Journal of Theoretical and Applied Information Technology

30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TBVA-GAN: A CLASS ACTIVATION MAPPING-GUIDED TUBERCULOSIS VISUAL ATTRIBUTION GENERATIVE ADVERSARIAL NETWORK

DING ZEYU[1,2], RAZALI YAAKOB[1], AZREEN AZMAN[1], SITI NURULAIN MOHD RUM[1], NORFADHLINA ZAKARIA[1], AZREE SHAHRIL AHMAD NAZRI[1]

[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, 43400, Malaysia

[2]Department of Computer Science, Changzhi University, Changzhi, 046000, China.

E-mail: dzy6489@163.com, {∗razaliy, azreenazman, snurulain, n_fadhlina, azree}@upm.edu.my

ABSTRACT

Visual attribution (VA) methods play a crucial role in tuberculosis (TB) research by providing valuable insights into disease patterns and aiding in diagnostic interpretation. The advent of generative adversarial network (GAN)-based VA methods has gained significant attention from researchers due to their ability to generate fine-grained feature maps that accurately reflect the location of lesions. These methods localize lesions by converting chest X-ray (CXR) images containing lesions into normal CXR images and analyzing the differences between the two. However, current methods only perform surface-level transformations, neglecting the vital information of whether lesions are present. Moreover, the transformation process assigns equal weights to the entire image, without specifically prioritizing the regions with a higher probability of lesions occurrence. In this study, a novel framework is proposed, namely the class activation mapping-guided tuberculosis visual attribution generative adversarial network (TBVA-GAN). This innovative model leverages the informative regions derived from class activation mapping to effectively guide the GAN in prioritizing the transformation of these crucial areas. Moreover, to guarantee the precision of TB localization, an auxiliary TB detection model is incorporated, ensuring that the converted CXR images are devoid of TB pathology. By employing this additional verification step, the accuracy of TB localization is significantly enhanced. The proposed TBVA-GAN in this study achieves promising VA results on the TBX11K dataset, surpassing existing GAN-based TB VA models.

Keywords: *Visual attribution, Tuberculosis, Deep learning, GAN*

1. INTRODUCTION

Deep learning techniques have demonstrated a remarkable ability to automatically learn and extract complex patterns from medical images, leading to significant breakthroughs in various fields such as skin cancer diagnosis [1], lung nodule detection [2], and hypertrophic cardiomyopathy recognition [3]. Progress has also been achieved in the field of tuberculosis diagnosis with this technology, surpassing even radiologists in detecting active pulmonary tuberculosis [4]. This advancement holds the potential to assist medical practitioners more effectively in making diagnostic decisions [5]. However, the opacity of deep learning models, often referred to as the black box nature, presents a significant obstacle to their acceptance and adoption by radiologists [6]. The lack of interpretability and explain ability hampers radiologists' confidence in the decision-making process and raises concerns regarding clinical accountability.

Visual attribution (VA) techniques have emerged as a promising tool in the realm of medical imaging [7][8][9], affording researchers and radiologists invaluable discernment into the intrinsic patterns and distinguishing characteristics that guide diagnostic determinations. VA is a task that involves localizing and visualizing evidence of a specific category within an image, thereby highlighting the areas that contribute most significantly to the final diagnostic outcome. This technology is widely applied in medical image analysis, particularly in weakly supervised localization and segmentation [10][11][12][13][14].

The traditional methods of VA are based on gradient-based approaches [15], where the gradients

of the model's output with respect to a specific class are computed to identify the regions in the image associated with that class. This approach is known as class activation mapping (CAM)-based methods which have been widely used in medical image analysis applications, such as lung cancer detection [16], diabetic retinopathy classification [17], covid-19 prediction [18], tuberculosis (TB) detection and visual explanation [19]. However, the heat maps generated by this method are based on lower-resolution feature maps and can suffer from miss alignments, resulting in poor performance in fine-grained localization tasks. As shown in Figure 1(a), the red region indicates areas with a high probability of TB presence. However, due to the low resolution of its feature map, this red region extends beyond the lung region, resulting in an imprecise localization of the TB location. To address this issue, the researchers put forward GAN-based methods [7][8].

GAN-based methods transform abnormal images into healthy images and analyze the differences between the two to locate the lesion regions. Figure 1(b) showed the changemap, which has the same resolution as the input raw image, indicating the differences between abnormal and healthy images. Brighter regions in the changemap correspond to the lesion areas. However, when applying this technology to TB detection [9], the contrast between light and dark areas in the changemap is not clearly discernible, posing challenges for researchers and physicians in accurately identifying the lesion area with the naked eye. Besides, selecting an optimal threshold value to extract the lesion region becomes a challenging task. In addition to the above problems, the GAN-based TB detection method can only ensure that the transformed images appear healthy at a superficial level, with the presence of TB remaining unknown. As a result, the resulting changemap lacks some critical TB information, making it inaccurate to use for TB localization. Moreover, a further limitation of this method is that when converting TB-infected images to normal ones, the transformation weight is uniformly distributed across the entire image, without any special consideration given to areas containing TB. This deficiency could significantly impact the quality and reliability of the transformed images, particularly in areas where the concentration of TB is high.

In response to the aforementioned issues, this study introduces a class activation mapping-guided tuberculosis visual attribution generative adversarial network (TBVA-GAN) model. This model exhibits the following advantages:

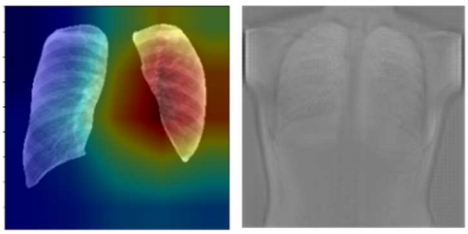

Figure 1: Two Commonly Used VA Methods

(a) Compared to CAM-based methods, the proposed TBVA-GAN model is capable of generating fine-grained changemaps that are consistent with the resolution of the original image, enabling more precise localization of TB regions.

(b) In the process of transforming TB-infected CXRs to normal CXRs, this model enhances the weight of TB region conversion guided by CAM. As a result, the generated changemap exhibits a more pronounced contrast between light and dark areas, resulting in clearer and well-defined contours of the TB region. Consequently, the TB-affected area can be visually observed with the naked eye.

(c) In contrast to the manual threshold setting required in MDVA-GAN [9], this study proposes a threshold estimation algorithm guided by CAM, which can automatically extract the TB mask.

(d) When employing conventional methods to convert TB-infected CXRs to normal CXRs, the resulting normal CXRs may appear normal superficially but cannot guarantee the absence of TB within. The algorithm proposed in this study ensures that the converted normal CXR is free from TB, thereby ensuring the accuracy of the changemap.

2. LITERATURE REVIEW

Visual analysis plays a crucial role in automated diagnosis, providing significant support for medical professionals' diagnostic procedures. Currently, two mainstream approaches dominate VA research: CAM-based methods and GAN-based methods.

2.1 Class Activation Mapping (CAM)-based Methods

CAM enables the visualization of image regions contributing significantly to specific classification outcomes within convolutional neural networks (CNNs). Grad-CAM [15] and its variants [20][21][22] have made improvements by employing gradient-based computations and have found successful applications in medical image analysis. For example, research [23] has integrated deep Bayesian optimization and Grad-CAM to develop an interpretable AI framework for diagnosing COVID-19. This framework assists radiologists in swiftly identifying lesion locations. However, Grad-CAM exhibits limitations. Due to the computational efficiency requirements, convolutional neural networks often employ pooling operations, which reduce the resolution of the extracted feature maps. Consequently, heat maps based on low-resolution feature maps suffer from low resolution and alignment issues, leading to suboptimal performance in fine-grained image localization. Similar challenges are encountered in tuberculosis diagnosis using CAM-based methods. For instance, research [24][25] employed CAM-based methods to locate lesion areas in tuberculosis cases. However, due to the diverse nature of tuberculosis types, especially in the case of cavitary pulmonary tuberculosis, lesions exhibit small spatial extents in CXR, and the low resolution of heat maps often results in out-of-range localizations, which can disrupt radiologist diagnoses. Hence, in tuberculosis diagnosis, CAM-based methods are constrained by the limitations of heat map resolution and can only provide approximate lesion localization, falling short in achieving fine-grained precision.

2.2 Generative Adversarial Network (GAN)-based Methods

Due to the limitations of CAM, researchers have attempted to use GAN to address VA problems. GAN-based methods transform images containing lesions into healthy images, highlighting the differences between the two to generate a change-map that accurately reflects the location of the lesions. Baumgartner developed a novel VA technique based on Wasserstein GAN (WGAN), which was able to accurately localize lesion regions in real data from patients with mild cognitive impairment (MCI) and alzheimer's disease (AD) [26]. Similarly, VANT-GAN [8] utilized a similar approach to transform abnormal images into normal images and localize lesion regions by highlighting the differences between the two. However, both of these methods do not specifically emphasize the regions with a higher probability of containing lesions during image transformation. Instead, they assign uniform weights to the entire image. This results in excessive noise in regions without lesions. Recently, Nawaz proposed a multi-domain VA GAN (MDVA-GAN) [9] that can perform VA tasks for multiple diseases and achieve good results on CheXpert and TBX11K datasets. Nevertheless, the study did not specify how to extract the corresponding disease mask from the changemap. Furthermore, ANT-GAN [7] can transform abnormal images into normal images while focusing on the lesion regions without affecting the healthy parts. Nevertheless, the resulting "normal" image may visually resemble a typical image but does not guarantee the absence of lesion-related information within it. Fortunately, TUNA-NET [27] has introduced a solution to detect the presence of lesion-related information in the transformed images. TUNA-NET, by incorporating an auxiliary function for lesion detection, preserves lesion-related information when transforming adult pneumonia images into pediatric pneumonia images. Research [28] accomplishes image transformations in the latent space, thereby avoiding pixel-level conversions and maintaining superior semantic consistency.

Based on the analysis of the aforementioned literature, there is still a need for improvement in the following areas: (a) In the process of image transformation, specific attention should be given to regions with a high likelihood of containing lesions, rather than applying uniform weighting across the entire image. (b) To ensure the accuracy of the changemap, it is essential to guarantee the absence of lesion-related information in the transformed normal images. (c) A method capable of automatically extracting masks from a changemap should be designed.

3. TUBERCULOSIS VISUAL ATTRIBUTION GENERATIVE ADVERSARIAL NETWORK (TBVA-GAN)

This section will discuss the proposed methodology for this work. Figure 2 shows the architecture of TBVA-GAN and each part will discuss in the next subsection.

3.1 Definition of Elements in TBVA-GAN

This section will be explained the definitions in figure 2. TBVA-GAN is improved based on Cycle-

GAN [29] and Grad-CAM [15]. Based on figure 2, real_x_T is the CXR with TB (The red regions in real_x_T indicate the TB areas). G_{T2N} is a generator to produce a changemap that depicts the location of

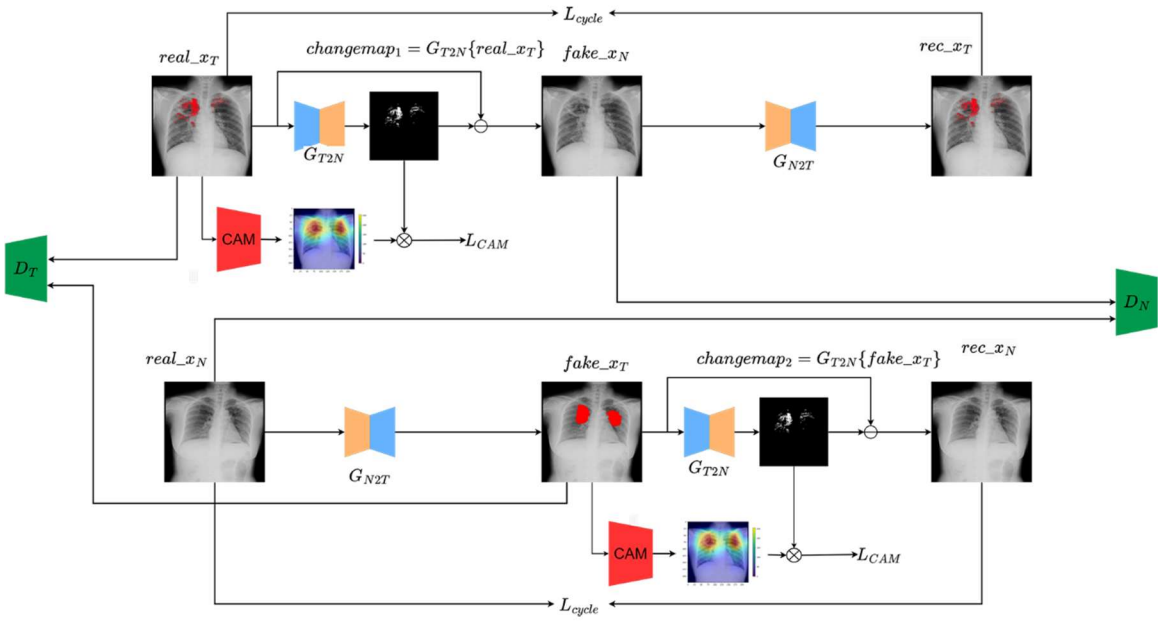

Figure 2: The Overall Framework of TBVA-GAN

TB from the CXR image which contains TB. The fake_x_N is obtained by subtracting $changemap_1$ from real_x_T. The purpose of this process is to minimize the presence of TB information in fake_x_N, thereby generating a normal-looking CXR image. G_{N2T} is another generator to generate a TB-looking CXR image from a normal CXR image. D_T is the discriminator of the GAN, capable of distinguishing whether the input CXR is real_x_T or fake_x_T. Similarly, D_N serves as another discriminator to determine whether the input CXR is real_x_N or fake_x_N. CAM refers to the heat map produced by Grad-CAM [15]. It is used to guide the generated changemap to focus more on the regions with the highest probability of TB occurrence.

3.2 Loss function of TBVA-GAN

The objective of this architecture is to obtain a changemap that is capable of effectively visualizing the precise location of TB. This changemap can be represented as $changemap_1 =$ real_x_T − fake_x_N. To obtain this changemap, a comprehensive objective loss function was devised in this study, consisting of five components, as expressed by the following equation:

$$Loss = \mathcal{L}_{\text{GAN}} + \mathcal{L}_{\text{Cycle}} + \mathcal{L}_{\text{Identity}} + \lambda_1\mathcal{L}_{\text{Cls}} + \lambda_2\mathcal{L}_{\text{CAM}} \quad (1)$$

Among them, the role of $\mathcal{L}_{\text{GAN}}$ is to generate a visually similar image fake_x_N that resembles real_x_N. $\mathcal{L}_{\text{Cycle}}$ ensures that the only difference between fake_x_N and real_x_N is the style, while the overall content remains consistent. $\mathcal{L}_{\text{Identity}}$ guarantees that if the input image is real_x_N, the resulting transformed image remains real_x_N, thereby ensuring an empty changemap that does not reflect the presence of TB. $\mathcal{L}_{\text{Cls}}$refers to the classification loss, which ensures that fake_x_N not only visually resembles real_x_N but also maintains consistency in terms of the presence or absence of TB information. $\mathcal{L}_{\text{CAM}}$ aims to guide the generated changemap to focus more on the regions with the highest probability of TB occurrence. The following sections provide detailed explanations for each of the aforementioned components.

3.3 Generative Adversarial Network (GAN) Loss

The full name of GAN is generative adversarial network, which consists of a generator (G_{T2N}) and

a discriminator (D_N). For example, in Figure 2, when taking a CXR containing TB (real_x_T) as input, the role of the generator G_{T2N} is to generate a changemap ($changemap_1 = G_{T2N}(real\_x_T)$), and it can reflect the position of TB. The generated normal CXR is fake_x_N. The role of the discriminator D_N is to distinguish the real normal CXR real_x_H from the generated normal CXR fake_x_N. The generator and the discriminator fight against each other and evolve with each other to ensure that the generator G_{T2N} can obtain a more realistic normal CXR fake_x_N and a more accurate changemap. This paper is based on Cycle-GAN and uses a bidirectional conversion model, in addition to G_{T2N} and D_N, there are also G_{N2T} and D_T to accomplish the conversion of normal CXR real_x_N to CXR containing TB fake_x_T. We assume that the observed samples are drawn from their respective distributions, with $real_{x_T} \sim p_T(x)$ and $real_{x_N} \sim p_N(x)$. $\mathbb{E}$ represents the mathematical expectation. The GAN loss function can be expressed as:

$$\begin{aligned}\mathcal{L}_{\text{GAN}} &= \mathbb{E}_{p_T}\left[\ln \mathcal{D}^T(real_{x_T})\right] + \mathbb{E}_{p_N}\left[\ln \mathcal{D}^N(real_{x_N})\right] \\ &+\mathbb{E}_{p_N}\left[\ln\left(1-\mathcal{D}^T\left(G_{\text{N2T}}(real_{x_N})\right)\right)\right] \\ &+\mathbb{E}_{p_T}\left[\ln\left(1-\mathcal{D}^N\left(real_{x_T}-G_{\text{T2N}}(real_{x_T})\right)\right)\right]\end{aligned}$$

(2)

3.4 Cycle Loss

To ensure that the only difference between real_x_T and fake_x_N is the style, while maintaining consistent overall content, it is required that the rec_x_T generated by fake_x_N closely resembles real_x_T. The cycle loss is introduced as below:

$$\mathcal{L}_{\text{Cycle}} = \mathbb{E}_{p_N}\left[\left\|G_{\text{T2N}}\left(G_{\text{N2T}}(real_{x_N})\right) - real_{x_N}\right\|_1\right] + \mathbb{E}_{p_T}\left[\left\|G_{\text{N2T}}\left(real_{x_T} - G_{\text{T2N}}(real_{x_T})\right) - real_{x_T}\right\|_1\right]$$

(3)

3.5 Identity Loss

G_{T2N} is functional in the conversion of a CXR containing TB real_x_T to normal CXR fake_x_N. However, if the input of G_{T2N} is a normal CXR real_x_N, the output should be an empty changemap. Likewise, if the input of G_{N2T} is a CXR containing TB real_x_T, the output should remain unchanged. Therefore, identity loss is proposed with the following expression.

$$\mathcal{L}_{\text{identity}} = \mathbb{E}_{p_T}[\|G_{N2T}(real\_x_T) - real\_x_T\|_1] + \mathbb{E}_{p_N}[\|G_{T2N}(real\_x_N)\|_1]$$

(4)

3.6 Classification Loss

Using GAN, we can generate realistic fake_x_N that may appear to be similar to real_x_N, but crucial information may be missing, such as the presence of TB. For instance, the discriminator D_T can distinguish between real image real_x_T and generated image fake_x_T, but it cannot guarantee whether fake_x_T contains TB. Similarly, the discriminator D_N can differentiate between real image real_x_N and generated image fake_x_N, but the presence of TB in fake_x_N remains unknown. Creating similar-looking images is meaningless. We want the generated images not only to be visually similar but also to remain consistent in terms of the presence or absence of TB. To achieve this, we trained an auxiliary model $\mathcal{F}$ to detect TB on a labeled dataset. In the experimental section, the auxiliary model will be introduced.

$$\begin{aligned}\mathcal{L}_{\text{cls}} &= \mathbb{E}_{p_T}\ln\left[1-\mathcal{F}\left(real_{x_T} - G_{\text{T2N}}(real_{x_T})\right)\right] \\ &+\mathbb{E}_{p_N}\left[\ln\left(\mathcal{F}\left(G_{\text{N2T}}(real_{x_N})\right)\right)\right]\end{aligned}$$

(5)

3.7 Class Activation Mapping (CAM) Loss

Grad-CAM is capable of generating attention maps, as demonstrated in the CAM section of Figure 2. The intensity of red color in the attention map corresponds to a higher likelihood of TB occurrence. When the input is a CXR containing TB, G_{T2N} can generate a changemap, then the attention map generated by Grad-CAM can guide the transformation of G_{T2N}. G_{T2N} focuses on the red region of the attention map and tries to keep the rest of the region unchanged. Then CAM loss can be expressed as:

$$\begin{aligned}\mathcal{L}_{\text{CAM}} &= \mathbb{E}_{p_T}\left[\left\|(\mathbf{1}-\mathbf{M}_{CA\ \ _1})\odot\left(G_{\text{T2N}}(real_{x_T})\right)\right\|_2^2\right] \\ &+\mathbb{E}_{p_N}\left[\left\|(\mathbf{1}-\mathbf{M}_{CAM_2})\odot\left(G_{\text{T2N}}\left(G_{\text{N2T}}(real_{x_N})\right)\right)\right\|_2^2\right]\end{aligned}$$

(6)

The $\odot$ represents element-wise multiplication and $\mathbf{1}$ is an all-ones matrix of the same size as the input image. $\mathbf{M}_{CAM}$ denotes the attention map obtained by Grad-CAM. The region is represented by $\mathbf{1}-\mathbf{M}_{CAM}$ indicates the area outside the red region, and changes in this

region should be minimized as much as possible. If G_{T2N} tries to modify the part of the attention map outside the red area, it will be penalized with the corresponding L2 loss.

4. EXPERIMENT

This study was conducted on a computer system equipped with an Intel(R) Core(TM) i7-7700K CPU, 16GB RAM, and a NVIDIA GeForce RTX 3090 24GB GPU. The operating system was Ubuntu 20.04.3 LTS, with PyTorch version 1.12.1 and Python version 3.9.1 employed. After conducting extensive experiments, we observed that the optimal performance was achieved when the value of the parameter λ_1 in Equation 1 was set to 1.5, and the parameter λ_2 was set to 2.

4.1 Dataset

The dataset used in this paper is TBX11K, a large dataset containing a total of 11,200 CXR images with 4 categories: healthy, active TB, sick non-TB, and latent TB. Each CXR with TB was marked with a box as ground truth by an experienced radiologist.

TBX11K was divided into three datasets: training set, validation set, and test set. The test set was used for competition validation and no corresponding annotation was provided, so only data from the training set validation set were used in this study. In addition, only two types of data, healthy and active TB, were used in this study, with a total of 3800 healthy CXR and 630 active TB data.

4.2 The Auxiliary Tuberculosis Recognition Model

The auxiliary TB recognition model $\mathcal{F}$ can be employed to detect the presence of TB in the generated images. If the input image x contains TB, then $\mathcal{F}(x) = 1$, otherwise, $\mathcal{F}(x) = 0$. This model utilized the pre-trained VGG16 [30] as the baseline model, modified the classification layer, and fine-tuned it on a new dataset. In order to ensure the balance of the data, 630 active TB CXR and 630 healthy CXR were used in the TB recognition model. Among them, the 630 healthy CXRs were randomly selected from all 3800 healthy CXRs. To ensure the diversity of the data, preprocessing operations including random horizontal flipping and random angle rotation were applied to the pictures, with the rotation angles of 45º, 90º, 135º, 180º, 225º, 270º, 315º. All the pictures were resized to 224*224. The dataset was divided into a training set and a testing set with 70% and 30% respectively. After testing, the model achieved an accuracy of 99.21% and a recall rate of 99.21%. Meanwhile, the visualization of CXR using Grad-CAM also relies on the model.

4.3 Comparison of the Changemap of Each Model

Figure 3(a) displays the original CXR images with the TB region marked by a blue box as the ground truth. Figure 3(b-d) depicts the changemaps generated by MDVA-GAN, ANT-GAN, and TBVA-GAN, respectively. ANT-GAN was reproduced on the TBX11K dataset using the method provided by the author. The changemap represents the extent of change from TB CXR images to CXR images of healthy individuals, with brighter areas indicating more pronounced changes. If the brighter regions align with the TB region in the ground truth, it suggests that the model primarily focuses on the TB lesion. By employing a reasonable threshold to extract the brighter regions from the changemap, the approximate location of TB can be determined. This process is further elaborated in Figure 4, and detailed explanations will follow in subsequent experiments. The changemaps produced by MDVA-GAN and ANT-GAN exhibit subtle variations in brightness, necessitating attentive scrutiny to discern the brighter regions. Additionally, the outlines and boundaries of these brighter areas lack clarity. In contrast, the changemap generated by TBVA-GAN exhibits a distinct contrast in brightness, allowing the naked eye to readily perceive the outlines of the brighter regions.

Figure 3(e) shows the heatmaps generated by Grad-CAM, where the areas closer to red indicate a higher probability of containing TB. From the images, it is evident that the red regions possess well-defined outlines and clear boundaries, aligning closely with the TB locations indicated in the ground truth. Conversely, the green regions exhibit less distinct boundaries, making it challenging to establish a threshold for extracting that specific area. Furthermore, the green regions noticeably extend beyond the TB region marked in the ground truth, rendering the extraction of such regions less meaningful.

Therefore, from a qualitative analysis perspective, TBVA-GAN can provide more

precise and distinct localization of TB lesion areas compared to other models, thereby offering better diagnostic support for medical practitioners.

4.4 Determination of Threshold Value

The second row of images in Figure 4 depicts the binary mask extracted by Grad-CAM, which only extracts the red areas from the heatmap, as this region has the highest probability of TB occurrence and its boundary is more clearly defined. It can also be observed that although the mask generated by Grad-CAM misses some TB regions marked in the ground truth, it still exhibits a general consistency with the position of TB in the ground truth, and the overlapping areas occupy a significant portion of the mask. Therefore, the brighter pixels in the change map constitute a considerable proportion within the mask, and this characteristic can be leveraged to analyze the histogram and estimate an appropriate threshold. Figure 5 presents an example of the histograms for the Grad-CAM region and the ground truth region in the changemap. It can be observed from the histograms that their distributions are relatively consistent, with the brighter pixels occupying the majority. Extensive experiments have indicated that selecting a threshold at 97% of the pixel value at the peak of the histogram yields the best results.

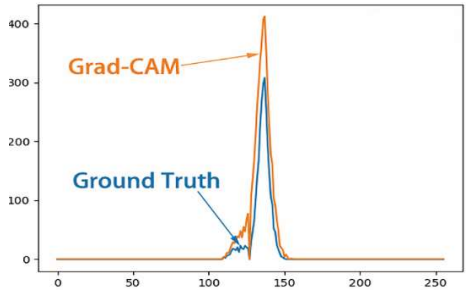

Figure 5: Determination of the threshold value

4.5 Comparison of the Binary Masks of Each Model

After obtaining the threshold, the changemap is binarized, resulting in a preliminary binary image mask. However, this mask often contains a significant amount of noise that requires further processing. Subsequently, by applying erosion (with a 2x2 kernel) and dilation (with a 5x5 kernel) operations, the final mask can be obtained. The second to fifth rows of Figure 4 respectively display the VA mask of Grad-CAM, MDVA-GAN, ANT-GAN, and TBVA-GAN. Comparing the masks obtained from these models reveals that the mask generated by MDVA-GAN exhibits the poorest performance. It fails to capture certain TB areas marked in the ground truth and inaccurately labels non-TB regions as TB. ANT-GAN shows an improvement compared to MDVA-GAN, as it no longer misses areas with TB presence. However, it still erroneously labels regions without TB as positive. TBVA-GAN demonstrates the best performance, with its generated mask showing a higher degree of alignment with the actual location of TB, and exhibiting the smallest error.

4.6 Quantitative Results

Based on the binary images, masks extracted by each model can be selected using a red bounding box. The TB region annotated in the ground truth is shown as a blue box. The ability of each model to locate the TB region can be measured using the intersection over union (IoU) and Dice score. The results of IoU and Dice for MDVA-GAN, Grad-CAM, ANT-GAN, and Grad-CAM are shown in Table 1.

We also evaluated the performance of TBVA-GAN models with and without the inclusion of $\mathcal{L}_{\text{Cls}}$ and $\mathcal{L}_{\text{CAM}}$ constraints, and the results are presented together in Table 1. The results show that both the $\mathcal{L}_{\text{Cls}}$ and $\mathcal{L}_{\text{CAM}}$ contributed to the improvement of TBVA-GAN's effectiveness. This demonstrates that TBVA-GAN is proficient at focusing on regions with a high probability of TB occurrence, and the resulting normal CXR does not contain TB-related information, thus yielding a more precise changemap. Moreover, this reaffirms the effectiveness of the proposed method for mask extraction from the changemap as described in this study. Incorporating both $\mathcal{L}_{\text{Cls}}$ and $\mathcal{L}_{\text{CAM}}$ concurrently, the TBVA-GAN achieved IoU and Dice scores of 38.3% and 55.2%, respectively, surpassing the performance of all the models listed in Table1. This suggests that TBVA-GAN outperforms other models in accurately localizing TB lesion regions, thereby offering improved support for radiologists in diagnosis.

Table 1: Performance of TB Localization for Each Model

Methods	IoU	Dice
MDVA-GAN [9]	25.4%	40.4%
Grad-CAM [15]	33.1%	49.6%
ANT-GAN [7]	35.8%	52.8%
TBVA-GAN ($\mathcal{L}_{\text{Cls}}$)	32.7%	49.3%
TBVA-GAN ($\mathcal{L}_{\text{CAM}}$)	36.1%	52.9%
TBVA-GAN ($\mathcal{L}_{\text{Cls}} + \mathcal{L}_{\text{CAM}}$)	38.3%	55.2%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. DISCUSSION

This study compared the performance of several models on the TBX11K dataset for VA. The results indicate that MDVA-GAN merely performs a superficial transformation between CXRs containing TB and normal CXRs, resulting in a poor conversion quality due to the loss of crucial TB presence information. ANT-GAN, on the other hand, focuses on the transformation of regions containing TB, leading to a significant improvement over MDVA-GAN. TBVA-GAN leverages the coarse positional information of TB provided by Grad-CAM and incorporates an auxiliary model for TB presence determination, achieving favorable results in TB VA. However, TBVA-GAN introduces multiple constraint terms, making training convergence difficult and requiring extensive parameter tuning for effective testing.

6. CONCLUSION

This study presents a TB VA generative adversarial network guided by Class Activation Mapping (TBVA-GAN). The performance of TBVA-GAN was tested on TBX11K dataset, surpassing current TB VA models with an IoU score of 38.3% and a Dice score of 55.2%.

Compared to traditional CAM-based models, this model is capable of generating fine-grained changemaps that reflect the precise locations of TB, enabling more accurate TB localization. In contrast to existing GAN-based models, this model leverages CAM to guide the conversion process from CXRs with TB to normal CXRs, by assigning higher weights to the region indicated by CAM. However, due to the lower resolution of CAM, the provided lesion localization information may exhibit misalignment and introduce potential errors. In the future, it could be worthwhile to explore the incorporation of multiple CAM resolutions in combination to enhance the accuracy of lesion localization by considering various scales. Furthermore, this research introduces an automatic mask extraction method from the changemap, which has been extensively validated for its effectiveness. In addition, TBVA-GAN leverages an auxiliary TB detection model to ensure that the transformed normal images are free of TB-related information, resulting in a more precise changemap. However, this reliance on the accuracy of the auxiliary TB detection model, especially when introducing new data, highlights the need to enhance the model's generalization capabilities to maintain favorable outcomes.

7. ACKNOWLEDGMENT

This work was funded by the Ministry of Higher Education under Fundamental Research Grant Scheme(FRGS/1/2020/ICT02/UPM/02/5). Thank you to Ms. Syeda Shaizadi Meraj for her contribution to this project.

REFERENCES:

[1] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, “Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer”, *Medical Image Analysis* , Vol. 68, 2021, pp. 101915

[2] A. Rey, B. Arcay, and A. Castro, “A hybrid CAD system for lung nodule detection using CT studies based in soft computing”, *Expert Systems with Applications*, Vol. 168, 2021, pp. 114259.

[3] Z.I. Attia, P.A. Noseworthy, F., Lopez-Jimenez, S.J. Asirvatham, A.J. Deshmukh, B.J. Gersh, ... & P.A. Friedman, “An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction”, *The Lancet*, Vol. 394, 2019, pp. 861–867.

[4] S. Kazemzadeh, J. Yu, S. Jamsh, “Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists”. *Radiology*, Vol. 306, 2023, pp. 124-137.

[5] V. Acharya , G. Dhiman, K. Prakasha, “AI-assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model”. *Computational Intelligence and Neuroscience*, 2022.

[6] A.J. London, “Artificial intelligence and black-box medical decisions: accuracy versus explainability”, *Hastings Center Report*, Vol. 49, 2019, pp. 15–21.

[7] L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan, and J. Paisley, “An adversarial learning approach to medical image synthesis for lesion detection”, *IEEE journal of*

biomedical and health informatics, Vol. 24, 2020, pp. 2303–2314.

[8] T. Zia, S. Murtaza, N. Bashir, D. Windridge, and Z. Nisar, VANT-GAN: adversarial learning for discrepancy-based visual attribution in medical imaging, *Pattern Recognition Letters*, Vol. 156, 2022, pp. 112–118.

[9] M. Nawaz, F. Al-Obeidat, A. Tubaishat, T. Zia, F. Maqbool, and A. Rocha, "MDVA-GAN: multi-domain visual attribution generative adversarial networks", *Neural Computing and Applications*, 2022, pp. 1–16.

[10] C. Yan, J. Yao, R. Li, Z. Xu, and J. Huang, "Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays", *in Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 103–110.

[11] Z. Tan, H. Madzin, and Z. Ding, "Semi-supervised Semantic Segmentation Methods for UW-OCTA Diabetic Retinopathy Grade Assessment", *in Mitosis Domain Generalization and Diabetic Retinopathy Analysis: MICCAI Challenges MIDOG 2022 and DRAC 2022, Held in Conjunction with MICCAI 2022*, Singapore, 2022, pp. 97–117.

[12] H. Qu, P. Wu, Q. Huang, J. Yi, Z. Yan, K. Li, G. M. Riedlinger, S. De, S. Zhang, and D. N. Metaxas, "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images", *IEEE transactions on medical imaging*, Vol. 39, 2020, pp.3655–3666.

[13] A. Chamanzar and Y. Nie, "Weakly supervised multi-task learning for cell detection and segmentation", *in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp 513–516.

[14] X. Ouyang, S. Karanam, Z. Wu, T. Chen, J. Huo, X. S. Zhou, Q. Wang, and J.-Z. Cheng, "Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis", *IEEE transactions on medical imaging*, Vol. 40, 2020, pp. 2698–2710.

[15] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization", *in Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[16] E.S.N. Joshua, M. Chakkravarthy, and D. Bhattacharyya, "Lung cancer detection using improvised grad-cam++ with 3d cnn class activation", *in Smart Technologies in Data Science and Communication: Proceedings of SMART-DSC* 2021, pp 55–69.

[17] H. Jiang, J. Xu, R. Shi, K. Yang, D. Zhang, M. Gao, H. Ma, and W. Qian, "A multi-label deep learning model with interpretable grad-CAM for diabetic retinopathy classification", *in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1560–1563.

[18] H. Moujahid, B. Cherradi, M. Al-Sarem, L. Bahatti, A.B.A.M.Y. Eljialy, A. Alsaeedi, and F. Saeed, "Combining CNN and Grad-Cam for COVID-19 Disease Prediction and Visual Explanation", *Intelligent Automation & Soft Computing*, Vol. 32, 2022.

[19] Z.S. Ameen, A.S. Mubarak, C. Altrjman, S. Alturjman, and R. Abdulkadir, "Explainable Residual Network for Tuberculosis Classification in the IoT Era", *2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT)*, 2021, pp. 9–12.

[20] A. Chattopadhay, A. Sarkar, P. Howlader, and V.N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks", *in 2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, pp. 839–847.

[21] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks", *in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 24–25.

[22] P.T. Jiang, C.B. Zhang, Q. Hou, M.M. Cheng, and Y. Wei, "Layercam: Exploring hier- archical class activation maps for localization", *IEEE Transactions on Image Processing*, Vol. 30, 2021, pp. 5875–5888.

[23] Hamza A, Attique Khan M, Wang S H, et al, "COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization". *Frontiers in Public Health*, 2022, Vol. 10, pp. 1046296.

[24] Bhandari M, Shahi T B, Siku B, et al. "Explanatory classification of CXR images into COVID-19", *Pneumonia and*

Tuberculosis using deep learning and XAI. Computers in Biology and Medicine, 2022, Vol. 150, pp. 106156.

[25] Sharma V, Gupta S K, Shukla K K, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images". *Intelligent Medicine*, 2023.

[26] C.F. Baumgartner, L.M. Koch, K.C. Tezcan, J.X. Ang, and E. Konukoglu, "Visual feature attribution using wasserstein gans", *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8309–8319.

[27] Y. Tang, Y. Tang, V. Sandfort, J. Xiao, and R.M. Summers, "Tuna-net: Task-oriented unsupervised adversarial network for disease recognition in cross-domain chest x-rays", *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference*, 2019, pp. 431–440.

[28] Zia T, Wahab A, Windridge D, et al, "Visual attribution using Adversarial Latent Transformations". *Computers in Biology and Medicine*, 2023, Vol. 166, pp. 107521.

[29] J.Y. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", *in Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Computer Science*, 2014.

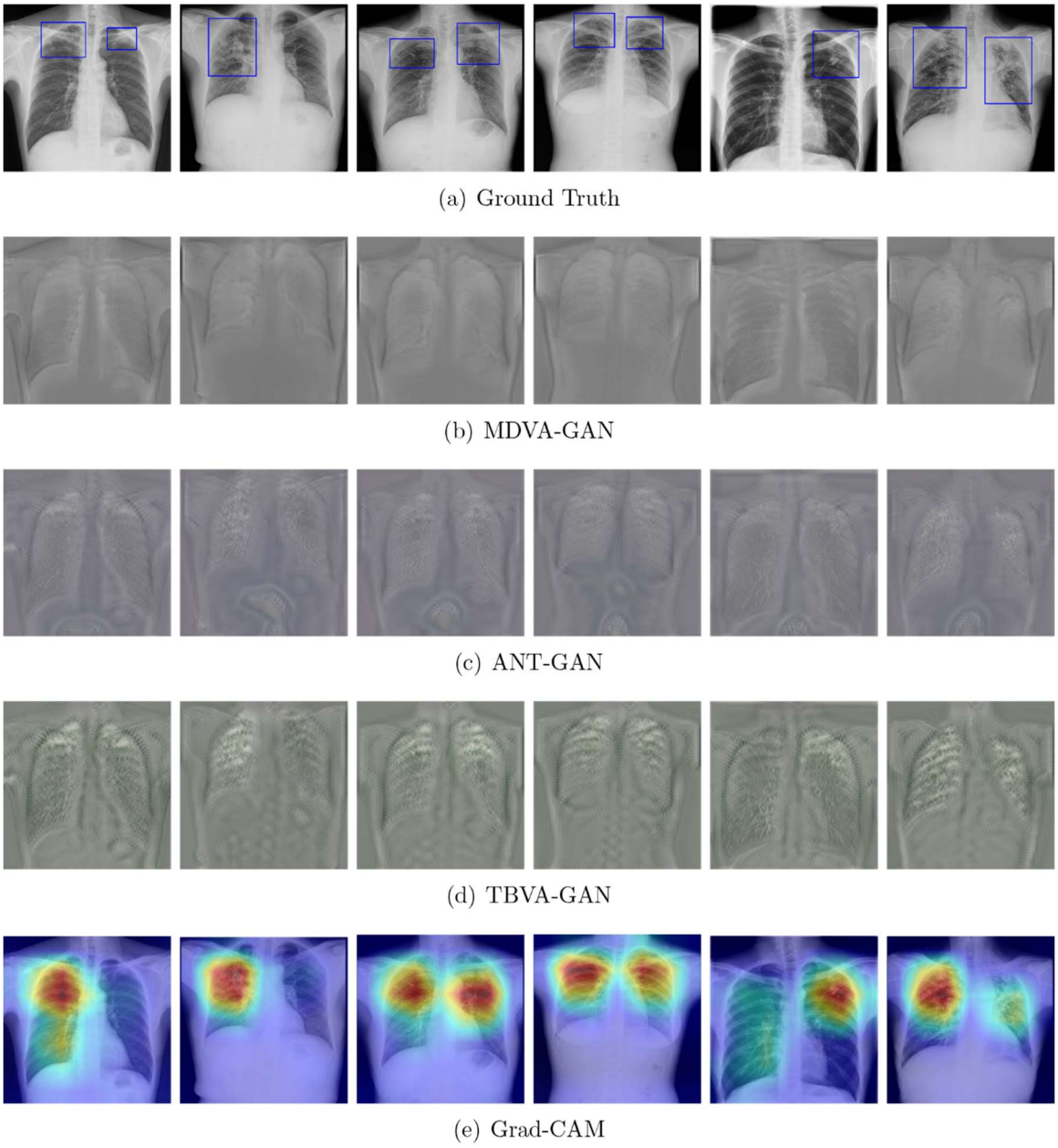

Figure 3: Comparison of The Changemaps Generated By Each Model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

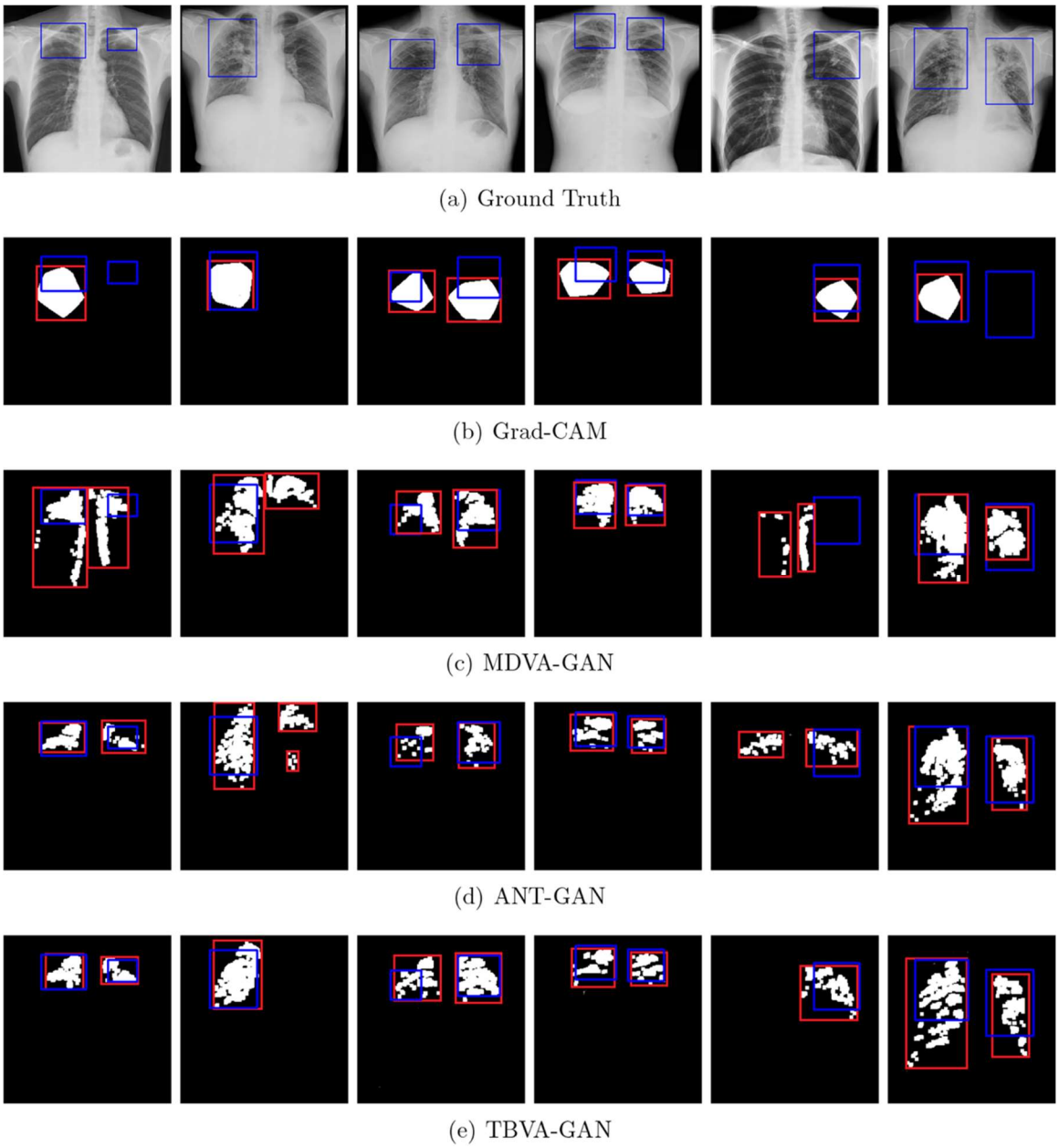

Figure 4: Comparison of The Binary Masks of Each Model