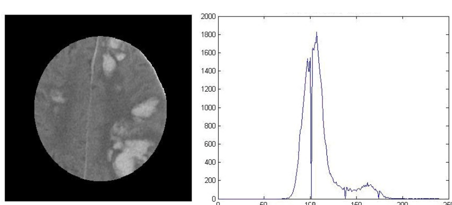

Fig. 2. Image $\boldsymbol{T}_0$, interior region; left: image; right: intensity distribution

[2] The intensity below 10 (background) and above 250 (skull) are not shown in the histogram so that the intensity of inner part of the brain is shown in more detail.

Step 2: Remove the gray matter

Most parts of the content inside the skull are the gray matter. In the histogram of intensity on $\boldsymbol{T}_0$, the peak corresponds to the gray matter. (Figure 2) Hence, a simple subtraction off the peak intensity from $\boldsymbol{T}_0$ will give us an image with the gray matter removed. We call it $\boldsymbol{T}_1$. (Figure 3)

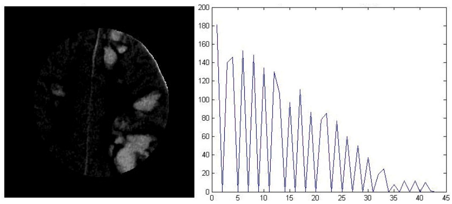

Fig. 3. Image $\boldsymbol{T}_1$, gray matter removed; left: the image; right: intensity distribution

Step 3: Reduce noise

There is much noise as white dots or tiny fragments produced in $\boldsymbol{T}_1$, because we subtract only a single intensity value from various parts of the gray matter. A second level 2D *Biorthogonal* wavelet transform is used to reduce the noise [5, 6]. We finally get the image with reduced noise but more distinguishable diseased parts. We denote the resultant image to be $\boldsymbol{T}_2$. (Figure 4)

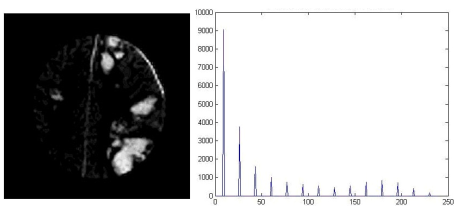

Fig. 4. Image $\boldsymbol{T}_2$, noise reduced; left: image; right: intensity distribution

Step 4: Generate a binary image of hemorrhage

After the preprocessing, we can define thresholds according to the intensity distribution of image $\boldsymbol{T}_2$. We set the hemorrhage threshold to be the median of the peaks obtained from the wavelet transform. Finally we get a binary image $\boldsymbol{T}_3$. (Figure 5)

Fig. 5. Image $\boldsymbol{T}_3$, each white pixel group represents a possible hemorrhage region of the image

2.2 Feature Extraction

As human doctors use size, shape and position of the potential hemorrhage region to classify them, we need quantifiable features that describe the size, shape and position for antomatic classification. For each potential hemorrhage region, we use the Matlab function *regionprops* [7] to extract the area, major and minor axis lengths, eccentricity, solidity and extent. Also, features for the skull and the background are extracted from the labeled skull and background regions [8]. These features describe the size, shape and position of the potential hemorrhage region; therefore, they are useful for classification. The class of each feature vector is one of the following values: EDH, SDH_ Acute, ICH and normal. All features are described in Table 1.

Table 1. Features extracted from each region

	Name	Description[7]	Example
1	Area	The actual number of pixels in the region.	607
2	Major axis length	The length (in pixels) of the major axis of the ellipse that has the same second-moments as the region.	38.1135
3	Minor axis length	The length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region.	23.5155
4	Eccentricity	The eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length.	0.7295
5	Solidity	The proportion of the pixels in the convex hull that are also in the region. Computed as Area/ConvexArea[3].	0.8772
6	Extent	The proportion of the pixels in the bounding box that are also in the region. Computed as the area divided by area of the bounding box.	0.6485
7	Skull	Whether the region is adjacent to skull or not.	false
8	Background	Whether the region is adjacent to background or not.	true

2.3 Classification

According to the features extracted in section 2.2, we classify the regions into five categories: EDH, SDH_ Acute, ICH, other and normal, where the first three classes refer to the three types of hemorrhages we focus on, the other refers to the remaining types of hemorrhage, and normal means that the region is not a hemorrhage. For example, the potential hemorrhage regions of Figure 5 classified as ICH are shown in Figure 6.

As there may be more than one type of hemorrhage present in a brain CT image, the class for each image cannot have only one of the class values as the regions have. Instead, the class for each image is a boolean vector <EDH, SDH_ Acute, ICH, normal>, where each boolean value indicates the presence of certain type of hemorrhage.

[3] The number of pixels in convex image, which is the convex hull, with all pixels within the hull filled in.

Fig. 6. Each white pixel group represents an ICH region of the image

The class of the image is classified according the classifications of its regions. If the regions are classified as some type(s) of hemorrhage (EDH, SDH_Acute, or ICH), the image is also classified as the same type(s) of the hemorrhage(s). Otherwise if all regions are classified as normal, the image itself is also classified as normal.

3 Experimental Results

We obtained 35 CT brain images (15 EDH, 9 SDH_Acute, 6 ICH and 5 normal) belonging to 12 patients from the National Neuroscience Institute, Tan Tock Seng Hospital, Singapore. After preprocessing, we obtained 818 potential hemorrhage regions (15 EDH, 19 SDH_Acute, 47 ICH and 737 normal).

3.1 Classification of Potential Hemorrhage Regions

We used J48 classifier, a decision tree classifier based on C4.5 [9], from WEKA [10] to train and test the region features. 10-fold cross validation was used. The average accuracy (correctly classified regions / all regions) is 93.0%. As there are many more normal class cases than the other classes, the data is highly imbalanced, which causes high accuracy for normal class and relatively lower accuracy for other classes. The detailed testing results for each class are reported as shown in Table 2.

Table 2. Detailed testing results for each class

	EDH	SDH_Acute	ICH	normal
Precision	60.0%	53.8%	60.0%	95.9%
Recall	60.0%	36.8%	44.7%	98.2%

The decision tree obtained from J48 is shown in Figure 7. The knowledge represented by the decision tree is actually very close to the doctor's knowledge in classifying potential hemorrhage regions. For example, if the region's area is less than or equal to 2891 pixels (6.89cm^2) and greater than 91 pixels (0.22cm^2), and the eccentricity is less than or equal to 0.9426 (the greater the eccentricity is, the elongated is the region), and the region is not adjacent to skull, then the region is ICH. This is also a typical rule for doctors to recognize ICH manually.

```
Area <= 2891
|   Eccentricity <= 0.9426
|   |   skull = false
|   |   |   Area <= 91
|   |   |   |   Extent <= 0.6485: normal
|   |   |   |   Extent > 0.6485: ICH
|   |   |   Area > 91: ICH
|   |   skull = true
|   |   |   Area <= 1263: normal
|   |   |   Area > 1263
|   |   |   |   Eccentricity <= 0.9322: EDH
|   |   |   |   Eccentricity > 0.9322: SDH_Acute
|   Eccentricity > 0.9426: normal
Area > 2891
|   Eccentricity <= 0.8579: ICH
|   Eccentricity > 0.8579
|   |   Area <= 7185
|   |   |   Extent <= 0.1852
|   |   |   |   MajorAxisLength <= 274.1822: EDH
|   |   |   |   MajorAxisLength > 274.1822: SDH_Acute
|   |   |   Extent > 0.1852: SDH_Acute
|   |   Area > 7185: EDH
```

Fig. 7. Decision tree obtained from the training data using the J48 classifier

3.2 Classification of Images

The classification of the image is considered as: 1. correct, if the predicted class(es) and the actual class(es) are exactly the same; 2. partially correct, if the actual class(es) is/are included in the prediction, but other class(es) is/are also predicted; 3. incorrect, if the predicted class(es) is different from the actual class. Among the 35 images, 18 are classified correctly, 6 are classified partially correctly, and 11 are classified incorrectly.

4 Conclusion

In this paper, we propose a method to classify CT brain images of head trauma automatically and quickly. The method consists of three phases: preprocessing, feature extraction and classification. In the preprocessing phase, we segment the hemorrhage regions from the CT brain image using ellipse fitting, background removal and wavelet decomposition technique. The segmented result is a binary image with potential hemorrhage regions in white and the rest in black. Then for each of the potential hemorrhage regions, we extract information about its size, shape and relative location, and create a feature vector. Lastly, we use machine learning algorithms to classify the potential hemorrhage regions into different hemorrhage types or normal regions according to the extracted features. The CT brain images are then classified according to the classification of its potential hemorrhage regions.

The fast and scalable automatic medical image classification can help to build a medical image search system according to the syndrome types (in our case of CT brain images, the syndrome types are head trauma types) and facilitate doctors' research on certain syndrome as well as education for medical profession.

In our future work, we will extend the classification types to include other head traumas. We will also explore other machine learning algorithms and compare their classification results. Finally, we will do text mining to extract further information from the text of neuroradiologists' report to find more features for classification.

Acknowledgement

This project is supported in part by the National University of Singapore under Academic Research Fund grant no. R252-000-290-112.

References

1. Huang, H.K.: PACS: Basic Principles and Applications. Wiley-Liss Inc., Canada (1999)
2. Downie, A.: Tutorial: CT in Head Trauma, http://www.radiology.co.uk/srs-x/tutors/cttrauma/tutor.htm
3. Pang, B.C., Yin, H.: Analysis of clinical criterion for "talk and deteriorate" following minor head injury using different data mining tools. Journal of Neurotrauma, National Neurotrauma Society, Gainesville, Florida, USA (in press)
4. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct Least Square Fitting of Ellipses. IEEE Pattern Analysis and Machine Intelligence 21(5) (May 1999)
5. Daubechies, I.: Ten Lectures on Wavelets. SIAM: Society for Industrial and Applied Mathematics (June 1, 1992)
6. Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.M.: Wavelet Toolbox 4 User's Guide. Matlab Wavelet Toolbox User's Guide, Version 4 (2007)
7. Image Processing Toolbox User's Guide. Matlab Image Processing Toolbox User's Guide Version 3 (2002)
8. Cosic, D., Loncaric, S.: Rule-Based Labeling of CT Head Image. In: Keravnou, E.T., Baud, R.H., Garbay, C., Wyatt, J.C. (eds.) AIME 1997. LNCS, vol. 1211, pp. 453–456. Springer, Heidelberg (1997)
9. Quilan, J.R.: C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco (1993)
10. Holmes, G., Donkin, A., Witten, I.H.: WEKA: a machine learning workbench. In: Proceedings Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia, pp. 357–361 (1994)