OTITIS MEDIA VOCABULARY AND GRAMMAR

Anupama Kuruvilla[1], *Jian Li*[2], *Pablo Hennings Yeomans*[1], *Pedro Quelhas*[3], *Nader Shaikh*[4], *Alejandro Hoberman*[4] *and Jelena Kovačević*[1,2]

[1]Dept. of BME and Center for Bioimage Informatics, [2]Dept. of ECE
Carnegie Mellon University, Pittsburgh, PA, USA
[3]Universidade do Porto, Faculdade de Engenharia (DEEC)
INEB - Instituto de Engenharia Biomedica, Porto, Portugal
[4]Division of General Academic Pediatrics, Children's Hospital of Pittsburgh
University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

ABSTRACT

We propose an automated algorithm for classifying diagnostic categories of otitis media (middle ear infection): acute otitis media, otitis media with effusion and no effusion. Acute otitis media represents a bacterial superinfection of the middle ear fluid and otitis media with effusion a sterile effusion that tends to subside spontaneously. Diagnosing children with acute otitis media is hard, leading to overprescription of antibiotics that are beneficial only for children with acute otitis media, prompting a need for an accurate and automated algorithm. To that end, we design a feature set understood by both otoscopists and engineers based on the actual visual cues used by otoscopists; we term this *otitis media vocabulary*. We also design a process to combine the vocabulary terms based on the decision process used by otoscopists; we term this *otitis media grammar*. The algorithm achieves 84% classification accuracy, in the range or outperforming pediatricians who did not receive special training, as well as state-of-the-art classifiers.

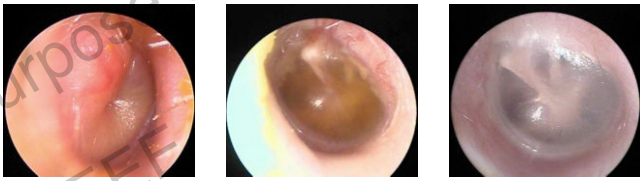
***Index Terms*—** otitis media, classification, vocabulary, grammar

1. INTRODUCTION

Otitis media is a general term for middle-ear inflammation and may be classified clinically as either acute otitis media (AOM) or otitis media with effusion (OME); AOM represents a bacterial superinfection of the middle ear fluid and OME a sterile effusion that tends to subside spontaneously. Although middle ear effusion is present in both cases, this clinical classification is important because antibiotics are generally beneficial only for AOM. However, proper diagnosis of AOM, as well as distinction from both OME and no effusion (NOE) require considerable training (see Figure 1 for example images).

AOM is the most common infection for which antimicrobial agents are prescribed in children in the US. By age seven, 93% of children will have experienced one or more episodes of otitis media [1]. AOM results in significant social burden and indirect costs due to time lost from school and work. Estimated direct costs of AOM in 1995 were $1.96 billion and indirect costs were estimated to be $1.02 billion, with a total of 20 million prescriptions for antimicrobials related to otitis media [2].

The authors gratefully acknowledge support from the NIH through award 1DC010283 and the CMU CIT Infrastructure Award. Pablo Hennings Yeomans performed the work while at CMU.



(a) AOM. (b) OME. (c) NOE.

Fig. 1. Sample (cropped) images from the three diagnostic classes.

The above considerations underscore the critical need for an accurate classification algorithm, able to discriminate images of tympanic membranes (TM) obtained with an otoendoscope into one of three stringent diagnostic categories: AOM, OME and NOE. To our knowledge, the only related work in this area is [3], where the authors investigate the influence of color on the classification accuracies of individual classes, with the conclusion that the color alone is not sufficient for accurate classification.

The problem at hand is a standard image-processing task, classification, and as is typical, we have a choice ranging from creating or using a fairly universal classifier, one that works well in most instances, or one tailored to the specific application at hand, or any in between. In this paper, we adopt the following guiding principles:

• *Vocabulary.* We aim to design a feature set understood by both otoscopists and engineers based on the actual visual cues used by otoscopists; we term this *otitis media vocabulary (OMV)*. To design OMV, we use otoscopic findings listed in Table 1.

• *Grammar.* We aim to design a process to combine the vocabulary terms based on the decision process used by otoscopists; we term this *otitis media grammar*. To design the grammar, we use the findings from [4], summarized in the next section.

We compare our algorithm designed following the above principles to a universal classifier WND-CHARM [5] and a multiresolution classifier originally designed for biomedical applications [6]. The ground truth is provided by a panel of expert otoscopists.

2. EPIDEMIOLOGY OF ACUTE OTITIS MEDIA

The number of AOM episodes has increased substantially in the past two decades, as have the associated costs. Approximately 25 million visits are made to office-based physicians in the US for otitis media yearly, resulting in a total of 20 million prescriptions for antimicro-

bials related to otitis media [1]. Accurate diagnosis is imperative to ensure that antimicrobial therapy is limited to the appropriate patients; this, in turn, increases the likelihood of achieving optimal outcomes and minimizing the risk of encouraging antibiotic resistance.

To design a feature set based on the visual cues used by otoscopists, we need to understand what those cues are; we list only those that were found most relevant and that we use to design the features, see Table 1. For example, the presence of middle-ear effusion is evidenced by TM abnormalities, such as white or yellow discoloration and opacification. A diagnosis of AOM can be established when distinct fullness or bulging of the TM is noted in addition to evidence of middle-ear effusion.

	AOM	**OME**	**NOE**
Color	White, pale yellow, markedly red	White, amber, grey, blue	Grey, pink
Position	Distinctly full, bulging	Neutral, retracted	Neutral, retracted
Translucency	Opacified	Opacified, semi-opacified	Translucent

Table 1. Otoscopic findings associated with clinical diagnostic categories on TM images.

To explore the diagnostic processes used, Drs. Shaikh and Hoberman asked 7 expert otoscopists to independently describe TM findings and assign a diagnosis (AOM/OME/NOE) on a collection of 135 randomly selected TM images from an image library (see Figure 1 for an example). To control for differences in color rendition between computers, they mailed color-calibrated laptops to each expert. Just by evaluating still images, with no information about mobility or ear pain, the diagnosis (AOM vs. no AOM) endorsed by the majority of experts was in agreement with the live diagnosis 88.9% of the time, underscoring the limited role that symptoms and mobility of the TM have in the diagnosis of AOM. Bulging of the TM was the finding judged best to differentiate AOM from OME [4].

3. OTITIS MEDIA VOCABULARY

The expert otoscopist uses his specialized knowledge when discriminating between the different diagnostic categories. The goal of our proposed methodology is to create a feature set—OMV, which will mimic the visual cues of trained otoscopists closely.

Methodology. To design OMV, we follow the process from [7]:

Formulation of initial set of descriptions. We obtain initial descriptions of those characteristics best describing a given diagnostic category from the summary of otoscopic findings in Table 1.

Computational translation of key terms. From this set, the key terms, such as *bulging*, are translated into their computational synonyms, creating a computational vocabulary (in our case, we construct a feature describing the opposite, *concavity*).

Computational translation of descriptions. Using the computational vocabulary, entire otoscopist's descriptions, such as *bulging and white*, are translated.

Verification of translated descriptions. Based on these translated descriptions, the otoscopist tries to identify the diagnostic category being described, emulating the overall system with translated descriptions as features and the otoscopist as the classifier.

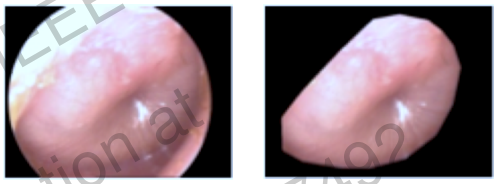
Refinement of insufficient terms. If the otoscopist is unable to identify a diagnostic category based on translated descriptions, or if a particular translation is not understandable, then that translation is refined and presented again to the otoscopist for verification.

Otitis media vocabulary. If the otoscopist is able to identify a diagnostic category based on translated descriptions, then the discriminative power of the key terms and their corresponding computational interpretations is validated, and these terms can be included as OMV terms to create features.

This feedback loop is iterated until a sufficient set of terms have been collected to formulate OMV:

$$\left\{ \begin{array}{lll} \text{concavity } f_c & \text{translucency } f_t & \text{amber level } f_a \\ \text{grayscale variance } f_v & \text{bubble presence } f_b & \text{light } f_\ell \end{array} \right\}.$$

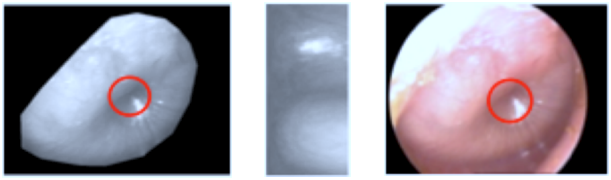
Automated segmentation. Segmentation is a crucial step to extract relevant regions on which reliable features for classification can be computed. We now briefly summarize an active-contour based algorithm we implemented. We compute a snake potential of the grayscale version of the input, and then a set of forces that outline the gradients and edges. Then, the active contour algorithm is initialized by a circumference in the center of the image, and the algorithm iteratively grows this contour. The algorithm stops at a predefined convergence criterion, which leaves an outline that covers the relevant region in the image. This outline is used to generate the final mask that is applied to the input image; Figure 2 shows an example.



(a) Original. (b) Segmented.

Fig. 2. Automated segmentation of TM images.

Concavity. To identify bulging, we design a feature detecting the opposite, the concave region located centrally in the TM; we call it *concavity* feature. The input is a grayscale version (Figure 3(a)) of the segmented original RGB image $X \in \mathbb{R}^{M \times N}$ as in Figure 2. We use a sliding window to extract a local circular neighborhood, $X_R(m, n)$, of radius R ($R = 60$ in our experiments). That circular neighborhood is then transformed into its polar coordinates to obtain $X_R(r, \theta)$, with $r \in \{1,\ 2,\ \ldots,\ R\}$, $\theta \in [0, 2\pi]$, and



(a) Grayscale. (b) Polar. (c) Labeled.

Fig. 3. Computational steps for the concavity feature.

$$r \ = \ \sqrt{(m - m_c)^2 + (n - n_c)^2}, \qquad \theta \ = \ \arctan \frac{(n - n_c)}{(m - m_c)},$$

where (m_c, n_c) are the center coordinates of the neighborhood X_R. In Figure 3(b), the resulting image has r as the horizontal axis and θ as the vertical one. The concave region changes from dark to bright from the center towards the periphery of the concavity; in polar coordinates this change from dark to bright occurs as the radius

grows, see Figure 3(b). Defining the bright region $B = \{(r,\theta) \mid r > R'\}$ and the dark region $D = \{(r,\theta) \mid r \leq R'\}$, and with $R' \in [1/4R, 3/4R]$, we compute the ratio of the two means,

$$f_{c,R'} = \frac{\mathrm{E}[\, X_R(r,\theta) \mid_{(r,\theta)\in B}\,]}{\mathrm{E}[\, X_R(r,\theta) \mid_{(r,\theta)\in D}\,]},$$

As the concave region is always centrally located, we experimentally determine a square neighborhood I (here 151×151) to compute the concavity feature,

$$f_c = \max_{R'\in I} f_{c,R'}.$$

Translucency. Translucency of the TM is the main characteristic of NOE in contrast with opacity in AOM and semi-opacity in OME; it results in the clear visibility of the TM, which is primarily gray. We thus design the translucency feature to measure the grayness of the TM. We do that by using a simple color-assignment technique. As these images were taken under different lighting and viewing conditions, according to [8], at least $3-6$ images are needed to characterize a structure/region under all lighting and viewing conditions. We take the number of images to be $N_{tl} = 20$.

Then, we perform the following once to determine gray-level clusters in translucent regions: We extract N_t pixels from translucent regions ($N_t = 100$) of N_{tl} RGB images by hand segmentation, to obtain a total of $N_{tl}N_t$ pixels from images (here 2000). We then cluster these $N_{tl}N_t$ pixels using k-means clustering to obtain K cluster centers $c_k \in \mathbb{R}^3$, $k = 1, 2, \ldots, K$, capturing variations of gray in the translucent regions.

To compute the translucency feature for a given image X, for each pixel (m,n), we compute K Euclidean distances of $X(m,n)$ to the cluster center c_k, $k = 1, 2, \ldots, K$,

$$d_k(m,n) = \sqrt{\sum_{i=1}^{3}(X_i(m,n) - c_{k,i})^2},$$

where $i = 1, 2, 3$ denotes the color channel. If any of the computed K distances falls below a threshold T_t (found experimentally), the pixel is labeled as translucent and belongs to the region $R_t = \{(m,n) \mid \min_k d_k(m,n) < T_t\}$, otherwise it is not translucent. The binary image X_t is then simply the characteristic function of the region R_t, $X_t = \chi_{R_t}$. We then define the translucency feature as the mean of X_t,

$$f_t = \mathrm{E}[\, X_t \,].$$

Amber Level. We use that OME is predominantly amber or pale yellow to distinguish it from AOM and NOE. We apply a color-assignment technique similar to that used for computing X_t to obtain a binary image X_a, indicating amber and nonamber regions, and define the amber feature as the mean of X_a,

$$f_a = \mathrm{E}[\, X_a \,].$$

Grayscale Variance. Another discriminating feature is the variance of the intensities across the grayscale version of the image X_g,

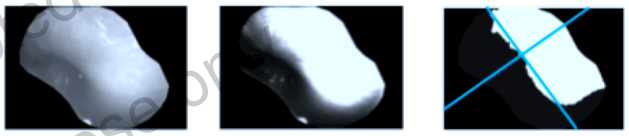
$$f_v = \mathrm{var}(X_g);$$

for example, OME has a more uniform appearance than AOM and NOE, and has consequently a much lower variance that can be used to distinguish it from the rest.

Bubble Presence. The presence of visible air-fluid levels, or bubbles, behind the TM is an indication of OME. The algorithm takes in red and green channels of the original RGB image and performs Canny edge detection [9], to place parallel boundaries on either sides of the real edge, creating a binary image X_b in between. This is followed by filtering and morphological operations to enhance edge detection and obtain smooth boundaries. We then define the bubble feature as the mean of X_b,

$$f_b = \mathrm{E}[\, X_b \,].$$

Light. Examination of the TM is performed by an illuminated otoendoscope. The distinct bulging in AOM results in nonuniform illumination of the TM, in contrast to the uniform illumination in NOE. Our aim is to construct a feature that will measure this nonuniformity as the ratio of the brightly-lit to the darkly-lit regions.



(a) Grayscale. (b) Contrast-enh. (c) Dominant orient.

Fig. 4. Computational steps for the light feature.

We start by performing contrast enhancement on the grayscale image in Figure 4(a) to make the nonuniform lighting prominent. The resulting image in Figure 4(b) is thresholded at T_ℓ (found experimentally) to obtain a mask of the brightly-lit binary image X_b in Figure 4(c). To find the direction ($\theta_{\max}$) perpendicular to the maximum illumination gradient, we look at lines passing through (m_c, n_c) (the pixel coordinates at which f_c is obtained) at angle θ with the horizontal axis. Defining the bright region $B = \{(m,n) \mid n \geq \tan(\theta)(m - m_c) + n_c\}$ and the dark region $D = \{(m,n) \mid n < \tan(\theta)(m-m_c)+n_c\}$, we compute the ratio of the two means,

$$r(\theta) = \frac{\mathrm{E}[\, X_b(m,n) \mid_{(m,n)\in B}\,]}{\mathrm{E}[\, X_b(m,n) \mid_{(m,n)\in D}\,]}.$$

Then, the direction perpendicular to the maximum illumination gradient is $\theta_{\max} = \arg\max_\theta r(\theta)$, and we define the light feature as

$$f_l = r(\theta_{\max}).$$

4. OTITIS MEDIA GRAMMAR

Inspired by [4], we design a decision process to combine the OMV terms based on the decision process used by the otoscopists and term it otitis media grammar. The decision process has a hierarchical tree scheme wherein we use the OMV to discriminate AOM/OME/NOE. The hierarchy consist of two levels shown in Figure 5:

First Level. At the first level, we perform a coarse separation based on bulging (concavity feature), translucency and light. While ideally, if there is bulging present, the image should be classified as AOM, concavity feature alone cannot accomplish the task; we use the light feature as an aid as AOM will be nonuniformly-lit unlike OME/NOE, as we explained earlier. In the second split, we use translucency to discriminate NOE from the rest. Unfortunately, some of the OME images will show up in the same category due to semi-translucency observed in mild infection. This process results in a separation into two superclasses: AOM/OME (acute/mild infection) and NOE/OME (no/mild infection).

Second Level. At the second level, we use a weighted combination of four features, amber level, bubble presence, translucency

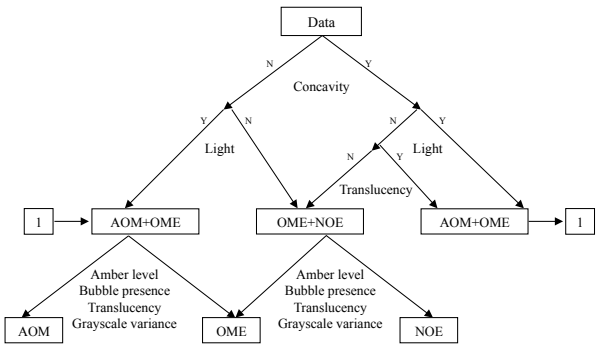


Fig. 5. A hierarchical classifier implementing otitis media grammar.

and grayscale variance, $w_a f_a + w_b f_b + w_\ell f_\ell + w_v f_v$, to help separate superclasses into individual classes. During the training phase, we determine weights w that maximize the classification accuracy of training data; these are then used in the testing phase to classify.

5. RESULTS AND DISCUSSION

Data Set. As part of a separate clinical trial, 826 TM images were collected from children with AOM, OME and NOE. A panel of three experienced otoscopists examined these images and provided the ground-truth labels. As mentioned before, these images pose challenges even for experienced otoscopists; thus a rather poor agreement in labeling the set. As having accurate ground-truth labels is crucial, we asked the panel to provide a diagnosis confidence level for each image; levels between 80-100 indicated an almost perfect example of its diagnostic class, while levels below 30 indicated almost no confidence. Based on these, we selected a subset of 181 images (48 of AOM, 63 of OME and 70 of NOE), by selecting those for which all three experts agreed with confidence of over 60.

Results. We compare our algorithm, otitis media classifier (OMC), to four other classifiers: (1) In the correlation filter classifier (CFC), the image is transformed into the polar domain, correlation is performed on concentric annular regions extracted from the polar-transformed image, and the class label is assigned based on the correlation measure. (2) WND-CHARM (WCM) [5] is a universal classifier that extracts a large number (4,008) of generic image-level features, then classifies with a nearest neighbor algorithm. (3) MRC (multiresolution classifier), originally designed for biomedical applications [6], decomposes the image into subbands using a multiresolution decomposition (for example, wavelets or wavelet packets), followed by feature extraction and classification in each subband and a global decision based on weighted individual subband decisions. We ran MRC with 2 levels and 26 Haralick texture features on the grayscale image and each of the 20 subbands (546 in total). (4) SSC (SIFT and shape descriptors classifier) extracts SIFT features and shape descriptors for the image and uses bag-of-words model, then classifies using support vector machine. We used a 5-fold cross validation setup. The CFC and SSC are classifiers we developed in the course of this work.

	CFC	WCM	MRC	SSC	OMC
AOM	76.6	68.2	53.5	85.6	81.3
OME	72.9	60.8	66.3	71.3	85.7
NOE	75.6	63.4	75.1	71.8	81.4
Total	62.1	64.1	68.3	69.1	**84.0**

Table 2. Classification accuracies [%].

Table 2 compares the performance of the five classifiers. The OMC outperforms the four other classifiers by a fair margin. This validates our methodology in that a small number of targeted, physiologically-meaningful features, vocabulary, is what is needed for accurate classification.

6. CONCLUSIONS

We created an automated system for identification of three diagnostic classes of otitis media. Our guiding principle was to design and use a vocabulary of features that mimics the actual visual cues used by the otoscopists in their diagnostic process. Results demonstrate that our simple and concise 6-feature OMV is effective on the problem, underscoring the importance of using targeted, physiologically-meaningful features instead of a large number of general-purpose features. The classification process, grammar, is a hierarchical process mimicking in part the diagnostic process used by otoscopists. Our future work will focus on designing new features to prevent misclassification, as well as refining the hierarchical classifier.

7. REFERENCES

[1] D. W. Teele, J. O. Klein, and B. Rosner, "Epidemiology of otitis media during the first seven years of life in children in greater Boston: A prospective, cohort study," *The Journ. Infectious Disease*, vol. 160, no. 1, pp. 83–94, 1989.

[2] American Academy of Pediatrics, "Diagnosis and management of acute otitis media," *Pediatrics*, vol. 113, no. 5, pp. 1451–1465, 2004.

[3] C. Vertan, D. C. Gheorghe, and B. Ionescu, "Eardrum color content analysis in video-otoscopy images for the diagnosis support of pediatric otitis," in *Int. Symp. on Signals, Circuits and Systems*, Bucharest, Romania, July 2011.

[4] N. Shaikh, A. Hoberman, P. H. Kaleida, H. E. Rockette, M. Kurs-Lasky, H. Hoover, M. E. Pichichero, O. F. Roddey, C. Harrison, J. A. Hadley, and R. H. Schwartz, "Otoscopic signs of otitis media," *The Pediatric Infectious Disease Journ.*, vol. 30, no. 10, pp. 822–826, 2011.

[5] L. Shamir, N. Orlov, D. M. Eckley, T. Macura, J. Johnston, and I. G. Goldberg, "WND-CHARM: Multi-purpose image classification using compound image transforms," *Pattern Recogn. Letters*, vol. 29, pp. 1684–1693, 2008.

[6] A. Chebira, Y. Barbotin, C. Jackson, T. E. Merryman, G. Srinivasa, R. F. Murphy, and J. Kovačević, "A multiresolution approach to automated classification of protein subcellular location images," *BMC Bioinformatics*, vol. 8, no. 210, 2007.

[7] R. Bhagavatula, M. C. Fickus, J. W. Kelly, C. Guo, J. A. Ozolek, C. A. Castro, and J. Kovačević, "Automatic identification and delineation of germ layer components in H&E stained images of teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Rotterdam, The Netherlands, Apr. 2010, pp. 1041–1044.

[8] P. N. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions?," in *Proc. IEEE Int. Conf. Comp. Vis. and Patt. Recogn.*, June 1996, pp. 270–277.

[9] J. Canny, "A computational approach for edge detection," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 8, no. 6, pp. 1293–1299, 1986.