Lumbar Spine Discs Labeling using Axial View MRI Based on the Pixels Coordinate and Gray Level Features

Ala S. Al Kafri[1], Sud Sudirman [1], Abir J. Hussain[1], Paul Fergus[1], Dhiya Al-Jumeily[1], Mohammed Al-Jumaily[2], Wasfi Al-Rashdan[3] and Mohammad Bashtawi[3]

[1] Faculty of Engineering and Technology, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK
a.s.alkafri@2015.ljmu.ac.uk,
{s.sudirman, a.hussain, p.fergus, d.aljumeily}@ljmu.ac.uk
[2]Dr Sulaiman Al Habib Hospital, Dubai Healthcare City, Dubai, UAE
maljumaily@yahoo.fr
[3]Irbid Speciality Hospital, Irbid, Jordan
drwasfi1@hotmail.com,mohdbakhiet@yahoo.com

Abstract. Disc herniation is a major reason for lower back pain (LBP), it cost the United Kingdom (UK) government over £1.3 million per day. In fact a very high proportion of the UK population will complain from their back pain. Furthermore, Magnetic Resonance Imaging (MRI) is one of the main diagnosing procedure for LBP. Automatic disc labeling in the MRI to detect the herniation area will reduce the required time to issue the report from the radiologist. We present a method for automatic labeling for the lumbar spine disc area using the axial view MRI based on the pixels coordinate and gray level features. We use a clinical MRI for the training and testing. Moreover, the accuracy and the reconstructed images was the main indicator for our result. The highest achieved accuracy was 98.9 and 91.1 for Weighted KNN and Fine Gaussian SVM respectively.

Keywords: LBP, MRI, Lumbar Spine Disc, Disc Herniation

1   Introduction

Low back pain considered as the second most popular illness after the common cold. More than half of the world population were affected by the lower back pain once in their live [1]. Sixty to eighty percent of the UK population will suffer from their back pain once in their lives [2]. As a result back pain cost the UK government £1.3 million per day [3]. One of the main functions for diagnosing the cause of lower back pain is the MRI examination. It has been reported that the number of MRI examination in 2014 has increased in the UK by 11.3% to reach 2.61 million in compared with the number of imaging tests in 2013 [4]. However, the number of radiologist in the UK is insufficient for the clinical demand made by the radiology services as there an increased of 26% for the MRI request in comparison to only 3% increase in the consultant radiology workforce [5]. The gap between the increasing rate of the radiology

services compared to the number of radiologist is too high which justifies the need of automating the diagnosing process which usually include two steps that start with labeling the inter vertebral discs area then diagnosing the disc abnormality. The focus of this paper is on labeling and localizing the disc area in the lumbar spine using the axial view MR images. In our previous work [6] we have developed a framework for detecting the disc herniation in the lumbar spine which require labeling the disc area to be able to detect the herniation.

Currently diagnosing the lower back pain is done by a visual observation and analysis of the lumbar spine MR images and this process could take up much of a physician time and effort. Moreover, it can increase the possibility of misdiagnosis. In some other disease, a computer aided diagnosing systems (CAD) were developed to help the physician in the diagnosing process as an example of these systems, CAD system for detecting colonic polyp, CAD system for detecting breast cancer in mammography and CAD system for detecting prostate cancer using MR images [7] [8] [9]. Unfortunately, this type of CAD does not exist to diagnose the back pain unlike other diseases as the CAD systems are used in the diagnosing process. At the same time, there is a pressing need for this type of application to help radiologists and orthopedists in their tasks. This makes one of the motivations of this research. Such an application will need to employ sophisticated algorithms to overcome a number of technical challenges due to the wide range of imaging characteristics and resolutions [10] as well as technical limitations in detecting and highlighting areas of interest.

2 Related Research

Image analysis and comparison are performed by means of classifying its features. This is done by comparing the features from the test image in question with those from training data. A brute force approach for comparing two sets of image features would compare every feature in one set to every feature in the other and keep track of the "best so far" match. This results in a heavy computational complexity in the order of $O(N^2)$ where N is the number of features in each image. A number of proposed algorithms have been proposed to improve the computational complexity, including the popular kd-tree technique [11]. This technique uses exact nearest neighbour search and works very well for low dimensional data but quickly loses its effectiveness as dimensionality increases. The popularity of the kd-tree technique has seen a number of proposals to further improve the algorithm including [12], [13]. Jiang *et. al.* [14] proposed a visualization and quantitative analysis framework using image segmentation technique to derive six features that are extracted from patients MR images, which have a close relationship with Lumbar Disc Herniation score [14]. The six features are the distribution of the protruded disc, the ratio between the protruded part and the dural sacs, and its relative signal intensity. Alomari *et. al.* [7], [15] proposed a probabilistic model for automatic herniation detection work by combining the appearance and shape features of the lumbar intervertebral discs. The technique models the shape depending on both the T1-weighted and T2-weighted co-registered sagittal views for building a 2D feature image. The disc shape feature is modelled using Ac-

tive Shape Model algorithm while the appearance is modelled using the normalised pixel intensity. These feature-pairs are then classified using Gibbs-based classifier. The paper reported that 91% accuracy is achieved in detecting the herniation. More details about the related research are available in our previous publication [6].

3   Proposed Approach

The main goal of for our work is to detect the disc herniation automatically. To reach this goal we need firstly to automatic label the disc area via machine learning and artificial intelligence. In our experiment we will concentrate on the gray level for each pixel in the area of interest of the lumbar spine MRI and the coordinate value for that pixel.

3.1   Feature Definition

Our feature set covers the coordinate value for each pixel in the disc area in addition to the intensity value for that pixel. Those features will be used to predict the disc area in the axial view MRI. Our feature set covers the coordinate value for X and Y coordinates and the intensity characteristic of the image content based on having different intensity level for each part of the lumbar spine in the MR image.

Training Data. One of the main steps for using the machine learning is to prepare a training data to train the classifiers. In our experiments, the training data is the data that contain the disc and non-disc area clearly labelled. This training data is provided by manually labelling each disc image to create a mask image. An example of such mask images with the corresponding MRI can be seen in **Fig. 1** .
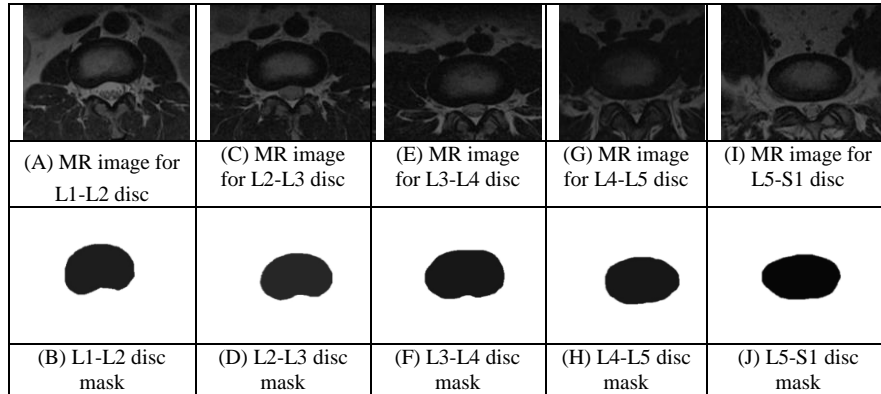

| (A) MR image for L1-L2 disc | (C) MR image for L2-L3 disc | (E) MR image for L3-L4 disc | (G) MR image for L4-L5 disc | (I) MR image for L5-S1 disc |
| (B) L1-L2 disc mask | (D) L2-L3 disc mask | (F) L3-L4 disc mask | (H) L4-L5 disc mask | (J) L5-S1 disc mask |

Fig. 1. Discs MRI and the developed Mask for each disc.

The intensity level for the disc on the mask is used for this comparison. The disc area on the mask filled in black colour with intensity level 0 and the remaining area of the mask image have the white colour with intensity level 255. The comparison will

work pixel by pixel. For the same disc, each two pixels with the same coordinate in the MRI and the mask will be compared. If the pixel on the original MR image matches a pixel with intensity 255 the output will be zero indicating a non-disc area and if the original image pixel matches a pixel with intensity zero in the mask then the output will be one indicating a disc area. At the end of this comparison, an extra column of data will be produced indicating the disc area which will be used for system training process.

System Classifier. There are a number of algorithms that could be used in the knowledge-based/artificial intelligence system. Our approach is to experiment with a number of classifiers and perform the training and classification process. The best classifier will be selected based on the achieved accuracy (high true positive and false negative rates as well as low false positive and true negative rates). To illustrate the training and classification process, the training will use the truth data which have been developed using a contrast weighted MR images as discussed earlier. Contrast weighted images are used to emphasis different types of tissues within the same MR images. The trained system will then be able to produce labelled images of the affected areas if a disc herniation is detected in the input MR images.

3.2 Performance Measurement

In order to evaluate the capability of our classifiers, we introduce a framework of performance measures, posed in conjunction with a series of comparator trials. In the comparator trials, we use models that are purposely selected from different theoretical classes. We reason that to be justified, our test classifiers should significantly outperform both linear classes of models and also simplistic models such as K-nearest neighbour (weak learners). Additionally, to demonstrate that the data we use as input contains true dependency, we compare the performance of all of the models presented with a random guessing baseline, showing that uninformed decisions are insufficient to produce significant results. Additionally, since generalisation is the goal of our classifier, we apply the described performance framework to the training (70 percent of the sample data), testing (20 percent of the sample data) and validation (10 percent of the data). A comparison of results between the reconstructed images from the classifiers with the original images is used to provide an indication of the classifier performance as described in section 5.2. To furnish the classifier responses with objective measures of performance, we utilise the scalar metrics method, which we now describe.

Scalar Metrics. To characterise the capability of the classifiers simulated in our experimental trials, we use a number of scalar measures, each of which provides a different summary of the deviation between our classifier outputs and the corresponding ground truth values. Such statistical measures are necessary, since complete discrimination between sets of outcomes is often unrealistic, prompting a trade-off between the types of errors committed. A listing of the measures and their derivations is presented in **Table 1** . We use accuracy which represents the most general correct classification proportion, grouping correct and incorrectly classified outcomes without reference to the underlying error types for each classifier then we depend on other statistical measures.

Metric	Abbreviation	Computation	Range
Area Under Curve	AUC	$0 <= area(ROC) <= 1$	[0,1]
Sensitivity	SEN	TP/(TP+FN)	[0,1]
Specificity	SPEC	TN/(TN+FP)	[0,1]
Accuracy	ACC	(TP+TN)/(TP+FN+TN+FP)	[0,1]
TP = True Positive Count, TN = True Negative Count FP = False Positive Count, FN = False Negative Count			

Table 1. Performance Metrics.

4 Experiment Result and Analysis

To evaluate the capability of our proposed classifiers, we conducted a series of empirical simulations, using the extracted feature from the MRI as our sample. The models used in our experiment are listed in **Table 2** .

4.1 Feature Origination

Three feature are extracted from the selected five MR images and the produced masks for 101 patients using Matlab as defined in section 3.1. Two experiments have been performed using two data sample. In the first set of experiments, we train the system using one patient data by selecting a random sample contain 7500 pixels from the five discs and use this sample in the system training. **Table 2** shows the list of classifiers and their result whereas, **Fig. 3Error! Reference source not found.** show the reconstructed output from the Weighted KNN model as a model with the highest accuracy. In the second set of expirements, we train the system by using 7000 pixels as a random sample from seven patients with 1000 pixel data selected from each one of them. **Table 2** shows the list of classifiers and their result while **Fig. 4** represents the reconstructed output from SVM Fine Gaussian as a model with the highest accuracy among the benchmarked models. The trials presented in this work were conducted using a common dataset partitioning scheme. We reserved 70% of the data for model training and the remaining 30% was divided 20 for testing and 10% for validation. The results of our experiment are divided into two parts the first part is the classifier training result and this will be explained in section 4.2 whereas the second part is the reconstructed image of predicted data from the classifier which has the highest accuracy and this explained in section 4.3.

4.2 Classifier Training Result

The results from our experimental procedure are presented and organised for each respective classifier for the first experiment in **Table 2** when the data source was one patient details with 7500 pixels for the first experiment whereas, the second experiment the data source used seven patient details with 7000 pixels. We then proceed to present our evaluation of the classifiers according to the accuracy. We have experiment all the 23 available classifiers in Matlab. Only the best 12 classifiers with the

highest accuracy have been discussed. **Fig. 2** shows the classifiers for the two experiments and the achieved accuracy for each one of them. In our experiments, we concentrate on the disc area hence a positive value has been given when the classifier predict a disc area while a negative value is given when the non-disc area have been predicted. However, if the classifier is correctly predicted the area either disc or non-disc a true value is given in this case and a false value is given for the incorrect prediction.

Model	The first experiment Accuracy	The second experiment Accuracy
Complex Tree [16] [17]	98.4	94.0
Medium Tree [16] [17]	98.3	92.0
Fine KNN [18] [19]	98.4	89.1
Medium KNN [18] [19]	98.9	91.1
Coarse KNN [18] [19]	98.6	90.7
Cosine KNN [18] [19]	93.4	88.7
Cubic KNN [18] [19]	98.9	91.0
Weighted KNN [18] [19]	98.9	90.8
Quadratic SVM [20]	98.8	90.6
Medium Gaussian SVM [20]	98.7	90.8
Fine Gaussian SVM [20]	98.8	91.1
Ensemble RUSBoosted Trees [21] [22]	98.0	84.1
Key: *SVM Support Vector Machine* *KNN K-Nearest Neighbour*		*Note: The selected classifier is highlighted in the table*

Table 2. First and Second Experiment results based on the evaluation of 12 models
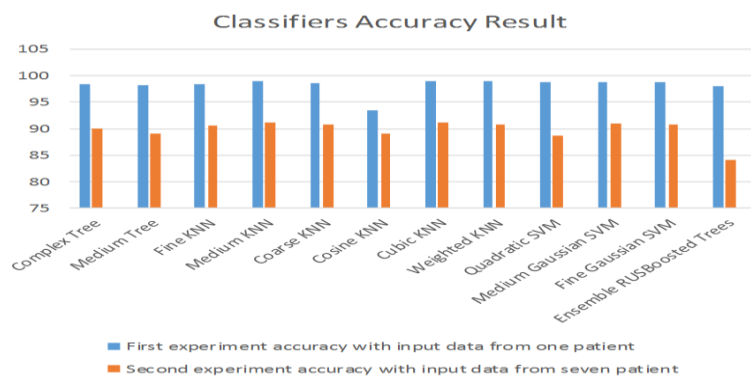

Fig. 2. The classifiers for the two experiments and the accuracy for each of them.

In our experiments, we concentrate on classifiers' accuracy as the quality of the constructed images depends on the value of the accuracy compared to the other performance measures. The best classifier for the first experiment was for the Weighted KNN, Medium KNN and Cubic KNN classifiers with accuracy = 98.9. Whereas, for the second experiment SVM Fine Gaussian classifier and Medium KNN got the same accuracy which is 91.1. However, Weighted KNN is the selected classifier in the first and SVM Fine Gaussian classifier for the second experiments respectively as they got the highest true positive rate. The selected classifiers will be used in subsequent steps and analysis.

4.3 Image Reconstruction

In this section, we have reconstructed the images based on the classifiers result as discussed in Section4.2. From the first experiment, we took the result of the Weighted KNN classifier for each x and y coordinate as an input to Matlab codewhich converts this matrix to image as shown in **Fig. 3** . On the other hand, the result of the SVM Fine Gaussian classifier has been selected in the second experiment and used as an

Fig. 3. The Original and Predicted Images by Weighted KNN
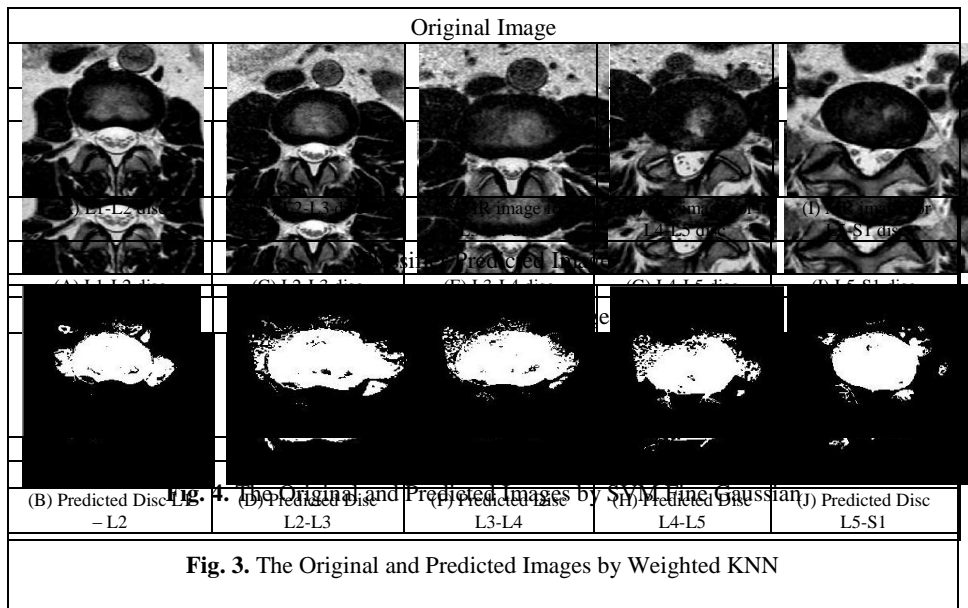
input with x and y coordinate to the Matlab code to reconstruct the output image. **Fig. 4** shows the predicted image for SVM Fine Gaussian classifier.

5 Discussion and Conclusion

Overall it can be seen from the two experiments that the highest accuracy for the first experiment using one patient data is 98.9% which is higher than the highest accuracy

in the second experiment using seven patients' data which reached 91.1%. Looking at the detail, the Medium, Cubic and Weighted KNN classifiers achieved the highest accuracy of 98.9% and as a second factor for selecting one of these classifiers we looked at the true positive rate for all of them to find a slight difference between them as medium and Cubic KNN classifiers got 93% while Weighted KNN got 94% and this classifier has been selected for the rest of work. In the second experiment, the Fine Gaussian SVM and Medium KNN achieved 91.1% as an accuracy rate but the ROC for the true positive rate for Fine Gaussian SVM was 62% and 60% for Medium KNN. Hence the Fine Gaussian SVM classifier was selected for the rest of the work. The second part of the experiment was to visualise the output from the selected classifier using a new patient data. The selected classifiers used to predict the disc area based on a new patient input for each disc in the lumbar spine in addition to the X and Y coordinates for the corresponding pixel. The classifier output and the X and Y coordinate have been used as an input to a Matlab revisualisation code to produce the output mask for the disc area from disc (L1 − L2) to (L5 − S1) as shown in **Fig. 3** from the first experiment while **Fig. 4** shows the output from the second experiment.

The reconstructed images in **Fig. 3** and **Fig. 4** show the discs area in the lumbar spine for the same patient. In general, the predicted images from the Weighted KNN classifier has better accuracy than SVM Fine Gaussian predicted images and we relate that to the difference in the classifier accuracy. However, the re-visualized images for both classifiers have some deformation if we compare it with the original image for that disc even with a very high accuracy rate which reached 98.9 percent in Weighted KNN. It is very important to have a very good quality mask from these experiments as this mask will be used to detect the disc herniation as mentioned earlier. The deformation in the output mask may relate to the data sample used to train the classifiers which need to be enhanced in a matter of the balancing the ratio between the two classes to improve the overall classifier performance outcomes. Additionally, it is also the possibility that the current features set are not sufficient to give the required output. We, therefore, set a plan for future work relating to this experiment in using new features in training the classifiers and test use the image processing technique to automatically remove the noise from the produced image in addition to the data balance between the disc and the non-disc area in the input data.

References

[1]   J. Koh, R. S. Alomari, V. Chaudhary, and G. Dhillon, "Lumbar Spinal Stenosis CAD from ClinicalMRMand MRI Based on Inter-and Intra-Context Features with a Two-Level Classifie," vol. 7963, pp. 796304-796304–8, 2011.

[2]   G. Waddell and A. K. Burton, "Occupational health guidelines for the management of low back pain at work: Evidence review," *Occup. Med. (Chic. Ill).*, vol. 51, no. 2, pp. 124–135, 2001.

[3]   R. Gordon and S. Bloxham, "A Systematic Review of the Effects of Exercise and Physical Activity on Non-Specific Chronic Low Back Pain.," *Healthc. (Basel, Switzerland)*, vol. 4, no. 2, p. 22, 2016.

[4] S. Dixon, "Diagnostic Imaging Dataset Annual Statistical Release 2013/14," pp. 1–27, 2014.

[5] Royal College of Radiologists, "2015 Clinical Radiology UK Workforce Census," *R. Coll. Radiol.*, no. September, 2016.

[6] A. S. Al Kafri, S. Sudirman, A. J. Hussain, P. Fergus, D. Al-Jumeily, M. Al-Jumaily, and H. Al-Askar, "A Framework on a Computer Assisted and Systematic Methodology for Detection of Chronic Lower Back Pain Using Artificial Intelligence and Computer Graphics Technologies," in *International Conference on Intelligent Computing*, 2016, pp. 843–854.

[7] A. Raja'S, J. J. Corso, V. Chaudhary, and G. Dhillon, "Automatic diagnosis of lumbar disc herniation with shape and appearance features from MRI," in *SPIE Medical Imaging*, 2010, p. 76241A--76241A.

[8] F. Taher, N. Werghi, and H. Al-ahmad, "Computer Aided Diagnosis System for Early Lung Cancer Detection," *Syst. Signals Image Process. (IWSSIP), 2015 Int. Conf.*, pp. 5–8, 2015.

[9] H. Arimura, T. Magome, Y. Yamashita, and D. Yamamoto, "Computer-aided diagnosis systems for brain diseases in magnetic resonance images," *Algorithms*, vol. 2, no. 3, pp. 925–952, 2009.

[10] M. Lootus, T. Kadir, and A. Zisserman, "Vertebrae Detection and Labelling in Lumbar MR Images," *Lect. Notes Comput. Vis. Biomech.*, vol. 17, no. i.

[11] J. H. Freidman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 209–226, 1977.

[12] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *Proc. 5th ACM-SIAM Symp. Discret. Algorithms*, vol. 1, no. 212, pp. 573–582, 1994.

[13] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1000–1006, 1997.

[14] H. Jiang, W. Qi, Q. Liao, H. Zhao, W. Lei, L. Guo, and H. Lu, "Quantitative evaluation of lumbar disc herniation based on MRI image," in *Abdominal Imaging. Computational and Clinical Applications*, Springer, 2012, pp. 91–98.

[15] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Lumbar spine disc herniation diagnosis with a joint shape model," in *Computational Methods and Clinical Applications for Spine Imaging*, Springer, 2014, pp. 87–98.

[16] E. Alpaydin, "Introduction to Machine Learning, third edition," p. 640, 2014.

[17] P.-N. Tan, M. Steinbach, and V. Kumar, "Classification : Basic Concepts , Decision Trees , and," *Introd. to Data Min.*, vol. 67, no. 17, pp. 145–205, 2006.

[18] X. Wu, V. Kumar, Q. J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.

[19] N. Bhatia and C. Author, "Survey of Nearest Neighbor Techniques," *IJCSIS)*

Int. J. Comput. Sci. Inf. Secur., vol. 8, no. 2, pp. 302–305, 2010.

[20]   C. C. Aggarwal, *Data classification : algorithms and applications*. 2015.

[21]   C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and   a. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," *2008 19th Int. Conf. Pattern Recognit.*, no. March 2016, pp. 8–11, 2008.

[22]   C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 40, no. 1, pp. 185–197, 2010.