

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

# Detecting tympanostomy tubes from otoscopic images via offline and online training



Computers in Biology and Medicine

毘 相

## Xin Wang<sup>a</sup>, Tulio A Valdez<sup>b</sup>, Jinbo Bi<sup>a,\*</sup>

<sup>a</sup> Department of Computer Science and Engineering, University of Connecticut, 371 Fairfield Way, Unit 4155, Storrs, CT 06269-4155, USA <sup>b</sup> Department of Ototaryngology, Connecticut Children's Medical Center, Hartford, CT 06106 USA

#### ARTICLE INFO

Article history: Received 24 November 2014 Accepted 23 March 2015

Keywords: Object detection Tympanostomy tube Otoscopic image Cascaded classifier Support vector machine

## ABSTRACT

Tympanostomy tube placement has been commonly used nowadays as a surgical treatment for otitis media. Following the placement, regular scheduled follow-ups for checking the status of the tympanostomy tubes are important during the treatment. The complexity of performing the follow up care mainly lies on identifying the presence and patency of the tympanostomy tube. An automated tube detection program will largely reduce the care costs and enhance the clinical efficiency of the ear nose and throat specialists and general practitioners. In this paper, we develop a computer vision system that is able to automatically detect a tympanostomy tube in an otoscopic image of the ear drum. The system comprises an offline classifier training process followed by a real-time refinement stage performed at the point of care. The offline training process constructs a three-layer cascaded classifier with each layer reflecting specific characteristics of the tube. The real-time refinement process enables the end users to interact and adjust the system over time based on their otoscopic images and patient care. The support vector machine (SVM) algorithm has been applied to train all of the classifiers. Empirical evaluation of the proposed system on both high quality hospital images and low quality internet images demonstrates the effectiveness of the system. The offline classifier trained using 215 images could achieve a 90% accuracy in terms of classifying otoscopic images with and without a tympanostomy tube, and then the real-time refinement process could improve the classification accuracy by 3-5% based on additional 20 images.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Tympanostomy tube placement is the most common outpatient surgical procedure in the United States. Each year more than 650,000 children younger than 15 years of age receive tympanostomy tubes [1]. Tympanostomy tubes are commonly inserted because of persistent middle ear fluid affecting hearing, frequent ear infections or ear infections that persist after antibiotic therapy [2]. All of these conditions are composed of the term otitis media, which is second in frequency only to upper respiratory infection as the most common illness diagnosed in children by pediatric health providers. Most children will experience at least one episode of acute otitis media by the age of 3 years and by the age of 6 years, nearly 40% have experienced three or more infections [3]. Otitis media with effusion (OME) can resolve spontaneously, however for patients in which, there is persistent middle ear fluid for at least 3 months with the decrease in hearing, further treatment in the form of tympanostomy tubes

http://dx.doi.org/10.1016/j.compbiomed.2015.03.025 0010-4825/© 2015 Elsevier Ltd. All rights reserved. insertion may be recommended [2]. Fig. 1 illustrates the placement of the tympanostomy tube on the ear drum.<sup>1</sup>

Since first described in 1954 by Armstrong, tympanostomy tube placement has become the surgical treatment of choice for otitis media [4]. Placement of tympanostomy tubes improves hearing significantly in the presence of otitis media with effusion, reduces the incidence of recurrent acute otitis media (AOM), and provides a mechanism for drainage and administration of topical antibiotic therapy for acute otitis media. The latter has gained significant importance since it allows localized treatment for ear infections rather than systemic antibiotics use. The insertion of tympanostomy tubes involves aspiration of the middle ear fluid leading to instant improvement of hearing thresholds. The tympanostomy tubes are designed to extrude naturally from the tympanic membrane normally within 6 months to 1 year following the procedure.

Following tympanostomy tube placement, regular scheduled follow-ups are recommended by the American Academy of Otolaryngology every 6 months to check the status of the ear tubes [5]. This follow up care is currently performed by an ear, nose and throat

<sup>\*</sup> Corresponding author. Tel.: +1 860 486 1458. *E-mail address:* jinbo@engr.uconn.edu (J. Bi).

<sup>&</sup>lt;sup>1</sup> The left graph of Fig. 1 is downloaded from www.kidshealth.org.nz.

specialist resulting in increased cost compared to the cost of a visit to a general practitioner [6]. The overall complexity of the procedure is low because it mostly identifies the presence and patency of the tympanostomy tube. If this procedure can be performed by an automated computer program, it can significantly reduce cost as well as enhance the clinical efficiency of both ear nose and throat specialists and general practitioners.

Computer vision techniques have been widely used to automatically detect an object in natural images such as human faces [7-10], cars [11-13] and pedestrians [14-16]. The related techniques are also used to analyze medical images which have largely shown to facilitate the diagnosis and treatment decisions [17–23]. In this study we have developed a computer vision system to predict on an image if a tympanostomy tube is in place and collect feedback to adaptively improve the system classification performance. In this system, we extract a set of image features including RGB intensity features, edge-map based features and other advanced features such as Scale-invariant Feature Transform (SIFT) [24], and Histograms of Oriented Gradients (HOG) features [25]. We use them to learn a discriminative model to determine if an otoscopic image presents a tympanostomy tube. A cascaded classifier is constructed by the support vector machine (SVM) [26] algorithm from labeled image patches. The prediction of this classifier, when applied to a test image, is visualized. A refinement process is designed to refine this trained classifier at the point of patient care according to user feedbacks. Extensive experimental results demonstrate the effectiveness and efficiency of the proposed approach.

## 2. Related works

Most object detection techniques are composed of two major aspects, feature extraction and model learning, for each of which various methods have been proposed. Readers can consult with [27,28] for an overview and challenges of the field. In the subsequent paragraphs, we briefly review several of the most relevant methods.

Region based and edge-map based features are the two types of widely used features for object detection. For the first type, the features are generated from colors [29–32], or the varying distribution of intensities [33,34,24,25]. Among these features, the SIFT [24] and HOG [25] features are the most widely used nowadays. Other algorithms have considered texture information, such as gray level co-occurrence matrix (GLCM) [35], local binary patterns (LBP) [36] and wavelet texture [37,12]. We explored these features in our system. The edge-map based features are used to capture contour shapes by computing the most representative edge fragments [38–41]. These features are robust to occlusion but are not invariant to illumination conditions. Geometrical shapes and the structure of lines and arcs in an edge map were also studied in [42–44], which facilitated to obtain more reliable features. It has been shown that combining the region based features and edge-map features may lead to robust detection [45,46]. We hence implemented feature extraction methods in both of the categories.

Various machine learning algorithms have been applied to build probabilistic models for prediction of object classes. Typical methods include Bayesian classifiers [47], expectation maximization [48], *k*-Nearest Neighbor [49], logistic regression [50] and support vector machines (SVM) [26,51]. These methods perform comparably although some may serve certain specific purposes, such as selecting features for use in the model. In this work, an SVM algorithm with linear classifiers was used and served as a good learning model for detecting the tubes. The SVM algorithm solves a quadratic programming optimization problem for a predictive classifier that has an optimal margin to separate different classes of examples.

Besides the two major components discussed above, online learning methods that sequentially take user feedback in building a classifier have been used to detect and track moving objects [52–55]. Our system also consists of a real-time refinement process. Although similar to an online learning method, our refinement targeted at the unseen examples that the offline models fail to predict. After the system highlights the predictions on an image, the users can correct the prediction results on the image. Our system is capable of recording this feedback and retraining the related classifiers.

## 3. Material and methods

In this section, we describe our approach to automatically predicting if an otoscopic image of ear drum has a tympanostomy tube mounted. Fig. 2 provides an overview of our detection system. In this system, an image is first converted into a feature vector in the feature extraction process. Then a classifier that has been pre-trained and stored in a database (which we call knowledge database) is applied to the image features to predict the presence and location of a tube. The results are then visualized and highlighted in the image. Users can give feedback by correcting the predictions if they are inaccurate. This feedback message is then recorded and used in an automatic process to refine the classifiers stored in the knowledge database. This system is constructed



Fig. 1. Illustration of a tympanostomy tube in place and patency in the anterior inferior quadrant of the tympanic membrane: (left) a graph of ear structure and tube and (right) an otoscopic image.



Fig. 2. The system overview.

based on two processes: an offline learning process to construct the classifiers in the knowledge database; and an online refinement process that takes in user correction information to refine the classifiers.

The construction of a knowledge database aims to incorporate domain knowledge and data-driven classifiers derived from labeled images. During the offline classifier training, we collected otoscopic images of patients from Connecticut Children's Medical Center (CCMC). These images are labeled manually with respect to whether or not it shows a tympanostomy tube. We build classifiers, as functions of image features, to distinguish images with a tube from images without. The usefulness of the various features extracted from the images is examined by building classifiers with different combinations of features. We propose a new cascaded classifier which is composed of three lavers. These lavers are designed according to the domain knowledge. For instance, we have observed that many tympanostomy tubes are in green color. It is hence important to design an efficient classifier to detect green tubes which might also be easier given the target is more characterized and focused on the green channel of the image. Each layer aims to detect tympanostomy tubes of a different characteristic. For each layer of the cascaded classifier, the sample of labeled images was split to have training and test sets.

During the real-time system refinement, a new image is given to the system, which first converts it to a feature vector. The classifiers in the database can make a prediction based on the feature vector. Then the system yields a marked image that highlights the results. If the classifier fails to point out the presence of a tube (false negative) or predicts as it contains a tube but gives a wrong location, a user may mark the correct location of the tube. If the classifier claims a tube in the image but there is actually not (false positive), a user can remove the incorrectly located markers. Then this image (actually its feature representation) will be added to the training set stored in the knowledge database and all the three layers of the cascaded classifier will be retrained.

#### 3.1. Feature extraction

We describe the three sets of features that are extracted from the otoscopic images: color-based features, edge-map based features and a set of advanced features.

#### 3.1.1. RGB intensity features

Tympanostomy tubes are built with conspicuous colors that are different from the tissues normally seen in the ear. For example, green, purple, blue or white are often used for the tympanostomy tubes. For an image of true color, each pixel is typically represented by the component intensities of red, green and blue. We compute the intensity of the three channels in the image and then measure the "directional" difference of the intensities between any two channels. For instance, the difference from the red channel to the green channel is calculated by subtracting the intensity of the red channel from that of the green channel. If the difference along a direction is a negative value, we re-set it to 0. We exclude the pixels in the completely dark regions in the redundant margin of the image. We hence obtain 6 quantitative values of the difference at each pixel. The mean, maximum and standard deviation of each of the 6 differences are computed over all valid pixels in an image, which gives us  $3 \times 6 = 18$  features for each image. These features are used to predict tympanostomy tubes on the basis that the images containing a tympanostomy tube have different RGB intensity distributions than the normal ear images.

The RGB intensity features may be most useful for distinguishing a specific kind of tubes. For instance, we notice that most of our clinical images show tympanostomy tubes of green color. Hence, we could consider detecting the green tubes in the first layer of the cascaded classifier. Fig. 3 illustrates the RGB features by an example of the original image and two images of the difference computed using the intensities of the green channel minus those of the red and blue channels, respectively.

#### 3.1.2. Edge-map features

An obvious characteristic of the tympanostomy tube is its circular shape if it is viewed from the side, which is also the view of the tube in the ear images. We hence design features to capture circular structures presented in an image. For each pixel, we calculate the entropy value based on its  $9 \times 9$  neighborhood, which measures the "business" of that region. The pixels along an edge are expected to have higher entropy values than those far from an edge. All entropy values are normalized to the range of [0, 1]. Since we care about the pixels that have high entropy values, we set a threshold of 0.8. More precisely, if a pixel has an entropy of lower than 0.8, the pixel's entropy will be set to 0; or otherwise set to 1. This step converts an ear image into a binary image showing the edges, which is called an edge map. We then compute the Hough-transformation [56] of the edge map and apply the Phase-Coding algorithm [57] to detect circles.

For detecting the center of a circle, we restrict the search to pixels at the center area of the image that is within 70% of its length and width. We search for circles of varying radius. Let the length of the short side of the image be *L*. The maximal limit of the radius is set to  $L \times [2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}]$  and the corresponding minimal length is one-third of the maximal limit. Empirical experiments show that the circles with radius in the range of [L/10, L/3] can match the size of tympanostomy tubes in most of



Fig. 3. Features computed from the RGB intensities.



Fig. 4. A demonstration of the process of transforming the original image into the edge map from (a) to (c), and the application of Hough transformation to detect circles.

the images. By setting up these limits, we can screen out most of the unwanted circles. Fig. 4 shows an exemplar image (Fig. 4(a)) that is transformed to a edge map (Fig. 4(c)) via entropy computation (Fig. 4(b)) and a circle shows the detected location of a tympanostomy tube.

To further prune circles that do not correspond to a tympanostomy tube, we crop image patches within each detected circle, and extract the complete set of RGB intensity features from each patch. Let  $T_i$  denote the number of detected circles in the *i*-th image, and  $\alpha_i^j$  contains the RGB features for the *j*-th patch in the *i*-th image. Let  $\beta_i^j$  denote the RGB features of the rest area in the image. We quantify a numerical feature for each image *i* based on the circle detection as max {*Euclidean*( $\alpha_i^j, \beta_i^j$ ),  $j = 1, ..., T_i$ }, which is to compute the Euclidean distance between the RGB feature vectors of each cropped patch and the complement of the patch, and then take the maximum distance across all detected circles. This feature reflects that the RGB intensity of the tube area should differ the most from its ambient surrounding. For the circle that attains this maximum distance, its center location and radius are also used as numerical features for the image. If no circles are detected, the distance feature is set to 0, the center coordinates are set to (-1, -1), and the radius is set to 0.

## 3.1.3. Other features

Besides the features computed from RGB intensities and edgemaps to capture the color and shape of a tube, HOG [25] and SIFT [24] features are used by adopting the Bag of Visual Words method [58]. For HOG features, we extract 31 default HOG features from each cell of  $8 \times 8$  pixels. Then, a cluster analysis via *k*-means partitions the cells of all images into 20 clusters. The centroids of the clusters form the words in a vocabulary. By counting the number of cells in each of the 20 clusters, each image is converted to 20 features. For SIFT features, we first compute 128 SIFT features from patches of  $20 \times 20$  pixels using the default setting of SIFT. Then the same cluster analysis procedure is used to compute 20 SIFT-based features.

We extract texture-based features using the gray level cooccurrence matrix (GLCM). Four features are computed to characterize the GLCM including "Contrast", "Correlation", "Energy" and "Homogeneity" [35]. Local Binary Patterns (LBP) [36] and wavelet texture might also be useful for the detection of tubes. Fifty eight LBP features are extracted. For wavelet features, at first we perform a single-level two-dimensional wavelet decomposition of an image, which yields an approximation coefficient matrix, as well as horizontal, vertical, and diagonal coefficient matrices. We extract 8 wavelet features by computing the average values and standard deviations of the four matrices.

#### 3.2. Learning system

Besides the feature vectors of the training images, the knowledge database also stores a cascade of classifiers, which are first trained in an offline process based on collected training data, and then refined by an online process based on user feedback at the point of care.

#### 3.2.1. Offline classifier training

The SVM algorithm is employed to train each layer of the cascaded classifier. Given *n* images, each characterized by a data point  $(x_i, y_i)$  where  $x_i \in \mathbb{R}^m$  is the feature vector of image *i* and  $y_i$  is the binary label indicating whether it shows a tube. A linear model  $w^T x + b$  is used to classify images, where  $w \in \mathbb{R}^m$  and *b* are the weight vector and the bias term of the model, respectively. The SVM method minimizes the regularized loss function  $\lambda ||\mathbf{w}||^2 + \sum_i \xi_i$  for the best  $(\mathbf{w}, b)$  where the losses  $\xi_i$  are defined by constraints  $y_i(\mathbf{w}^T x_i + b) \ge 1 - \xi_i$ , and  $\xi_i \ge 0$ , i = 1, 2, ..., n and  $\lambda$  is the regularization parameter. This optimization problem formulates a quadratic program, and can be solved using CPLEX solvers [59]. After the linear model is constructed, it can be used to determine the label of a new image represented by  $\hat{x}$  by computing  $\operatorname{sign}(\mathbf{w}^T \hat{x} + b)$ .

The cascaded classifier consists of three layers as shown in Fig. 5. Classifiers of the three layers are trained separately with different sets of features. The first layer is designed to detect the green-color tubes. Since the green tubes are widely used and commonly seen in otoscopic images, a simple classifier capable of detecting them would facilitate the detection task and improve the efficiency. RGB features related to the green channel are extracted from the original images and used in the SVM training. In our

evaluation, the resultant classifier proved to be effective and detected all green tubes in the otoscopic images that were collected at Connecticut Children's Medical Center (CCMC).

Images downloaded from the internet contained tubes of many other colors. We use all internet images and CCMC images with non-green tubes to train the second layer of the cascaded classifier. The second layer is designed to not only utilize the color property of the tympanostomy tubes but also their shape information. We train a SVM model using all RGB intensity features and the shape features from edge-maps, and hence each image is represented by a vector of 18+4=22 features.

In the third layer of the classifier, additional features as described in Section 3.1.3 are extracted. We have also experimented with each type of the advanced features by combining them



Fig. 6. The procedure for testing the cascaded classifier on a newly obtained image.



Fig. 5. The offline procedure to train the cascaded classifier.

individually with the basic RGB and shape features in training the third classifier.

Fig. 6 shows the test process after the cascaded classifier is constructed. If the first classifier classifies the image as a positive example, meaning that it contains a tube (likely a tube of green color), then the system outputs the prediction, and directly goes to the visualization step. Otherwise, this image is moved to the second classifier, and then to the third. If any of the classifiers accepts this image as containing a tube, the system will come to a final decision. If all three classifiers predict the image negative, then the system will report it as not containing a tube.

#### 3.2.2. Real-time classifier refinement

The real time refinement step allows the user to adjust the classifier by providing feedbacks to the system. The motivation of designing this refinement step comes from the fact that otoscopic images may show tympanostomy tubes with significantly varying colors, shapes, illuminance and scales when they are taken by different clinicians, or collected from different healthcare organizations. Clinicians may have different gestures when shooting the pictures. As a result, tubes are presented distinctly in angles and scales in the images from different practices. It is hence desirable to have a system that is able to adapt the classifiers to the specific setting of an organization.

However, this refinement step is not a necessary component for the system to perform. The refinement will be triggered only when a user corrects the system prediction on a test image. Hence, it runs in a semi-automatic mode in the sense that as soon as a user makes correction and saves it, the system will store the corrected data into the database and retrain the classifiers.

Fig. 7 shows the refinement procedure. If the system misclassifies an image that actually contains a tube, the user can change the label of the image to be "+1", and specifies the center and radius of a circle to properly mark out the tube region. If the system predicts correctly but fails to mark the tube location correctly, there exist two cases. The first case is that the tube is largely visible on the image but the system marks a wrong position. The user can correct those markers. The second case is that the tube is obscured badly by wax or other structures. In this case, no circles may be seen on the image, but the user may still leave a marker to the area that he perceives to contain the tube. If

an image does not contain a tube, but the system makes a false positive prediction, the user can just correct the label and remove all the markers on the image. Based on the feedbacks, the system will automatically refine the extracted features by adjusting the parameters used in feature extraction algorithms.

The three layers of the cascaded classifier will be retrained using the augmented training set that includes the test image. Note that to retrain classifiers, the system does not need to hold the raw images in the knowledge database but only the extracted feature vectors of all training images and the specific test image. The database may increase over time when more and more test images receive corrections, but the storage of the image features only requires a moderate amount of space. The size and model parameters of the classifiers will be updated in the knowledge database once retrained.

## 4. Empirical evaluation

The proposed system was validated on otoscopic images collected from two sources: 235 images of real patients collected from CCMC and 40 images downloaded from internet. The CCMC images were taken with high quality and high resolution, and 77 of them contained a tube. The internet images from Google or YouTube had overall low quality: low resolutions with higher level of artifacts and distortion, and 27 of them showed a tube.

## 4.1. Experimental setting

We randomly selected 40 images to form a hold-out set in order to evaluate the system performance, including evaluating both the offline training and the online refinement. We excluded 20 images from offline classifier training and used them to simulate the online refinement process. The rest of 215 images were partitioned into 3 even subsets for use in a three-fold cross validation (CV) procedure. We stratified the partition so that each of the above sets of images (totally, 5 of them) had equal ratios of positive versus negative images and CCMC versus internet images.

To train a classifier, the regularization parameter  $\lambda$  used in the SVM algorithm was tuned using the three-fold CV process where two of the CV subsets were used to train a classifier with the



Fig. 7. The real-time refinement procedure to retrain the system at the point of care.

pre-chosen value of  $\lambda$ , then the remaining subset was used to test the resultant classifier. We ran the CV process with each choice of  $\lambda$  from the pre-chosen values  $[2^{-14}, 2^{-8}, ..., 2^0, 2^1, ..., 2^{14}]$ . The choice of  $\lambda$  that gave the best test classification performance was used in the SVM algorithm to train the final classifier with the full training data. The three classifiers in the cascade were all trained separately using the above CV process. Note that the first-layer classifier was trained with positive images of only green tubes whereas the second-layer classifier was trained with those positive images excluded. All 215 images were used to train the thirdlayer classifier.

In the classifier cascade, the first layer used 6 RGB intensity features, and the second layer used 22 features including 18 RBG features and 4 edge-map based features. For the third layer, 20 HOG, 20 SIFT, 58 LBP, 4 GLCM and 8 Wavelet features were extracted. Together with the 22 baseline features used in the second layer, we considered totally 132 features in this layer. However, we observed that using all of the advanced features did not necessarily improve the detection accuracy. Hence, we experimented with each set of the advanced features by combining the 22 baseline features with one set at a time to find the best classifier for the third layer.

## 4.2. Results

In the first two layers of the classifier cascade, final decisions were made only for those images that were classified as positive by any of the two classifiers. Hence, sensitivity (true positive rate) was evaluated at the end of the second layer, but the final specificity (true negative rate) would have to be evaluated until the full cascade was applied. We observed that the 22 baseline features were useful because the first two classifiers, by themselves, already achieved a sensitivity of 78% as estimated in the CV (averaged over the test sets of the three CV folds), and also an overall accuracy 87% if treating all non-classified images as predicted negative at the second layer.

Table 1 shows the CV performance of the full cascaded classifier where we tested the classifier separately on the set of high quality CCMC images and the enlarged set including low quality internet images. The best validation error rates were in bold fonts. From Table 1, we see that each of the advanced feature sets can improve the classification accuracy from the baseline features (from the sensitivity 78%). The classifiers using additional HOG and Wavelet features achieved the same validation error on the CCMC image set but one had slightly better sensitivity and the other slightly better specificity. The classifier using additional Wavelet features had slightly better overall performance when tested on the enlarged image set.

#### Table 1

Cross validation performance of the offline learning system.

Data sets	3rd layer features	Error rate	Sensitivity	Specificity
High quality CCMC image set	HOG SIFT LBP GLCM Wavelet	<ul> <li>0.03</li> <li>0.04</li> <li>0.04</li> <li>0.05</li> <li>0.03</li> </ul>	0.94 0.88 0.92 0.92 0.92	0.98 0.99 0.98 0.97 0.99
CCMC + low quality internet image set	HOG SIFT LBP GLCM Wavelet	0.06 0.06 0.07 0.08 <b>0.05</b>	0.89 0.91 0.85 0.88 0.92	0.96 0.94 0.98 0.95 0.95

We also tested classifiers that were based only on the advanced features with a variety of feature combinations. The classification errors ranged from 30% to 45%. This showed that although these advanced features improved our classifiers, they were by themselves not as effective as the baseline features that we proposed to use (because the first two classifiers using the proposed features achieved an overall accuracy of 87%).

To further measure the classifier performance, the receiver operating characteristic (ROC) curves were drawn to show validation performance in the CV in Fig. 8. The ROC plot is widely used to characterize the performance of a classifier where the area under the curve (AUC) is a statistic indicating the overall classification performance. We observed that the classifier with SIFT features showed comparable performance on the high quality images (shown in Fig. 8(a)), but dropped its sensitivity on the enlarged image set at the region of low false positive rate (shown in Fig. 8(b)). Other classifiers performed similarly on the enlarged data set.

To tune the hyper-parameter in the SVM algorithm, Fig. 9 shows how the classification performance varies when the regularization parameter changes in the range from  $2^{-14}$  to  $2^{14}$  for each layer of the classifier. We used the parameter choices that gave the best CV performance for each feature set in the experiments and also used them in the refinement system. After a proper value of  $\lambda$  was chosen, a classifier was obtained by running SVM on the full training data, and then tested on the hold-out set with the test performance shown in Table 2 (top section).

The classifier was then refined using the 20 images we held for the refinement purpose. We conducted a simulation of how the system will be refined when deployed into a clinical use. The 20 images were randomly ordered and shown to the system in sequence. After each image was tested, and if a user (we) made correction, the system adjusted the features based on the user feedback on the image and retrained the classifiers. After all 20 images were used, the resultant classifier was tested on the hold-out set again. Table 2 (bottom section) shows the refinement performance.

Table 2 shows that the offline-trained classifier with additional HOG features achieved the best classification error rate and best specificity. Then, the refinement step based only on 20 test images could already improve the detection performance by 3–5%. The best classification error rate was obtained by the classifier using the baseline and HOG features. This classifier's accuracy was increased by 7% in sensitivity and 1% in specificity by the refinement step.

Fig. 10 visualizes exemplar images after the system makes predictions for an image. The identified tubes were marked by the red circles detected from edge-maps. It can be seen that our system was able to detect the tubes of a variety of colors, as well as tubes in different orientations or partially covered by obstructs. Two images that were errors of the offline-trained system are also shown in Fig. 10 and the errors were corrected during the refinement phase. Fig. 11 shows an image used in the refinement step. The offline-trained system was not able to correctly locate the tube (shown in Fig. 11(a)), but a user marked out the tube area with a blue circle (in Fig. 11(b)). After the system retrained the classifier, the tube was correctly identified in this image.

Fig. 12 reports the run time of the proposed system when running it in a Dell Precision T3500 machine. Fig. 12(a) shows the average run time of feature extraction for an image in which we compared the run time of extracting features from RGB intensities and from edge maps. Since our empirical experiments show that the system achieves the best performance while integrating HOG features, we also included the averaged run time for computing HOG features in Fig. 12(a). Fig. 12(b) shows the run time for training each layer of the cascaded classifier when all 215 training images are used. The refinement step takes a similar cost to that of training due to its incremental nature. It can be seen that classifier



Fig. 8. ROC comparison of the cascaded classifier trained with different choices of advanced features. Note that the axis ranges are adjusted (zoomed in) for clear illustration.



Fig. 9. The plots of validation error rate versus the regularization parameter for each classifier in the cascade. (a) First layer classifier, (b) second layer classifier, and (c) third layer classifier.

Table 2					
Detection	performance of	on the	hold-out	images.	

Data sets	3rd layer feature	Error rate	Sensitivity	Specificity
Hold-out set using offline classifier	HOG SIFT LBP GLCM Wavelet	<b>0.10</b> 0.15 0.12 0.15 0.13	0.85 0.82 0.81 0.80 0.92	0.92 0.87 0.92 0.88 0.85
Hold-out set using refined classifier	HOG SIFT LBP GLCM Wavelet	0.07 0.10 0.10 0.13 0.08	0.92 0.94 0.87 0.87 0.92	0.93 0.87 0.92 0.88 0.92

training only takes a small portion of the total run time, which also ensures the efficiency of the real-time refinement. From Fig. 12, the most time-consuming step lies in the feature extraction, and specifically in the edge-map feature computation. This is because we repeatedly detect circles in an image using a variety of radius ranges.

Overall, our system was able to effectively detect tympanostomy tubes out of different sets of images with varying presence of tympanostomy tubes, and with varying degrees of image qualities. When presented with a test image, the system would mark around the area where it thought the tube was located (see Figs. 10 and 11 for examples). On the high quality images we obtained a specificity above 95% using all three classifiers. Our sensitivity was also high with only SIFT features having a sensitivity below 90% at 88%. There was a drop in both the sensitivity and specificity when the classifier was tested on a mixed set of high quality and internet



**Fig. 10.** (a)–(d) are exemplar images that show tympanostomy tubes of different colors and are correctly detected by the offline learning system. The system also correctly locates the tubes. (e, f) show the errors made by the offline system. (e) shows a false negative where a green-color tube is obscured severely. Although the system correctly locate two circles in the tube area, the system misclassifies the image as not containing a tube. (d) shows a false positive where a bright area presents a circle-like structure. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)



**Fig. 11.** (a) shows an image used for real-time refinement. The classifier fails to label it with a tube although several circles are detected. In (b), a human user can mark out the center location of the circle for the tube area and specify the radius (the additional circle). The label of this image and its features will be updated based on the feedbacks. The offline-trained classifiers will be retrained using this image together with other images stored in the knowledge database. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



Fig. 12. Run time of the system in feature extraction and in classifier training. (a) shows the average run time of extracting features for an image, and the computation costs vary among the three groups of features. (b) shows the run time for training each layer of the cascaded classifier.

images with sensitivities and specificities reducing to 89% and 95.5% on average. The classifier using the baseline plus the HOG features achieved the best performance on both the high quality CCMC image set and the enlarged image set after classifier refinement. This performance was boosted to 92% and 93%, respectively, in sensitivity and specificity on the enlarged image set. Based on the experimental results, the HOG features (as shown in Table 2) and Wavelet features (as shown in Table 1) may be more useful in detecting the tympanostomy tube than other advanced features.

## 5. Discussion

Tympanostomy tube placement is the primary surgical intervention for otitis media, which is a pediatric health problem worldwide. In a cross-sectional questionnaire study of 40,000 Norwegians, the estimated life-time prevalence of tympanostomy tube surgery was about 12% [60]. Children under the age of 7 years of age are at increased risk of otitis media due to the combination of an immature immune system and a poor function of the Eustachian tube [61]. The Eustachian tube is a slim connection between the back of the nose and the middle ear space and equalizes pressure with the external environment. Tympanostomy tubes bypass the Eustachian tube by allowing equalization of middle ear pressures with the outside environment. Placement of tympanostomy tubes significantly improves hearing, reduces the prevalence of middle ear effusion and may reduce the incidence of recurrent acute otitis media.

With well over 650,000 performed each year in the United States at an approximate cost of \$2700 per surgery [62] the contribution to health care costs is approximately 1.8 billion [2]. This, however, does not account for follow up care. Follow-up of a tympanostomy tube surgery is normally performed to verify the position of the tympanostomy tube in the tympanic membrane and assess the patency of the tympanostomy tube. This examination is performed using an otoscope which is an instrument that allows illumination of the tympanic membrane with magnification which is interpreted in succession by a physician.

The American Academy of Otolaryngology and Head and Neck Surgery recommends the initial control within one month after tube placement and then at least once every six months until the tubes extrude. Current check ups are performed by ear nose and throat specialists at a significant financial cost for a low complexity visit. Austad et al. [6] studied the possibility of utilizing primary care physicians for tympanostomy tube check-up as a cost saving measure with no difference in audiological and subjective hearing outcomes two years after the surgery. There are however complications that can arise from tympanostomy tube surgery such as otorrhea (drainage coming through the tympanostomy tube) and tympanic membrane perforation, which physicians must be aware and vigilant not to miss [63].

Although computer aided diagnosis has been around for well over a decade, its role has been mostly limited to the fields of radiology and pathology with little participation in the process of real time medical decision making with patients. We believe there is the potential of computer image interpretation to perform simple tasks allowing direct participation in patient care. Identifying the presence of tympanostomy tubes may offer an alternative to a physician follow-up visit in the future for patients with tympanostomy tubes. Recent studies have revealed a low up rate of post operative follow up compliance after placement of tympanostomy tubes [64] and new commercially available technologies including adaptable otoscopes to cell phones cameras can provide those images from the comfort of the patient's home. A clinical trial is currently being considered to evaluate the clinical usability and efficiency of our computer vision system. In the trial, we are not proposing complete replacement of the physician by machine image recognition algorithms but for patients with an uncomplicated course it may obviate an unnecessary trip to the physician's office. It can be an intermediate step in between visits and reducing the cost of the follow up and it may help with the overall follow up compliance.

We have identified several factors that make the identification of tympanostomy tubes a good fit for computer vision. First tympanostomy tubes consistently have an easily identifiable circular shape and second they are usually made of bright color plastic compounds usually green, blue, or black, which are distinguishable from any human tissue. In this study we have successfully implemented a computer vision system to approach the problem of identifying tympanostomy tubes from otoscopic images. We incorporated a variety of ear tubes of various colors and in various locations within the tympanic membranes achieving adequate detection rate in up to 90% of cases. We have also used images of various quality to test reliability of our methodology even with less than ideal images.

This study represents the first attempt at incorporating computer vision as a low cost alternative for a low acuity common visit of tympanostomy tube check-up. There are, however, limitations to our technique which are common to otoscopic examinations such as limited visualization due to cerumen. The current realtime refinement step retrains the system's classifier as long as a user feedback is received. A more robust system may need to select among the test images corrected by users to evaluate if inclusion of such an image in classifier training will enhance system performance. Another limitation would be to adequately identifying patency of the tympanostomy tubes since adequate visualization of the lumen of the ear tube may not be satisfactorily achieved without multiple images from various angles which we have not attempted in this study. Future directions may involve utilization of multiple wavelengths to achieve different depth of penetration of light through the tympanic membrane and contrast enhancement as well as multiple images of the same ear at different angles to be able to assess for patency.

### 6. Conclusions

We have proposed and developed a computer vision system for automatically detecting a tympanostomy tube in an otoscopic image of the ear drum. The proposed system comprises an offline classifier training process and a real-time system refinement process. The offline learning process creates a cascaded classifier with 3 layers that is designed according to our domain knowledge and observation on the various characteristics of tympanostomy tubes. A state-of-the-art supervised learning method, support vector machine, has been used to train the cascaded classifier that is then stored in a knowledge database of the system. The refinement process allows users to interact with the system aimed at system adaptation over time to a user's specific setting. It seeks user feedback on the prediction result of a test image during the point of check-up. Based on the user feedback, the system can automatically augment the training set by including the features extracted from the test image to refine the classifiers stored in the knowledge database. We propose the method to extract the RGB intensity features and the combination with edge-map based features to improve the quality of image representation. Other advanced but commonly used features are also examined in our system. Our empirical results show that the offline learning system can achieve a detection accuracy reaching 90% and the real-time refinement process can improve the performance of the classifier by another 3-5%.

#### **Conflict of interest statement**

None declared.

#### Acknowledgments

This work was supported by the Connecticut Institute for Clinical and Translational Science (CICATS). Jinbo Bi and her student Xin Wang were also supported by the research grants from National Science Foundation of United States (IIS-1320586, DBI-1356655, IIS-1407205 and IIS-1447711).

## References

- K.A. Cullen, M.J. Hall, A. Golosinskiy, et al., Ambulatory Surgery in the United States, 2006, National Health Statistics Reports, vol. 11, 2009, pp. 1–25.
- [2] R.M. Rosenfeld, S.R. Schwartz, M.A. Pynnonen, D.E. Tunkel, H.M. Hussey, J.S. Fichera, A.M. Grimes, J.M. Hackell, M.F. Harrison, H. Haskell, et al., Clinical practice guideline tympanostomy tubes in children, Otolaryngology—Head Neck Surg. 149 (1 Suppl) (2013) S1–S35.
- [3] R.M. Rosenfeld, Evidence Based Otitis Media, vol. 1, PMPH-USA, 2003.
- [4] B. Armstrong, A new treatment for chronic secretory otitis media, AMA Arch. Otolaryngol. 59 (6) (1954) 653–654.
- [5] C.S. Derkay, J.D. Carron, B.J. Wiatrak, S.S. Choi, J.E. Jones, Postsurgical follow-up of children with tympanostomy tubes: results of the American Academy of Otolaryngology—Head and Neck Surgery Pediatric Otolaryngology Committee National Survey, Otolaryngology—Head Neck Surg. 122 (3) (2000) 313–318.
- [6] B. Austad, I. Hetlevik, V. Bugten, S. Wennberg, A.H. Olsen, A.-S. Helvik, Can general practitioners do the follow-ups after surgery with ventilation tubes in the tympanic membrane? Two years audiological data, BMC Ear Nose Throat Disord. 14 (1) (2014) 2.
- [7] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on

Computer Vision and Pattern Recognition, vol. 1, IEEE, Piscataway, NJ, 2001, p. I-511.

- [8] K. Levi, Y. Weiss, Learning object detection from a small number of examples: the importance of good features, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, IEEE, Piscataway, NJ, 2004, p. II-53.
- [9] J.-G. Wang, J. Li, W.-Y. Yau, E. Sung, Boosting dense sift descriptors and shape contexts of face images for gender recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Piscataway, NJ, 2010, pp. 96–102.
- [10] A.A. Mohammed, R. Minhas, Q. Jonathan Wu, M.A. Sid-Ahmed, Human face recognition based on multidimensional PCA and extreme learning machine, Pattern Recognit. 44 (10) (2011) 2588–2597.
- [11] H. Schneiderman, T. Kanade, A statistical method for 3D object detection applied to faces and cars, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, 2000, pp. 746–751.
- [12] C. Papageorgiou, T. Poggio, A trainable system for object detection, Int. J. Comput. Vis. 38 (1) (2000) 15–33.
- [13] Y. Chen, J. Bi, J.Z. Wang, Miles: multiple-instance learning via embedded instance selection, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 1931–1947.
- [14] P. Viola, M.J. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: IEEE International Conference on Computer Vision, IEEE, Piscataway, NJ, 2003, pp. 734–741.
- [15] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 743–761.
- [16] J. Gall, V. Lempitsky, Class-specific Hough forests for object detection, in: Decision Forests for Computer Vision and Medical Image Analysis, Springer, New York, NY, 2013, pp. 143–157.
- [17] T. McInerney, D. Terzopoulos, Deformable models in medical image analysis: a survey, Med. Image Anal. 1 (2) (1996) 91–108.
- [18] B. Solaiman, B. Burdsall, C. Roux, Hough transform and uncertainty handling. application to circular object detection in ultrasound medical images, in: International Conference on Image Processing, IEEE, Piscataway, NJ, 1998, pp. 828–831.
- [19] T. Behrens, K. Rohr, H.S. Stiehl, Robust segmentation of tubular structures in 3-D medical images by parametric object detection and tracking, IEEE Trans. Syst. Man. Cybern. Part B: Cybern. 33 (4) (2003) 554–561.
- [20] A. Andreopoulos, J.K. Tsotsos, Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI, Med. Image Anal. 12 (3) (2008) 335–357.
- [21] Y. Zheng, B. Georgescu, H. Ling, S.K. Zhou, M. Scheuering, D. Comaniciu, Constrained marginal space learning for efficient 3d anatomical structure detection in medical images, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2009, pp. 194–201.
- [22] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, S. Tanaka, Computer-aided colorectal tumor classification in NBI endoscopy using local features, Med. Image Anal. 17 (1) (2013) 78–100.
- [23] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J.V. Hajnal, D. Rueckert, Multiple instance learning for classification of dementia in brain MRI, Med. Image Anal. 18 (5) (2014) 808–818.
- [24] D.G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157.
- [25] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 886–893.
- [26] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
- [27] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Comput. Surv. 38 (4) (2006) 13.
- [28] D.K. Prasad, Survey of the problem of object detection in real images, Int. J. Image Process. 6 (6) (2012) 441.
- [29] G. Somanath, M. Rohith, D. Metaxas, C. Kambhamettu, D-clutter: Building object model library from unsupervised segmentation of cluttered scenes, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2009, pp. 2783–2789.
- [30] C. Gu, J.J. Lim, P. Arbeláez, J. Malik, Recognition using regions, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2009, pp. 1030–1037.
- [31] D.A. Klein, S. Frintrop, Center-surround divergence of feature statistics for salient object detection, in: IEEE International Conference on Computer Vision (ICCV), IEEE, Piscataway, NJ, 2011, pp. 2214–2219.
- [32] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
- [33] H.P. Moravec, Rover visual obstacle avoidance, in: International Joint Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 1981, pp. 785– 790.
- [34] C. Harris, M. Stephens, A combined corner and edge detector., in: Alvey Vision Conference, vol. 15, Manchester, UK, 1988, p. 50.
- [35] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. 6 (1973) 610–621.
- [36] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1) (1996) 51–59.

- [37] R. Strickland, H.I. Hahn, Wavelet transform methods for object detection and recovery, IEEE Trans. Image Process. 6 (5) (1997) 724–735.
- [38] J. Winn, J. Shotton, The layout consistent random field for recognizing and segmenting partially occluded objects, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, Piscataway, NJ, 2006, pp. 37–44.
- [39] A. Opelt, A. Pinz, A. Zisserman, Learning an alphabet of shape and appearance for multi-class object detection, Int. J. Comput. Vis. 80 (1) (2008) 16–44.
- [40] J. Shotton, A. Blake, R. Cipolla, Multiscale categorical object recognition using contour fragments, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1270–1281.
- [41] O. Barinova, V. Lempitsky, P. Kholi, On detection of multiple object instances using Hough transforms, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1773–1784.
- [42] Y. Chi, M.K. Leung, Part-based object retrieval in cluttered environment, IEEE Trans. Pattern Anal. Mach. Intell. 29 (5) (2007) 890–895.
- [43] D.K. Prasad, M.K. Leung, S.-Y. Cho, Edge curvature and convexity based ellipse detection method, Pattern Recognit. 45 (9) (2012) 3204–3221.
- [44] D.K. Prasad, M.K. Leung, C. Quek, Ellifit: an unconstrained, non-iterative, least squares based geometric ellipse fitting method, Pattern Recognit. 46 (5) (2013) 1449–1465.
- [45] V. Ferrari, T. Tuytelaars, L. van Gool, Simultaneous object recognition and segmentation from single or multiple model views, Int. J. Comput. Vis. 67 (2) (2006) 159–188.
- [46] P. Schnitzspan, M. Fritz, S. Roth, B. Schiele, Discriminative structure learning of hierarchical representations for object detection, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ, 2009, pp. 2238–2245.
- [47] B. Ommer, J.M. Buhmann, Learning the compositional nature of visual object categories for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (3) (2010) 501–516.
- [48] G. Bouchard, B. Triggs, Hierarchical part-based visual object categorization, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, Piscataway, NJ, 2005, pp. 710–715.
- [49] A. Bosch, A. Zisserman, X. Muoz, Scene classification using a hybrid generative/discriminative approach, IEEE Trans. Pattern Anal. Mach. Intell. 30 (4) (2008) 712–727.
- [50] V.F. van Ravesteijn, C. van Wijk, F.M. Vos, R. Truyen, J.F. Peters, J. Stoker, L.J. van Vliet, Computer-aided detection of polyps in CT colonography using logistic regression, IEEE Trans. Med. Imaging 29 (1) (2010) 120–131.
- [51] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
- [52] K.-C. Lee, D. Kriegman, Online learning of probabilistic appearance manifolds for video-based recognition and tracking, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, Piscataway, NJ, 2005, pp. 852–859.
- [53] O. Javed, S. Ali, M. Shah, Online detection and classification of moving objects using progressively improving detectors, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, Piscataway, NJ, 2005, pp. 696–701.
- [54] B. Babenko, M.-H. Yang, S. Belongie, Robust object tracking with online multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1619–1632.
- [55] B. Liu, J. Huang, C. Kulikowski, L. Yang, Robust visual tracking using local sparse appearance model and k-selection, IEEE Trans. Pattern Anal. Mach. Intell. 35 (12) (2013) 2968–2981.

- [56] H. Yuen, J. Princen, J. Illingworth, J. Kittler, Comparative study of Hough transform methods for circle finding, Image Vis. Comput. 8 (1) (1990) 71–77.
- [57] T. Atherton, D. Kerbyson, Size invariant circle detection, Image Vis. Comput. 17 (1999) 795–803.
- [58] L. Fei-Fei, P. Perona, A Bayesian hierarchical model for learning natural scene categories, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, IEEE, Piscataway, NJ, 2005, pp. 524–531.
- [59] IBM ILOG CPLEX Division, New York, NY, IBM ILOG CPLEX Callable library 12.1 Reference Manual, 2009.
- [60] K.J. Kværner, H.A. Kristiansen, M.B. Russell, Otitis media history, surgery and allergy in 60-year perspective: a population-based study, Int. J. Pediatr. Otorhinolaryngol. 74 (12) (2010) 1356–1360.
- [61] C.D. Bluestone, J.D. Swarts, Human evolutionary history: consequences for the pathogenesis of otitis media, Otolaryngology—Head Neck Surg. 143 (6) (2010) 739–744.
- [62] S. Mui, B.M. Rasgon, R.L. Hilsinger Jr., B. Lewis, G. Lactao, Tympanostomy tubes for otitis media: quality-of-life improvement for children and parents, Ear Nose Throat J. 84 (7) (2005) 418–420.
- [63] D.J. Kay, M. Nelson, R.M. Rosenfeld, Meta-analysis of tympanostomy tube sequelae, Otolaryngology-Head Neck Surg. 124 (4) (2001) 374–380.
- [64] R. Kao, D.J. Kirse, A.K. Evans, Compliance with recommendations for tympanostomy tube followup Patient characteristics, Otolaryngology–Head Neck Surg. 151 (3) (2014) 489–495.

Xin Wang received a Master degree in Control Theory and Control Engineering from Dalian University of Technology, China. He is currently studying at the University of Connecticut toward his Ph.D. degree in Computer Science. His research interests include machine learning, data mining and intelligent systems.

**Tulio A. Valdez** received a M.D. degree from Pontificia Universidad Javeriana School of Medicine, Colombia, and his training in otolaryngology/head and neck surgery from Tufts University School of Medicine, USA. He was a Research Fellow at New England Medical Center and Pediatric Otolaryngology at Texas Children's Hospital. He is currently an Assistant Professor of Otolaryngology, University of Connecticut School of Medicine and also a Pediatric Surgeon at Connecticut Children's Medical Center. His research interests include pediatric otolaryngology.

**Jinbo Bi** received a Ph.D. degree in mathematics from Rensselaer Polytechnic Institute, USA, and a Master degree in Electrical Engineering and Automatic Control from Beijing Institute of Technology, China. She is an Associate Professor of Computer Science and Engineering at the University of Connecticut. Prior to her current appointment, she worked with Siemens Medical Solutions on computer aided diagnosis research and Partners Healthcare on clinical decision support systems. Her research interests include machine learning, data mining, bioinformatics and biomedical informatics.