

Region of Interest Based Image Categorization

Ashraf Elsayed¹, Frans Coenen¹, Marta García-Fiñana², and Vanessa Sluming³

¹ Department of Computer Science, University of Liverpool,

Ashton Building, Ashton Street, Liverpool L69 3BX, United Kingdom

² Centre for Medical Statistics and Health Evaluation, University of Liverpool,
Shelley's Cottage, Brownlow Street, Liverpool L69 3GS, United Kingdom

³ School of Health Sciences, University of Liverpool,
Thompson Yates Building, The Quadrangle, Brownlow Hill, Liverpool L69 3GB,
United Kingdom

{a.el-sayed, coenen, m.garciafinana, vanessa.sluming}@liv.ac.uk

Abstract. Region Of Interest Based Image Classification (ROIBIC) is a mechanism for categorising images according to some specific component or object that features across a given image set. This paper describes and compares two such approaches. The first is founded on a weighted graph mining technique whereby the ROI is represented using a tree structure which allows the application of a weighted graph mining technique to identify features of interest, which can then be used as the foundation with which to build a classifier. The second approach is founded on a time series analysis technique whereby the ROI are represented as time series which can then be used as the foundation for a Case Based Reasoner. The presented evaluation focuses on MRI brain scan data where the classification is focused on the *corpus callosum*, a distinctive region in MRI brain scan data. Two scenarios are considered: distinguishing between musicians and non-musicians and epilepsy patient screening.

Key words: Image mining, Image categorisation

1 Introduction

Image categorisation is concerned with the labelling of images into one or more predefined classes. The principal challenge of image categorisation is the capture of the significant features within images that facilitate the desired classification. Edge detection, segmentation and registration all have a significant part to play in this process. One method of simplifying the image categorisation process is to focus on some particular feature or Region Of Interest (ROI) within the image set. The advantage offered is that the remainder of the image can be ignored and thus computational advantages gained. Alternatively, the representation can be more detailed. We refer to this approach as ROI Based Image Categorisation (or ROIBIC). Of course ROIBIC is not suited to every image categorisation application; not all such applications include a significant and identifiable ROI that appears across the data set. The most appropriate applications for ROIBIC are those where the data set includes a common feature whose size and shape

strongly influences the categorisation, i.e. a set of images that can be categorised according to the shape of some object that consistently appears across the image set. The focus of the work described in this paper is the categorisation of Magnetic Resonance Imaging (MRI) brain scan data according to a specific feature within the data called the *corpus callosum*.

This paper describes and compares two approaches to ROIBOC: graph based and time series based. Both approaches, although operating in very different manners, are essentially supervised learning mechanisms where by a pre labelled training set is used to build a “classifier” which can be applied to unseen data. The first approach uses a tree based representation for the common feature, one tree per image. A graph mining technique is then applied to identify frequently occurring sub-graphs (sub-trees). The identified set of trees are then used to describe the image set in terms of a set of attributes each of which equates to a frequently occurring sub-tree. A classification algorithm is then applied to this attribute set to build a classifier to be applied to “unseen” data. The second approach is founded on a time series representation coupled with a Case Based Reasoning (CBR) technique. The features of interest, when identified, are represented as time series, one per image. These time series are then stored in a Case Base (CB) which can be used to categorise unseen data. The unseen data is compared with the categorisation on the CB using a Dynamic Time Warping (DTW) based similarity checking mechanism, the categorisation associated with the most similar time series (case) in the CB is then adopted as the categorisation for the unseen data. Both approaches require the application of a registration process and segmentation, and this will entail the established difficulties encountered when conducting this process (i.e. poor image contrast, intensity inhomogeneities and partial-volume effects). Both approaches are intended to preserve the size and shape of the feature of interest (ROI).

The rest of this paper is structured as follows. In section 2 a brief overview of related previous work is presented, followed in section 3 with an overview of the application that acts as the focus for this paper. The advocated approaches are described in Sections 4 and 5 respectively. A complete evaluation of these approaches is reported in Section 6, followed by some conclusions in Section 7.

2 Previous Work

Current image categorisation techniques can be divided into two groups according to the image features used for the classification: global approaches and ROI based approaches. The first use features that reflect all the information contained within image sets. One such technique is the use of colour histograms to represent images. For example in [21] a k-nearest neighbour classifier is applied to colour histogram represented images to discriminate between “indoor” and “outdoor” images, in [12] a time series technique is applied to classify histogram represented retina images, and in [22] a Bayesian classifiers is applied to edge direction histograms to categorise city and landscape images. Support Vector Machines (SVMs) built on colour histograms were applied to classify im-

ages containing a generic set of objects in [3]. Although the global features can usually be computed with little cost and are effective for certain classification tasks, a significant drawback is that structural and relative spatial information is lost. Furthermore, for many applications (such as medical applications) image attributes such as colour and intensity have limited discriminative power.

A number of ROI-based approaches have been proposed to maintain local and spatial properties of an image using the concept of *regions* or *blocks*. In [9] cityscape images are divided into 16 non-overlapping equal-sized blocks. Dominant orientations are then computed for each block, and the images classified as city or suburb as determined by the majority orientations of blocks. In [24] graph and photograph images are divided into blocks and each block assigned to one of two categories. If the percentage of blocks classified as photograph is higher than a threshold, the image is marked as a photograph; otherwise, the image is marked as a graph. A disadvantage of this approach is that a rigid partition of an image into fixed-size blocks often breaks an object into several blocks or puts different objects into a single block. Thus visual information about objects may be destroyed by a rigid partition. In [20] an alternative approach is described where images are classifying according to spatial orderings of regions where each region is represented by a symbol corresponding to an entry in a pattern library. Image segmentation is one way to extract object information whereby an image is decomposes into a collection of regions, each corresponding to an object. There are many examples where image segmentation has been applied successfully to the image categorisation problem. ROIBIC, as described in this paper, advocates an approach where the focus is on a single ROI common across the image set.

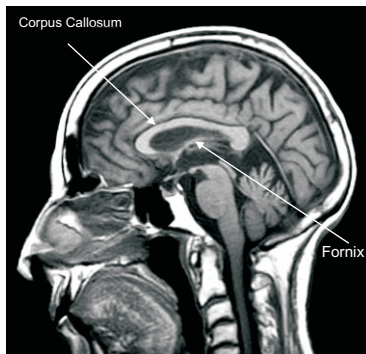


Fig. 1. corpus callosum in a midsagittal brain MR image.

3 Application Domain

Although generally applicable, the ROIBIC approaches described in this paper are directed at MRI brain scan data, more specifically the categorisation of MRI data according to a specific object contained in these images called the corpus callosum. The corpus callosum is a highly visible structure in MRI scans whose function is to connect the left and right hemispheres of the brain, and to provide

the communication conduit between these two hemispheres. Figure 1 gives an example MRI scan, the corpus callosum is located in the centre of the image. A related structure, the *fornix* is also indicated. The fornix often “blurs” into the corpus callosum and thus presents a particular challenge in the context of the segmentation of these images so as to isolate the corpus callosum ROI.

The corpus callosum is of interest to medical researchers for a number of reasons. The size and shape of the corpus callosum has been shown to be correlated to sex, age, neurodegenerative diseases and various lateralized behaviour in people. It is also conjectured that the size and shape of the corpus callosum reflects certain human characteristics (such as a mathematical or musical ability). Several studies indicate that the size and shape of the corpus callosum, in humans, is correlated to sex [1, 6, 19], age [19, 25], brain growth and degeneration [11, 16], handedness [5], epilepsy [4, 18, 23] and brain disfunction [7, 13].

4 Graph Based Approach

In this and the following section the two proposed techniques, graph based ROIBIC and time series based ROIBIC, are described commencing with graph based ROIBIC. A schematic of the graph based process is given in Figure 2. The process commences with segmentation and registration to isolate the ROI. The details of the identified ROI are then acquired by tessellating the images into homogeneous sub-regions, according to (say) colour or intensity, and then storing the result in a *quad-tree data structure*. A *weighted sub-graph mining* approach is then applied to the tree represented image set (one tree per image) to identify frequent sub-graphs. The identified sub-trees (graphs) then form the fundamental elements of a *feature space*, i.e. a set of attributes with which to describe the image set. Experiments conducted by the authors have revealed that, for many image sets, the graph mining process can identify a great many frequent sub-graphs; more than required for the desired categorisation. Therefore a feature selection strategy is applied so that only those sub-graphs that serve as the best discriminators are retained. Each image is then described in terms of a binary-valued feature vector indicating the selected attributes (sub-graphs) that appear in each image. Once the image set has been represented in this manner any appropriate classifier generator may be applied; for the corpus callosum application Quinlan’s C4.5 algorithm was used [17].

4.1 Tessellation

The tessellation process comprises the recursive decomposition of the identified ROI, for each image, into quadrants. The tessellation proceeds until either sufficiently homogeneous *tiles* are identified or some user specified level of granularity is reached. The result is then stored in a quadtree data structure such that each root node represents a tile in the tessellation. Nodes nearer the root of the tree represent larger tiles than nodes further away from the root. Thus the tree is “unbalanced” in that some root nodes cover larger areas of the ROI than others.

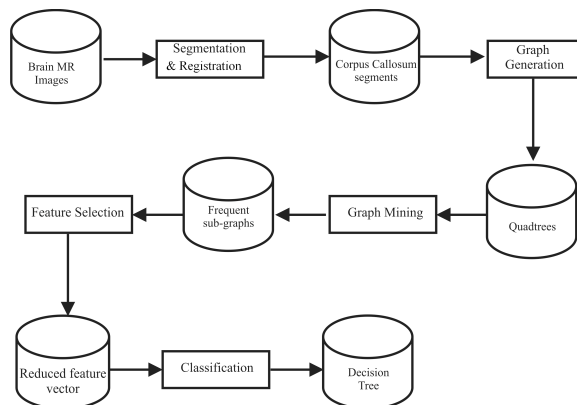


Fig. 2. Framework of graph mining ROIBIC.

Note also that the relative location of the each tile is maintained in the structure of the tree. Thus the advantage of the representation is thus that information about the relative location and size of groups of pixels is maintained.

The tessellation can be conducted according to a variety of image features such as colour or intensity. With respect to the corpus callosum application a binary encoding was used, the tiles included in the corpus callosum were allocated a “1” (black) and tiles not included a “0” (white). Sufficiently homogeneous was defined as a tile that was 95% black or white. The research team experimented with a number of “granularity” settings as reported in the evaluation described in Section 6. Interested readers may like to refer to a previously published work by the authors [8] regarding further details of the tessellation process.

4.2 Weighted Graph Mining

As noted above, in the quad-tree representation nodes nearer the root can be considered to be more significant than others (because they cover a larger area). A weighted frequent sub-graph mining algorithm was therefore developed to identify commonly occurring features across the tree represented image set. The weightings were calculated according to the proximity of individual nodes to the root node in each tree. This weighting concept was built into a variation of the well known gSpan algorithm [26]. The algorithm operates in a depth first search manner, level by level, following a “generate, calculate support, prune” loop. Candidate sub-graphs are pruned if their *support* (frequency of occurrence across the graph set) is below a user defined “support threshold”. The lower the threshold the greater the number of frequent sub graphs that will be identified. Space restrictions preclude further discussion of this weighted sub-graph mining algorithm here, however, interested readers are referred to [14].

Experimentation with respect to the Corpus Callosum application indicated that, to capture the necessary level of detail, a low support threshold was required. However this produced a large number of frequent sub-graphs many of which were redundant. A feature selection operation (discussed in the following subsection) was thus applied to the identified frequent sub-graphs.

4.3 Feature Selection and Classifier Generation

Feature selection is a well understood process, in the context of Data Mining, for removing irrelevant data from a feature space so as to enhance computational efficiency. Feature selection has attracted a great deal of attention from the data research community, especially in the context of classification and prediction where the aim is to identify features that are “strong discriminators”. For the corpus callosum application described here, a straightforward wrapper method was adopted whereby a decision tree generator was applied to the feature set (an approach also advocated by other practitioners, see for example [10]). The advantage of decision tree algorithms, with respect to feature selection, is that they inherently estimate the suitability of features for the separation of objects representing different classes. Features that are included as “choice points” in the decision tree were thus selected, while all remaining features were discarded. For the work described here, the well established C4.5 algorithm [17] was used.

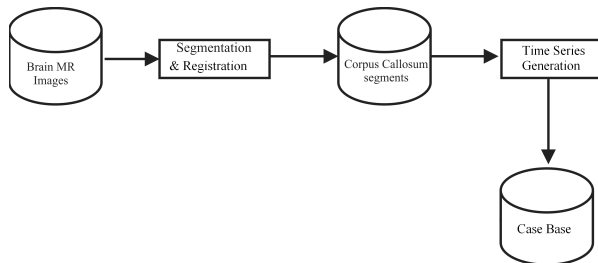


Fig. 3. Framework of time series based ROIBIC.

5 Time Series Based Approach

In this section the time series based ROIBIC approach is described; a schematic is presented in Figure 3. As for the graph based approach the process commences with the segmentation and registration of the input images. The identified ROI are then encoded as time series. Each time series is conceptualised as a *prototype* or case contained in a Case Base (CB), to which a Case Based Reasoning (CBR) mechanism may be applied. Thus an unseen record is classified according to the “best match” discovered in the CB. The CBR community has proposed many techniques to identify the desired best match. In this paper the authors advocate a Dynamic Time Warping (DTW) time series comparison mechanism that operates regardless of the length of the individual time series [2].

5.1 The Time Series Representation

After the identification of the individual ROI, using segmentation, the registration process was undertaken by fitting each ROI into a Minimum Bounding Rectangle (MBR). Individual time series is then derived, according to the boundary line circumscribing the ROI, using an ordered sequence of N vectors radiating

out from a reference point. The time series was then expressed as a series of values (one for each of the N vectors) describing the size (length) of intersection of the vector with the ROI. It should be noted that the representation maintains the structural information (shape and size) of the ROI. It should also be noted that N is variable due to the differences of the shape and size of the individual ROI within the image data set.

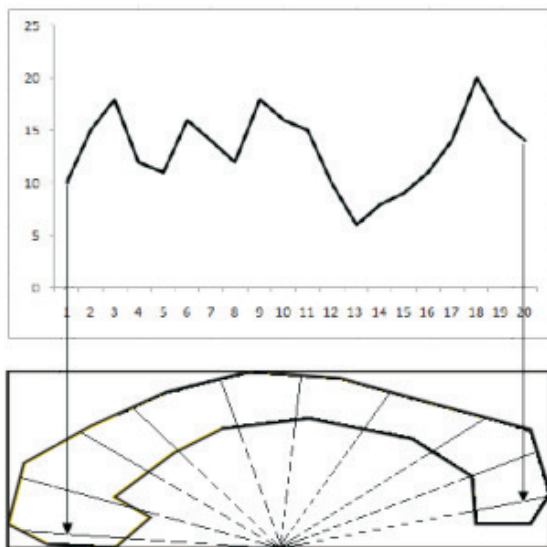


Fig. 4. Conversion of corpus callosum into time series.

With respect to the corpus callosum application the time series generation process is illustrated in Figure 4. The midpoint of the lower edge of the MBR was selected as the reference point. The vectors were derived by rotating an arc about the reference point pixel by pixel, thus the value of N will vary across the image set. In this manner time series curves were generated of the form described in the top half of Figure 4 where the X-axis represents the vector (arc) number, and the Y-axis the “pixel-distance” where the vector intersects the corpus callosum.

5.2 The Dynamic Time Warping Algorithm

DTW [2] is a time series analysis technique for comparing curves. The advantage offered is that DTW is able to find the optimal alignment between two time series Q and C , of length n and m respectively. It is often used: to determine time series similarity for classification, or to find corresponding regions between two time series. The DTW-distance between the two time series Q and C is $D(M, N)$, which we calculate in a dynamic programming approach using:

$$D(i, j) = d(q_i, c_j) + \min \{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (1)$$

Backtracking along the minimum cost index pairs $w(i, j)_k$ starting from (M, N) yields the DTW *warping path*.

An example is given in Figure 5 where the warping path between two time series Q and C of different length is presented. Note that given two identical curves the warping path would be the straight line connecting the two opposite corners of the grid. The degree of similarity can thus be determined by comparing the calculated warping path with the “ideal path”.

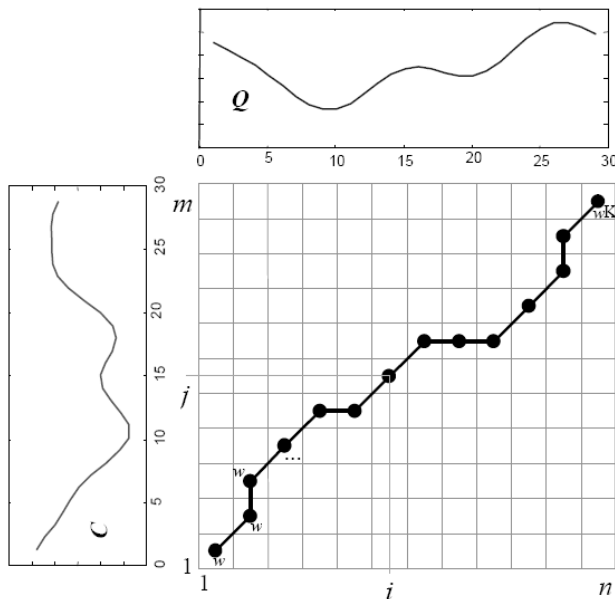


Fig. 5. An example warping path [15].

6 Evaluation

The two advocated approaches to ROIBIC are evaluated and compared in this section with respect to the corpus callosum application. This section describes the evaluation of the proposed technique using an appropriate MRI image set. The evaluation was undertaken in terms of classification accuracy, sensitivity and specificity. Two studies have been used for the investigation: (i) a comparison between musician and non-musician MRI scans, and (ii) an epilepsy screening process. The studies are discussed in detail in Sub-sections 6.1 and 6.2 below.

6.1 Musicians v. Non-Musicians

For the musicians study a data set comprising 106 MRI scans was used, 53 representing musicians and 53 non-musicians (i.e. two equal classes). The study was of interest because of the conjecture that the size and shape of the corpus callosum reflects human characteristics (such as a mathematical or musical ability). Table

1 shows the Ten Cross Validation (TCV) results obtained. The G-ROIBIC and T-ROIBIC columns indicate the results using the graph based and time series based approaches respectively. For the graph based approach a quad tree depth of six coupled with a 30% support threshold was used. For comparison purposes the results using other support threshold and depth settings are given in Table 2. The best result for each level is indicated in **bold** font. Inspection of Tables 1 and 2 demonstrate that the overall classification accuracy of the ROIBIC time series based approach significantly improves over that obtained using the graph based approach. In many TCV cases the time series based approach obtains 100% accuracy although visual inspection of the corpus callosums in the data set does not allow for the clear identification of any defining feature.

Table 1. TCV Classification Results for Musicians Study

Test set ID	G-ROIBIC			T-ROIBIC		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1	92.45	94.12	90.91	91	100	85.71
2	96.23	94.55	98.04	100	100	100
3	95.28	96.15	94.44	91	100	85.71
4	93.40	94.23	92.59	100	100	100
5	97.17	96.3	98.08	100	100	100
6	94.34	96.08	92.73	100	100	100
7	97.17	96.3	98.08	100	100	100
8	95.28	96.15	94.44	100	100	100
9	96.23	94.55	98.04	100	100	100
10	95.28	96.15	94.44	100	100	100
Average	95.28	95.458	95.179	98.2	100	97.14
SD	1.54	0.95	2.7	3.8	0.0	6.03

Table 2. TCV Classification accuracy (%) using graph based ROIBIC

Levels	Support Threshold (%)							
	20	30	40	50	60	70	80	90
4	70.75	69.81	68.87	71.70	68.87	61.32	52.83	50.94
5	90.57	83.96	80.19	85.85	80.19	81.13	80.19	70.75
6	85.85	95.28	84.91	83.96	90.57	83.96	77.36	75.47
7	83.80	85.85	89.62	86.79	87.74	75.47	76.42	78.30

6.2 Epilepsy Screening

For the epilepsy study three data sets were used. The first comprised the control group from the musicians study together with 53 MRI scans from epilepsy patients. The Second data set used all 106 MRI scans from the musicians study and the 53 epilepsy scans. The third data set comprised the 106 MRI scans from the musicians study augmented with 106 epilepsy cases. The objective was to seek support for the conjecture that the shape and size of the corpus callosm is influence by conditions such as epilepsy ([4, 18, 23]). Tables 3 and 4 show the Ten Cross Validation (TCV) classification results for the three epilepsy data sets.

Again the G-ROIBIC and T-ROIBIC columns indicate the results obtained using the graph based and time series based ROIBIC approaches respectively. A quad-tree depth of six was again used, coupled with a 30% support threshold, as this had been found to give the best results in the case of the musicians study. Inspection of Tables 3 and 4 indicates that the graph based approach significantly out-performs the time series based approach. Best results are obtained using the large, 212 MRI scan data set, because this includes many more training examples.

Table 3. TCV Classification results for Epilepsy Study (Data Sets 1 and 2)

Test set ID	106 MR scans						159 MR scans					
	G-ROIBIC			T-ROIBIC			G-ROIBIC			T-ROIBIC		
	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.
1	88.68	87.27	90.2	72.73	80.00	66.67	86.79	76.67	92.93	75.00	70.00	83.33
2	85.85	86.54	85.19	81.82	83.33	80.00	79.25	65.63	88.42	81.25	85.71	77.78
3	85.85	86.54	85.19	72.73	80.00	66.67	82.39	69.84	90.63	75.00	70.00	83.33
4	82.08	82.69	81.48	81.82	83.33	80.00	84.28	72.58	91.75	81.25	85.71	77.78
5	76.42	75.93	76.92	81.82	83.33	80.00	83.65	72.13	90.82	81.25	85.71	77.78
6	85.85	86.54	85.19	81.82	83.33	80.00	86.79	76.67	92.93	75.00	70.00	83.33
7	72.64	74	71.43	63.64	66.67	60.00	78.62	64.62	88.3	81.25	85.71	77.78
8	85.85	86.54	85.19	81.82	83.33	80.00	82.39	69.84	90.63	68.75	66.67	71.43
9	88.68	87.27	90.2	72.73	80.00	66.67	86.79	76.67	92.93	68.75	66.67	71.43
10	82.08	82.69	81.48	63.64	66.67	60.00	79.25	65.63	88.42	81.25	85.71	77.78
Average	83.40	83.60	83.25	75.46	79.0	72.0	83.02	71.03	90.78	76.88	77.19	78.18
SD	5.25	4.88	5.75	7.48	6.67	8.78	3.22	4.73	1.89	5.15	9.06	4.37

Table 4. TCV Classification results for Epilepsy Study (Data Set 3)

Test set ID	212 MR scans					
	G-ROIBIC			T-ROIBIC		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1	85.38	85.71	85.05	81.82	88.89	76.92
2	84.91	86.27	83.64	77.27	80.00	75.00
3	87.74	89.22	86.36	81.82	88.89	76.92
4	89.15	89.52	88.79	77.27	80.00	75.00
5	86.32	87.38	85.32	68.18	70.00	66.67
6	85.85	86.54	85.19	72.73	77.78	69.23
7	84.91	86.27	83.64	77.27	80.00	75.00
8	84.43	84.11	84.76	81.82	88.89	76.92
9	83.96	83.33	84.62	72.73	77.78	69.23
10	87.74	89.22	86.36	81.82	88.89	76.92
Average	86.04	86.76	85.38	77.27	82.11	73.78
SD	1.68	2.12	1.52	4.79	6.51	3.89

6.3 Discussion

With respect to classification accuracy both algorithms performed well although the time series approach produced the best results for the musicians study while

the graph based approach produced the best results for the epilepsy study. There is no obvious reason why this might be the case, visual inspection of the MRI scans does not indicate any obvious distinguishing attributes with respect to the size and shape of the corpus callosum. With respect to computational complexity image segmentation and the application of DTW for classification are both computationally expensive processes. The time complexity for the image segmentation was about 30 seconds per image. For the given data sets the application of DTW took 90 seconds on average to categorise the test set. The graph based approach was significantly faster.

7 Conclusion

Two approaches to Region of Interest Based Image Classification (ROIBIC) have been described. The first was founded on a graph representation to which graph mining techniques could be applied to obtain a feature space. The second used a time series based approach. The work was directed at the classification of MRI scans according to the nature of the corpus callosum featured within these images. Two studies were used for the evaluation: (i) distinguishing musicians from non-musicians, and (ii) epilepsy screening. However, the approach has more general applicability. The research team are also interested in alternative methods of pre-processing MRI data, and mechanism for the post-processing of results to provide explanations for specific classifications. The latter is seen as particularly significant in the context of medical research involving MRI scan data.

References

1. Allen, L., Richey, M., Chain, Y. and Gorski, R. (1991). Sex differences in the corpus callosum of the living human being. *Journal of Neuroscience*, 11, pp 933-942.
2. Berndt, D. and Clifford, J. (1994) Using dynamic time warping to find patterns in time series. AAAI-94 workshop on Knowledge Discovery in Databases, Seattle, Washington, pp 359-370.
3. Chapelle, O., Haffner, P. and Vapnik, V. (1999). Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5), pp 1055-1064.
4. Conlon, P. and Trimble, M. (1988). A study of the corpus callosum in epilepsy using magnetic resonance imaging. *Epilepsy Res*, 2, pp 122-126.
5. Cowell, P., Kertesz, A. and Denenberg, V. (1993). Multiple dimensions of handedness and the human corpus callosum. *Neurology*, 43, pp 2353-2357.
6. Davatzikos, C., Vaillant, M., Resnick, S., Prince, J., S. Letovsky, S. and Bryan, R. (1996). A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography*, 20, pp 88-97.
7. Duara, R., Kushch, A., Gross-Glenn, K., Barker, W., Jallad, B., Pascal, S., Loewenstein, D., Sheldon, J., Rabin, M., Levin B. and Lubs, H. (1991). Neuroanatomic differences between dyslexic and normal readers on magnetic resonance imaging scans. *Archives of Neurology*, 48, pp 410-416.
8. Elsayed, A., Coenen, F., Jiang, C., García-Fiñana, M. and Sluming, V. (2009). Corpus Callosum MR Image Classification. *Proc. AI'2009*, Springer, pp 333-346.

9. Gorkani, M. and Picard, R. (1994). Texture Orientation for Sorting Photos "at a glance", Proc. 12th Int'l Conf. on Pattern Recognition, pp 459-464.
10. Grabczewski, K. and Jankowski, N. (2005). Feature selection with decision tree criterion. Proc 5th Int. Conf. on Hybrid Intelligent Systems (HIS'05), pp 212-217.
11. Hampel, H., Teipel, S., Alexander, G., Horwitz, B., Teichberg, D., Schapiro, M. and Rapoport, S. (1998). corpus callosum atrophy is a possible indicator of region and cell type-specific neuronal degeneration in Alzheimer disease. *Archives of Neurology*, 55, pp 193-198.
12. Hijazi, M.H.QA., Coenen, F. and Zheng, Y. (2009). A Histogram Based Approach to Screening of Age-related Macular Degeneration. Proc. Medical Image Understanding and Analysis (MIUA'09), pp154-158.
13. Hynd, G., Hall, J., Novey, E., Eliopoulos, D., Black, K., Gonzalez J., Edmonds, J., Riccio, C. and Cohen, M. (1995). Dyslexia and corpus callosum morphology. *Archives of Neurology*, 52, pp 32-38.
14. Jiang, C. and Coenen, F. (2008). Graph-based Image Classification by Weighting Scheme. Proc. AI'2008, Springer, pp 63-76.
15. Keogh, E. and Pazzani, M. (1999). Scaling up dynamic time warping to massive datasets. Proc. of the 3rd Europ. Conf. on Principles of Data Mining and Knowl. Discovery, pp. 1-11.
16. Lyoo, I., Satlin, A., C. K. Lee, C. and Renshaw, P. (1997). Regional atrophy of the corpus callosum in subjects with Alzheimer's disease and multi-infarct dementia. *Psychiatry Research*, 74, pp 63-72.
17. Quinlan R. (1993). C4.5: A program for machine learning, Morgan Kaufmann.
18. Riley, J.D., Franklin, D.L., Choi, V., Kim, R.C., Binder, D.K., Cramer, S.C. and Lin, J.J. (2010). Altered white matter integrity in temporal lobe epilepsy: Association with cognitive and clinical profiles. To appear in *Epilepsia*.
19. Salat, D., Ward, A., Kaye, J. and Janowsky, J. (1997). Sex differences in the corpus callosum with aging. *Journal of Neurobiology of Aging*, 18, pp 191-197.
20. Smith, J. and Li, C. (1999). Image Classification and Querying Using Composite Region Templates, Int'l J. Computer Vision and Image Understanding, 75(1/2), pp 165-174.
21. Szummer, M. and Picard, R. (1998). Indoor-Outdoor Image Classification, Proc. IEEE Int'l Workshop on Content-Based Access of Image and Video Databases, pp 42-51.
22. Vailaya, A., Figueiredo, M., Jain, A. and Zhang, H. (2001). Image Classification for Content-Based Indexing, IEEE Transactions on Image Processing, 10(1), pp. 117-130.
23. Weber, B., Luders, E., Faber, J., Richter, S., Quesada, C.M., Urbach, H., Thompson, P.M., Toga, A.W., Elger, C.E. and Helmstaedter, C. (2007). Distinct regional atrophy in the corpus callosum of patients with temporal lobe epilepsy. *Brain*, 130, pp 3149-3154.
24. Wang, J., Li, J. and Wiederhold, G. (2001b). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9), pp 947-963.
25. Weis, S., Kimbacher, M., Wenger, E. and Neuhold, A. (1993). Morphometric analysis of the corpus callosum using MRI: Correlation of measurements with aging in healthy individuals. *American Journal of Neuroradiology*, 14, pp 637-645.
26. Yan, X. and Han, J. (2002). gspan: Graph-based substructure pattern mining. In ICDM'02: 2nd IEEE Conf. Data Mining, pp 721-724.