IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 March 2024, accepted 15 March 2024, date of publication 20 March 2024, date of current version 28 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3379886

RESEARCH ARTICLE

Multihead Neural Network for Multiple Segmented Images-Based Diagnosis of Thyroid-Associated Orbitopathy Activity

SANGHYUCK LEE[1], JEONG KYU LEE[2], AND JAESUNG LEE[1]
[1]Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, Republic of Korea
[2]Department of Ophthalmology, Chung-Ang University College of Medicine, Chung-Ang University Hospital, Seoul 06973, Republic of Korea

Corresponding authors: Jeong Kyu Lee (lk1246@cau.ac.kr) and Jaesung Lee (curseor@cau.ac.kr)

This work was supported in part by the Chung-Ang University Research Scholarship Grants in 2023, in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) under Grant NRF-2021R1A2C1011351, and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by MSIT [Artificial Intelligence Graduate School Program (Chung-Ang University)] under Grant 2021-0-01341.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Thyroid-associated orbitopathy is an autoimmune disease that causes changes in various structures close to the eye. Medical images, such as three-dimensional computed tomography scans, can be used by medical experts to diagnose thyroid-associated orbitopathy. Meanwhile, image segmentation has been widely used in medical imaging owing to its significant impact on improving model performance by filtering out unnecessary pixel values. In this study, a neural network specialized in processing multiple segmented images was proposed to evaluate thyroid orbitopathy activity, focusing on the fact that multiple segmented images can be extracted from orbital computed tomography scans. The proposed neural network consists of multiple convolutional embedding heads, a group squeeze-and-excitation block, and a classifier stage. Our empirical study shows that the proposed model outperforms four baseline models on a thyroid-associated orbitopathy activity dataset obtained from a cohort of 1,068 patients at Chung-Ang University Hospital between January 2008 and October 2019. The proposed model achieved an average area under the receiver operating characteristic curve of 0.800, accuracy of 0.721, F1 score of 0.416, sensitivity of 0.728, and specificity of 0.720 across 50 replicate experiments. The source code for the proposed model is available at https://github.com/tkdgur658/MultiheadGroupSENet.

INDEX TERMS Thyroid-associated orbitopathy, computed tomography, multihead neural network.

I. INTRODUCTION

Thyroid-associated orbitopathy (TAO) is an autoimmune disorder characterized by changes in several structures near the eyes [1]. Specifically, TAO is characterized by orbital inflammation, adipose tissue expansion, and an upregulated synthesis of hydrophilic glycosaminoglycans [2]. TAO activity can be evaluated using the clinical activity score, which reflects symptoms including redness of the conjunctiva and swelling of the eyelid [3]. However, TAO diagnosis based on computed tomography (CT) or magnetic resonance imaging often requires considerable time in practice because clinical decisions must be made after observing changes in various TAO-related structures [4]. Consequently, neural network (NN)-based TAO diagnostic systems have been widely considered [5].

Meanwhile, image segmentation is an effective image preprocessing method in machine learning-based diagnosis because it enables subsequent learning algorithms, such as NN, to avoid learning unessential information [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

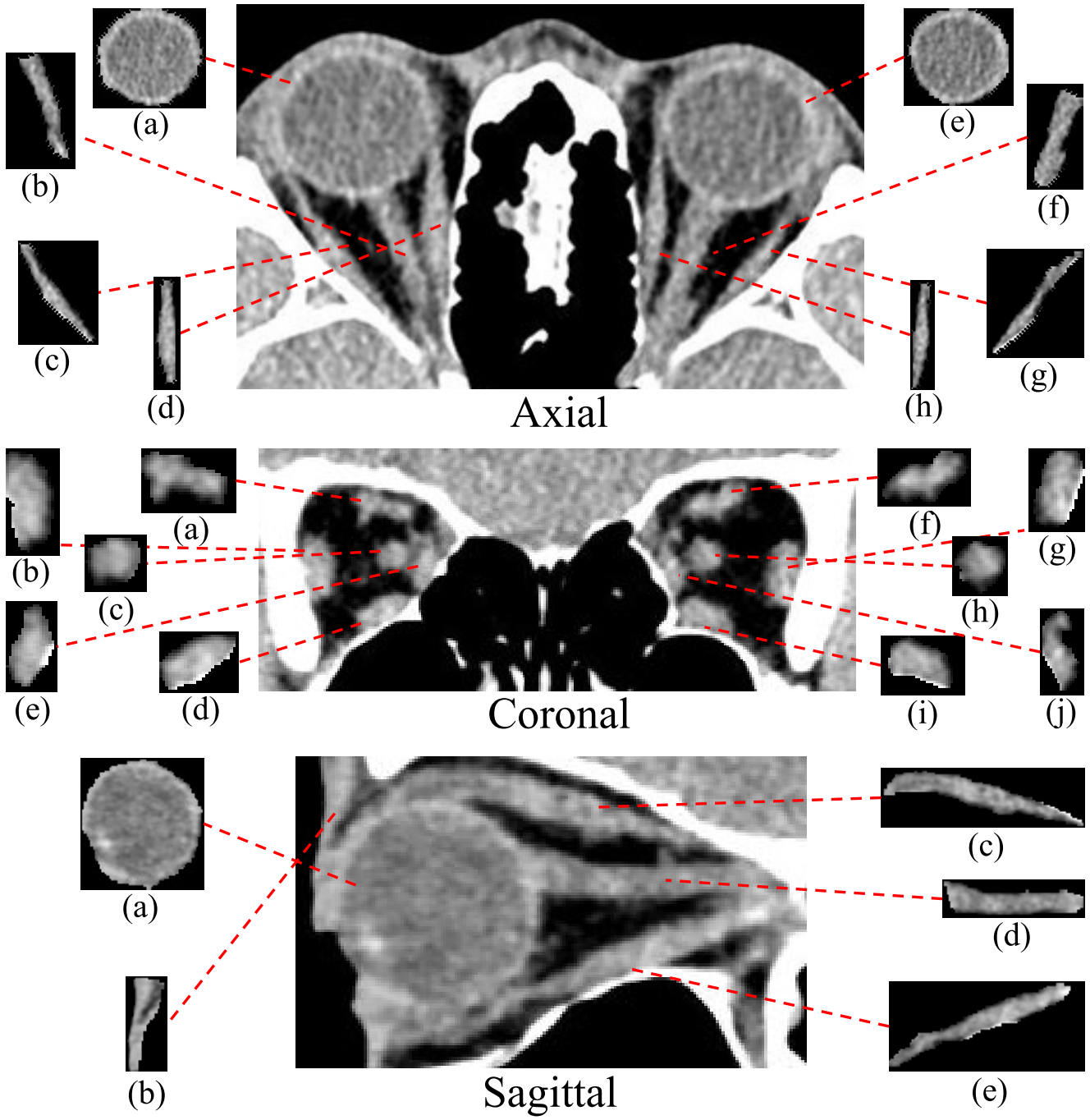


FIGURE 1. Different segmentation images from an orbital CT image. In the axial view, (a) right eyeball, (b) right optic nerve, (c) right lateral rectus muscle (LR), (d) right medial rectus muscle (MR), (e) left eyeball, (f) left optic nerve, (g) left LR, and (h) left MR are depicted. In the coronal view, (a) right superior rectus muscle (SR), (b) right LR, (c) right optic nerve, (d) right inferior rectus muscle (IR), (e) right MR, (f) left SR, (g) left LR, (h) left optic nerve, (i) left IR, and (j) left LR are shown. In the sagittal view, (a) eyeball, (b) upper eyelid, (c) SR, (d) optic nerve, and (e) IR are presented.

Specifically, image segmentation distinguishes the targeted structure from other elements within a medical image [7]. Therefore, the segmentation process in medical imagery enables NNs to concentrate on particular objects pertinent to diagnosis, assisting the predictive model in disregarding features of ancillary structures irrelevant to the diagnostic process [8]. In the case of orbital CT images, a large number of segmented images can be generated according to the number of chosen structures around the eye [9]. Moreover, the number of CT slices of a patient and the view direction, such as axial, coronal, and sagittal views, can increase the total number of segmented images [10]. Figure 1 shows the different segmented images extracted from the orbital CT image of a patient; details of the

segmented images used in this study are described in the CT preprocessing of the proposed method section. Since TAO patients can be characterized by changes in multiple structures near the eye, training an NN based on these different segmented images could be a promising approach in terms of predictive performance [11]. However, most conventional studies have adopted the strategy of simply cropping the orbital area where the TAO-related structures are gathered, which limits their diagnostic performance improvements [12], [13], [14], [15], [16].

Our previous work [17] showed promising performance for TAO activity prediction based on a large number of segmented images. However, the predictive model adopted an intermediate fusion strategy, resulting in low predictive performance due to unnecessary information interaction between different segmented images.

This study aims to devise a novel NN that assesses the activity of TAO patients effectively and to analyze important design aspects for processing multiple segmented CT images. The proposed NN consists of multiple convolutional embedding heads, a group squeeze-and-excitation (SE) block, and a classifier stage. Each convolutional head extracts a segmented image embedding. Then, the extracted embeddings from different convolutional heads are assigned importance by the group SE block. Finally, these importance-assigned embeddings are used for TAO activity classification in the classifier stage.

The main contributions of our TAO diagnosis study are as follows:

1) This study introduces a multihead NN for diagnosing TAO activity based on multiple segmented images, demonstrating superior performance over four existing models across five evaluation metrics in experiments involving a cohort of 1,068 patients at Chung-Ang University Hospital.
2) A multiple segmented-images input strategy, a less-explored approach in the field of NN-based TAO diagnosis, was extended by our study. Moreover, the proposed method was compared with well-studied crop image input strategies, including their advantages, disadvantages, and experimental results.
3) Future directions for NN-based TAO activity diagnosis are presented through multiple in-depth analyses, including a fusion strategy comparison, group SE ablation, and attention score sorting, which suggest designing a predictive network based on a late-fusion strategy and improving the less-important image-filtering approach.

The remainder of this paper is organized as follows: Section II describes our proposed NN. Section III describes the experimental setup and results. Section IV presents the analysis of the training and inference results of the proposed NN. Finally, Section V concludes the paper and discusses future research directions.

II. RELATED WORK

In recent years, NN-based diagnostic methods have been developed using various foundational architectures. Among them, CNNs are the most popular because of their exceptional capability for local feature extraction [18]. For example, ConvNeXt, which was initially developed for natural images, has been extended to predict the severity of lung damage [19]. In addition, three traditional CNNs were utilized to diagnose acute lymphoblastic leukemia, with their extracted spatial features fed into either an extreme gradient boosting classifier [20]. The versatility of CNNs is extended to various diagnostic objectives, including breast cancer diagnosis [21], facial skin condition identification [22], and brain tumor detection [23]. The scalability of CNN is promising with the combination of various machine learning methods [24].

The successes of CNNs have similarly manifested in studies related to TAO [12], [13], [14], [15], [16], [17]. For TAO activity diagnosis, MRI images of 108 patients were included to train and validate the VGG and ResNet variants [12]. Orbital parts in MRI images were intercepted with fixed size, and randomly cropped again to augment training data. MRI provides better contrast resolution than CT; however, its use may be limited owing to the high cost and the limited number of available samples. In the same year, a 3D-ResNet variant was proposed to differentiate between patients with TAO and normal subjects based on 1,435 CT scans [13]. The cropping process was performed using a rectangular boundary and a criterion that the output images should contain the orbital bone and eyeball on the sagittal axis. However, this cropping process still outputs a wide range of less-important regions for TAO diagnosis, resulting in low predictive performance. For assessment of severity of TAO, multiple convolutional blocks were used for learning information of axial, coronal, and sagittal planes [14]. The features extracted from three different views help the model to learn 3D information successfully; but still have difficulty to focus on individual anatomic structures. ResNet-VGG pipeline was used to detect enlarged extraocular muscles in CT images [15], which can be useful in the management of patients with TAO. Similar to previous studies, they use a rectangular cropped area, so unnecessary features contained within the boundary interfere with learning. Finally, two recent studies on the diagnosis of TAO activity based on CT images have been conducted [16], [17]. The activity diagnostic performance of a multichannel CNN can be improved by concurrently inputting orbital and single-photon emission CT images [16]. However, they still input different structures concurrently on a single cropped image; the interaction of different anatomic structures makes it difficult to focus on individual structures. An intermediate-level fusion model for differently segmented CT images was proposed to iteratively find the most important CT cuts among a large number of preselected CT cuts [17]. This approach can effectively extract anatomy structure-independent features based on large numbers of segmented

TABLE 1. Summary of the literature review of NN-based TAO diagnosis.

Works	Tasks	Techniques Preprocessing	Classification Model	Advantages	Limitations
[12]	Activity diagnosis	Intercepting the orbital part with fixed size in MRI images	ResNet variant and VGG variant	Low preprocessing costs based on a simple cropping process	Difficulty obtaining large number of samples due to high data preparation costs of MRI data
[13]	Differentiation between TAO patients and normals	Making Cropped CT images containing the orbital bone and eyeball on the sagittal axis	3D ResNet	Low preprocessing costs based on a simple cropping process	During training, suffered from disturbance of TAO-unrelated objects
[14]	Severity diagnosis	Cropping the CT images to include orbital area for different planes	Multiview CNN	Learning representation from three different views to enhance predictive performance	Limitation in focusing on individual structures due to including multiple structures together in single image
[15]	Evaluating enlarged extraocular muscles	Cropping CT images into the region of interest automatically using ResNet	VGG-16	Low preprocessing costs based on the preprocessing model	Limitation in focusing on individual structures due to including multiple structures together in single image
[16]	Activity diagnosis	Automatic segmentation based on V-Net for CT and SPECT	Multichannel 3D CNN	Performance improvements based on SPECT information	High data preparation costs in acquiring both CT and SPECT images
[17]	Activity diagnosis	Segmenting 78 orbital structure images	Early fusion Multihead CNN	Performance improvments based on extracting spatial features for preselected CT slices and anatomic structure	High data preprocessing costs from the segmentation process
This work	Activity diagnosis	Segmenting 78 orbital structure images	Late fusion Multihead CNN	Performance improvements based on extracting high-level spatial features for each anatomic structure on CT Slices	High data preprocessing costs from the segmentation process

images and the same number of different convolution blocks. However, the intermediate fusion can make classifying of the fully connected (FC) layer hard by not sufficiently extracting high-level features for each structure. Table 1 summarizes the literature review of NN-based TAO diagnosis.

Our brief review indicates that recent NN-based TAO studies have limitations in that the predictive models often suffer from interference of TAO-irrelevant objects during the training phase. One approach to tackle this issue is to segment each anatomic structure [17]. By filtering out TAO-irrelevant structures, the predictive model can focus on TAO-relevant structures during training. However, this approach has not been deeply studied yet. In this study, we proposed a new NN architecture for processing multiple segmented images to diagnose TAO activity more accurately.

III. PROPOSED METHOD

A. CT PREPROCESSING

The Institutional Review Board of Chung-Ang University Hospital approved this study (IRB No, 2312-003-19499), and the informed consent requirement was waived due to its retrospective design. This study was conducted in accordance with the ethical standards outlined in the Declaration of Helsinki. The CT scans of 1,068 TAO patients used in this study were obtained from Chung-Ang University Hospital between January 2008 and October 2019 [17]. Each patient was classified as active or inactive using a seven-point modified clinical activity score, with any scoring equal to or more than three considered active. The number of active and inactive are 144 (52 men, 92 women) and 924 (226 men, 698 women) TAO patients, respectively ($p<0.005$). The mean active TAO patient age was 46.14 ± 13.61 years, and the mean inactive TAO patient age was 35.43 ± 12.18 years ($p<0.001$). The demographic information for the patients is described in Table 2. Two ophthalmologists with more than five years of experience in oculoplasty and blinded to patient information evaluated clinical inflammatory activity by analyzing CT images. Then, the Hounsfield Unit windowing process for better structure identification was performed.

TABLE 2. Subject characteristics.

Characteristics	Active TAO	Inactive TAO	*p*-value
Number of subjects	144	924	
Age (years, mean ± standard deviation)	46.14 ± 13.61	35.43 ± 12.18	<0.001
Sex (male/female)	52/92	226/698	<0.005

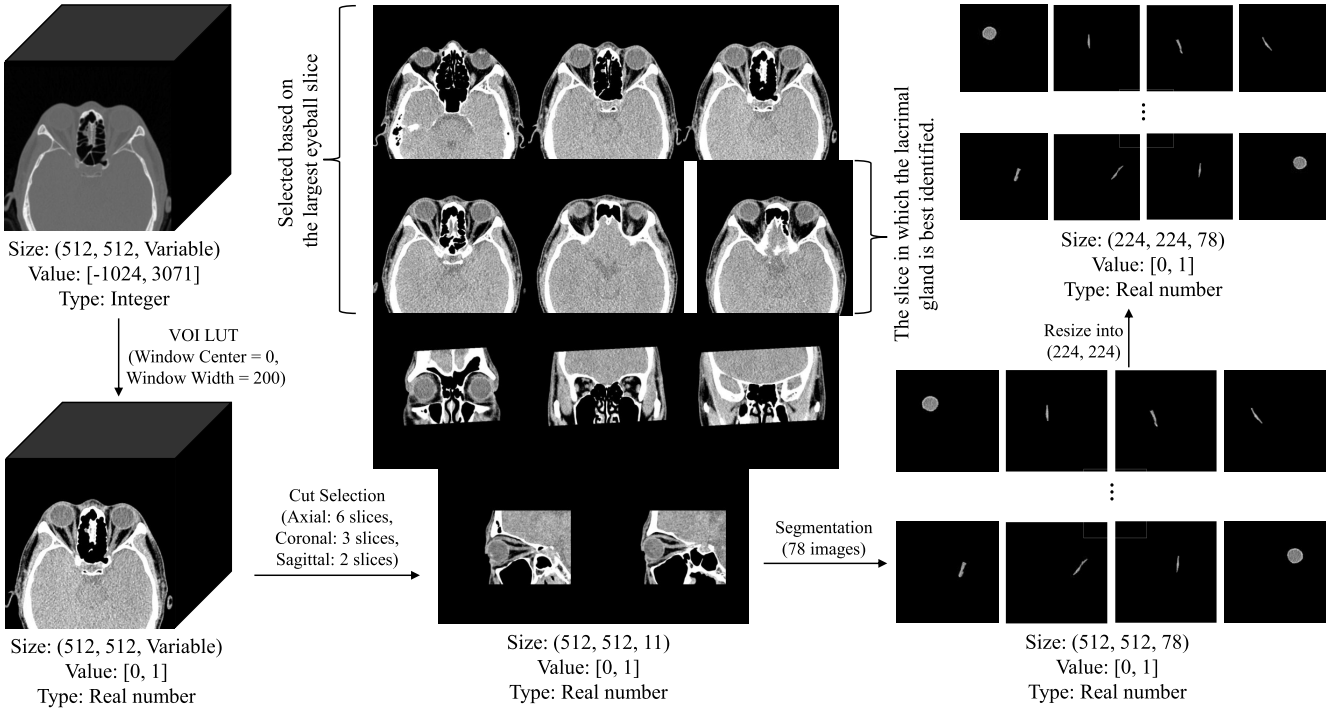


FIGURE 2. CT data preprocessing. First, our CT images were normalized to values from zero to one by VOILUT function. Then, six axial slices were selected based on the eyeball and lacrimal gland, three coronal slices were selected based on the eyeball and orbit, and two sagittal slices were selected based on the eyeball. Finally, 78 segmented images were extracted from the selected slices and resized into the size of (224, 224).

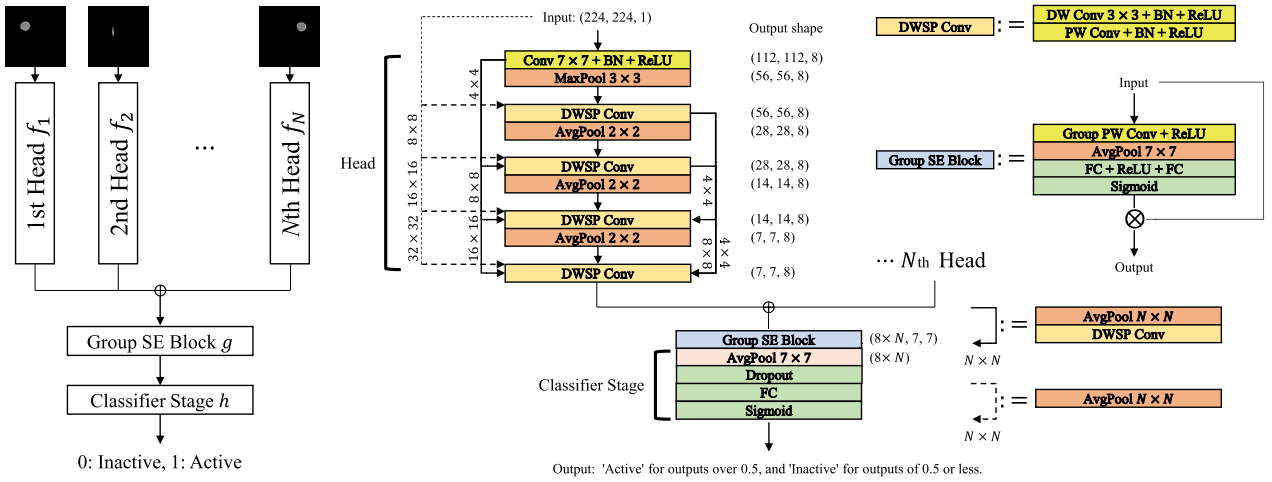
Value of Interest Look Up Table (VOI LUT) function in Pydicom library was used to convert the original CT pixel values into values ranging from 0 to 1. Window Center and the Window Width were set to 0 and 200, respectively. Then, 11 specific CT slices were chosen, encompassing axial, coronal, and sagittal planes for each patient. This selection process was designed to focus on the anatomical features most pertinent to TAO, aiming to improve the performance of TAO activity evaluation and avoid confusion due to unnecessary information. Specifically, the orbital CT of each patient had 80 to 400 image slices. Among the total slices for a patient, only a few CT slices were selected from axial, coronal, and sagittal planes. We selected the slice with the largest lens in the axial plane (AX1) first, then slices 3mm above (AX2) and below (AX3) and 7mm above (AX4) and below (AX5) AX1. For identification of the lacrimal gland, the slice with the largest lacrimal gland was selected (AX6). For the coronal plane, the slice with the largest eyeball was selected first (CO1). Then, we picked slices 1/2 and 2/3 of the distance between CO1 and the orbit exit (CO2, CO3). Finally, for the sagittal plane, slices with the largest eyeball in both eyes were chosen (SA-L, SA-R). Consequently, we obtained 11 CT slices for each patient: six axial, three coronal, and two sagittal plane slices. Then, identifiable structures, including the eyeball, the optic nerve, four rectus muscles, and the orbital fat, were segmented from 11 CT slices. Table 3. IR, SR, MR, and LR stand for inferior, superior, medial, and lateral rectus muscles, respectively, and OR means optic nerve. As a result, we acquired 78 segmentation images from 11 selected slices. The overall preprocessing procedure is demonstrated in Figure 2.

TABLE 3. Segmentation criteria. IR, SR, MR, and LR means inferior, superior, medial, and lateral rectus muscles, respectively. OR stands for optic nerve.

CT cut	Criteria
AX1	Orbit, Eyeball, MR, LR, ON, Orbital fat (left and right)
AX2	Orbit, Eyeball, MR, LR, ON, Orbital fat (left and right)
AX3	Orbit, Eyeball, MR, LR, ON, Orbital fat (left and right)
AX4	Orbit, Eyeball, SR, Orbital fat (left and right)
AX5	Orbit, Eyeball, IR, Orbital fat (left and right)
AX6	Lacrimal gland (left and right)
CO1	Eyeball, Orbital fat (left and right)
CO2	MR, LR, SR, IR, ON, Orbital fat (left and right)
CO3	Orbit (left and right)
SA-L	Eyeball, SR, IR, ON, Upper eyelid (left only)
SA-R	Eyeball, SR, IR, ON, Upper eyelid (right only)
Total	78 segmented images

B. PROPOSED MODEL

For the set of input segmented images $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ per patient, the proposed NN extracts the high-level spatial features using the set of convolutional heads $F = \{f_1, f_2, \ldots, f_N\}$, where N is the number of input segmented images. For ith segmented image $\boldsymbol{x}_i$, ith convolutional head



(a) Schematic Model Overview (b) Detailed Model Diagram

FIGURE 3. An overview of the proposed model. (a) and (b) represents a schematic overview of the model and a detailed diagram of the proposed model, respectively. DSWP Conv stands for depthwise separable convolution [25].

f_i extracts the corresponding high-level feature without interfering with features extracted from other segmented images. Thus, the output of convolutional heads can be defined as $\boldsymbol{X}' = \{\boldsymbol{x}'_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_N | \boldsymbol{x}'_i \text{ is } f_i(\boldsymbol{x}_i)\}$. Then, group SE block g recalibrates each feature by modeling the correlation between features extracted from different segmented images. Therefore, the output of group SE block can be defined as $\boldsymbol{X}'' = g(\boldsymbol{X}')$. Finally, the final output of the proposed model can be generated as $\hat{y} = h(\boldsymbol{X}'')$, where h is the classifier stage function and $\hat{y} \in [0, 1]$ is the active score. Figure 3 illustrates the proposed model, and the details of each function (i.e., f, g, and h) are explained in the following paragraphs.

Each convolutional head processes the corresponding segmented images to generate a small spatial embedding with the following steps. First, a stem convolutional layer processes a corresponding segmented image with 7×7 kernels. All convolutional layers are succeeded by batch normalization (BN) [26] and rectified linear unit (ReLU) activations [27]; these details have been excluded for brevity and clarity. Then, a 3×3 max pooling operation was performed with a stride of two and padding of one. After extracting the feature map using the stem layer, four convolution layers and three average pooling layers processed the feature map with multiple skip connections. Specifically, depthwise separable convolution (DWSP Conv), consisting of depthwise convolution, pointwise convolution [25], and 2×2 average pooling, was repeated three times, followed by one DWSP Conv to generate the corresponding segmented image embedding. DSWP Conv is one of the most widely used convolution types due to its stable performance. Skip connections connect all the convolutional layers in each convolutional head and perform DWSP Conv after averaging the pooling into the input spatial size of the corresponding DWSP Conv layer [28]. The input segmented image is also connected to all convolutional layers in the head via average pooling of different sizes to ensure the flow of information from the original segmented image to subsequent layers. Mathematically, the output of jth DWSP Conv layer for ith image can be defined as $\boldsymbol{z}_i^j = \mathbf{DWSP}_i^j([\boldsymbol{x}_i, \boldsymbol{z}_i^0, \ldots, \boldsymbol{z}_i^{j-1}])$, where $[\cdot]$ refers to the global average pooling operation that reduces the elements of any input size into the $1/2^2$ size of $\boldsymbol{z}_i^{j-1}$ and concatenation operation, $\boldsymbol{z}_i^0$ is the stem layer output of ith image, and $\mathbf{DWSP}_i^j$ is the jth DSWP Conv for ith image. Thus, the output of ith head can be represented as $\boldsymbol{x}'_i = \boldsymbol{z}_i^4 = \mathbf{DWSP}_i^4([\boldsymbol{x}_i, \boldsymbol{z}_i^0, \boldsymbol{z}_i^1, \boldsymbol{z}_i^2, \boldsymbol{z}_i^3])$. Then, the proposed NN recalibrates N embeddings using an attention block named group SE block. The group SE block, inspired by the SE block [29], assigns explicit importance to each embedding, thereby emphasizing the more important segmented images while suppressing the less important ones. After passing through the convolutional heads, N embeddings $\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N$ are processed by the group SE block g, which consists of group squeeze and excitation stages. The squeeze stage comprises a group pointwise convolution layer and a global average pooling layer. In the group pointwise convolution layer, N pointwise convolutions are applied between the corresponding channels generated from the same embeddings, which aggregate the information for the channel axis of each embedding. The spatial information of each embedding is then squeezed through global average pooling. In the excitation stage, a sequence consisting of FC-ReLU-FC learns the correlation relationship between different squeezed embeddings. Finally, the sigmoid function normalizes the N embeddings to the N scores $s_1, .., s_N$ between zero and one. The scores are multiplied by each head output $\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N$. Thus, the output of group SE block can be represented as $\boldsymbol{X}'' = g(\boldsymbol{X}') = \{\boldsymbol{x}''_1, \boldsymbol{x}''_2, \ldots, \boldsymbol{x}''_N | \boldsymbol{x}''_i \text{ is } s_i \boldsymbol{x}'_i\}$. In the classifier stage, the embeddings recalibrated using group SE block $\boldsymbol{x}''_1, \boldsymbol{x}''_2, \ldots, \boldsymbol{x}''_N$ are spatially aggregated by 7×7 average pooling. Then, the dropout layer, with a probability of 0.5,

removes several nodes randomly to avoid co-adaptation of the network during training. Finally, the FC layer classifies the activity of the patient using a sigmoid function. Thus, final output of the model can be represented as $\hat{y} = h(\boldsymbol{X}'') = \textbf{Sigmoid} \circ \textbf{FC} \circ \textbf{Dropout} \circ \textbf{AvgPool}(\boldsymbol{X}'')$, where **Sigmoid** is a sigmoid function, **FC** is a FC layer, and **Dropout** is a dropout layer, and **AvgPool** is a 7×7 average pooling layer.

IV. EXPERIMENTS

Four comparison models were used to evaluate the predictive performance of the proposed model [13], [16], [30], [31]. The models by Song et al. [13] and Yao et al. [16] were used as representative models for NN-based TAO diagnosis based on CT images. For these two models, three-dimensional input images were used because they are developed as three-dimensional convolution-based models. Woo et al. [31] and Zhang et al. [30] were used as the representative model for general image processing. These two models are widely used in various computer vision tasks and can be easily adapted for multiple two-dimensional image processing by modifying the number of input channels.

All the experiments were implemented using Python 3.6 and the PyTorch library 1.10. The training and testing were conducted using two NVIDIA GeForce RTX 3090 GPUs in a data-parallel environment. The weights of each model were optimized using focal loss [32] with a gamma value of 2.0. AdamW [33] was used as an optimizer. The weight decay was set to 1e-4, and the learning rate was set to 1e-3. Each model was trained for a maximum of 30 epochs. An early stopping strategy was employed to terminate the training if there was no improvement in the loss value over three epochs. The batch size was set to 32. In each experiment, the data were stratified and randomly sampled into three sets: 60% for training, 20% for validation, and 20% for testing. Each model was trained and tested 50 times.

The activity diagnosis can be considered as a binary classification dealing with active and inactive classes. Thus, the following five metrics, which are widely used in binary classification problems, were used to evaluate the model performance.

Accuracy (ACC): ACC quantifies the proportion of correct predictions made by the model over all predictions made. In activity diagnosis, ACC can evaluate how many correct diagnoses are among total diagnoses. In binary classification, ACC can be mathematically defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

F1 Score (F1): F1 is the harmonic mean of precision and recall, and provides a balanced measure between these two characteristics. F1 is particularly effective in imbalanced datasets because it assigns equal weights to both false positives and false negatives. F1 can be mathematically represented as:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \tag{2}$$

where Precision = TP / (TP+FP) and Recall = TP / (TP+FN).

Sensitivity (SEN): SEN, also known as recall or true positive rate, measures the proportion of actual positives that are identified correctly. SEN evaluates how many positive cases the model detects. SEN is defined as follows:

$$SEN = \frac{TP}{TP + FN}, \tag{3}$$

Specificity (SPE): Also known as the true negative rate, measures the proportion of correctly identified actual negatives. SPE evaluates how many negative cases the model detects. SPE is defined as:

$$SPE = \frac{TN}{TN + FP}, \tag{4}$$

Area under the receiver operating characteristic curve (AUC): The receiver operating characteristic (ROC) curve plots SEN against (1 - SPE) for different threshold values. AUC measures the entire two-dimensional area underneath the ROC curve from (0,0) to (1,1). Thus, AUC evaluates the predictive performance independent of the threshold.

To evaluate the experimental results statistically, a paired t-test was performed with both 0.05 and 0.01 significance levels. The null hypothesis states that the mean difference between paired observations is zero. All the possible comparison model pairs were tested. All statistical tests were performed using SciPy 1.5.4, an open Python library.

Table 4 lists the experimental results. The proposed model outperformed all comparison models in terms of the average of all evaluation metrics. Moreover, statistical tests for the AUC, ACC, F1, and SEN rejected the null hypothesis, indicating the superiority of the proposed model. Specifically, in terms of AUC, the proposed model achieved an average value of 0.800, surpassing that of the nearest competitor by more than 0.050. In terms of ACC, the proposed model exhibited an average score of 0.721 with a standard deviation of 0.045, which is superior to that of the second-best model by approximately 0.024. Similarly, the proposed model led in terms of F1, achieving an average of 0.416, with a standard deviation of 0.046. This was approximately 0.037 higher than that of the second-best model. The proposed model also exhibited the highest average SEN of 0.728, which was significantly higher than those of the other models. Finally, the SPE of the proposed model was 0.720. Although the score does not exceed that of the second-best model by a large margin, it suggests that our model is superior in terms of identifying true negatives.

V. DISCUSSIONS AND FUTURE SCOPE

This study differs from previous NN-based TAO diagnostic studies [12], [13], [14], [15], [16], [17]. Previous studies predominantly used region-of-interest (RoI) cropping datasets as

TABLE 4. Experimental results. The mean of the corresponding evaluation metric is represented in each cell with its standard deviation. The value in parentheses indicates the average ranking of the corresponding model. If the best scoring model for each metric rejects the null hypothesis in a *t*-test with all other comparison models, the asterisk (*) emphasizes the performance. ** and * indicate statistical significance at the 0.01 and 0.05 levels, respectively.

Model	AUC	ACC	F1	SEN	SPE
Proposed	**0.800 ± 0.047 (1.20)****	**0.721 ± 0.045 (1.93)***	**0.416 ± 0.046 (1.34)****	**0.728 ± 0.087 (2.09)****	**0.720 ± 0.055 (2.16)**
Song et al. (2021)	0.742 ± 0.060 (2.24)	0.697 ± 0.057 (2.38)	0.379 ± 0.052 (2.03)	0.670 ± 0.089 (2.66)	0.702 ± 0.068 (2.48)
Yao et al. (2023)	0.646 ± 0.091 (3.36)	0.583 ± 0.180 (3.43)	0.280 ± 0.091 (3.39)	0.612 ± 0.239 (3.30)	0.579 ± 0.239 (3.32)
Zhang et al. (2022)	0.640 ± 0.100 (3.48)	0.598 ± 0.095 (3.70)	0.286 ± 0.080 (3.55)	0.596 ± 0.167 (3.46)	0.598 ± 0.116 (3.64)
Woo et al. (2023)	0.494 ± 0.080 (4.72)	0.527 ± 0.240 (3.56)	0.181 ± 0.083 (4.69)	0.472 ± 0.334 (3.49)	0.535 ± 0.327 (3.40)

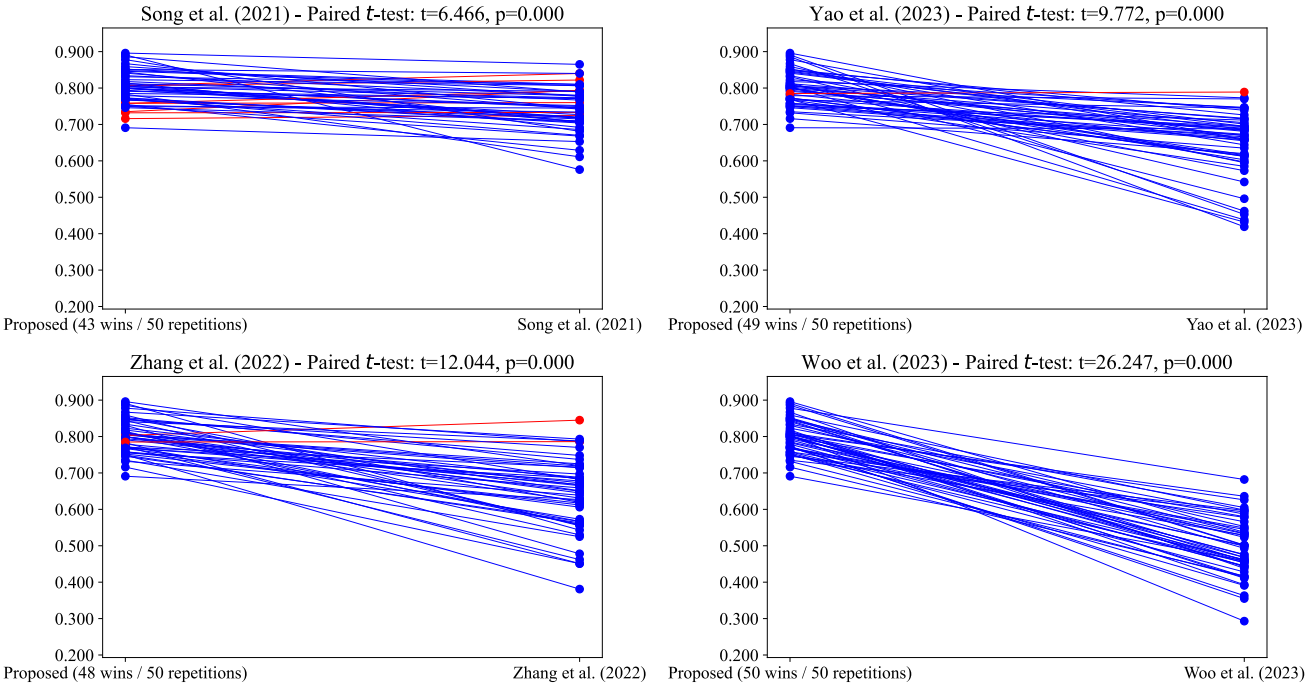


FIGURE 4. *t*-test dot plot of the main experiments in terms of AUC. Blue and red lines mean the proposed model wins and losses, respectively. The proposed model won 190 times among 200 experiments, rejecting four null hypotheses at a significance level of 0.01.

the major input type in NN-based TAO diagnosis [12], [13], [14], [15], [16]. RoI cropping typically involves selecting a rectangular area encompassing relevant medical image structures. Thus, previous works usually crop to cover orbit or a larger area. However, predictive model can suffer from loss of the information of more detailed-level structures by rough cropping and convolution in early steps. In contrast, this study used a large number of segmented images as inputs to the NN, similar to [17] rather than the RoI cropping dataset. Our strategy can extract high-level features for each structure within the orbit through the corresponding convolution head. Then, structure-level importance can be assigned by the Group SE block. Assigning importance to the features makes the classification performed by the last FC layer easier. As shown in Figures 4 and 5, the proposed method consistently outperforms Song et al. and Yao et al., which are the cropping-based methods. In Figure 4, the blue and red lines represent the performance difference between the proposed and the compared models. The blue line means that the proposed model shows better performance in that iteration. The red line means that the comparative model performs better than the proposed model at that iteration. As shown in Figure 4, the proposed model wins 190 times out of 200 experiments. The t-test rejects the null hypotheses all four times. In addition, Figure 5 shows the performance difference between the proposed model and each comparable model. The x-axis represents the performance difference of each iteration, and the y-axis represents the number of experiments. Compared to Song et al., the proposed method won in most iterations, achieving 0.058 mean difference. Compared to Woo et al., the proposed model won by a margin of more than 0.1 in all iterations.

In our previous work [17], a large number of segmented images were used, similar to this study. The main difference is that a late-fusion NN architecture was proposed in this study to aggregate multiple segmented image information at a higher level, in contrast to the intermediate-fusion approach of [17]. The proposed model compresses each segmented image to (7, 7, 8) dimensions and then performs information fusion. This strategy improves the extraction of unique fea-

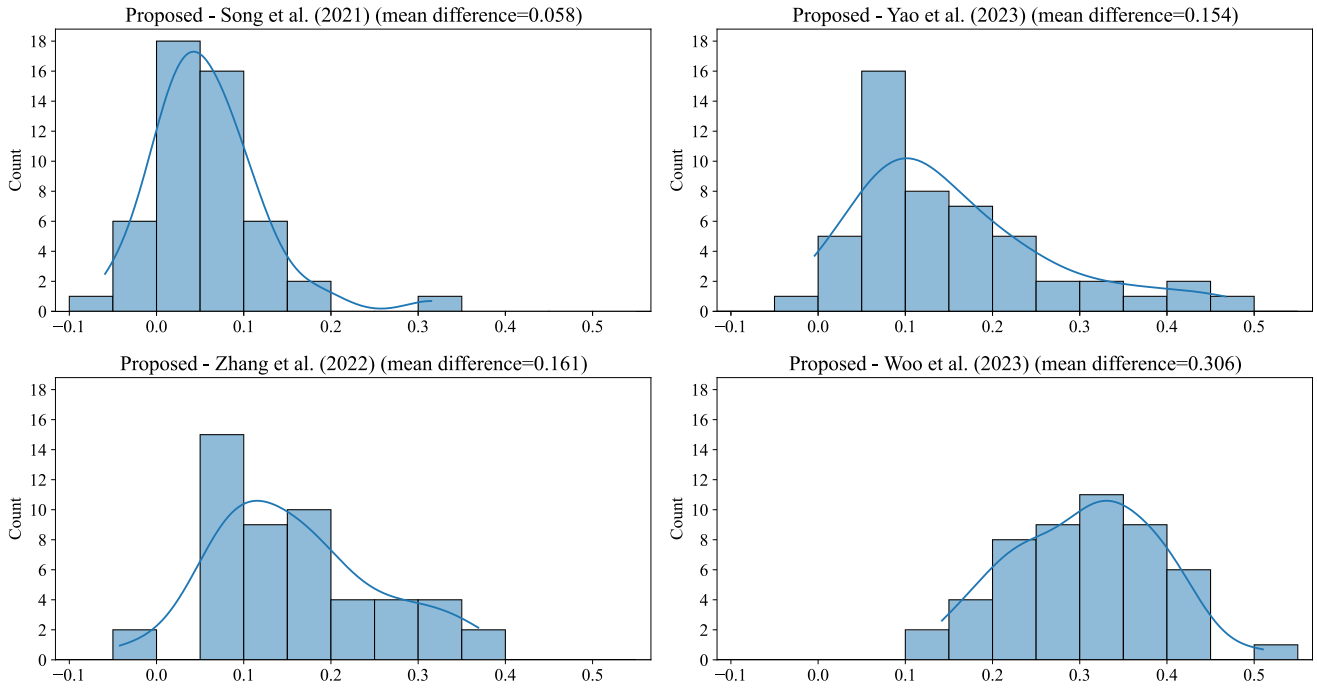


FIGURE 5. *t*-test histogram plot of the main experiments in terms of AUC. The x-axis represents the difference between the AUC of the proposed model and the comparison model for each experiment, and the y-axis shows the number of experiments. The curves were generated based on the data using kernel density estimation to smooth the distributions of the experimental results.

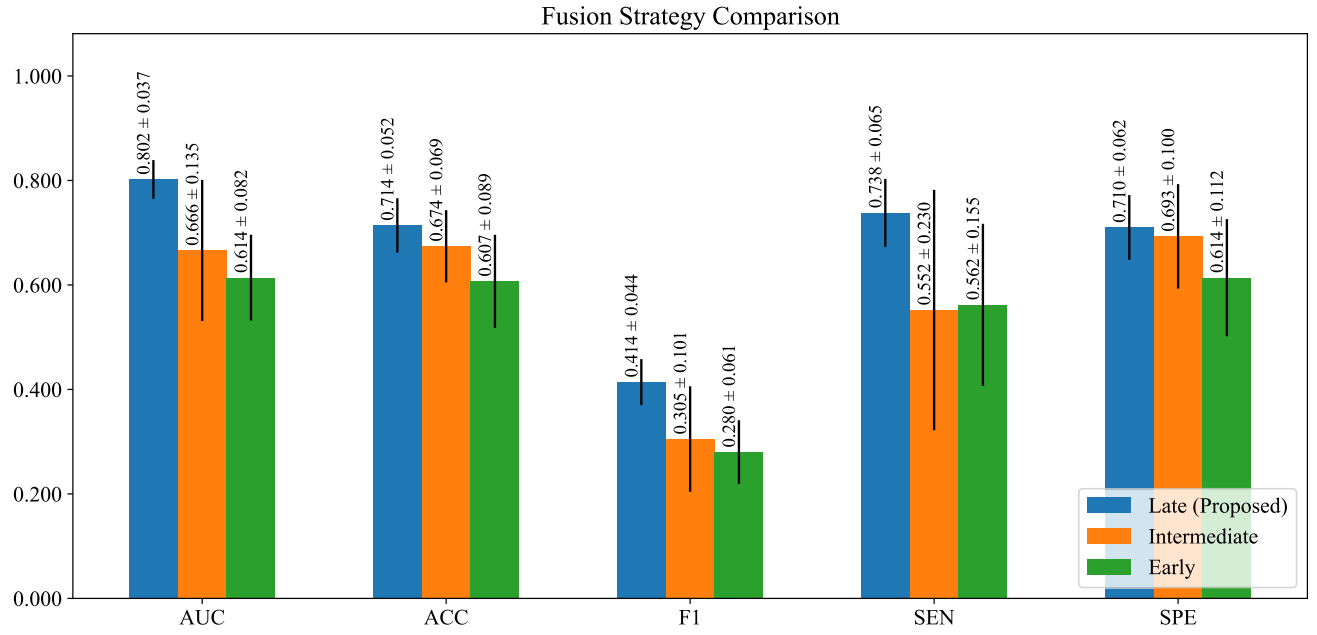


FIGURE 6. Bar plot for fusion strategy comparison. Early-fusion [30], intermediate-fusion [17], and late-fusion (proposed) were compared in terms of AUC, ACC, F1, SEN, and SPE. The late-fusion approach achieved the best performance in all metrics. In particular, the late-fusion model outperforms the second-best model by approximately 0.088 in AUC and 0.186 in SEN, respectively.

tures of each TAO-related structure. Figure 6 shows that late fusion is more promising than early or intermediate fusion. For all five metrics, late fusion consistently outperformed the other two methods. Specifically, the potential for late fusion is evident in SEN, which outperforms the other methods by approximately 0.180. In terms of the experimental settings in Figure 6, Zhang et al. [30] used an early fusion model, and Lee et al. [17] used an intermediate fusion model. The experimental results for Figure 6 were calculated from ten repetitions of the same settings as in the main experiments.

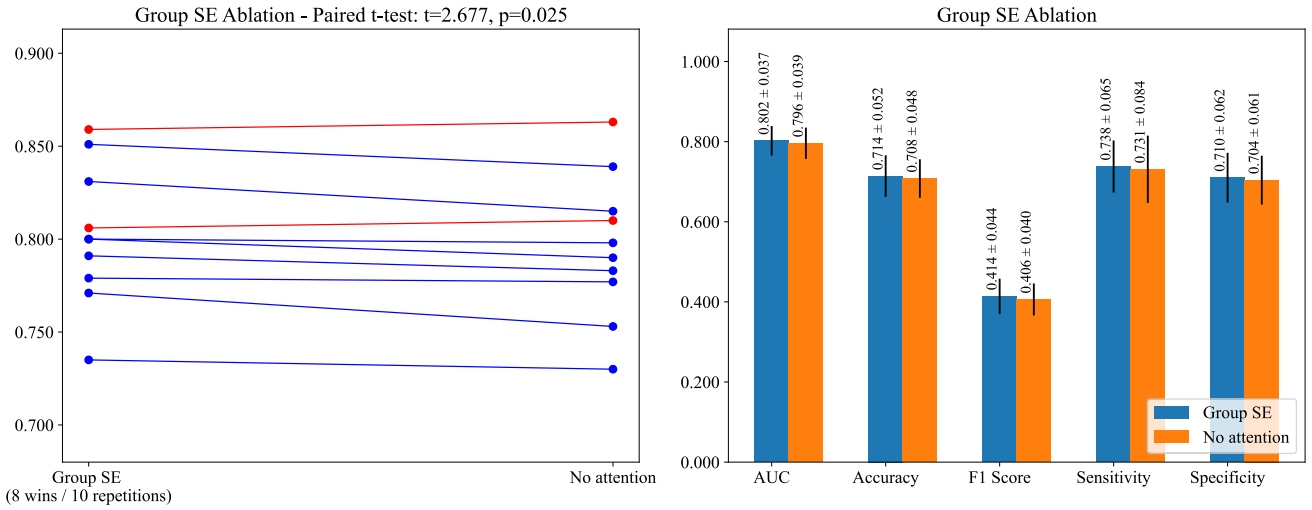


FIGURE 7. Bar plot and dot plot for group SE ablation. A comparative experiment was conducted by removing the group SE block from the proposed model. The bar plot shows that the overall predictive performances were improved by adding the group SE block in terms of the five evaluation metrics. In the dot plot, blue and red lines mean the group SE block wins and loses in terms of AUC, respectively. The dot plot demonstrates that the case of adding the group SE block won in eight among ten experiments with *p*-value of 0.025.

As demonstrated in Figures 1 and 2, this study uses a large number of segmented images explained in the CT preprocessing section. All structures included in the 78 segmented images were mentioned for their relevance to TAO in the existing literature [34], [35], [36]. In particular, volumetric changes in extraocular muscles are more characteristic than other orbit structures in CT images of TAO patients [37]. The proposed method considers the relative importance of each structure based on the group SE block. The attention scores for each structure can be calculated based on the group SE block of the proposed model. Table 5 lists the top 15 average attention scores with standard deviations generated from the group SE block of the proposed model for 50 tests in the main experiments. Although 32 were extraocular muscles among a total of 78 segmented images, the top 15 were all ranked by extraocular muscles. Moreover, 20 of the top 23 segmented images were extraocular muscles. Although TAO patients often experience eyeball pain, the segmented images of eyeballs were positioned at the low ranks. Among the total 14 eyeball images, the highest rank is 57. The fact that the TAO primarily affects the soft tissues surrounding the eye, such as the extraocular muscles and orbital fat, leading to changes that are more significant and diagnostic than changes in the eyeball itself is consistent with the attention scores [36]. The performance of the proposed model is slightly improved by the group SE block, as shown in Figure 7. The proposed can be used in clinical settings described in [38]. The output of the model can be provided to medical experts along with the attention score, which can provide reference data for inexperienced medical experts.

Despite the promising outcomes of this study, it had several limitations. The proposed method requires a substantial number of segmented images, is labor-intensive, and requires significant expertise. On the other hand, the cropping strategy is relatively simple as it only requires finding RoI and annotating it as a rectangle [16]. Thus, collaboration with recently developed segmentation models can help advance our approach. A recent work [10] on orbital tissue segmentation in CT images reported average Dice coefficients of 90%. In terms of NN architecture, existing NN-based diagnostics have been integrated with useful architectures, such as vision transformers [39] or graph-NNs [40]. Future research could consider new NN design and optimization approaches for global dependencies or graph structures [41], [42]. While the group SE block yielded a little performance improvement, a previous TAO diagnosis study has demonstrated more pronounced performance boosting by filtering out less important images [17]. Therefore, filtering out network nodes may be a promising approach. Finally, considering the fusion of information from various modalities may improve predictive performance. For example, TAO symptoms of the eye may be well identified on a facial photograph [43].

TABLE 5. Top 15 scores among the 78 attention scores generated from the group SE block. IR, SR, MR, and LR means inferior, superior, medial, and lateral rectus muscles, respectively.

Ranking	View	Structure	Left or Right	Score
1	Coronal	IR	R	0.640 ± 0.122
2	Coronal	SR	L	0.632 ± 0.131
3	Sagittal	IR	L	0.632 ± 0.123
4	Coronal	SR	R	0.630 ± 0.117
5	Sagittal	IR	R	0.622 ± 0.114
6	Axial	MR	L	0.615 ± 0.109
7	Sagittal	SR	R	0.603 ± 0.106
8	Coronal	IR	L	0.601 ± 0.118
9	Coronal	MR	R	0.594 ± 0.115
10	Coronal	MR	L	0.594 ± 0.107
11	Coronal	LR	L	0.591 ± 0.114
12	Axial	MR	L	0.586 ± 0.086
13	Sagittal	SR	L	0.583 ± 0.116
14	Coronal	LR	R	0.582 ± 0.108
15	Axial	MR	R	0.581 ± 0.103

VI. CONCLUSION

TAO is an autoimmune disease that leads to alterations in the structures close to eyes of the patient. This study leveraged the potential of image segmentation in medical imaging, emphasizing its ability to refine diagnostic performance. This paper introduces a new NN adept at processing a large number of segmented images to assess TAO activity. The proposed model incorporated multiple convolutional embedding heads, a group SE block, and a classifier stage. The proposed model exhibited the best performance across five evaluation metrics compared to the four comparative models. Consequently, the Discussion section presents future research directions related to data preprocessing and model architecture.

ACKNOWLEDGMENT

The funding organizations has no role in the design or conduct of this research.

REFERENCES

[1] F. Shen, J. Liu, L. Fang, Y. Fang, and H. Zhou, "Development and application of animal models to study thyroid-associated ophthalmopathy," *Experim. Eye Res.*, vol. 230, May 2023, Art. no. 109436.

[2] A. C. H. Lee and G. J. Kahaly, "Pathophysiology of thyroid-associated orbitopathy," *Best Pract. Res. Clin. Endocrinol. Metabolism*, vol. 37, no. 2, 2023, Art. no. 101620.

[3] H. Li et al., "T cell subsets are associated with clinical activity and efficacy of 4.5G intravenous glucocorticoid for moderate-to-severe thyroid-associated ophthalmopathy," *Endocrine Res.*, vol. 48, nos. 2–3, pp. 55–67, Jul. 2023.

[4] X. Liao, F. M. A. A. Aljufairi, K. K. H. Lai, K. K. W. Chan, R. Jia, W. Chen, Z. Hu, Y. Wei, W. C. W. Chu, C. C. Y. Tham, C. P. Pang, and K. K. L. Chong, "Clinical significance of corneal striae in thyroid associated orbitopathy," *J. Clin. Med.*, vol. 12, no. 6, p. 2284, Mar. 2023.

[5] J. Diao, X. Chen, Y. Shen, J. Li, Y. Chen, L. He, S. Chen, P. Mou, X. Ma, and R. Wei, "Research progress and application of artificial intelligence in thyroid associated ophthalmopathy," *Frontiers Cell Develop. Biol.*, vol. 11, Jan. 2023, Art. no. 1124775.

[6] P.-H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, "Current and emerging trends in medical image segmentation with deep learning," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 7, no. 6, pp. 545–569, Apr. 2023.

[7] A. Tragakis, C. Kaul, R. Murray-Smith, and D. Husmeier, "The fully convolutional transformer for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3649–3658.

[8] I. Qureshi, J. Yan, Q. Abbas, K. Shaheed, A. B. Riaz, A. Wahid, M. W. J. Khan, and P. Szczuko, "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Inf. Fusion*, vol. 90, pp. 316–352, Feb. 2023.

[9] W. Li, H. Song, Z. Li, Y. Lin, J. Shi, J. Yang, and W. Wu, "OrbitNet—A fully automated orbit multi-organ segmentation model based on transformer in CT images," *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106628.

[10] S. H. Lee, S. Lee, J. Lee, J. K. Lee, and N. J. Moon, "Effective encoder–decoder neural network for segmentation of orbital tissue in computed tomography images of Graves' orbitopathy patients," *PLoS ONE*, vol. 18, no. 5, May 2023, Art. no. e0285488.

[11] L. Paniagua, M. F. Bande, J. M. Abalo-Lojo, and F. Gonzalez, "Computer aided volumetric assessment of orbital structures in patients with Graves' orbitopathy: Correlation with serum thyroid antiperoxidase antibodies and disease activity," *Int. Ophthalmol.*, vol. 43, no. 9, pp. 3377–3384, Jun. 2023.

[12] C. Lin, X. Song, L. Li, Y. Li, M. Jiang, R. Sun, H. Zhou, and X. Fan, "Detection of active and inactive phases of thyroid-associated ophthalmopathy using deep convolutional neural network," *BMC Ophthalmol.*, vol. 21, no. 1, pp. 1–9, Dec. 2021.

[13] X. Song, Z. Liu, L. Li, Z. Gao, X. Fan, G. Zhai, and H. Zhou, "Artificial intelligence CT screening model for thyroid-associated ophthalmopathy and tests under clinical conditions," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 16, no. 2, pp. 323–330, Feb. 2021.

[14] J. Lee, W. Seo, J. Park, W.-S. Lim, J. Y. Oh, N. J. Moon, and J. K. Lee, "Neural network-based method for diagnosis and severity assessment of Graves' orbitopathy using orbital computed tomography," *Sci. Rep.*, vol. 12, no. 1, p. 12071, Jul. 2022.

[15] K. Hanai, H. Tabuchi, D. Nagasato, M. Tanabe, H. Masumoto, S. Miya, N. Nishio, H. Nakamura, and M. Hashimoto, "Automated detection of enlarged extraocular muscle in graves' ophthalmopathy with computed tomography and deep neural network," *Sci. Rep.*, vol. 12, no. 1, p. 16036, 2022.

[16] N. Yao, L. Li, Z. Gao, C. Zhao, Y. Li, C. Han, J. Nan, Z. Zhu, Y. Xiao, and F. Zhu, "Deep learning-based diagnosis of disease activity in patients with graves' orbitopathy using orbital SPECT/CT," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 50, no. 12, pp. 1–9, 2023.

[17] J. Lee, S. Lee, W. J. Lee, N. J. Moon, and J. K. Lee, "Neural network application for assessing thyroid-associated orbitopathy activity using orbital computed tomography," *Sci. Rep.*, vol. 13, no. 1, p. 13018, Aug. 2023.

[18] F. Yuan, Z. Zhang, and Z. Fang, "An effective CNN and transformer complementary network for medical image segmentation," *Pattern Recognit.*, vol. 136, Apr. 2023, Art. no. 109228.

[19] D. Kienzle, J. Lorenz, R. Schön, K. Ludwig, and R. Lienhart, "Covid detection and severity prediction with 3D-convnext and custom pretrainings," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 500–516.

[20] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features," *Diagnostics*, vol. 13, no. 6, p. 1026, Mar. 2023.

[21] Y. Zhang, Y.-L. Liu, K. Nie, J. Zhou, Z. Chen, J.-H. Chen, X. Wang, B. Kim, R. Parajuli, R. S. Mehta, M. Wang, and M.-Y. Su, "Deep learning-based automatic diagnosis of breast cancer on MRI using mask R-CNN for detection followed by ResNet50 for classification," *Academic Radiol.*, vol. 30, pp. S161–S171, Sep. 2023.

[22] M. Kim and M. H. Song, "High performing facial skin problem diagnosis with enhanced mask R-CNN and super resolution GAN," *Appl. Sci.*, vol. 13, no. 2, p. 989, Jan. 2023.

[23] K. R. Reddy and R. Dhuli, "A novel lightweight CNN architecture for the diagnosis of brain tumors using MR images," *Diagnostics*, vol. 13, no. 2, p. 312, Jan. 2023.

[24] Y. Akkem, S. K. Biswas, and A. Varanasi, "Smart farming using artificial intelligence: A review," *Eng. Appl. Artif. Intell.*, vol. 120, Apr. 2023, Art. no. 105899.

[25] Y. Kaya and E. Gürsoy, "A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection," *Soft Comput.*, vol. 27, no. 9, pp. 5521–5535, May 2023.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[27] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, Sep. 2022.

[28] P. Zou and J.-S. Wu, "SwinE-UNet3+: Swin transformer encoder network for medical image segmentation," *Prog. Artif. Intell.*, vol. 12, no. 1, pp. 99–105, Mar. 2023.

[29] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[30] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. Smola, "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2736–2746.

[31] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. So Kweon, and S. Xie, "ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16133–16142.

[32] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2999–3007.

[33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, New Orleans, USA, 2019, pp. 1–19.
[34] R. Li, J. Li, and Z. Wang, "Quantitative assessment of the intraorbital segment of the optic nerve in patients with thyroid orbitopathy using diffusion tensor imaging," *Acta Radiologica*, vol. 64, no. 2, pp. 725–731, Feb. 2023.
[35] J. Shao, X. Huang, T. Gao, J. Cao, Y. Wang, Q. Zhang, L. Lou, and J. Ye, "Deep learning-based image analysis of eyelid morphology in thyroid-associated ophthalmopathy," *Quant. Imag. Med. Surgery*, vol. 13, no. 3, pp. 1592–1604, Mar. 2023.
[36] X. Huang, W. Tang, Y. Shen, L. He, F. Tong, S. Liu, J. Li, P. Li, Y. Zhang, X. Ma, R. Wei, and W. Yang, "The significance of ophthalmological features in diagnosis of thyroid-associated ophthalmopathy," *Biomed. Eng. OnLine*, vol. 22, no. 1, p. 7, Feb. 2023.
[37] K. Rana, D. Garg, L. S. S. Yong, C. Macri, J. Y. Tong, S. Patel, J. Slattery, W. O. Chan, G. Davis, and D. Selva, "Extraocular muscle enlargement in dysthyroid optic neuropathy," *Can. J. Ophthalmol.*, Dec. 2023.
[38] J. H. Choi, J. Lee, S. H. Lee, S. Lee, A.-S. Moon, S.-H. Cho, J. S. Kim, I. R. Cho, W. H. Paik, J. K. Ryu, and Y.-T. Kim, "Analysis of ultrasonographic images using a deep learning-based model as ancillary diagnostic tool for diagnosing gallbladder polyps," *Digestive Liver Disease*, vol. 55, no. 12, pp. 1705–1711, Dec. 2023.
[39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–22.
[40] J Zhou, G Cui, S Hu, Z Zhang, C Yang, Z Liu, L Wang, C Li, and M Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
[41] Y. Dai, Y. Gao, and F. Liu, "TransMed: Transformers advance multi-modal medical image classification," *Diagnostics*, vol. 11, no. 8, p. 1384, Jul. 2021.
[42] X. Song, F. Zhou, A. F. Frangi, J. Cao, X. Xiao, Y. Lei, T. Wang, and B. Lei, "Multicenter and multichannel pooling GCN for early AD diagnosis based on dual-modality fused brain network," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 354–367, Feb. 2023.
[43] J. H. Moon, K. Shin, G. M. Lee, J. Park, M. J. Lee, H. Choung, and N. Kim, "Machine learning-assisted system using digital facial images to predict the clinical activity score in thyroid-associated orbitopathy," *SSRN Electron. J.*, p. 22085, 2022.

SANGHYUCK LEE is currently pursuing the Ph.D. degree majoring in artificial intelligence with Chung-Ang University. His research interests include computer vision and medical imaging.

JEONG KYU LEE received the B.S. degree from the College of Medicine, Korea University, and the M.S. and Ph.D. degrees in ophthalmology from Korea University. He trained in advanced eye surgery with the University of California at San Diego. He is an Ophthalmologist, specializing in oculoplastics, with expertise in eyelid and tear duct diseases, tumors, and thyroid eye diseases. He is a Professor with the Ophthalmology Department, Chung-Ang University College of Medicine, Chung-Ang University Hospital. He was an Editor of the *Korean Journal of Ophthalmology*.

JAESUNG LEE received the B.S., M.S., and Ph.D. degrees in computer science from Chung-Ang University, Seoul, South Korea, in 2007, 2009, and 2013, respectively. Currently, he is the Head and an Associate Professor with the Department of Artificial Intelligence, Chung-Ang University, where he is also the Chief of the AI/ML Innovation Research Center. His research primarily focuses on classification, feature selection, and multilabel learning with information theory. His research interests include broad and diverse, encompassing machine learning, multilabel learning, model selection, and neural architecture search.