Deep Neural Network With Structural Similarity Difference and Orientation-Based Loss for Position Error Classification in the Radiotherapy of Graves' Ophthalmopathy Patients

Wenjie Liu, Lei Zhang, *Senior Member, IEEE*, Guyu Dai, Xiangbin Zhang, Guangjun Li, and Zhang Yi, *Fellow, IEEE*

***Abstract*—Identifying position errors for Graves' ophthalmopathy (GO) patients using electronic portal imaging device (EPID) transmission fluence maps is helpful in monitoring treatment. However, most of the existing models only extract features from dose difference maps computed from EPID images, which do not fully characterize all information of the positional errors. In addition, the position error has a three-dimensional spatial nature, which has never been explored in previous work. To address the above problems, a deep neural network (DNN) model with structural similarity difference and orientation-based loss is proposed in this paper, which consists of a feature extraction network and a feature enhancement network. To capture more information, three types of Structural SIMilarity (SSIM) sub-index maps are computed to enhance the luminance, contrast, and structural features of EPID images, respectively. These maps and the dose difference maps are fed into different networks to extract radiomic features. To acquire spatial features of the position errors, an orientation-based loss function is proposed for optimal training. It makes the data distribution more consistent with the realistic 3D space by integrating the error deviations of the predicted values in the left-right, superior-inferior, anterior-posterior directions. Experimental results on a constructed dataset demonstrate the effectiveness of the proposed model, compared with other related models and existing state-of-the-art methods.**

***Index Terms*—Deep neural network, volumetric modulated radiation therapy, position error, SSIM analysis, EPID dosimetry.**

Manuscript received June 7, 2021; revised November 5, 2021 and December 13, 2021; accepted December 19, 2021. Date of publication December 23, 2021; date of current version June 6, 2022. This work was supported in part by the National Natural Science Fund for Distinguished Young Scholar under Grant 62025601, in part by the General Program of National Natural Science Foundation of China under Grant 61772353, in part by the National Major Science and Technology Projects of China under Grant 2018AAA0100201, and in part by the Sichuan University Innovation Spark Project Library under Grant 2018SCUH0040. *(Corresponding author: Lei Zhang.)*

Wenjie Liu, Lei Zhang, and Zhang Yi are with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: liuwj@stu.scu.edu.cn; leizhang@scu.edu.cn; zhangyi@scu.edu.cn).

Guyu Dai, Xiangbin Zhang, and Guangjun Li are with the Department of Radiation Oncology, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610044, China (e-mail: 418098640@qq.com; rtcheungx@163.com; gjnick829@sina.com).

Digital Object Identifier 10.1109/JBHI.2021.3137451

I. Introduction

GRAVES' ophthalmopathy (GO) is a potentially sight-threatening ocular disease whose most common clinical features are upper eyelid retraction, edema, and erythema of the periorbital tissues and conjunctivae, and proptosis [1]. Radiation therapy (RT) is an established treatment modality for GO, and is beneficial for patients who do not respond to initial RT or experience symptom recurrence without an apparent risk of increased morbidity [2]. Intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) are often used for GO patients [3]–[5] because they could achieve steeper dose gradients between the target and normal structures, thus reducing the dose to surrounding normal tissues without compromising the planning target coverage [3], [6]. However, the problem occurs that errors during treatment may affect the treatment results, such as position errors [7]. Cone beam computed tomography (CBCT) can be used to correct errors, but it introduces additional radiation dose to the patient, and intra-fractional changes can remain after the CBCT scanning [8], [9]. Therefore, in vivo verification of radiotherapy treatment has become increasingly important [7], [10].

One effective in vivo verification method is the electronic portal imaging device (EPID) dosimetry [11]. It is capable of recording radiation dose deviations and converting them into 2D images [12], [13]. Gamma analysis is then used to compare the quantitative agreement and distance-to-agreement between the measured EPID dose and the designed RT plan dose [14], [15]. But the performance of models for detecting position errors in this way is hardly satisfactory [16].

With the development of neural networks, some researchers have tried to apply convolutional neural networks (CNN) to RT plan quality assurance and have achieved results beyond previous studies [17], [18]. On the one hand, CNNs are able to extract deeper features from medical images that may have been overlooked in previous studies. On the other hand, the CNN

model is an end-to-end model, without adding an additional classifier at the end like traditional machine learning methods. Although these CNN models have improved the accuracy of error detection in radiotherapy to a new level, some practical problems remain and are summarized as follows:

1) Most models simply use the dose difference (DD) images as model input, which are obtained by subtracting two EPID maps. However, these processed images can not express all the position error features of the original EPID dose images.
2) Existing studies of radiotherapy delivery errors mostly identify the type of treatment errors from a macroscopic perspective. When refined to classify positional errors, the model performance is not satisfactory.
3) The identification of position errors in VMAT delivery is characterized by three-dimensional spatiality, which has not been taken into account by previous CNN-based models. So a new model needs to be designed to capture the spatiality features of this task.

The purpose of this study is to solve the above problem and to improve the accuracy of CNN-based models in identifying position errors for GO patients using EPID dosimetry in an in vivo scenario. To this end, a deep neural network model with structural similarity difference and orientation-based loss is proposed to extract image features fully. Three types of Structural SIMilarity (SSIM) sub-index maps are computed to enhance the luminance, contrast, and structural features between two EPID images, respectively. An orientation-based loss function is proposed to train the model to capture the spatial features of errors in different directions. It simulates the spatiality of the position error by calculating the errors in the left-right (LR), superior-inferior (SI), and anterior-posterior (AP) directions. The results are compared with traditional machine learning methods and advanced CNN models.

The significant contributions of this work are summarized as follows:

1) Three input images were calculated from EPID transmission fluence maps using SSIM sub-indexes, which enhance images' luminance, contrast, and structural features.
2) An end-to-end neural network model is designed to capture the features of DD images and SSIM sub-index images. This model contains a feature extraction network and a feature enhancement network.
3) An orientation-based loss function is proposed to optimize the training of the model. It captures the three-dimensional spatial characteristics of the position errors and makes the prediction results more consistent with real-world scenarios.

II. Related Work

Due to the complexity of radiotherapy treatment, in vivo verification of radiotherapy treatment has become increasingly important. Traditional machine learning methods have achieved significant improvements in the task of identifying treatment errors [19]–[21]. For these approaches, the PORTS code or the open-source radiomics library package PyRadiomics [22] is first used to extract the radiomics features of the processed images. Then, these features are fed to a classifier to classify errors. The mainstream classifiers include logistic regression, k-nearest neighbor (KNN), supporting vector machine (SVM), linear discriminant classifier (LDC), and random forest (RF). However, the traditional methods achieve limited performance as they depend solely on hand-crafted features. Other researchers explored more flexible approaches such as Markov models [23], but the results are still unsatisfactory.

With the development of deep neural networks (DNNs) [24]–[28], more and more researchers have tried to apply them in the medical field and achieved breakthroughs, including breast cancer diagnosis [29]–[33], thyroid diagnosis [34], and pancreas segmentation [35]. Compared with traditional machine learning methods, DNN models are able to extract more features and automatically filter useful information, thus improving prediction performance.

Therefore, Nyflot *et al.* used CNN models to identify systematic agility multileaf collimator (MLC) mis-positioning as well as random MLC mis-positioning [17]. The whole model includes three feature extractors, each consisting of four convolutional layers and two linear layers. The experiments are compared with the manual feature extraction approach, and the results show that the CNN-based model significantly outperforms the traditional machine learning approach. Potter *et al.* designed an artificial neural network model and a CNN model to detect and classify six types of dose delivery errors and seven types of spatial errors, respectively [36]. The results showed that both the dose difference maps and the distance-to-agreement maps are suitable features for error classification in IMRT QA. Kimura *et al.* used a CNN model to detect MLC positional errors from dose difference maps in patient-specific QA for VMAT [18]. The results of the five-fold cross-validation showed that the proposed method was superior to those based on gamma analysis and provided an effective solution for detecting MLC errors.

The above study shows that CNNs can be effectively used in EPID delivery error analysis for radiotherapy treatment. In order to explore the types of dose delivery errors more comprehensively, Wolfs *et al.* trained a CNN model to detect and identify error type and magnitude of simulated treatment errors in lung cancer patients [7]. They provided a proof-of-concept of CNNs for error identification using EPID dosimetry in an in vivo scenario.

Previous studies illustrate the potential of CNN-based models for analyzing EPID images of identifying position errors, but the performance of these models is unsatisfactory. Furthermore, most of these models extract features from DD images for classification, but such images can express limited information, which also limits the performance of the model. Most importantly, none of these studies considered the spatial nature of the position error, so a new model urgently needs to be designed to solve the above problem.

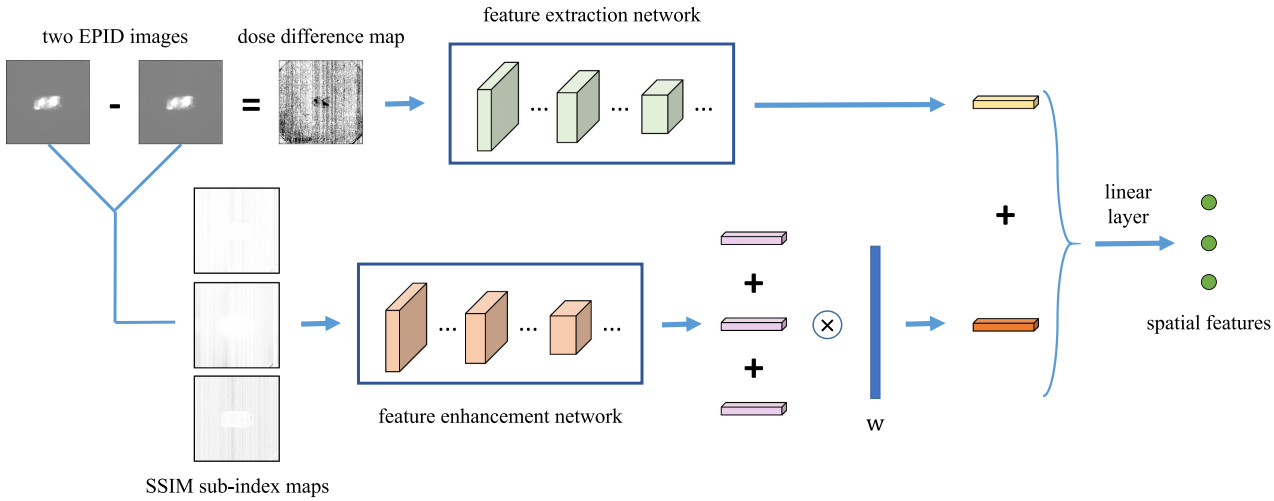

Fig. 1. Diagram of the overall architecture of the proposed model. Firstly, the dose difference map obtained by subtracting the two EPID images is fed into the feature extraction network to extract the dose features. After that, three SSIM sub-index maps are computed and input to the feature enhancement network to extract luminance, contrast, and structural features, respectively. A linear layer is used to integrate them. Finally, all features are combined together to obtain three spatial error values.

TABLE I
OVERVIEW OF THE CATEGORIES OF THE SIMULATED POSITION ERRORS

Categories	LR	SI	AP	number of images
1	$0-3mm$	$0-3mm$	$0-3mm$	640
2	$0-3mm$	$0-3mm$	$>3mm$	320
3	$0-3mm$	$>3mm$	$0-3mm$	320
4	$0-3mm$	$>3mm$	$>3mm$	160
5	$>3mm$	$0-3mm$	$0-3mm$	320
6	$>3mm$	$0-3mm$	$>3mm$	160
7	$>3mm$	$>3mm$	$0-3mm$	160
8	$>3mm$	$>3mm$	$>3mm$	80

LR: left-right; SI: superior-inferior; AP: anterior-posterior.

III. MATERIALS AND METHODS

A. Dataset

The dataset used in this paper includes 2240 EPID transmission fluence maps collected from the Department of Radiation Oncology of the West China Hospital of Sichuan University. Specifically, 40 VMAT plans (P1A1: $240^\circ - 120^\circ$ CW; P1A2: $120^\circ - 240^\circ$ CCW) of GO patients who received radiotherapy between November 2019 and October 2020 were selected, and their prescription dose was 20 Gy in 10 fractions within two to three weeks. These plans were then delivered to a head phantom on an EDGE linac (Varian Medical Systems, Palo Alto, CA), and the beam transmission fluence maps were measured by as1200 EPID (Varian Medical Systems, Palo Alto, CA) using dosimetry mode.

For each plan, a baseline transmission fluence map was obtained first by conducting one measurement without position errors. Then, EPID transmission fluence maps with position errors were measured after position errors were simulated by translating the head phantom along left-right, superior-inferior, and anterior-posterior directions. To obtain more accurate results, 3 mm was chosen as the threshold for position errors instead of 1 cm as in previous studies [7]. Table I shows the simulation position error values in each direction and the total number of fluence maps for each category.

In in vivo verification of radiotherapy treatment, the effect of random mechanical error on fluence maps is negligible compared with the effect of position errors [21]. For GO patients, the rigid anatomical structure around the target volumes also makes the effect of anatomical changes on the fluence map negligible. Therefore, this study can focus on improving the accuracy of CNN-based models in identifying position errors for GO patients using EPID transmission fluence maps.

B. Overall Architecture

The impact of position error on treatment is difficult to estimate quantitatively, so previous studies have treated positional error as a two-class classification problem. In order to better assist clinical treatment, this paper classifies position errors by refining them into eight categories based on their spatial characteristics. So the model structure is designed with the idea of extracting the features that represent this spatiality. The overall architecture of the proposed model is shown in Fig. 1. It contains a feature extraction network as well as a feature enhancement network. Based on previous studies, dose difference maps generated by subtracting two EPID dose images can be used to predict whether there is an error in the patient's position at the time of treatment. So the feature extraction network is designed to extract the features of the dose difference maps. These features represent the pixel difference features between EPID images with and without position errors. However, a previous study shows that the dose difference maps can not fully reflect the features of the position errors [37]. To improve the performance of the model, a feature enhancement network is proposed to extract features from three additional types of images. Specifically, three structural similarity difference maps will be computed from two EPID dose images using different SSIM sub-indexes. They can reflect the luminance, contrast, and

structure differences between dose maps, respectively, allowing the network to extract more useful information from higher dimensional data.

Formulaically, the training data come from the dataset $D = \{(x_n, x'_n, y_n); n = 1, 2, \ldots, N\}$ where N is the total number of EPID images, x_n and x'_n are the EPID dose maps measured with and without positional errors, respectively, and y_n is its corresponding label. For convenience and brevity, the subscript n has been removed in the latter part. So given the images x and x', the dose difference map x^d is first obtained by $x - x'$ ($-$ is a pixel-level subtraction operation). Then features $z^d = f_e(x^d \mid \phi)$ is extracted by the feature extraction network, where $z^d \in \mathbb{R}^C$ and C is the channel dimension (C = 512 in this study). f_e stands for a pre-trained DNN with parameters ϕ. Unlike the radiomics features, the values in z^d are not a specific metric but a set of values containing the dose difference features of the images x^d. On the other hand, the three types of images x^l, x^c, and x^s are computed by SSIM sub-indexes. How to get these images will be described in detail in the next subsection. Then similarity features z^l, z^c, and z^s are extracted by another network, i.e., $z^l = f_a(x^l \mid \varphi)$, $z^c = f_a(x^c \mid \varphi)$, and $z^s = f_a(x^s \mid \varphi)$. f_a stands for another pre-trained DNN with parameters φ.

To integrate these three similarity features, a linear layer is added for better dimensionality reduction. The final similarity features z^f are calculated from (1):

$$z^f = \sigma((z^l + z^c + z^s) \cdot w + b), \tag{1}$$

where w and b are the weights and bias in the linear layer, σ is the RELU activation function, and $\cdot$ is the inner product of the matrix w and the calculated vector. Finally, the two types of features (z^d, z^f) are added together and a linear layer with a sigmoid activation function is used to calculate the three prediction values (p^{LR}, p^{SI}, p^{AP}). Unlike the traditional method, the proposed model does not predict the category to which the image belongs, but the value of the position error in three directions. Because the commonly used classification loss function can not reflect the spatial characteristics of the position errors, an orientation-based loss is proposed to improve the model performance, which will be introduced in section III-D.

The overall end-to-end training process of the proposed method is shown in Algorithm 1. All weights will be updated by training, and no layer is frozen. By predicting the position error in each direction by analyzing the EPID fluence maps, this method can be combined with CBCT in the clinical treatment of GO patients. The fluence maps of the first fraction should be used as a baseline, and fluence maps should be measured after using CBCT to correct the position errors at the first few fractions. Then the proposed model can be used to compare the baseline maps and other fluence maps. The frequency of using CBCT can be reduced if the predictions show no apparent errors, and the proposed model can be used to monitor the next fractions.

Algorithm 1: Training Process of the Proposed Model.

Input:
The input dataset: $D = \{(x_n, x'_n, y_n); n = 1, 2, \ldots, N\}$
Output:
The probability of position error (p^{LR}, p^{SI}, p^{AP}) in three directions

```
Ending epochs = 200
Initializing the networks with pre-trained parameters
while training epoch < ending epochs do
  for n = 1 : N do
    Calculating dose difference maps x_n^d ← x_n − x'_n
    Calculating SSIM sub-index maps x_n^l, x_n^c, x_n^s by (2-4)
    Extracting features z_n^d ← f_e(x_n^d | φ), z_n^l, z_n^c, z_n^s ← f_a(x_n^l, x_n^c, x_n^s | φ),
    Integrating similarity features z_n^f by (1)
    Computing position error probability (p^LR, p^SI, p^AP)
    Computing the orientation-based loss by (5)
    Updating gradients with back propagation algorithm
  end for
end while
while training epochs = ending epochs do
  Save the model and parameters
end while
```

C. SSIM Sub-Index Maps

This section will describe how to obtain structural similarity difference images for two EPID dose images. Most previous position error studies have focused on extracting information from DD images. Although these images contain some position error features, the biggest problem with this approach is that some features such as luminance or contrast may be overlooked or weakened because of simple pixel subtraction. SSIM index has been proposed more than a decade ago to quantify the visibility of errors between a distorted image and a reference image using a variety of known properties of the human visual system [38]. Similarly, position error of patient treatment in radiotherapy can be determined from the difference between two EPID dose images. Although this task is not quite the same as the quality assessment task of distorted images, the similarity between these two tasks is that both explore the correlation between two images. So establishing the relationship between EPID images by SSIM coefficients may allow the model to extract more useful features.

In order not to lose as much information as possible, three sub-index maps are generated instead of only calculating one SSIM index map. Specifically, given a pair of images (x, x'), patches m and n are chosen from the same spatial location of x and x'. Then three sub-indexes of the two patches are computed by:

$$\boldsymbol{l(m,n)} = \frac{2\mu_m\mu_n + C_1}{\mu_m^2 + \mu_n^2 + C_1}, \tag{2}$$

$$\boldsymbol{c(m,n)} = \frac{2\sigma_m\sigma_n + C_2}{\sigma_m^2 + \sigma_n^2 + C_2}, \tag{3}$$

$$\boldsymbol{s(m,n)} = \frac{\sigma_{mn} + C_3}{\sigma_m\sigma_n + C_3}, \tag{4}$$

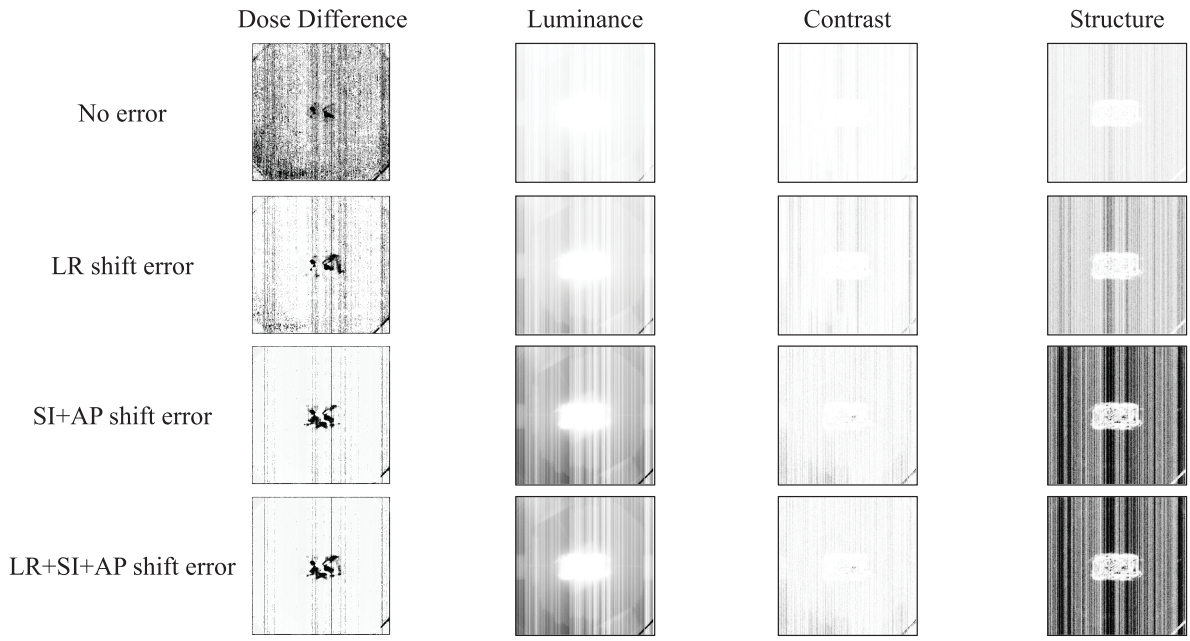

Fig. 2. The dose difference maps and SSIM sub-index maps generated from the EPID images of one patient. The first column indicates the dose difference images, and the next three columns indicate the images generated by the SSIM sub-indexes for luminance, contrast, and structure, respectively. The first row shows the images without error, and the next three rows show the images under the presence of single or two or three directional errors, respectively.

where $\boldsymbol{l(m,n)}$, $\boldsymbol{c(m,n)}$ and $\boldsymbol{s(m,n)}$ are the luminance, contrast and structure index, respectively. μ_m and μ_n, σ_m and σ_n, and σ_{mn} are the local means, standard deviations and cross-covariance of m and n, respectively. C_1, C_2 and C_3 are the regularization constants. Three new images can be generated by sliding the window to calculate the coefficient values between the two patches. Fig. 2. shows the dose difference images and SSIM sub-index images generated from two EPID images of one patient.

In the original paper [38], $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, and $C_3 = C_2/2$ are very important parameters to avoid numerical instability, where K_1 and K_2 are small constants. L is the range of pixel values of the images ($L = 255$ in this paper). As suggested by Peng *et al.* [37], $K_1 = 0.01$ and $K_2 = 0.03$ are used in this experiment. Similar to their work, the patch size is set to 11×11 pixels and the stride is one pixel. In order not to change the size of the images, the local window will add 10 pixels to the edges with zero-padding.

D. Orientation-Based Loss

This section focuses on the design and ideas related to the orientation-based loss function. There is a certain tolerance for positional errors in the clinical treatment of radiotherapy, and errors exceeding this threshold may seriously affect the normal tissues in the vicinity of the site to be treated. At the same time, due to the difficulty of data collection, existing models treat this task as a classification task, i.e., the presence or absence of positional errors. Unfortunately, these models are far from meeting the clinical standard by simply predicting the presence or absence of positional error. The position error is characterized by three-dimensional spatiality, which has never been studied in previous work. To better monitor treatment, this study considers the prediction of positional error as a multi-classification task. This design allows radiotherapists to obtain more accurate position error prediction results.

According to Table I, all three directions can be predicted whether the error is greater than a threshold, which is a standard 8-classification problem. Cross-entropy loss is often used as the loss function for training in such a task. However, the problem occurs that errors between certain ranges are all considered as one class, which will make the class boundaries more difficult to determine. For example, an error of 0 mm and an error of 1 mm belong to one category, but the difference between them becomes difficult to reflect from the loss value. To solve the above problem, an orientation-based loss is proposed to better determine the bounds of the classification. To the best of our knowledge, our work is the first one in which the spatial nature of the radiotherapy position error features has been taken into account. So given a set of comparison images (x, x'), the label y will be redefined as $y = (y^{LR}, y^{SI}, y^{AP})$ instead of just the category number. $y^{LR}, y^{SI}, y^{AP} \in \{1, 0\}$ represent the labels of patients in the three directions left-right, superior-inferior and anterior-posterior, respectively. 1 represents an error greater than the threshold (3 mm) while 0 represents an error less than the threshold. The orientation-based loss is defined as:

$$loss = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{3}[(p_n^{LR} - y_n^{LR})^2 + (p_n^{SI} - y_n^{SI})^2 + (p_n^{AP} - y_n^{AP})^2], \tag{5}$$

where p^{LR}, p^{SI}, p^{AP} are the predicted result of the proposed model. This design captures the spatial distance between predicted and actual errors, thus reducing the intra-class distance between data of the same class. With the ability to predict features in all three directions simultaneously, the proposed method provides more practical monitoring support compared to simply classifying whether errors exist.

IV. Experiment

A. Training Details

The size of the original dose image is 1180×1180. Due to the nature of EPID images, the position error features are all concentrated in the middle region of the image. So the center crop 512×512 is used to eliminate the redundant background and all images are normalized to [0,1]. The Adam optimizer is used with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-4}. The batch size equal to 4 and the feature extraction network is the pre-trained VGG16 model initialized on ImageNet [39], which can also be replaced by any other DNN model. A GPU of GeForce RTX 3090 is used to train the model. The dataset is randomly divided in the ratio of training:validation:test = 3:1:1. The training stops at the 200th epochs and the parameters of the networks with the best performance on the test data will be saved. Accuracy will be used as the evaluation metric for model performance in this paper. It is defined as follows:

$$\text{Accuracy(Acc.)} = \frac{right}{all}, \tag{6}$$

where *right* represents the number of correct position error predictions in the test set and *all* represents the total number of EPID fluence maps in the test set.

B. Results and Analysis

Most previous studies have mainly used traditional machine learning methods for error classification of EPID images. To compare with these methods, commonly used models are tested on the constructed dataset. As in previous studies, radiomic features are first calculated by the functions provided by the Pyradiomics library. Afterward, features with a problematic range of values and features with low correlation are excluded and 369 features from one dose difference image and three structural similarity images are obtained. Finally, four machine learning methods (XGBoost, KNN, SVM, LDC) are used as a classifier to get the final results. For these four models, GridSearchCV was used to find hyperparameters to allow these models to obtain the best performance. For XGBoost, n_estimators = 240, and the maximum tree depth of 3 were searched. For KNN, the suitable number of neighbors was 9. For SVM, the suitable hyperparameters for the kernel, C, gamma were rbf, 10, 0.01, respectively. For LDC, the suitable solver (svd) was searched. Several state-of-the-art DNN models are also reproduced for comparison with the proposed model, and the results are shown in Table II. For the model of Nyflot *et al.* [17], a linear layer with a sigmoid activation function is added at the last layer to obtain the classification results. For the input combination of DD + SSIM images, the features of different types of images are extracted separately by the same network and concatenated together for prediction.

TABLE II
Comparison of Different Models for the Eight-Class Positional Errors Classification

model	Acc.
XGBoost	0.597
KNN	0.500
SVM	0.685
LDC	0.676
Nyflot et al. [17] (DD images alone)	0.560
Kimura et al. [18] (DD images alone)	0.597
Nyflot et al. [17] (DD + SSIM images)	0.453
Kimura et al. [18] (DD + SSIM images)	0.574
Ours	**0.722**

The proposed model achieves the highest accuracy values among all models. Compared with traditional machine learning methods, DNNs can automatically extract features from images and are not limited to radiomic features, which enriches feature representation. Fig. 3 shows the confusion matrix of the prediction results of different models on the test set. It can be seen that the prediction results of machine learning methods are not stable enough and tend to predict the results to a certain category and lead to poor performance in classifying other categories. In contrast, the proposed method has no obvious tendency and the prediction accuracy for all classes is close and higher than 0.5.

The existing state-of-the-art DNN-based models were tested separately using two different types of input data. The results show that the input combination of DD + SSIM images is not suitable for the previous model. Still, the proposed model can easily combine different image features, which also shows the strong robustness and adaptability of our model.

To illustrate that the proposed method can indeed extract deeper features, the features extracted by the different DNN models are reduced to two dimensions to compare the category distributions and are shown in Fig. 4. The dimensionality reduction method used in this paper is the t-distributed Stochastic Neighbor Embedding (t-SNE) method [40]. It can be seen that the intra-class distance is longer and there is no clear boundary between classes when the model of Nyflot *et al.* [17] or Kimura *et al.* [18] is used. At the same time, there are a large number of points scattered and overlapping together, which increases the difficulty of discriminating classes. The proposed method reduces the inter-class distance and there is a more obvious boundary line between classes. However, there are a few points that almost overlap and are far from the other part of the points. To investigate the reason for this problem, the predicted results for each patient were taken out separately for the analysis. It is found that most of these overlapping points originated from the simulated error EPID maps of two patients' RT plans. It means that our model is not so sensitive to some patients that it does not classify them correctly. In the future, we will collect more data to improve the classification performance of the model.

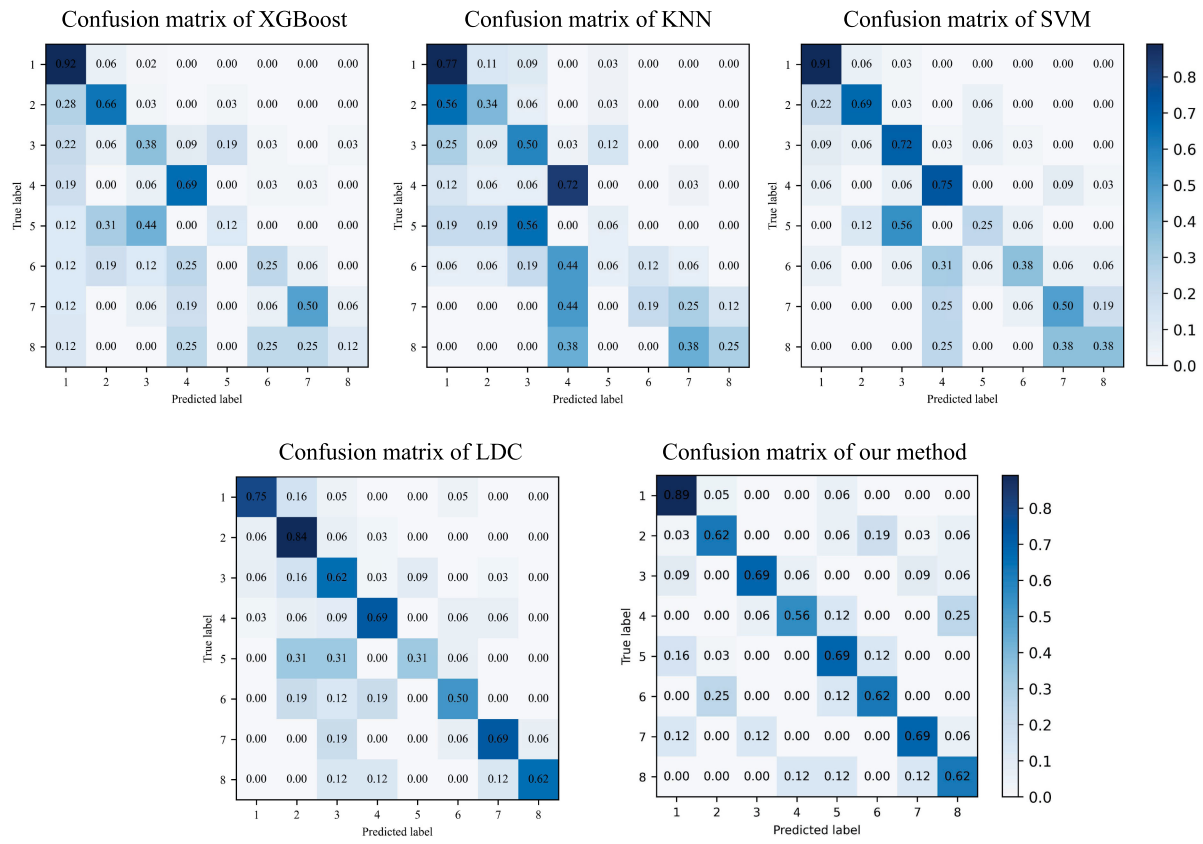

Fig. 3. The confusion matrices of XGBoost, KNN, SVM, LDC, and the proposed method on testing data. Each of the 8 classes corresponds to Table I. All models are trained with DD images and SSIM sub-index images. The proposed model has more uniform prediction results in each category and outperforms other models.

TABLE III
COMPARISON OF DIFFERENT PRE-TRAINED NETWORKS FOR EIGHT-CLASS POSITIONAL ERRORS CLASSIFICATION

model	Acc.
pre-trained ResNet101 [41]	0.643
pre-trained GoogLeNet [42]	0.671
pre-trained DenseNet169 [43]	0.708
pre-trained VGG16 [44]	**0.722**

TABLE IV
ABLATION STUDY FOR EIGHT-CLASS POSITIONAL ERRORS CLASSIFICATION

model	Acc.
VGG16 + CrossEntropyLoss	0.398
pre-trained VGG16 + CrossEntropyLoss	0.643
pre-trained VGG16 + OrbLoss	0.652
pre-trained VGG16 + SSIM +CrossEntropyLoss	0.671
pre-trained VGG16 + SSIM + OrbLoss	**0.722**

OrbLoss: Orientation-based loss CrossEntropyLoss: Cross-entropy loss.

C. Ablation Study

Different pre-trained networks will have different focuses in extracting features. To select the most suitable network for predicting position errors, an ablation experiment is designed to test the performance of different models. Since the input size is different from that of the pre-trained models, we replaced the last layer (max pooling or average pooling) of all pre-trained models with an adaptive averaging pooling layer to ensure that the dimensionality of the acquired features does not change. The comparison of different pre-trained networks for eight-class positional errors classification is shown in Table III. The highest accuracy is achieved by the pre-trained VGG16 network, so it is used as the feature extraction network in this paper. Although the prediction results vary among DNN models, most of them perform better than traditional machine learning methods, which proves the superiority of neural networks.

In this paper, the SSIM sub-index maps are computed and the orientation-based loss function is proposed. To explore whether these two modules can improve the model performance, comparison experiments are performed and the results are shown in Table IV. The prediction performance of the model was significantly reduced when the SSIM sub-index maps are not used for training. As shown in Fig. 2, the SSIM sub-index images do not contain the same information as the DD images. Hence, the performance of the model is further improved when richer information is introduced during training. Contrasted with the orientation-based loss function is the cross-entropy loss function, which is the most commonly used loss function in classification. The prediction performance of the model is also reduced when the orientation-based loss function is not used for training.

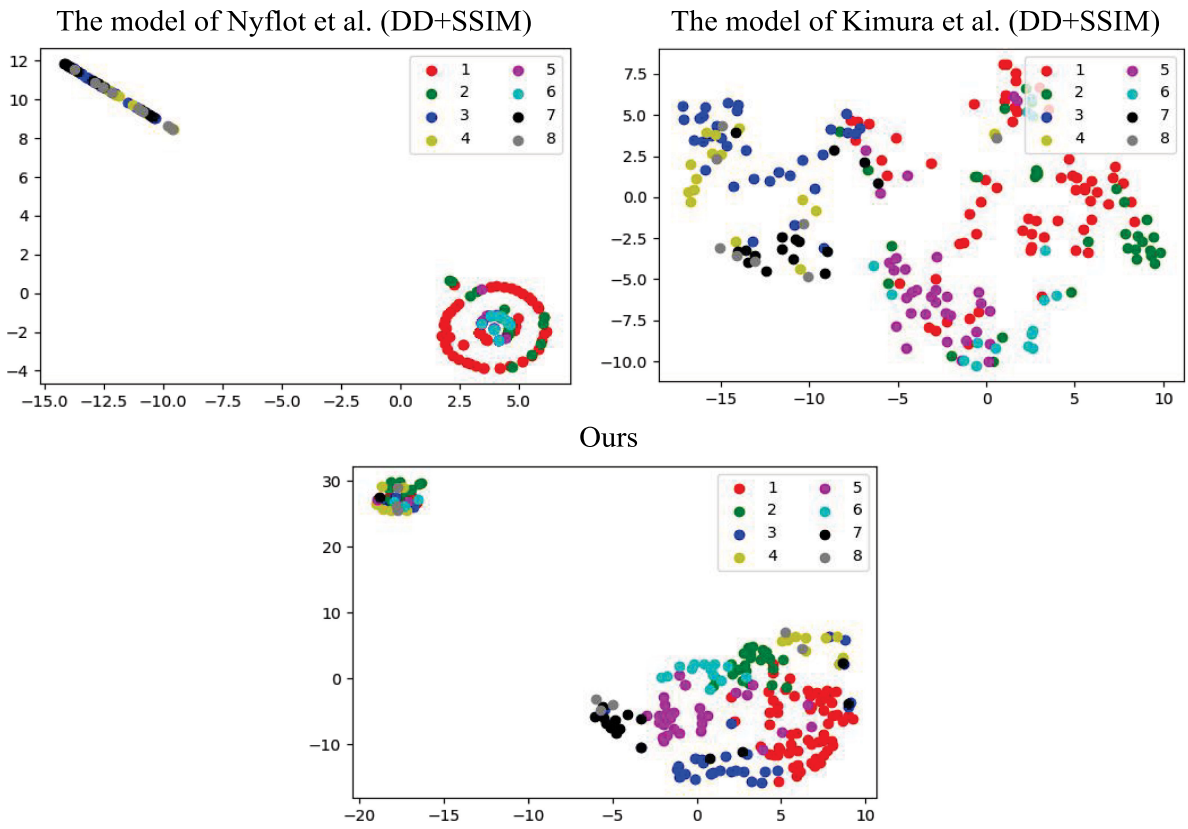

Fig. 4. Scatter plot of features reduced to two dimensions. The horizontal and vertical axes represent the first and second dimensions, respectively. Each of the eight classes corresponds to the eight position errors in Table I.

The position error is characterized by three-dimensional spatiality in the actual treatment, and the proposed orientation-based loss function can capture this spatial feature to improve the performance. The feature enhancement network introduces feature extraction for SSIM sub-index maps, an approach that allows the network to learn richer features. The combination of these two techniques allows the model to extract deeper features from more information to achieve state-of-the-art performance.

V. Conclusion

This paper proposes a DNN model with structural similarity difference and orientation-based loss for position error classification of GO patients using EPID transmission fluence maps. The DD maps and three types of SSIM sub-index maps of EPID images are fed into different networks to capture deeper position error features. The orientation-based loss function is proposed for training the model, which obtains the spatial characteristics of the position errors by calculating the differences of the predicted values in the left-right, superior-inferior, and anterior-posterior directions. The two-dimensional feature scatters plot illustrates that this loss function allows the model to better learn the bounds between classes. The experimental results on the constructed dataset indicate that the proposed model can be effectively used for the position error classification of GO patients, compared with other related models and existing state-of-the-art methods.

Most existing studies on radiotherapy error identification used dose delivery information, and information from the RT plan itself like CT images, monitor unit (MU) dose, etc. are not used. One possible solution is to combine neural networks and multimodal data fusion techniques. This information will be added and explored to enhance model performance in future work.

References

[1] R. S. Bahn, "Mechanisms of disease Graves' ophthalmopathy," *New England J. Med.*, vol. 362, no. 8, pp. 726–738, 2010.
[2] M. Matthiesen *et al.*, "The efficacy of radiation therapy in the treatment of Graves' orbitopathy," *Int. J. Radiat. Oncol., Biol., Phys.*, vol. 82, no. 1, pp. 117–123, 2012.
[3] Y.-J. Li *et al.*, "The efficacy of intensity modulated radiation therapy in treating thyroid-associated ophthalmopathy and predictive factors for treatment response," *Sci. Rep.*, vol. 7, no. 1, pp. 17533–9, 2017.
[4] I. San-Miguel *et al.*, "Volumetric modulated arc therapy (VMAT) make a difference in retro-orbital irradiation treatment of patients with bilateral Graves' ophthalmopathy: Comparative analysis of dosimetric parameters from different radiation techniques," *Rep. Practical Oncol. Radiotherapy*, vol. 21, no. 5, pp. 435–440, 2016.
[5] S.-C. Wang *et al.*, "Comparison of IMRT and VMAT radiotherapy planning for Graves' ophthalmopathy based on dosimetric parameters analysis," *Eur. Rev. Med. Pharmacological Sci.*, vol. 24, no. 7, pp. 3898–3906, 2020.
[6] K. Chao *et al.*, "A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: Initial results," *Int. J. Radiat. Oncol., Biol., Phys.*, vol. 49, no. 4, pp. 907–916, 2001.
[7] C. J. Wolfs, R. A. Canters, and F. Verhaegen, "Identification of treatment error types for lung cancer patients using convolutional neural networks and epid dosimetry," *Radiotherapy Oncol.*, vol. 153, pp. 243–249, 2020.

[8] P. Alaei and E. Spezi, "Imaging dose from cone beam computed tomography in radiation therapy," *Physica Medica*, vol. 31, no. 7, pp. 647–658, 2015.
[9] S. T. Heijkoop *et al.*, "Quantification of intra-fraction changes during radiotherapy of cervical cancer assessed with pre- and post-fraction cone beam CT scans," *Radiotherapy Oncol.*, vol. 117, no. 3, pp. 536–541, 2015.
[10] C. Bojechko and E. Ford, "Quantifying the performance of in vivo portal dosimetry in detecting four types of treatment parameter variations," *Med. Phys.*, vol. 42, no. 12, pp. 6912–6918, 2015.
[11] B. Mijnheer *et al.*, "Error detection during vmat delivery using EPID-based 3D transit dosimetry," *Physica Medica*, vol. 54, pp. 137–145, 2018.
[12] S. Nijsten *et al.*, "A global calibration model for EPIDs used for transit dosimetry," *Med. Phys.*, vol. 34, no. 10, pp. 3872–3884, 2007.
[13] M. Podesta, S. M. J. J. G. Nijsten, L. C. G. G. Persoon, S. G. Scheib, C. Baltes, and F. Verhaegen, "Time dependent pre-treatment EPID dosimetry for standard and FFF VMAT," *Phys. Med. Biol.*, vol. 59, no. 16, p. 4749, 2014.
[14] D. A. Low and J. F. Dempsey, "Evaluation of the gamma dose distribution comparison method," *Med. Phys.*, vol. 30, no. 9, pp. 2455–2464, 2003.
[15] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Med. Phys.*, vol. 25, no. 5, pp. 656–661, 1998.
[16] E. S. Hsieh, K. S. Hansen, M. S. Kent, S. Saini, and S. Dieterich, "Can a commercially available EPID dosimetry system detect small daily patient setup errors for cranial IMRT/SRS," *Practical Radiat. Oncol.*, vol. 7, no. 4, pp. e283–e290, 2017.
[17] M. J. Nyflot, P. Thammasorn, L. S. Wootton, E. C. Ford, and W. A. Chaovalitwongse, "Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks," *Med. Phys.*, vol. 46, no. 2, pp. 456–464, 2019.
[18] Y. Kimura, N. Kadoya, S. Tomori, Y. Oku, and K. Jingu, "Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated ARC therapy," *Physica Medica*, vol. 73, pp. 57–64, 2020.
[19] L. S. Wootton, M. J. Nyflot, W. A. Chaovalitwongse, and E. Ford, "Error detection in intensity-modulated radiation therapy quality assurance using radiomic analysis of gamma distributions," *Int. J. Radiat. Oncol.* Biol.* Phys.*, vol. 102, no. 1, pp. 219–228, 2018.
[20] M. Sakai *et al.*, "Detecting MLC modeling errors using radiomics-based machine learning in patient-specific QA with an EPID for intensity-modulated radiation therapy," *Med. Phys.*, vol. 48, no. 3, pp. 991–1002, 2021.
[21] C. Ma *et al.*, "The structural similarity index for IMRT quality assurance: Radiomics-based error classification," *Med. Phys.*, vol. 48, no. 1, pp. 80–93, 2021.
[22] J. J. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, 2017.
[23] C. J. Wolfs *et al.*, "External validation of a hidden markov model for gamma-based classification of anatomical changes in lung cancer patients using EPID dosimetry," *Med. Phys.*, vol. 47, no. 10, pp. 4675–4682, 2020.
[24] L. Zhang, Z. Yi, and J. Yu, "Multiperiodicity and attractivity of delayed recurrent neural networks with unsaturating piecewise linear transfer functions," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 158–167, Jan. 2008.
[25] L. Zhang, Z. Yi, S. L. Zhang, and P. A. Heng, "Activity invariant sets and exponentially stable attractors of linear threshold discrete-time recurrent neural networks," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1341–1347, Jun. 2009.
[26] L. Zhang, Z. Yi, and S.-i. Amari, "Theoretical study of oscillator neurons in recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5242–5248, Nov. 2018.
[27] L. Wang, L. Zhang, X. Qi, and Z. Yi, "Deep attention-based imbalanced image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3051721.
[28] L. Wang *et al.*, "A dual simple recurrent network model for chunking and abstract processes in sequence learning," *Front. Psychol.*, vol. 12, pp. 587405–587405, 2021.
[29] X. Shu, L. Zhang, Z. Wang, Q. Lv, and Z. Yi, "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2246–2255, Jun. 2020.
[30] Y. Feng, L. Zhang, and J. Mo, "Deep manifold preserving autoencoder for classifying breast cancer histopathological images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 1, pp. 91–101, Jan./Feb. 2020.
[31] L. Xie, L. Zhang, T. Hu, H. Huang, and Z. Yi, "Neural networks model based on an automated multi-scale method for mammogram classification," *Knowl.-Based Syst.*, vol. 208, 2020, Art. no. 106465.
[32] Z. Wang, L. Zhang, X. Shu, Q. Lv, and Z. Yi, "An end-to-end mammogram diagnosis: A new multi-instance and multi-scale method based on single-image feature," *IEEE Trans. Cogn. Devel. Syst.*, vol. 13, no. 3, pp. 535–545, Sep. 2021.
[33] T. Hu, L. Zhang, L. Xie, and Z. Yi, "A multi-instance networks with multiple views for classification of mammograms," *Neurocomputing*, vol. 443, pp. 320–328, 2021.
[34] L. Wang, L. Zhang, M. Zhu, X. Qi, and Z. Yi, "Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks," *Med. Image Anal.*, vol. 61, 2020, Art. no. 101665.
[35] J. Mo, L. Zhang, Y. Wang, and H. Huang, "Iterative 3D feature enhancement network for pancreas segmentation from CT images," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12535–12546, 2020.
[36] N. J. Potter, K. Mund, J. M. Andreozzi, J. G. Li, C. Liu, and G. Yan, "Error detection and classification in patient-specific IMRT QA with dual neural networks," *Med. Phys.*, vol. 47, no. 10, pp. 4711–4720, 2020.
[37] J. Peng *et al.*, "Implementation of the structural similarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection," *Med. Phys.*, vol. 47, no. 4, pp. 1907–1919, 2020.
[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[39] J. Deng, W. Dong, R. Socher, L. J. Li, and F. F. Li, "ImageNet: A large-scale hierarchical image database," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
[40] L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2625, 2008.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.
[43] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely connected convolutional networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, 2014, [Online]. Available: http://search.proquest.com/docview/2081521649/