# Assessment of Computational Visual Attention Models on Medical Images

Varun Jampani*
CVIT, IIIT-Hyderabad
Hyderabad, India.
varunjampani@research.iiit.ac.in

Ujjwal
CVIT, IIIT-Hyderabad
Hyderabad, India.
ujjwal@research.iiit.ac.in

Jayanthi Sivaswamy
CVIT, IIIT-Hyderabad
Hyderabad, India.
jsivaswamy@iiit.ac.in

Vivek Vaidya
Medical Image Analysis Lab
GE Global Research
Bangalore, India.
vivek.vaidya@ge.com

## ABSTRACT

Several computational visual saliency models have been proposed in the context of viewing natural scenes. We aim to investigate the relevance of computational saliency models in medical images in the context of abnormality detection. We report on two studies aimed at understanding the role of visual saliency in medical images. Diffuse lesions in Chest X-Ray images, which are characteristic of Pneumoconiosis and high contrast lesions such as 'Hard Exudates' in retinal images were chosen for the study. These approximately correspond to *conjunctive* and *disjunctive* targets in a visual search task. Saliency maps were computed using three popular models namely Itti-Koch [7], GBVS [3] and SR [4]. The obtained maps were evaluated against gaze maps and ground truth from medical experts.

Our results show that GBVS is seen to perform the best ($Mdn.\ ROC\ area = 0.77$) for chest X-Ray images while SR performs the best ($ROC\ area = 0.73$) for retinal images, thus asserting that searching for conjunctive targets calls for a more local examination of an image while disjunctive targets call for a global examination. Based on the results of the above study, we propose extensions for the two best performing models. The first extension makes use of top down knowledge such as lung segmentation. This is shown to improve the performance of GBVS to some extent. In the second case the extension is by way of including multiscale information. This is shown to significantly (by 28.76%) improve abnormality detection. The key insight from these studies is that bottom saliency continues to play a predominant role in examining medical images.

*Corresponding author


*ICVGIP* '12, December 16-19, 2012, Mumbai, India


## Keywords

Visual Attention, Saliency Models, Chest X-rays, Retinal Images

## 1. INTRODUCTION

Visual Search is a common human activity. Searching for a friend in a crowd is an everyday example. It is also important in diagnosing diseases as radiologists search for lesions and other abnormalities in medical images before writing reports. This visual search activity is mediated by a cognitive process called *visual attention* which is the process of selectively attending to an area of visual field while ignoring the surrounding visual areas. The area of the image attended to visually is deemed as salient. Finding salient regions in an image is of interest to computer vision as well since visually salient features in an image are generally invariant to many image transformations and carry important image information [2]. Attempts have been made to use saliency to address problems such as object detection [22], image compression [5], tracking and image retrieval [11].

Several psychophysical and computational models of visual attention have been proposed in the literature. Their main objective is to simulate the behavioral data and to better understand human perception. Most of these models have been studied and validated in the context of viewing natural scenes. The goal of visual search in medical images is generally one of evidence gathering about the possibility of any 'abnormality' in the condition of a patient. The search therefore involves disregarding the 'usual' and detecting the 'unusual' visual elements. This is akin to searching for anything that is out of ordinary in a given face image. In medical images, abnormalities are wide ranging in terms of appearance. They can be glaring to subtle. We aim to investigate the performance of computational saliency models in medical images in the context of abnormality detection. Specifically, we are interested in investigating whether the computational saliency models can detect the abnormalities (subtle as well distinct) in medical images. Developing saliency models for medical images may aid in the development of computer aided diagnostic (CAD) tools. Such CAD

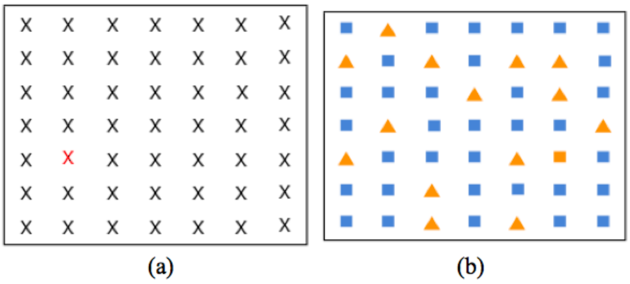


**Figure 1: Sample stimulus showing (a) disjunctive and (b) conjunctive target searches.**

tools have been argued to not only aid in diagnosis but also provide training for resident radiologists [14, 15].

Taking Treisman's 'Feature integration theory' (FIT) [21] as a basis, we carry out two studies to look at the role of visual saliency in medical images. FIT is one of the most influential theories on visual attention which mainly posits that visual attention is responsible for binding different features into consciously experienced whole. Information from different feature maps is collected in a master map (also called *saliency map*). This theory has been the basis for the development of many saliency models.

While FIT explains the saliency of different locations of a visual input, what makes these locations salient is partially addressed by visual search paradigms. A visual search task involves identifying *targets* from surrounding *distractors*. The most efficient of searches are those in which a single basic feature (e.g. orientation) defines the target and distractors are homogeneous. Such targets are called *disjunctive* targets. In general, pre-attentive visual processes are sufficient to identify such targets, which is called *pop-out* phenomenon [21, 13]. If targets and distractors share common features, search becomes inefficient and focused or serial attention is required to identify such targets [23]. Such targets are called *conjunctive* targets. Figure 1 shows sample search tasks involving disjunctive and conjunctive targets. In stimulus shown in figure 1(a), identifying target ('X in red color)') pops-out among distractors ('X in black color') and can be identified by pre-attentive processes. Such targets are called disjunctive targets. In figure 1(b), identifying orange square (target) among blue squares and orange triangles requires serial attention and such targets is called conjunctive targets.

Following the above-mentioned concepts, we chose two types of medical imaging abnormalities that approximately correspond to conjunctive and disjunctive targets in a visual search to study the relevance of computational visual saliency in medical images: abnormalities such as diffuse lesions in chest X-Ray images, which are characteristic of pneumoconiosis and high contrast lesions such as hard exudates in retinal images.

Pneumoconiosis is a lung disease caused by prolonged inhalation of industrial dust like coal, silica dust etc. It is mainly diagnosed by reading chest x-rays. Figure 2 shows sample chest x-ray segments showing diffuse abnormalities corresponding to different disease stages in Pneumoconiosis. Reading x-rays of pneumoconiosis requires lot of expertize [17, 9] as abnormal regions are confounded with anatomical structures such as blood vessels. Hence, searching for

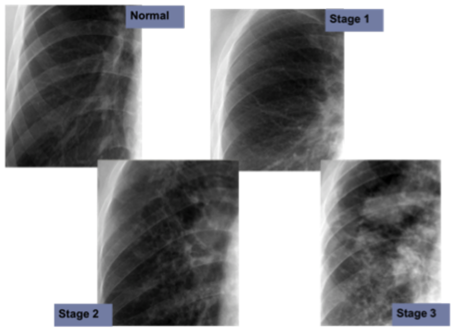


**Figure 2: Sample X-ray segments showing different disease stages in Pneumoconiosis.**

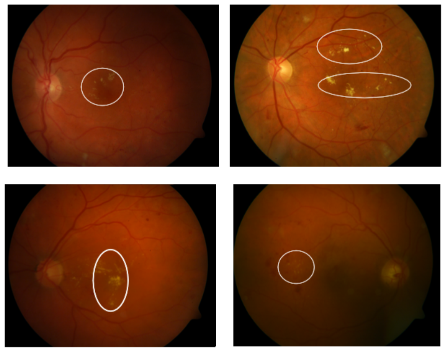

**Figure 3: Sample retinal images showing Hard Exudates (enclosed in white circles).**

pneumoconiosis abnormalities in a chest x-ray requires focused serial search and the abnormalities form roughly conjunctive targets as they share the intensity of the ribs and texture of the tissues between the ribs.

Hard Exudates (HEs) are whitish and yellowish deposits on the retinal surface due to the lipids leaked from the damaged blood capillaries in the retina. HEs are a common sign of Diabetic Retinopathy in patients. Figure 3 shows sample retinal images with HE locations circled. The images show a good variability in the lesions as well as the normal tissues. However, in general, HEs appear to have a high contrast and can generally be identified without focused attention. Thus, these abnormalities are disjunctive targets.

The relevance of computational saliency models in detecting abnormalities in these medical images is studied by computing saliency maps using three popular models: Itti-Koch (IK) [7, 6], Graph Based Visual Saliency (GBVS) [3] and Spectral Residual (SR) [4] saliency models. The saliency maps computed using the above models are evaluated against gaze maps in the case of chest x-rays and against expert markings in case of retinal images. The rationale for choosing the above saliency models and the evaluation strategies is explained latter.

Human visual attention is influenced by both stimulus driven *bottom-up* influences and goal/task driven *top-down* influences. In the context of medical images, bottom-up influences correspond to image features whereas top-down influences correspond to the knowledge and expertize of radiologists. The above-mentioned computational models (IK,

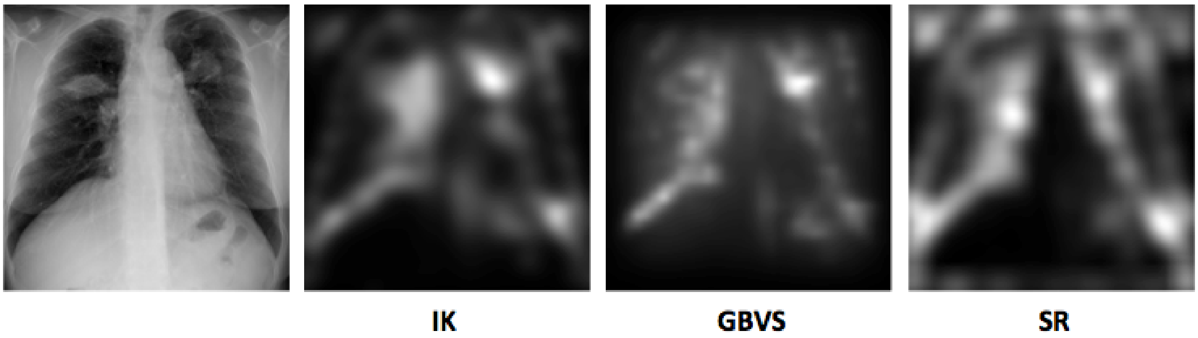


**Figure 4: A sample chest x-ray and the corresponding saliency maps computed using different saliency models.**

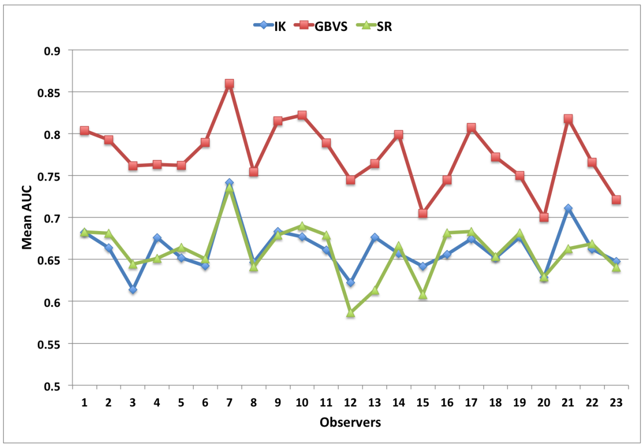


**Figure 5: Mean AUCs corresponding to different saliency models for all the observers.**

GBVS and SR) are bottom-up saliency models and they try to predict saliency solely based on image features. In the present study, we also investigated the role of top-down knowledge of anatomical features in predicting saliency in medical images.

The rest of the paper is organized as follows. In section-2, we discuss some related work. In section-3, we first explain the methodology used in present study followed by results in our studies on chest x-rays and retinal images. Important results are discussed in section-4 followed by conclusion and future work in section-5.

## 2. RELATED WORK

Since interpreting medical images is highly task dependent, it is generally expected that top-down mechanisms play a very important and significant role in guiding the observers' attention, whereas bottom-up processes might not play an important role. But, a recent study on brain CT images [16] showed that bottom-up mechanisms also play a significant role in guiding the eye movements of neurologists looking for stroke lesions on brain CT images. To the best of our knowledge, there are no other study investigating the role of bottom up saliency in medical images.

Based on some psychophysical experiments in the context of searching tumors in chest x-rays, Nodine and Kundel [18, 19] have developed a model of visual search and detection that has three main components: overall pattern recognition (global impression), focal attention to image detail; and decision-making. According to this model, visual search begins with a global response involving the entire retina, in which the context is established and gross deviations from normal are detected. This response initiates a series of checking fixations, using the fovea to resolve ambiguity and fill in detail. Since this Nodine and Kundel's model of visual search in medical images is very similar to that of FIT, we can expect that bottom up saliency models based on FIT would detect salient regions in medical images also. The present study is aimed in investigating whether this is true or not.

### 2.1 Saliency models studied

We used three popular computational bottom-up saliency models: IK [7, 6], GBVS [3] and SR [4]. IK and GBVS are biologically motivated saliency models, which closely follow FIT. They are one of the most popular models for the *focused attention stage* of visual attention and they predict human fixation points well. Both GBVS and IK use the same set of basic image features at multiple scales to compute saliency maps. Compared to IK, GBVS model makes long-range pixel comparison of feature values to compute final saliency maps and has more center bias [3]. It is reported [3] that GBVS outperforms IK model in predicting the saliency of observers while viewing natural images. SR model, unlike IK and GBVS, model the *pre-attentive* stage of visual attention. SR model tries to identify the disjunctive targets in a visual scene, which pops-out automatically. SR model uses spectral features and analyze log-spectrum of an image to compute saliency map. These three saliency models make a good representative set of saliency models modeling important aspect of human visual attention, and hence chosen for our study. Next, we discuss our study on chest x-rays of pneumoconiosis followed by our study on retinal images.

## 3. ASSESSMENT OF SALIENCY MODELS ON MEDICAL IMAGES

We now describe the two studies undertaken for assessing the 3 saliency models and present the results obtained. For each type of image, saliency maps were computed using the saliency models and they were compared with the ground truth. We also investigated whether incorporating top-down anatomical knowledge would improve the performance of saliency models.

## 3.1 Study on Chest X-ray Images

Since pneumoconiosis abnormalities form *conjunctive* targets for radiologists, we evaluated the role of bottom-up saliency in chest x-rays of pneumoconiosis by comparing the saliency maps against the eye fixations of observers of different expertize levels. In other words, we investigated how well the bottom-up saliency models can predict the fixations of observers reading chest x-rays of pneumoconiosis.

### 3.1.1 Eye Movement Recordings

Eye movement recordings were done in a room dedicated to eye tracking experiments. 23 observers of various expertize levels (from novices to staff radiologists) volunteered in this experiment. They were asked to read 17 good quality chest x-rays of pneumoconiosis, while their eye movements were recorded.

The eye movement recordings were done for earlier perception studies [9, 8] and the same recordings were used in the present study also i.e. to assess the relavence of computational bottom-up saliency models in chest x-rays. Experimental details are omitted here as the focus of present paper is not on perceptual experiments. Refer to [9, 8] for details regarding experimental design and eye movement recordings.

### 3.1.2 Results for chest X-rays

An ROC metric is used to study the role of bottom-up saliency models in predicting the eye fixations of the observers. ROC curves are drawn by considering the saliency maps extracted using saliency models as classifiers, and considering observer eye fixations as ground truth. For one threshold saliency value, the bottom-up saliency map is treated as a binary classifier on every pixel in the image [3, 10]. Saliency maps are thresholded such that a given percent of image pixels are classified as fixated and the rest are classified as not fixated. The fixations of the observers are treated as the ground truth. Varying over different thresholds yields different classifiers and thus an ROC curve. The area under this ROC (AUC) is considered as an indication for how well the bottom-up saliency map can predict the ground truth fixations of observers. For each observer, AUC values obtained for all images are averaged to get a single mean AUC value.

Figure 4 shows saliency maps extracted by using IK, GBVS and SR models on a sample chest x-ray used in our study. Matlab codes for all these saliency models are available on their respective author's webpages. The default parameter settings used by the authors were used in all the saliency map computations.

Figure 5 shows the plot of AUCs for all the observers in our study. The plot indicates that GBVS model outperforms other saliency models in predicting the fixation of the observers. Wilcoxon signed rank test showed that the AUCs corresponding to GBVS saliency maps ($Mdn = 0.77$) are significantly higher ($Z = 0.0, p < .001$) than those corresponding to both IK ($Mdn = 0.67$) and SR ($Mdn = 0.67$) saliency maps, when all the observers are considered. A median AUC value of 0.77 suggests that GBVS saliency model can be used to a reasonably good accuracy to predict the fixations of the observers.

### 3.1.3 An extended saliency model for chest x-rays

Our analysis showed that, on an average, around 84% of fixations are inside the lung regions whereas these regions cover only around 40-50% of the area in a chest x-ray. This clearly points to the importance given to lung regions which is due to the fact that pneumoconiosis is a lung disease. We wished to explore if it is possible to get better prediction of the observer saliency by combining the bottom-up saliency as predicted by GBVS model and the top-down saliency, which is the importance of lung regions. This was done by modifying the GBVS saliency maps with different weighting for the lung regions. The modified maps referred to as *Extended Graph Based Visual Saliency* (EGBVS) maps are derived using a method similar to that in [20] for combining top-down and bottom-up influences.

Figure 6 shows the proposed schema for computing EGBVS saliency map. The lung regions are segmented from the chest x-ray, using the procedure in [1] and a segmented lung mask is created as shown in figure 6. The EGBVS saliency map is obtained by combining GBVS saliency map and lung mask as follows:

EGBVS Saliency Map =
(GBVS Saliency map). $*$ (Segmented lung mask $+ K$)

Where $.*$ represents point wise multiplication and $K$ represents a positive real constant between 0 and 1. In segmented lung mask images, lung regions are represented by pixel values of 1 (white region) and remaining regions are represented by 0 (black region). The value of $K$ determines the relative importance given to lung and non-lung regions in final EGBVS saliency map. Thus, suppressing all information from non-lung regions is achieved by setting $K = 0$. Increasing $K$ increases the contribution from the non-lung regions but only relative to the lung region as this region's contribution is still higher ($(1 + K)$; $K < 1$). Empirically, $K = 0.5$ was found to be optimal. Figure 7 shows some EGBVS saliency maps computed from sample chest x-ray images.

Wilcoxon signed rank test showed that the AUCs for EGBVS ($Mdn = 0.81$) are significantly higher ($Z = 2.0, p < .001$) than those of GBVS ($Mdn = 0.77$), for all the observers. We have already seen that GBVS model predicts the observers' fixations to a good accuracy. EGBVS saliency model was found to perform better (5.4% increase in median AUC) than the GBVS model in predicting the eye fixations of the observers.

## 3.2 Study on Retinal Images

Since Hard Exudates (HEs) are approximately *disjunctive* targets, they can be identified through a pre-attentive process. Salient image regions corresponding to HE identification in an image need not correspond to the fixation locations of human readers. Hence, we did not collect eye-tracking data on retinal images for our study. Instead, we evaluated the bottom-up saliency models by comparing the saliency maps against the ground truth markings by medical experts.

### 3.2.1 Data set

The publicly available DIARETDB1 dataset [12] was used for our study. Out of total 89 images in this dataset, only 48 images had HEs and thus were considered for analysis. Ground truth markings provided by 4 medical experts are also available with this dataset. Figure 8 shows sam-

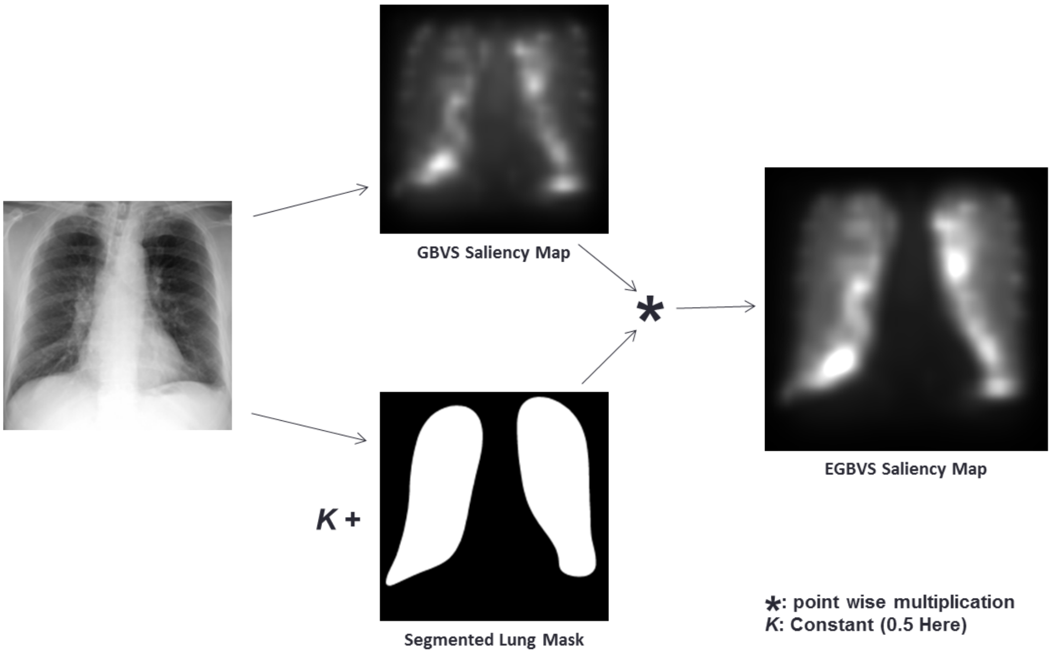


**Figure 6: Different steps in extracting EGBVS saliency map from a sample chest x-ray.**

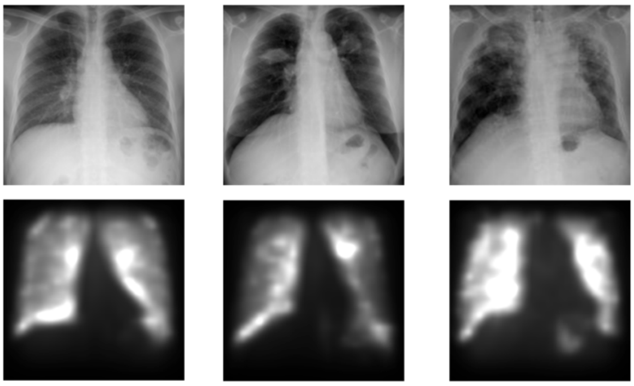

**Figure 7: Sample chest x-rays (above) and corresponding EGBVS saliency maps (below)**

ple retinal image and ground truth HE markings provided by experts. In the latter, agreement level is indicated by the grayscale value of the pixel: white indicating complete agreement and darker shades of grey indicating decresing level of agreement. In our work markings that are agreed upon by all medical experts (white regions) were considered as ground truth. A lesion is considered detected if there is more than 50% overlap between thresholded saliency map and the ground truth marking.

### 3.2.2 *Results*

The assessment of the computed saliency maps was done using the ROC curves as in [10]. The maps were thresholded to obtain different percent of saliency levels. These thresholded maps are compared with the ground truth to obtain a true positive rate (TPR) which is the ratio of correctly detected lesions to the total number of lesions. By varying the different thresholds an ROC is obtained. For each saliency model, an average ROC curve is computed by averaging ROC curves for all the images. The area under the ROC (AUC) is a metric used to rate the performance of a given bottom-up saliency model. A AUC value close to 1 indicates that the corresponding model is good at detecting HEs.

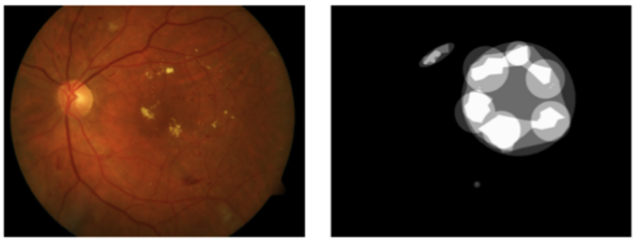

**Figure 8: (Left) A sample retinal image and (right) ground truth markings provided by four medical experts. Taken from DIARETDB1 dataset [12].**

Figure 10 shows saliency maps extracted by using IK, GBVS and SR models on a sample retinal image used in our analysis. Figure 11 shows the ROC curves for various saliency models considered. The corresponding average AUC values were found to be: 0.72 for IK, 0.70 for GBVS and 0.73 for SR. From these figures and the ROC plots we conclude that all three models perform roughly at the same level. In terms of lesion detection, all three models appear to be good at detecting lesions. However, this does not include detection of any false positives.

The optic disk in a retinal image (see figure 1) share similar characteristics with the HEs. Hence the OD is a fre-

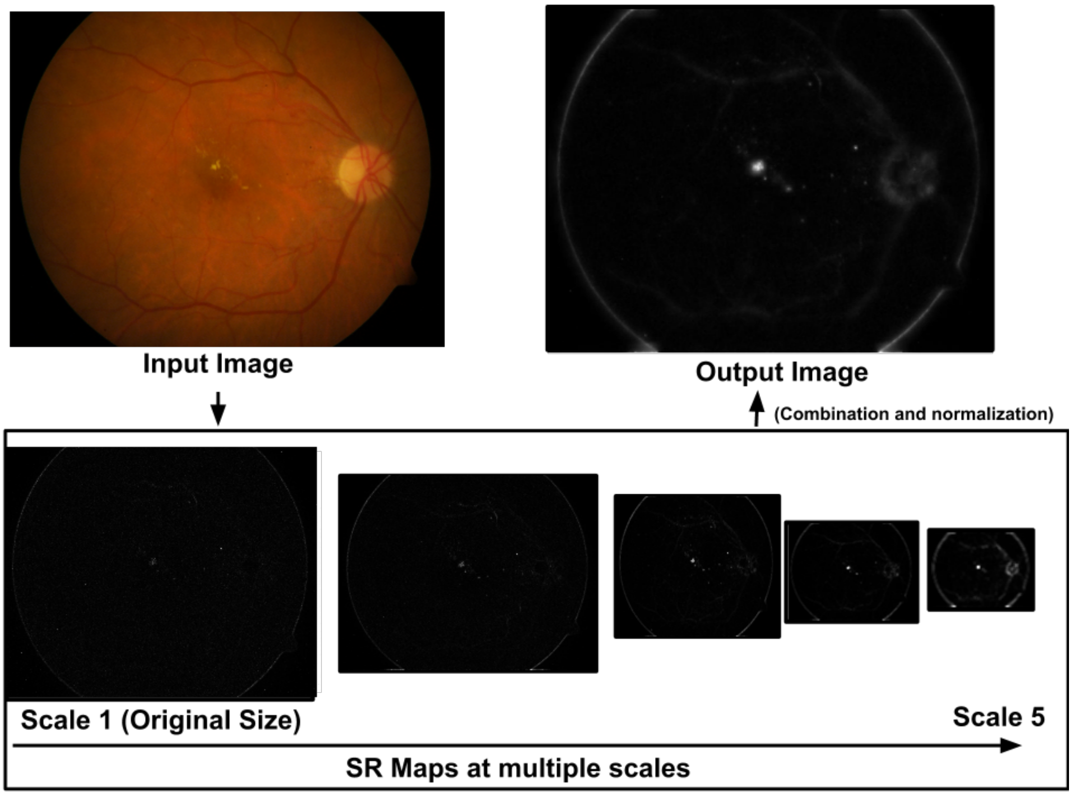


**Figure 9: Steps for deriving the ESR saliency map.**

quently detected false positive in the saliency maps of all the models. Suppressing the optic disk should theoretically not affect the performance of the models. This was confirmed experimentally: the average AUC rose to 0.74 with suppression for SR saliency maps. Thus, the improvement to be had is only 1.5% and it was also found that this improvement is not statistically significant.

### 3.2.3 Extended SR model

The SR model was originally designed to detect proto objects in given images [4]. Since the lesion of interest is a proto object, we wished to explore the possibility of improving the performance of this model in HE detection. The model is sensitive to the scale at which the map is computed. In [4], all given images were downsampled to $64 \times 64$ prior to computation of saliency maps and the final results were shown on the upsampled (to the original size) image. This standard size is argued to be the appropriate for "normal visual conditions".

However, our observation is that the SR saliency maps computed at different scales capture different proto objects as they contain different levels of details. The human visual system also has feature detectors which are sensitive to different scales which is modelled by IK [6, 7] by including a pyramid-based representation for feature extraction.

We take cue from the IK approach and extend the SR model by combining SR saliency maps computed at multiple scales of to get a single saliency map. We call such extended model as 'Extended SR' (ESR)model. Figure 9 shows a basic scheme for the proposed ESR model. Given an image, a dyadic scale representation is derived and the SR map is computed at each scale and normalized. Finally all the SR maps are resized to the original image size and added together and normalized to derive the ESR saliency map. In our implementation, 5 dyadic scales were used, starting with scale 1 which was the original image.

Figure 12 shows the computed ESR saliency map for some sample retinal images. The inclusion of information at different scales appears to detect and greatly improve the localisation of the lesions. Figure 11 shows the average ROC plot for ESR. This is well above the other curves. The average AUC values for ESR is greatly improved (0.94) which is a significant rise over the SR case (28.76% increase in AUC) thus validating the importance of including information from multiple scales.

## 4. DISCUSSION AND CONCLUSIONS

In this paper, we have analyzed the relevance of saliency models in detecting abnormalities in two types of medical images. In one study, we analyzed the role of bottom-up saliency in predicting the eye fixations of observers while diagnosing chest x-rays of pneumoconiosis. In another part of study, we analyzed the role of bottom-up saliency in detecting Hard Exudates in color retinal images.

Results obtained on chest x-rays indicate that the GBVS saliency model performs reasonably well in predicting the fixations of observers. Analyses on inter and intra-observer fixation consistency (not discussed here) showed that AUC values obtained for GBVS are comparable to AUCs obtained for analyses on inter and intra observer fixation consistency.

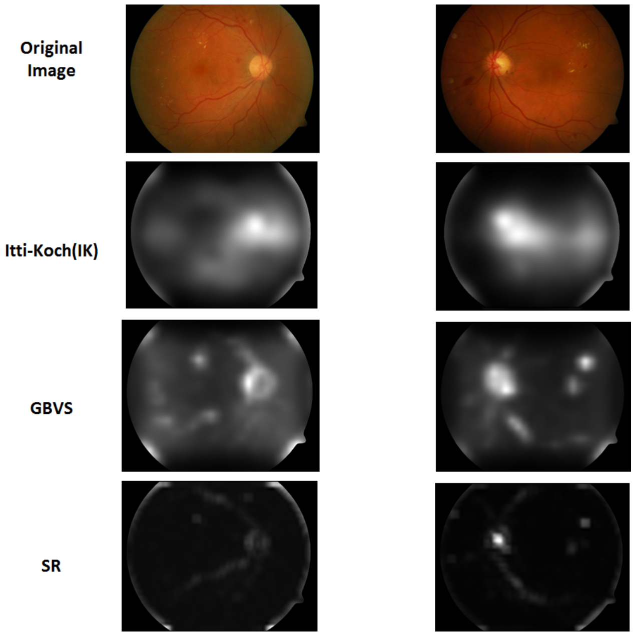


**Figure 10: Saliency maps extracted from 2 sample retinal images.**

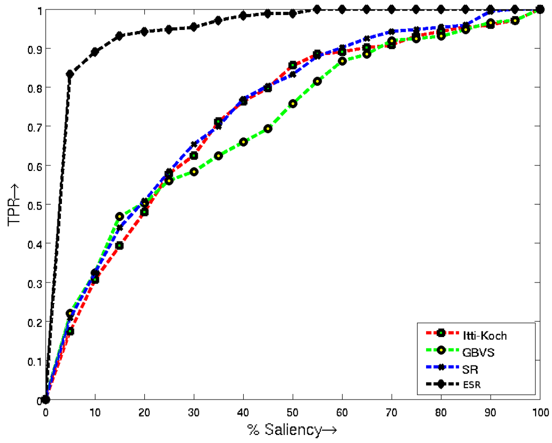


**Figure 11: Average ROC curves for different saliency models.**

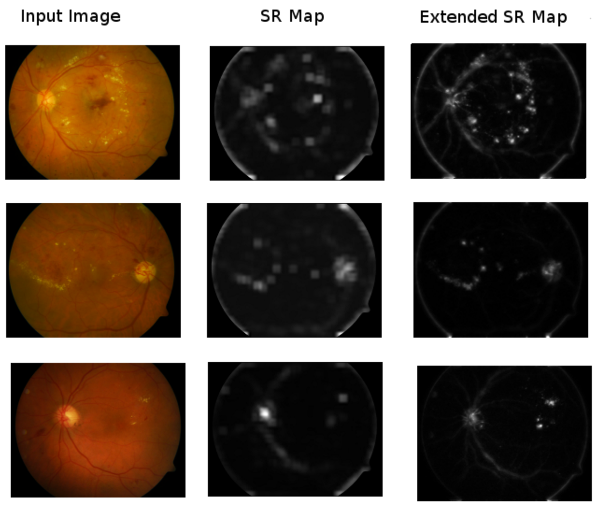


**Figure 12: Comparison of saliency maps computed with SR and ESR models.**

Thus GBVS performs well in predicting fixations. Since GBVS has been shown to be one of the best performing saliency model predicting the eye fixations on natural scenes [3], its current performance extends its suitability for X-ray images with subtle abnormalities. Thus, we can say that bottom-up image features, which attract fixations of observers in natural scenes, also play an important role in attracting the fixations of radiologists. The low performance of SR model can be attributed to the fact that it is designed to detect proto-objects via a pre-attentive type of process which is a reason it has not been assessed against gaze maps on natural scenes [4].

Analyses on retinal images with HE, indicate that the IK, GBVS and SR models are equally successful in picking up lesions as can be seen from their AUC values which are all significantly higher than 0.50. In terms of localisation of the detected lesions, IK is least preferrable as its saliency map is the most diffused while the ESR is most preferred. Poor localisation is a well known problem of IK saliency maps even for natural scenes and GBVS mitigates this problem but not significantly. It is generally expected that examining medical images involves top down knowledge to a great degree. However, our studies (with chest X-rays and retinal images) indicated that the role for bottom up knowledge is considerably high. Our attempt to incorporate top-down anatomical knowledge via a new saliency model (EGBVS) demonstrated that while there is an improvement, the margin of improvement ( $< 10\%$) is not as high as one would expect. This attempt used only anatomical knowledge (location of lungs). It is possible that further improvements in performance can be had by using other types of top-down influences such as the role of contralateral symmetry information, influences specific to expertize etc. These remain the part of future work.

In the case of retinal images, we found that suppressing the optic disc does not result in the significant improvement in HE detection accuracy of the SR model, which is to be expected as HEs and optic disc are similar in brightness and hence will be captured at the same (threshold) level of saliency. This reinforces our assumption that HE is a highly disjunctive feature. The finding on inclusion of multi-scale information improving the HE detection performance of the SR model demonstrates that rather than top down knowledge, the low level features are more important. The reason that this inclusion greatly improves the localisation of the detected lesion is that the saliency of features varies with scale. Alternate ways to combine the saliency information across scales needs further investigation. Likewise, there is a need to investigate if the false positive detection rate reduces with ESR as such a study can pave way for its employment as a candidate selection stage in a HE detection system.

Although the present study is specific to chest x-rays of pneumoconiosis and hard exudates in retinal images, we believe that many of these results can be extended to abnormalities with similar characteristics. For instance, many of the results obtained in the study of pneumoconiosis chest x-rays can be extended to other diffused lung diseases such as interstitial lung diseases. It is however, not clear to what extent these results can be extended to localized lung lesions such as tumors. However, the present study together with a previous study on stroke CT images [16] show that bottom up saliency plays an important role in medical images.

## 5. REFERENCES


[1] P. Annangi, S. Thiruvenkadam, A. Raja, H. Xu, X. Sun, and L. Mao. A region based active contour method for x-ray lung segmentation using prior shape and low level features. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 892–895. IEEE, 2010.

[2] L. Elazary and L. Itti. Interesting objects are visually salient. *Journal of Vision*, 8(3), 2008.

[3] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Advances in neural information processing systems*, 19:545, 2007.

[4] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[5] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *Image Processing, IEEE Transactions on*, 13(10):1304–1318, 2004.

[6] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.

[7] L. Itti and C. Koch. Computational modeling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[8] V. Jampani, V. Vaidya, S. Jayanthi, P. Ajemba, and L. Kishore. Effect of expertise and contralateral symmetry on the eye movements of observers while diagnosing pneumoconiosis. *XIV Medical Image Perception Society Conference*, Dublin, August, 2011.

[9] V. Jampani, V. Vaidya, S. Jayanthi, and L. Kishore. Role of expertise and contralateral symmetry in the diagnosis of pneumoconiosis: an experimental study. In *Proceedings of SPIE*, volume 7966, page 79661K, 2011.

[10] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2106–2113. IEEE, 2009.

[11] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[12] T. Kauppi, V. Kalesnykiene, J. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä. Diaretdb1 diabetic retinopathy database and evaluation protocol. *Proc. Medical Image Understanding and Analysis (MIUA)*, pages 61–65, 2007.

[13] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–27, 1985.

[14] E. Krupinski. The importance of perception research in medical imaging. *Radiation Medicine*, 18(6):329–334, 2000.

[15] D. Manning, A. Gale, and E. Krupinski. Perception research in medical imaging. *British journal of radiology*, 78(932):683–685, 2005.

[16] H. Matsumoto, Y. Terao, A. Yugeta, H. Fukuda, M. Emoto, T. Furubayashi, T. Okano, R. Hanajima, and Y. Ugawa. Where do neurologists look when viewing brain ct images? an eye-tracking study involving stroke cases. *PloS one*, 6(12):e28928, 2011.

[17] R. Morgan, M. Donner, B. Gayler, S. Marguiles, P. Rao, and P. Wheeler. Decision processes and observer error in the diagnosis of pneumoconiosis by chest roentgenography. *American Journal of Roentgenology*, 117(4):757–764, 1973.

[18] C. Nodine and H. Kundel. The cognitive side of visual search in radiology. *Eye Movements: From Psychology to Cognition. North Holland, Elsevier Science*, pages 572–582, 1987.

[19] C. Nodine, H. Kundel, et al. Using eye movements to study visual search and to improve tumor detection. *Radiographics*, 7(6):1241–1250, 1987.

[20] R. Peters and L. Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[21] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[22] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch. Attentional selection for object recognition: a gentle way. In *Biologically Motivated Computer Vision*, pages 251–267. Springer, 2002.

[23] J. Wolfe. Asymmetries in visual search: An introduction. *Attention, Perception, & Psychophysics*, 63(3):381–389, 2001.