IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 23, 2021, accepted January 31, 2021, date of publication February 2, 2021, date of current version February 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056641

A Comparative Study of Deep Learning Models for Diagnosing Glaucoma From Fundus Images

MANAL ALGHAMDI[1], (Member, IEEE), AND MOHAMED ABDEL-MOTTALEB[2], (Fellow, IEEE)
[1]Department of Computer Science, Umm Al-Qura University, Makkah 21421, Saudi Arabia
[2]Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA

Corresponding author: Manal Alghamdi (maalghamdi@uqu.edu.sa)

ABSTRACT Glaucoma is an eye disease that damages the optic nerve head (ONH) causing loss of vision. Therefore, early diagnosis and treatment are important in preventing possible blindness caused by glaucoma. Its current identification is based on the manual segmentation of the optic cup and disc to examine the cup-to-disc ratio (CDR). However, experts' annotation of these regions is a rather difficult and tedious task. Employing a convolutional neural network (CNN) for diagnosing glaucoma could be an alternative solution. However, its performance depends on the availability of a large number of labeled samples for the training phase. This paper presents an automatic glaucoma diagnosing framework based on three convolutional neural network (CNN) models with different learning methods, and compares the performance of these models with ophthalmologists. We use transfer and semi-supervised learning methods based on both labeled and unlabeled data. First, the transfer learning model starts with a pre-trained CNN model that was trained with non-medical data and fine-tunes it with our labeled data. Secondly, a semi-supervised framework is developed and trained using both labeled and unlabeled data based on two different unsupervised methods. The experimental results using two datasets, RIM-ONE and RIGA, demonstrate the efficacy of deep learning models when applied to glaucoma, which is a promising step towards automated screening for identifying individuals with early-stage glaucoma. Compared with annotations by two ophthalmologists, all the presented models achieve better performances, demonstrating the capability of artificial intelligence in diagnosing glaucoma with a high level of reliability.

INDEX TERMS Deep learning, semi-supervised learning, glaucoma, transfer learning, autoencoder.

I. INTRODUCTION

Glaucoma is one of the most widely recognized reasons of permanent blindness in the world. This eye disease is caused by changes in the structure of the retina, especially in the area of the optic nerve head (ONH) [1]. Signs of glaucoma may appear many years before patients recognize any changes in their visual field. Therefore, an automated screening system for diagnosing glaucoma is needed to facilitate the early diagnosis of the disease, as this is critical in preventing its progression. Nowadays, ophthalmologists around the world depend on the classic ophthalmoscope to visually assess the condition of the retina, which is time-consuming and barely reproducible. A fundus camera, which is faster, cheaper and generally more bearable for patients than other imaging modalities, is widely considered to be a viable option [2]. Modern fundus cameras with improved technology may produce high-quality images that contain useful information about the retinal structure.

The diagnosis of glaucoma could be achieved by investigating the retinal structures in fundus images; for instance, a pale OD which varies in color from orange to pink is a sign of the disease. These changes could only be observed by an expert examiner [3], however, it is a sensitive procedure and produces subjective judgments. One of the vital signs in fundus images used to diagnose glaucoma is the ratio of the size of the optic cup (OC) to the optic disc (OD), and the structure of these two areas. Therefore, previous research proposed to calculate parameters related to glaucoma, such as the cup-to-disc ratio (CDR), the inferior superior nasal temporal (ISNT) rule, the disc damage likelihood scale (DDLS) and the glaucoma risk index (GRI) [4]. However, the clinical evaluation involving manually labelling the cup and disc in each image is both labor-intense and

The associate editor coordinating the review of this manuscript and approving it for publication was Khin Wee Lai.

time-consuming. Access to screening services is becoming increasingly important, particularly with the increase of glaucoma patients. To facilitate access to screening, previous works presented automated computer systems to differentiate between the screened patients who should be examined by an ophthalmologist and those who should be asked for screening one year later. Such automated systems can help in reducing the workload for ophthalmologists while maintaining a high sensitivity of diagnosing glaucoma patients.

Developing automated diagnostic techniques that diagnose glaucoma is beneficial for the following reasons:

1) Manual examination of fundus images to diagnose glaucoma is a costly and time-consuming process that required extensively trained professionals.
2) Diagnosing glaucoma at its early stage requires frequent and regular visits to the ophthalmologists, which can be difficult in places that have shortage in ophthalmologists.
3) Assessing the effectiveness of treatment in preventing the disease progression is one of the key role of such automated diagnostic techniques.

Convolutional neural network (CNN) models have played a vital role in many computer vision applications including classification, detection and tracking. Such models can detect distinctive features from input images without pre-processing steps such as segmentation. However, developing these models and achieving the desired results requires large annotated training sets. Instead of the intensive and expensive training process, the pre-trained CNN models were used to solve different tasks from those that they were originally trained for [5]. These pre-trained models are usually developed using a very large dataset for a specific task, and then transferred to be used as a feature extractor for the problem at hand. The extracted features are then used to train another classifier, with a small dataset, for the new tasks. With limited amounts of labeled data, semi-supervised approaches tackle this issue by expanding the labeled data with a large number of unlabeled data to improve the performance in various medical imaging tasks, such as brain image segmentation [6], skin cancer diagnosis [7], dental X-ray image segmentation [8] and retina vessel segmentation [9]. With a limited amount of labeled data, the performance can be improved by using the unlabeled data to learn the feature representations. However, the previous semi-supervised approaches for glaucoma analysis focused on handcrafted features. Thus, the unlabeled data was not used for learning the features.

A. RELATED WORKS

Previous supervised deep learning methods have been successfully presented in ophthalmology [10]–[13]. However, most of these works trained their models using large but private datasets. Fu *et al.* [14] presented a disc-aware ensemble network based on an ensemble of four independent networks whose predictions were fused to obtain the final decision. Haleem *et al.* [15] proposed an approach for automatic glaucoma detection based on Regional Image Features Model (RIFM) that extract both geometric (*e.g.* , morphometric properties) and non-geometric based properties (*e.g.* , pixel appearance/intensity values, texture) from images. These features were then classified using a support vector machine (SVM) classifier. Chai *et al.* [16] developed a two-fold CNN for glaucoma detection. They fed the entire image to the first CNN to segment the OD that was fed to the second network. A concatenation of the CNN models was used followed by a fully-connected layer for the classification. Pal *et al.* [17] presented a multi-model network consisting of an autoencoder and a CNN classifier that shared the encoder framework. Zhao *et al.* [18] performed an optic disc segmentation using a weakly-supervised multi-task Learning model. Li *et al.* [19] detected glaucoma using a CNN with attention mechanism, which forces the network to pay more attention to a specific region of the image. The attention mechanism achieved good results but required extensive labeling by ophthalmologists to indicate where the network should focus when reading the images. Singh *et al.* [20] diagnoses the glaucoma by training multiple classifiers, like support vector machine (SVM), K-nearest neighbors (KNN), and Naive Bayes, using various features including inferior, superior, nasal, and temporal region area, and cup-to-disc ratio. They achieved good results but extracting all these features is a time and effort-consuming.

The lack of annotated data encourages approaches that extend the traditional supervised learning by using unlabeled data that might be more available. In retinal imaging, few works have followed this path by exploring semi-supervised learning [21] and transfer learning [22]. Transfer learning can be performed using one of two methods [5]. The first method involves using the new images as an input to the pre-trained network and using the outputs before the fully connected layers as feature vectors to train a specific classifier for the new domain. The second method performs backpropagation to fine-tune the weights of the pre-trained model. Either all the weights are fine-tuned or the weights of some of the early layers are frozen and the weights of the later layers are adjusted. This is based on the assumption that the earlier layers of the network consist of generic features that can be used by multiple applications (*e.g.* , edges or texture detectors), while the later layers contain more specific details of the classes contained in the training dataset [5].

Different eye diseases other than glaucoma, *e.g.* , diabetic retinopathy (DR), have been well-studied using fundus images and CNN with transfer learning [23], [24]. This is because the DR has a large annotated dataset available for researchers, which is not the case for the other diseases. Some of the research that examines glaucoma using CNNs on private datasets includes Chen *et al.* [25] and Chai *et al.* [16]. Chen *et al.* [25] extracted the features from pre-processed images and trained a CNN model on a large private dataset for glaucoma detection. In contrast, Chai *et al.* [16] developed a two-branch CNN model to analyze the whole image in one branch and the OD region in the

other branch. Their dataset contains around 3, 554 images. Perdomo *et al.* [26] presented a three-stage deep learning model for glaucoma detection using the publicly available RIM-ONE dataset. During the first stage, a segmentation of the OD and PC is performed using a deep CNN. The second stage used another CNN to predict multiple features including geometric, distance, axis and ratio features. Finally, the third stage applies a multilayer neural network to detect the glaucoma in the test images. Yu *et al.* [27] extended the U-Net network with the pre-trained ResNet-34 model used in the encoder path for OC and OD segmentation. More recently, Hemelings *et al.* [28] combined the ResNet-50 with the active learning technique to diagnose glaucoma from a collected private dataset with 2072 images. Gour and Khanna [29] proposed a dual CNN model based on the VGG-16 for classifying fundus images into eight ocular disease categories. The first CNN process the left eye, while the second CNN process the right eye of the patient. Features from both eyes are then concatenated for classification.

Instead of performing an expensive and dedicated training, researchers have used CNNs that were pre-trained using large datasets for specific problems and applied transfer learning to extract discriminative features in different domains. In the literature, transfer learning is used for glaucoma detection by [5], [30]–[33]. Orlando *et al.* [5] analyzed the effect of using different types of image pre-processing methods on the performance of the pre-trained CNNs for glaucoma detection. Two off-the-shelf CNN models known as OverFeat and VGG-S were employed as feature extractors followed by a regularized logistic regression model. Two small public datasets were used for experiments, Drishti-GS and DRIVE, which were manually annotated for detecting glaucoma. Al-Bander *et al.* [30] proposed a framework for glaucoma detection from fundus images using the publicly available dataset RIM-ONE. A pre-trained CNN model (AlexNet) was used as a feature extractor. These features were extracted from the last layer located before the output layer and then used to train SVM classifier. Their performance was affected by their use of the AlexNet model which was reported to be the least efficient model in the ILSVRC competition among the other available models [34]. Furthermore, they did not use any data augmentation techniques and did not perform fine-tuning for the AlexNet. Cerentini *et al.* [31] used GoogleNet, a pre-trained model, in two steps for feature extraction and image classification. In the first step, GoogleNet was combined with a sliding-window approach to detect the region of interest (ROI). The second step used the extracted ROI to train another GoogleNet for detecting glaucoma. They evaluated the performance of their approach on the publicly available, RIM-ONE, dataset. Their model uses only the ROI which is not the only important area in the retina to detect glaucoma. Gomez-Valverde *et al.* [35] compared the perofrmces of multiple pre-trained models for glaucoma detection including: VGG-19, RerNet, GoogLeNet and DeNet. The best performance was achieved by the VGG-19 on 2313 images collected from Drishti-GS, RIM-ONE and a privet dataset. Maheshwari *et al.* [33] diagnoses the glaucoma by re-training the Alexnet with features extracted using the local binary pattern (LBP) technique on the red, green and blue channels of the fundus images.

Recently, Jammal *et al.* [32] compared the grading performance of human vs ResNet model, which was further tuned with a pairs of fundus images and OCT scans to detect the retinal nerve fiber layer defect as a sign of glaucoma. Sreng *et al.* [36] presented a two-stage glaucoma screening system, of which the first stage segmented the optic disc region, while the second one classify the images using pre-trained models. The above models boosted their accuracy by combining the CNN with other steps such as segmentation, pre-processing, logistic regression or SVM. None of these models presented a complete CNN model to extract the underlying features of the fundus images, train the classifier and predict the likelihood of glaucoma from the test images. Moreover, none of these models performed fine-tuning for the pre-trained models.

Other researches [37], [38] presented promising results for glaucoma analysis by adopting a large number of unlabeled retinal images, which are more readily available than the labeled ones. Sedai *et al.* [37] developed a semi-supervised method to segment the OC with two autoencoders. The first autoencoder is used to learn the features from the unlabeled images. The second autoencoder uses the features learned by the first autoencoder to detect glaucoma. Their model was evaluated on a large private dataset. However, they only consider the OC area to diagnose the glaucoma while there are many signs of the disease that usually appear in different parts of the retina. Bechar *et al.* [38] presented a semi-supervised superpixel-by-superpixel OD and OC segmentation framework using three steps. The first step involved pre-processing the labeled and unlabeled data by applying the superpixel method followed by manually annotating the regions at these pixels. In the second step, features were extracted from the annotated regions. In the final step, a semi-supervised model was trained using a small number of labeled superpixels and a large number of unlabeled superpixels. The experiment was performed on the publicly available RIM-ONE dataset. These two methods [37], [38] depend on segmentation for their glaucoma detection, therefore, the performance is affected by the accuracy of the segmentation. In addition, they ignore the other important parts in the retina that usually hold signs of the glaucoma.

B. DIAGNOSING GLAUCOMA USING DEEP LEARNING

Majority of the recent methods in diagnosing glaucoma start with the segmentation of the region of interest, which can be examined later to identify the disease. Therefore, manual annotations and pre-processing steps are key parts of these methods. In addition, none of the previously developed deep learning models used the large number of unlabeled retinal images, which are more readily available than the labeled ones. This paper focuses on using deep convolutional neural networks (CNNs) for automated diagnosing glaucoma

from retinal images. These algorithms would be valuable in clinical evaluation of new treatments for glaucoma. Also, these algorithms can be used to automatically and remotely observe the progress of glaucoma in patients. Employing automated systems in clinical practice require extensive and regular evaluation [3]. One of our goals is to demonstrate that automated systems outperforms human experts in diagnosing glaucoma, which can help reduce the workload on ophthalmologists. Evaluation of these systems should be conducted using independent and publicly available datasets so that different research groups can compare the performance of their algorithms using the same dataset. The performance of human experts on the same dataset should also be available to evaluate the automated systems' performance against humans.

This paper is an extended version of our earlier work published in [39], which presented a semi-supervised learning model for diagnosing glaucoma from fundus images. In this paper, we present the performances of three different comprehensive automated glaucoma classifying systems and evaluate their performances on the multiple public datasets and compare their performances against two human experts. Ophthalmologists usually use other information in addition to fundus images when they make their decision, but in this work they were only provided with the fundus images to make a fair comparison with the automated approach. These systems are based on different deep learning techniques including supervised and semi-supervised training:

- **TCNN: Transfer Convolutional Neural Network model** which learns the signs of glaucoma by using an off-the-shelf CNN and tunes it in the context of diagnosing glaucoma with relatively a small size dataset.
- **SSCNN: Semi-supervised Convolutional Neural Network model with self-learning** which trains a neural network in a semi-supervised fashion in two stages [39]. The first stage uses the TCNN model as a glaucoma diagnosing classifier. The second stage is a self-learning method that is used to increase the training set with samples from the unlabeled dataset, and hence improve the performance of the pre-trained CNN. For each unlabeled sample, the self-learning method generates what is known as pseudo-label by selecting the class label with the highest prediction probability and considers it as the true label.
- **SSCNN-DAE: Semi-supervised Convolutional Neural Network model with autoencoder** presents a glaucoma classifier based on a convolutional denoising autoencoder (CNN-DAE) to extract discriminative features and then use these features to train fully connected layers in a supervised fashion. The CNN-DAE is used to learn the discriminative features in fundus images from the unlabeled data. Once the model is robust enough to represent the fundus images, the training data from the labeled set is then used to obtain the feature maps that are used later to train fully connected layers in a supervised fashion. The resulting classifier is then evaluated using the testing dataset.

The contributions of our paper are three-fold:

- We presented three convolutional neural networks for diagnosing glaucoma from fundus images using three learning methods (*i.e.*, supervised, transfer, and semi-supervised) chosen based on the availability of annotated data.
- We conducted several experiments using two public datasets to evaluate the performance of the presented models. Compared with the performance of two human experts, the presented model confirms the deep learning ability to perform better than human experts for this task.
- We make our three models publicly available[1] so that other researchers and practitioners in the community can easily build high performance disease diagnosing models that can be further trained on their datasets using our structure.

The remainder of the paper is organized as follows: the proposed CNN models are introduced in Section II. Datasets and the experimental setup are presented in Section III, followed by results and discussion. Finally Section IV concludes the paper.

II. CONVOLUTIONAL NEURAL NETWORK (CNN) MODELS

In this paper, the proposed models are based on CNNs. All the developed CNN models contain convolutional layers followed by fully connected layers. The output of the last fully connected layers is used as input to a softmax classifier for diagnosing glaucoma. The first model presented in Section II-A utilizes transfer learning which involves using a pre-trained CNN along with a limited-size labeled glaucoma dataset. The second model presented in Section II-B combines transfer learning (Section II-A) with a self-learning technique to perform semi-supervised learning. The final model in Section II-C combines a convolutional autoencoder (unsupervised learning) with a supervised learning stage to perform semi-supervised learning.

To make the paper self-contained, we start by defining the symbols used in the following sections. The training dataset X contains two parts: $X = [L, U] \in R^{d\times N}$, where $L = [x_1, x_2, \ldots, x_l] \in R^{d\times l}$ refers to the labeled samples and $U = [x_{l+1}, x_{l+2}, \ldots, x_{l+u}] \in R^{d\times u}$ refers to the unlabeled samples.

A. TCNN: TRANSFER CONVOLUTIONAL NEURAL NETWORK MODEL

We employ the VGG-16 model [40], which has been widely adopted for transfer learning in medical image processing tasks, including glaucoma diagnosis, and reported a good performance in previous studies [29], [41]–[43]. It consists of 16 layers where 13 convolution layers of size 3×3 are stacked followed by five 2×2 max-pooling layers with stride

[1]Codes are available here: https://github.com/manal-asdg/CNNs_medical.git

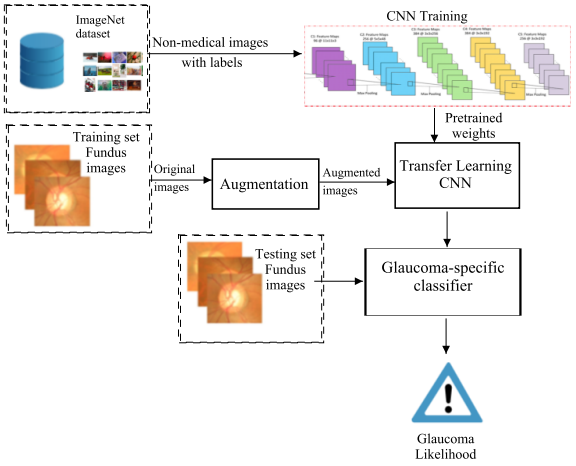

FIGURE 1. The structure of the transfer CNN (TCNN) model for diagnosing glaucoma. It consists of two stages: feature extraction using VGG-16 [40] and fine-tuning using the fundus images.

of 2 to reduce the spatial dimensions. After each convolution layer, a rectified linear unit (ReLU) activation function is applied. Then, three fully connected layers are used where the final layer has a softmax activation to produce the probabilities for each class. The weights of this model are obtained and used as input to our network, therefore, the full model (except the fully connected layers) is used to extract discriminative features.

Two stages of training are performed, as shown in Figure 1. The first stage extracted the features from the layer immediately before the output layer. Thus, the top fully connected layers and softmax layer were removed and new layers were added while the remainder of the layers were frozen to stop the weights from changing. The new layers consist of a flat layer, two batch normalization layers and a softmax layer to convert the output into two categories. For optimization, the Adam method [44] was used, which has the ability to perform well even with minimal tuning [45].

The second training stage performs fine-tuning by unfreezing and re-training the last convolutional block of the VGG-16 repeatedly to make it more specific to our dataset. The Adam optimizer was replaced at this stage by the Stochastic Gradient Descent (SGD) that empirically achieves better performance than Adam at this stage [45], [46]. This takes advantage of the fast convergence of Adam at the beginning of the training and later makes the model generalize better by using SGD. The learning was performed using the cross-entropy loss [47] on the softmax normalization score defined as:

$$J = -\frac{1}{L}\left(\sum_{i=1}^{L} y_i \cdot \log\left(\hat{y}_i\right)\right) \tag{1}$$

where y_i is the ground-truth for sample x_i, $\hat{y}_i$ is the estimated value and L is the size of the labeled samples.

In this model, data augmentation was used to overcome the overfitting by artificially enlarging the set of training samples with class-preserving transformations. This helps 'produce'

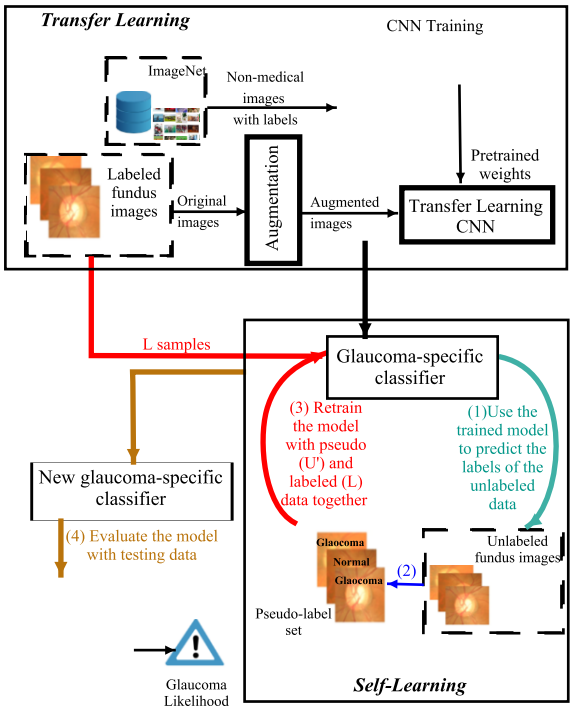

FIGURE 2. The structure of the semi-supervised transfer learning CNN (SSCNN) model for diagnosing glaucoma. It starts by using the TCNN for feature extraction followed by the self-learning method to increase the training set using the unlabeled data.

more training data, prevents overfitting and accounts for obvious invariances for the classification task at hand (*e.g.* , flipping an image should not affect the existence of the disease) [48]. In our experiment, four types of data augmentation were employed: vertical shift, which randomly shifts the images up and down using the nearest pixel to fill the blank; horizontal shift, which randomly shifts the images left and right using the nearest pixel to fill the blank; horizontal flip, which randomly flips the image horizontally; and random zoom-in of the image. These affine transformations are the most effective when applied to fundus images since they do not affect the signs of the disease in the image [25]. In addition, they help the model to have a better representation of the input image since it views the images in many transformed views including different scaling.

B. SSCNN: SEMI-SUPERVISED CONVOLUTIONAL NEURAL NETWORK MODEL BASED ON SELF-LEARNING

We combined the TCNN (Section II-A) with the self-learning to perform semi-supervised learning for diagnosing glaucoma [39]. The training process of the proposed method consisted of two stages, as shown in Figure 2. The first stage transfers and fine-tunes the pre-trained CNN model to learn the discriminative features of retinal images. The second stage extends the updated CNN model with the self-learning method to increase the training set by selecting the most reliable samples from the unlabeled data.

To improve the robustness of the classifier, especially with a limited-size labeled samples, it is necessary to expand the

initial training set [49]. Thus, the self-learning method [50] is employed to expand the initial training dataset L with samples from the unlabeled set that have high confidence when classified with the model presented in Section II-A. The obtained labels of these samples are called pseudo-labels. Adding them to the training dataset is performed in an iterative manner by modifying the classifier after each iteration and labeling the samples in the remaining unlabeled set. Given two datasets *i.e.*, labeled data L and an unlabeled data U, the self-learning process can be summarized as follows:

1) The unlabeled dataset U is firstly classified using the pre-trained CNN model from Section II-A.
2) The prediction scores from step (1) are ordered and the samples with the highest confidence scores $U' \subset U$, are added to L and dropped from U. As mentioned above, the labels for the samples in U' are called pseudo-labels.
3) The classifier is then re-trained by the new training set $L + U'$.

These steps are repeated until the algorithm converges. The unlabeled samples that remain unclassified at the end of the learning process are discarded because they are not reliable.

To avoid overfitting, increasing the size of the training set by samples U' was performed with a normal to glaucoma ratio which is the same as the distribution of labeled data [51].

The loss function in Eq 1, from the supervised learning, is also used for the self-learning in this model. Finally, the new glaucoma-specific classifier is evaluated with the testing dataset.

C. SSCNN-DAE: SEMI-SUPERVISED CONVOLUTIONAL NEURAL NETWORK MODEL BASED ON DENOISING AUTOENCODER

We combined both unsupervised (II-C1) and supervised (II-C2) learning to perform semi-supervised learning for diagnosing glaucoma. In both learning stages, we utilize convolutional layers instead of the fully connected layers because they provide stable latent representations at each network level, which preserves the localization. The semi-supervised learning consists of two stages as shown in Figure 3, and can be summarized as follows:

1) Unsupervised learning is performed using a CNN denoising autoencoder (CNN-DAE) to learn the representative features from the unlabeled data.
2) The encoder part of the CNN-DAE is used to extract features to perform supervised learning. The features are used to train fully connected layers and build the glaucoma-specific classifier.

The glaucoma-specific classifier is then evaluated using the testing data from the labeled dataset. More details about each stage are presented in the following sections.

1) UNSUPERVISED LEARNING

For the first stage of learning, we extended the standard denoising autoencoder (DAE) with convolutional encoding and decoding layers that share the weights between all input locations and discover strong spatial correlations [52]. Previous studies have shown that this extension improve the DAE performance in exploiting the image structures and detecting discriminative features [53], [54].

To make the paper self-contained, we begin by reviewing the classic autoencoders (AE), followed by the denoising autoencoder (DAE), the CNN autoencoder (CNN-AE) and finally the proposed convolutional denoising autoencoder (CNN-DAE).

AUTOENCODERS (AE)

are used to learn functions that can be used to reconstruct the inputs. These models are developed as a neural network with a single hidden layer that contains activations extracted as the new representation [55]. Mathematically, for a collection of unlabeled data $U = x_i,\ i \in [l+1, l+u]$, where l is the number of the labeled samples and u is the number of the unlabeled ones, the objective function of AE is defined as:

$$\begin{aligned} &min \sum_i D\left(\hat{x}_i, x_i\right) \\ &s.t.\ h_i = g\left(Wx_i + b\right), \\ &\quad \hat{x}_i = f\left(W'h_i + b'\right) \end{aligned} \tag{2}$$

where W, b, W', b' are the parameters to be optimized, D is a loss function, which can be a squared error or cross-entropy, g and f are the encoder and decoder nonlinear functions, h_i is the learned representation; and $\hat{x}_i$ is the reconstruction.

DENOISING AUTOENCODER (DAE)

extends the classic autoencoder by forcing the network to reconstruct the input given its noisy version [56]. The idea is that a robust network should be able to learn reconstructions of the input even with noises, which are caused by the inter-correlation between the features. Formally, given a predefined noise distribution $q(\bar{x}|x)$, and a noisy version, $\bar{x}$, of input x : $\bar{x} \sim q(\bar{x}|x)$ [55]. The objective function sums the expectations over all the noisy versions:

$$\begin{aligned} &min \sum_i E_{q(\bar{x}_i|x_i)} D\left(\hat{x}_i, x_i\right) \\ &s.t.\ h_i = g\left(W\bar{x}_i + b\right), \\ &\quad \hat{x}_i = f\left(W'h_i + b'\right) \end{aligned} \tag{3}$$

where $\hat{x}_i$ denotes the reconstruction calculated from the noisy input $\bar{x}_i$.

CNN AUTOENCODER (CNN-AE)

extends the standard DAE architecture with convolutional encoding and decoding layers. Although the architectures of both the CNN-AE and DAE are similar, the CNN-AE is more efficient because it allows the weights to be shared and it preserves the locality of the 2D image.

CNN DENOISING AUTOENCODER (CNN-DAE)

In this model, we extended the DAE as a CNN denoising autoencoder (CNN-DAE) for diagnosing glaucoma.

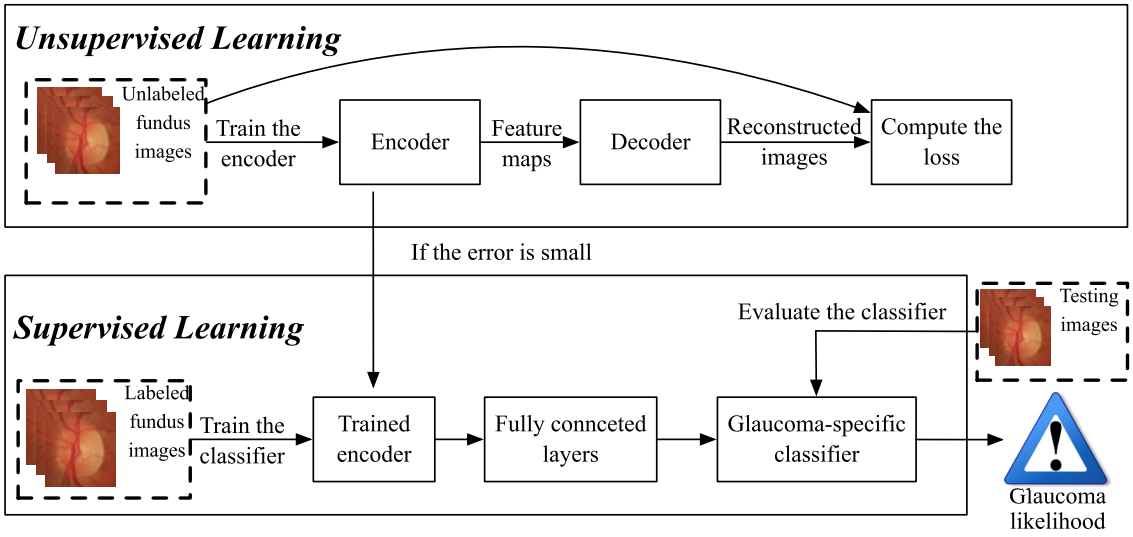

FIGURE 3. The structure of the semi-supervised learning CNN (SSCNN-DAE) model for diagnosing glaucoma. Two stages are performed: Unsupervised learning for feature extraction from the unlabeled data, followed by supervised learning to classify the labeled data.

It extends the standard DAE with convolutional encoding and decoding layers [57]. Compared with the classical DAE, the CNN-DAE is better suited to image processing. In addition, it improves on the CNN-AE by distorting the input image using noise, which makes it difficult to reconstruct the input. For that, the autoencoder will be forced to learn the representative features of the input image rather than just copying it.

The SSCNN-DAE model is built with a 6-layer CNN divided between the encoder and the decoder paths as illustrated in Figure 4. The encoder path contains three CNN layers with noise injected into each layer including the input layer, and the decoder contains three deconvolutional layers. Similar to [58], our model replaces the max-pooling layers in the encoder path with strided convolutional layers to downsample the feature space. Alternating 2×2 strided convolutional layers instead of 1×1 was enough to reduce the features space from preceding layers. The decoder path mirrors the encoder path; therefore, it contains three CNN layers with no upsampling layers. Constructing the original image at the decoder path was achieved using transposed convolution, which perform upsampling by swapping the forward and backward operations of the convolution [59]. In all layers, the rectified linear unit (ReLU) was applied as an activation function. The Adam method [44] was used as our optimizer to control the learning process [45].

Formally, let $q(\bar{x}|x)$ be a Gaussian noise with normal distribution, and $\bar{x}$ be a noisy version of the unlabeled sample $x : \bar{x} \sim q(\bar{x}|x)$. The objective function calculated as the sum of predictions over all the noisy versions:

$$\begin{aligned} &min \sum_i E_{q(\bar{x}_i|x_i)} D\left(\hat{x}_i, x_i\right) \\ &s.t.\ h_i = g\left(W * \bar{x}_i + b\right), \\ &\quad \hat{x}_i = f\left(W' * h_i + b'\right) \end{aligned} \tag{4}$$

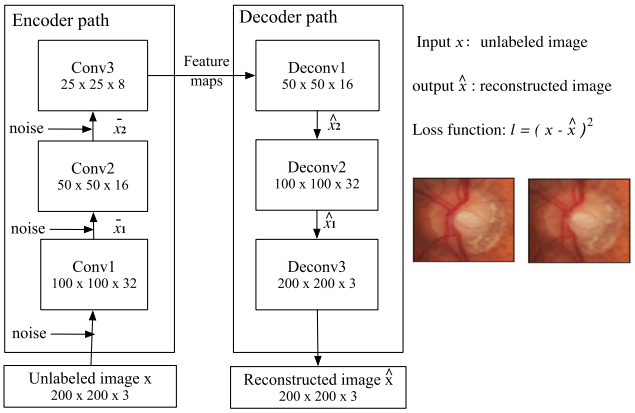

FIGURE 4. The structure of the 6-layer CNN denoising autoencoder (CNN-DAE) model, which is the first part of the the Semi-supervised Convolutional Neural Network model with autoencoder (SSCNN-DAE).

where $\hat{x}_i$ is the reconstructed version of the noisy input $\bar{x}_i$ and $*$ is the convolution operation.

The Gaussian noise is first injected into the unlabeled data to produce the noisy version which is then transformed to a more abstract representation, $\bar{x}_i$, through the encoder. After each decoding path, the cost is computed as the averaged square error between the reconstructed output $\hat{x}_i$ and the input x_i. The coefficients of the CNN-DAE are iteratively updated until the reconstructed image retains the most important features of the original image. The resulting encoder is used to generate feature maps that are used to define a classifier for the supervised learning in Section II-C2.

2) SUPERVISED LEARNING

Once the CNN-DAE is robust enough and the reconstruction error is as small as possible, the labeled data is used to generate the feature maps. At this stage, the labeled data is divided into a training set and a validation set. The training set is fed

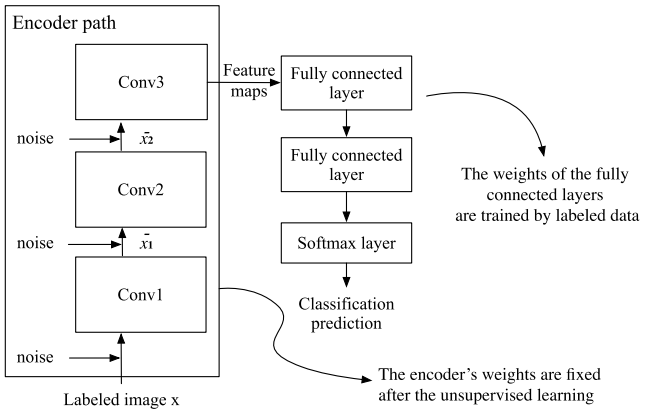

FIGURE 5. Structure of the CNN supervised learning, which is the second part of the the Semi-supervised Convolutional Neural Network model with autoencoder (SSCNN-DAE).

into the encoder path to generate the feature maps, which are then used to train fully connected layers and obtain a glaucoma-specific classifier. The predictions were performed using a softmax layer based on the same loss function defined in Eq 1. Only two fully connected layers were used, to avoid overfitting, followed by a softmax layer that maps the output into two categories. The structure of the supervised learning stage is presented in Figure 5.

III. EXPERIMENTS

To evaluate the glaucoma diagnosing performance of our proposed CNN models, we conducted experiments using two publicly available datasets as follow:

- RIM-ONE [60], which was utilized to fine-tune and evaluate TCNN model as well as to evaluate the SSCNN and SSCNN-DAE models. It contains 455 high-resolution retinal images where 255 images belong to normal individuals and 200 images belong to glaucoma patients.
- RIGA [61], which has been used as the unlabeled dataset in the SSCNN and the SSCNN-DAE models. This dataset consists of 750 retinal fundus images collected from three resources as follows:
  1) A subset of the MESSIDOR [62] dataset that consists of 460 retinal images with a resolution of $2,240 \times 1,488$ pixels.
  2) The Bin Rushd Ophthalmic Center dataset that contains 195 retinal images with a resolution of $2,376 \times 1,584$ pixels.
  3) The Magrabi Eye Center dataset that contains 95 retinal images with a resolution of $2,743 \times 1,936$ pixels.

  The RIGA dataset was released by the University of Michigan to evaluate their segmentation method for a glaucoma diagnosing application. Almazroa *et al.* [61] stated in their paper that the images belong to both healthy and glaucoma eyes, however, the ground truth was not provided. The images were cropped to be consistent with the labeled training samples.

A. EVALUATION CRITERIA

We used the classification accuracy to evaluate the performance of diagnosing glaucoma. The evaluation was based on 20% of the 455 labeled images. The achieved accuracy at each epoch is plotted as a curve which is computed from sensitivity (true positive rate) and specificity (true negative rate), defined as:

$$Accuracy = \frac{T_p + T_N}{T_P + F_p + F_N + T_N} \tag{5}$$

$$Sensitivity = \frac{T_P}{T_P + F_N} \tag{6}$$

$$Specificity = \frac{T_N}{T_N + F_P} \tag{7}$$

where T_P, T_N, F_P and F_N are the number of true positives, true negatives, false positives and false negatives, respectively. The area under the curve (AUC) of the sensitivity and specificity is also considered, because it is widely used for binary classification especially for medical images. It is also known as the receiver operation characteristic (ROC) curve [63]. Sensitivity means the ability of the method to correctly classify the image, while specificity is the inability of the method to classify the image correctly.

The models' performances were further assessed by comparing them with results obtained by two physicians. Specifically, two ophthalmologists independently classified the RIM-ONE 455 fundus images and their results were compared with the accuracies of the three models. For each image, the ophthalmologists were asked whether the presented image belonged to glaucomatous or to healthy eyes.

B. EXPERIMENTAL SETUP

For implementation, an NVIDIA GTX 1080TI 11GB GPU card with a 3584 CUDA parallel-processing core was used. All models were developed using Keras API [64] which was built on the top of the TensorFlow 2.0 [65].

For the transfer learning setup (Section II-A), the weights of VGG-16 were obtained from the Keras GitHub.[2] The first stage of the training includes 20 iterations to train the fully connected layers and the softmax layer. For the Adam optimizer used the learning rate was set to 0.001. The second stage of the training includes 50 iterations which are higher than the initial training to allow the top convolutional layers to detect more discriminative features. Performing fine-tuning needs a steady training process with smaller learning rate compared to the starting rate used in the first stage, otherwise the extracted features at the first stage will be destroyed and the optimization will destabilize [45]. Thus, the optimizer at this stage was chosen to be SGD and with a small learning rate set to 0.0001.

For the self-learning setup (Section II-B), the last learning rate and optimizer method from the TCNN model were used. The subset U', which represents the pseudo-labels, contains A samples with the highest scores as normal and B samples with

[2]https://github.com/fchollet

the highest scores as glaucoma. The ratio of $A : B$ is chosen to be $1 : 1.3$, which is the same as the ratio for the labeled data. The initial sizes of A and B were chosen to be 30 and 39 samples, respectively. With more iterations, the classifier performance improves, thus, these two sizes were increased during the learning process iteratively by $30 + 3 \times i$, where the number 30 is the initial size of A and i is the number of iterations. The number of epochs was firstly selected to be 12 to be suitable for the limited-size labeled dataset and then increased in an iterative manner by $epoch + 4 \times i$ [49], where i represents the iterations number.

In the last CNN model, all the unlabeled samples were used to train the autoencoder. The images were resized to 200×200. The strides and the size of the convolutional kernel in each layer were set to 2 and 3×3, respectively. The number of convolutional kernels in the 1st, 2nd and 3rd convolutional layers were 32, 16 and 8, respectively. The $l2$ regularizer was used in the training process. The learning rate for the Adam optimizer was set to 0.0001, the batch size to 40 and the epoch to 1600. The injected noise into the input to the encoder was chosen to be Gaussian noise with $\mu = 0$ and $\sigma = 1$. In its supervised learning stage, the labeled images were divided into 80% for training and 20% for validation. All the labeled images were fed into the encoder path of the CNN-DAE to extract the feature maps that were later used to train the fully connected layers in a supervised fashion. Two fully connected layers were used with 512 nodes at the first layer and two nodes representing the output at the second layer. The same regularizer, optimizer and learning rate that were used for the unsupervised learning were used. The batch size was set to 40 and the epoch to 100.

C. EXPERIMENTAL RESULTS

We performed various types of comparisons to evaluate the presented models. The first comparison in Section III-C1 was performed to compare the effectiveness of the proposed models using RIM-ONE dataset with two experts. The second comparison in Section III-C2 was performed to compare the presented performances with the related methods that were reviewed in Section I-A. The purpose of this comparison is to position our results with respect to the recent state-of-the-art techniques in the context of glaucoma diagnosing using deep learning. The third comparison in Section III-C3 aims to demonstrate the effectiveness of the self-learning method in the SSCNN, by comparing it with another method. The last comparison in Section III-C4 was performed to compare the effectiveness of using the CNN-DAE unsupervised learning in the SSCNN-DAE with that when using different types of autoencoder architectures.

1) THE PERFORMANCES OF THE MODELS

After fine-tuning, training the TCNN model and also expanding the initial labeled training set using the SSCNN and SSCNN-DAE, all three models were evaluated using the same testing set. The results are presented in Figure 6. It is clear that the semi-supervised learning models performed

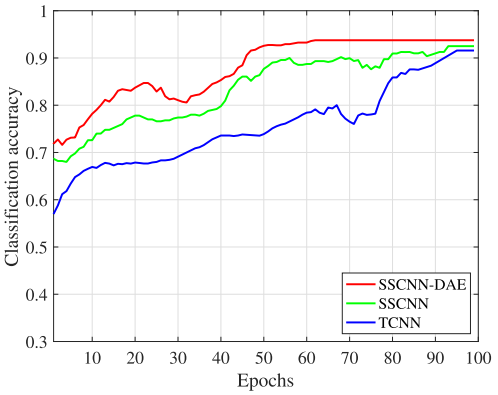

FIGURE 6. Comparing the classification accuracy of the presented deep learning models, where the semi-supervised learning models performed better than the transfer learning one.

better than the transfer learning one. The SSCNN and SSCNN-DAE models improved their classification accuracies by expanding their knowledge about the glaucoma by utilizing the information contained in the unlabeled samples. The best performance was achieved by the SSCNN-DAE model. This is because the CNN-DAE, with its added noise and convolutional layers, was more able to learn the discriminating features of the glaucoma from the unlabeled samples and produce a more accurate classifier than the other two models. The lowest performance was achieved by the TCNN model, because the transfer learning and the fine-tuning of the pre-trained network layers utilized a small set of labeled samples. Note that the curves for the three models have slight fluctuations, but they are nevertheless within a reasonable range. Almost all neural networks are trained with different forms of stochastic gradient decent and batch sizes, which determines how many samples are randomly used to make one update to the model parameters [66]. If we use all the samples for each update, we should see more stable curves.

Another performance measurement, the ROC, is presented in Figure 7. According to [67], an AUC value above 90% indicates an excellent diagnostic test, from 80% to 90% a good diagnostic test, from 70% to 80% a fair diagnostic test, from 60% to 70% a poor diagnostic test, and below 60% indicates a failed diagnostic test. On this basis, all presented models including; TCNN with AUC = 92.3%, SSCNN with AUC = 93.6% and SSCNN-DAE with AUC = 95% are effective in glaucoma diagnostic task.

We also compared the performance of the CNN models and that of two certified ophthalmologists in terms of glaucoma classification. Both ophthalmologists have been in practice for 17-20 years with experience in diagnosing and treating eye diseases, including vision loss, detached retinas, cataracts and glaucoma. The TCNN achieves 91.5%, the SSCNN achieves 92.4% and the SSCNN-DAE achieves 93.8% overall accuracy, while the two experts attain 59.2% and 55.4.0% accuracy. It is evident that the three deep learning CNN models outperform the performances of both ophthalmologists with clear margins. The margins were even wider for the semi-supervised learning models where the unlabeled

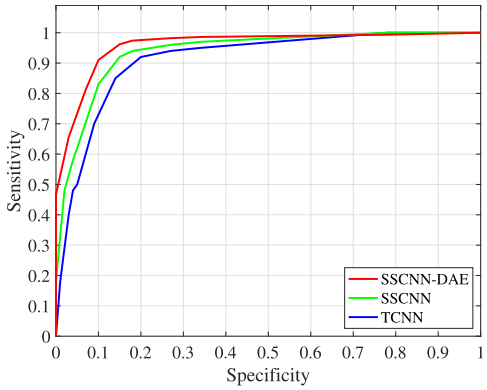

FIGURE 7. Comparing the ROC curves of the presented deep learning models with AUC=92.3% for TCNN, AUC=93.6% for SSCNN and AUC=95% for the SSCNN-DAE.

samples were used to automatically learn the retinal structures from the fundus images without any segmentations or pre-processing. Instead of the tedious process used by experts in a clinic to diagnose glaucoma, the CNN models were better able to capture the characteristics of glaucoma efficiently. This is an extremely promising outcome for the deployment of such models in clinical environments and to assess their use in the early diagnosing of glaucoma.

The obtained results are promising since the proposed models do not use any prior knowledge about the retinal image features such as OD or blood vessels structure. In addition, they are computationally simple since extracting the features from the region of interest is done automatically by the CNN from the images themselves without requiring any manual feature extraction. We evaluated the running time for all three models (*i.e.*TCNN, SSCNN and SSCNN-DAE), which include the training time per batch on a single NVIDIA GPU with 11GB memory. Note that, the testing time is around 1/3 of the training time [68]. All layers, consisting of convolution, pooling, fully connected and others, are included in the actual running time. The TCNN takes about 0.52 second for training per batch, with 1.5 days training in total. While, the SSCNN-DAE takes around 0.61 second for training per batch with 2 days training in total. Gradually increasing the training set in the SSCNN using the self-learning method increases the time to around 0.68 second for training per batch, with 2.5 days training in total. The achieved results, within these training times, validate the effectiveness of the presented models which can be integrated into a computer-aided diagnosis system and large scale screening of retinal images. Such a system can also be used to monitor the progression of disease under therapy in longitudinal studies.

2) COMPARISON WITH RELATED METHODS

An extensive comparison with the previous related methods is not practical because few studies have used the semi-supervised learning concept for diagnosing glaucoma, and also because most of the existing studies used their own private datasets. Therefore, we performed the following comparisons:

- To evaluate the performance of our two semi-supervised models against the related works, we compared with the only study that we could find by Bechar *et al.* [38] that used a semi-supervised learning model and evaluated it using RIM-ONE.
- To evaluate the performance of our transfer model against the related works, we compared it with the two studies by Al-Bander *et al.* [30] and Cerentini *et al.* [31] which were evaluated on RIM-ONE using transfer learning CNN.

We also compare our results with the models presented by Perdomo *et al.* [26] which use a supervised multi-stage CNN to diagnose glaucoma in the RIM-ONE dataset.

The results are presented in Table 1 and illustrate that our three models outperformed the other approaches by clear margins. The results reveal that the SSCNN-DAE model achieves the best performance. This was due to its ability to leverage the information from unlabeled samples by first learning the discriminative features using the CNN-DAE and then training simple fully connected layers with the generated features to find the disease. The SSCNN achieved the second best performance after the SSCNN-DAE. This was due to increasing the training set in an iterative fashion which causes the accuracy of the CNN model to increase gradually. Both our semi-supervised models outperformed the semi-supervised one presented by Bechar *et al.* [38]. Bechar *et al.* [38] focused only on the segmentation of the OD and OC to detect the disease, while SSCNN-DAE and SSCNN learned the glaucoma features from the entire structure of the fundus images. The TCNN model achieved the third best performance, outperforming the other related approaches by a clear margin. This is due to the fine-tuning of the weights of the pre-trained network that makes the later layers more relevant to the details in our dataset.

The results also demonstrate that the supervised methods [30], [31], even with transfer learning, are less accurate than all our presented models because they did not perform any fine-tuning of the pre-trained network layers. On the other hand, the supervised learning in [26] performed better than the transfer learning methods since it relies on detecting the disease from pre-extracted features. However, fine-tuning the pre-trained network layers, iteratively increasing the training set using unlabeled samples, and finally learning the distinctive features from the unlabeled data using the autoencoder increase the classification accuracies of our models.

3) THE EFFECTIVENESS OF THE SELF-LEARNING METHOD

The self-learning method was utilized in the SSCNN model to extend the training set iteratively. At each iteration, the predicted scores for the unlabeled samples were ordered by the self-learning method and the ones with the highest confidence were chosen to increase the training set. To show the effectiveness of the self-learning method, we compared it against the random selection method. At each iteration in the random selection, a fixed number of the unlabeled samples

TABLE 1. Comparison of the proposed glaucoma models' results (TCNN, SSCNN and SSCNN-DAE) with related methods, n/a means not available.

Method	Accuracy	Sensitivity	Specificity
SSCNN-DAE	93.8%	98.9%	90.5%
SSCNN	92.4%	91.7%	93.3%
TCNN	91.5%	90.5%	92.0%
Haleem *et al.* [15]	94.4	92.3 %	95.3
Sreng *et al.* [36]	92.1 %	n/a	n/a
Bechar *et al.* [38]	90.9%	n/a	n/a
Oscar *et al.* [26]	89.4%	89.5%	88.9%
Al-Bander *et al.* [30]	88.2%	85.0%	90.8%
Valverde *et al.* [35]	88.0 %	87.0%	89.0%
Cerentini *et al.* [31]	86.2%	n/a	n/a

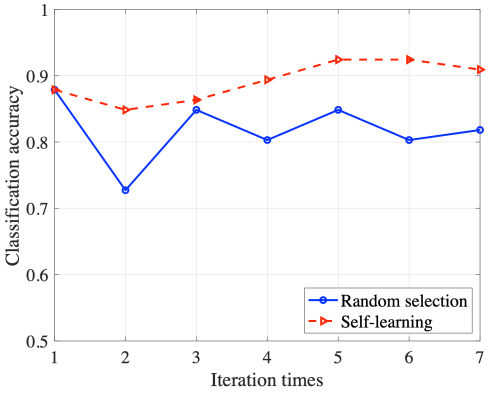

FIGURE 8. Comparing the performance of self-learning and the performance of random selection. The classification accuracies of these two methods are calculated per iteration.

were randomly chosen and the highest predicted class for each sample was selected as the true label. The classification accuracies of these two methods are calculated per iteration and presented in Figure 8. Clearly, the self-learning method performed better than the random selection method. Thus, this proves the efficacy of the self-learning method in utilizing the unlabeled samples.

4) THE EFFECTIVENESS OF THE CNN-DAE

The CNN-DAE method was used in the SSCNN-DAE model to generate representative feature maps for glaucoma from unlabeled data. These maps were then used to train fully connected layers as a glaucoma-specific classifier. To demonstrate the effectiveness of the CNN-DAE method, we compared it with the other unsupervised methods reviewed in Section II-C1, including autoencoders (AE), the denoising autoencoder (DAE) and the CNN autoencoder (CNN-AE). For each of these models, the autoencoders were used to learn the feature maps from the unlabeled data. These feature maps were then extracted from the labeled data and used to train the fully connected layers as a glaucoma-specific classifier. The same setup described in Section III-B was used for all these models.

The classification performance results are presented in Figure 9. The results highlight that the CNN-DAE method performed better than the other unsupervised methods. The DAE outperformed the AE because it corrupted the input with random noise at the training stage, which helped to learn

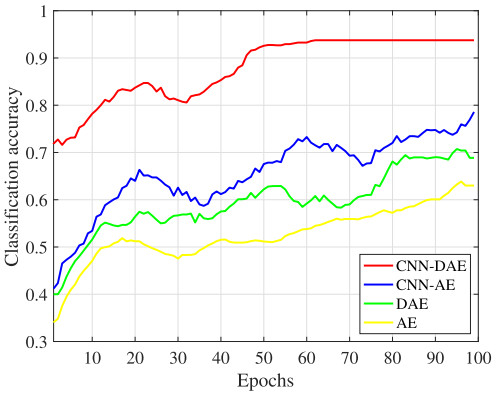

FIGURE 9. Evaluating the CNN-DAE against other unsupervised learning architectures, including autoencoders (AE), the denoising autoencoder (DAE) and the CNN autoencoder (CNN-AE).

more noise-resistant features. However, the CNN-AE outperformed both the DAE and the AE because its convolution structure helped the model to share weights and preserve the localization of features, thus preserving the spatial information of the input data. Therefore, combining both the DAE and the convolution structure helps our model to learn more robust and abstract features, which improves the performance of learning the representation.

IV. CONCLUSION

This paper presented an automatic glaucoma diagnosing framework based on convolutional neural network (CNN) models with different learning methods, and compared their performances with the performance of trained ophthalmologists. We have presented transfer and semi-supervised learning methods using both labeled and unlabeled data. The transfer learning model is based on CNN pre-trained with non-medical data and fine-tuned using domain specific labeled data. The semi-supervised models were constructed based on two different approaches and trained using both labeled and unlabeled data. Unlike the previous works where the optic disc features were handcrafted, the presented models automatically extract the key features of the disease from raw images without any enhancement or preprocessing steps. The experimental results on the RIM-ONE dataset demonstrated the efficacy of deep learning models when applied to glaucoma, which is clinically very significant as it helps to identify individuals with early-stage glaucoma.

Making improvements to generalize the models to diagnose more eye diseases is part of the authors' ongoing research. We believe, in addition to glaucoma, the diagnosis of other diseases from retinal images can benefit from the presented models, including for example diabetic retinopathy (DR) and age-related macular degeneration (AMD). More importantly, these models could be beneficial in not only assisting diagnosis, but also in classifying the severity of a specific retinal disease.

ACKNOWLEDGMENT

The authors would like to thank Mohamed Abou Shousha (from Bascom Palmer Eye Institute, Miami, USA) and

Lin Xu-chu (from the Department of Ophthalmology, the 303 Hospital of Chinese PLA, Guangxi Zhuang Autonomous Region, China) for their contribution in classifying the dataset used in the experiments of this study.

REFERENCES

[1] J. Flammer and E. Meier, *Glaucoma: A Guide for Patients—An Introduction for Care-Providers: A Quick Reference*. Bern, Switzerland: Hogrefe & Huber, 2003.

[2] F. Badalà, K. Nouri-Mahdavi, D. A. Raoof, N. Leeprechanon, S. K. Law, and J. Caprioli, "Optic disk and nerve fiber layer imaging to detect glaucoma," *Amer. J. Ophthalmol.*, vol. 144, no. 5, pp. 32–724, 2007.

[3] C. I. Sánchez, M. Niemeijer, A. V. Dumitrescu, M. S. Suttorp-Schulten, M. D. Abràmoff, and B. van Ginneken, "Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data," *Investigative Ophthalmol. Vis. Sci.*, vol. 52, no. 7, pp. 4866–4871, Jun. 2011.

[4] K. G. Gundersen, A. Heijl, and B. Bengtsson, "Sensitivity and specificity of structural optic disc parameters in chronic glaucoma," *Acta Ophthalmologica Scandinavica*, vol. 74, no. 2, pp. 120–125, May 2009.

[5] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *Proc. 12th Int. Symp. Med. Inf. Process. Anal.*, vol. 10160, 2017, Art. no. 101600U.

[6] S. Saha, A. K. Alok, and A. Ekbal, "Brain image segmentation using semi-supervised clustering," *Expert Syst. Appl.*, vol. 52, pp. 50–63, Jun. 2016.

[7] A. Masood, A. Al- Jumaily, and K. Anam, "Self-supervised learning model for skin cancer diagnosis," in *Proc. 7th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Apr. 2015, pp. 1012–1015.

[8] L. H. Son and T. M. Tuan, "A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation," *Expert Syst. Appl.*, vol. 46, pp. 380–393, Mar. 2016.

[9] X. You, Q. Peng, Y. Yuan, Y.-M. Cheung, and J. Lei, "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2314–2324, Oct. 2011.

[10] Y. Chai, H. Liu, and J. Xu, "Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models," *Knowl.-Based Syst.*, vol. 161, pp. 147–156, Dec. 2018.

[11] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, Aug. 2018.

[12] S. Phene, R. C. Dunn, N. Hammel, Y. Liu, J. Krause, N. Kitade, M. Schaekermann, R. Sayres, D. J. Wu, A. Bora, C. Semturs, A. Misra, A. E. Huang, A. Spitze, F. A. Medeiros, A. Y. Maa, M. Gandhi, G. S. Corrado, L. Peng, and D. R. Webster, "Deep learning and glaucoma specialists: The relative importance of optic disc features to predict glaucoma referral in fundus photographs," *Ophthalmology*, vol. 126, no. 12, pp. 1627–1639, 2019.

[13] K. D. Bojikian, C. S. Lee, and A. Y. Lee, "Finding glaucoma in color fundus photographs using deep learning," *JAMA Ophthalmol.*, vol. 137, no. 12, p. 1361, Dec. 2019.

[14] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Trans. Med. Imag.*, vol. 37, no. 11, pp. 2493–2501, Nov. 2018.

[15] M. S. Haleem, L. Han, J. V. Hemert, A. Fleming, L. R. Pasquale, P. S. Silva, B. J. Song, and L. P. Aiello, "Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (SLO) images," *J. Med. Syst.*, vol. 40, no. 6, p. 132, Jun. 2016.

[16] Y. Chai, L. He, Q. Mei, H. Liu, and L. Xu, "Deep learning through two-branch convolutional neuron network for glaucoma diagnosis," in *Smart Health*, H. Chen, D. D. Zeng, E. Karahanna, and I. Bardhan, Eds. Cham, Switzerland: Springer, 2017, pp. 191–201.

[17] A. Pal, M. R. Moorthy, and A. Shahina, "G-Eyenet: A convolutional autoencoding classifier framework for the detection of glaucoma from retinal fundus images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2775–2779.

[18] R. Zhao, W. Liao, B. Zou, Z. Chen, and S. Li, "Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis," in *Proc. AAAI*, 2019, pp. 809–816.

[19] L. Li, M. Xu, X. Wang, L. Jiang, and H. Liu, "Attention based glaucoma detection: A large-scale database and CNN model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10571–10580.

[20] L. K. Singh, H. Garg, M. Khanna, and R. S. Bhadoria, "An enhanced deep image model for glaucoma diagnosis using feature-based detection in retinal fundus," *Med. Biol. Eng. Comput.*, vol. 59, pp. 1–21, Jan. 2021.

[21] D. Mahapatra, "Combining multiple expert annotations using semi-supervised learning and graph cuts for medical image segmentation," *Comput. Vis. Image Understand.*, vol. 151, pp. 114–123, Oct. 2016.

[22] A. G. Roy and D. Sheet, "DASA: Domain adaptation in stacked autoencoders using systematic dropout," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*. IEEE, 2015, pp. 735–739.

[23] J. Benson, H. Carrillo, J. Wigdahl, S. Nemeth, J. Maynard, G. Zamora, S. Barriga, T. Estrada, and P. Soliz, "Transfer learning for diabetic retinopathy," *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105741Z.

[24] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–11.

[25] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 715–718.

[26] O. Perdomo, V. Andrearczyk, F. Meriaudeau, H. Müller, and F. A. González, "Glaucoma diagnosis from eye fundus images based on deep morphometric feature estimation," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 319–327.

[27] S. Yu, D. Xiao, S. Frost, and Y. Kanagasingam, "Robust optic disc and cup segmentation with deep learning for glaucoma detection," *Comput. Med. Imag. Graph.*, vol. 74, pp. 61–71, Jun. 2019.

[28] R. Hemelings, B. Elen, J. Barbosa-Breda, S. Lemmens, M. Meire, S. Pourjavan, E. Vandewalle, S. Van De Veire, M. B. Blaschko, P. De Boever, and I. Stalmans, "Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning," *Acta Ophthalmologica*, vol. 98, no. 1, pp. 94–100, Feb. 2020.

[29] N. Gour and P. Khanna, "Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network," *Biomed. Signal Process. Control*, vol. 66, Dec. 2020, Art. no. 102329.

[30] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Taee, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in *Proc. 14th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2017, pp. 207–210.

[31] A. Cerentini, D. Welfer, M. C. d'Ornellas, C. J. P. Haygert, and G. N. Dotto, "Automatic identification of glaucoma sing deep learning methods," *Stud. Health Technol. Informat.*, vol. 245, pp. 318–321, Jan. 2017.

[32] A. A. Jammal, A. C. Thompson, E. B. Mariottoni, S. I. Berchuck, C. N. Urata, T. Estrela, S. M. Wakil, V. P. Costa, and F. A. Medeiros, "Human versus machine: Comparing a deep learning algorithm to human gradings for detecting glaucoma on fundus photographs," *Amer. J. Ophthalmol.*, vol. 211, pp. 123–131, Mar. 2020.

[33] S. Maheshwari, V. Kanhangad, and R. B. Pachori, "CNN-based approach for glaucoma diagnosis using transfer learning and LBP-based data augmentation," 2020, *arXiv:2002.08013*. [Online]. Available: http://arxiv.org/abs/2002.08013

[34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[35] J. J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C. I. Sánchez, and M. J. Ledesma-Carbayo, "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning," *Biomed. Opt. Exp.*, vol. 10, no. 2, p. 892, 2019.

[36] S. Sreng, N. Maneerat, K. Hamamoto, and K. Y. Win, "Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images," *Appl. Sci.*, vol. 10, no. 14, p. 4916, Jul. 2020.

[37] S. Sedai, D. Mahapatra, S. Hewavitharanage, S. Maetschke, and R. Garnavi, "Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds. Cham, Switzerland: Springer, 2017, pp. 75–82.

[38] M. E. A. Bechar, N. Settouti, V. Barra, and M. A. Chikh, "Semi-supervised superpixel classification for medical images segmentation: Application to detection of glaucoma disease," *Multidimensional Syst. Signal Process.*, vol. 29, no. 3, pp. 979–998, Jul. 2018.
[39] M. Al Ghamdi, M. Li, M. Abdel-Mottaleb, and M. A. Shousha, "Semi-supervised transfer learning for convolutional neural networks for glaucoma detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3812–3816.
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–14.
[41] N. Bayramoglu and J. Heikkila, "Transfer learning for cell nuclei classification in histopathology images," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016, pp. 532–539.
[42] R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?" *ICT Exp.*, vol. 4, no. 4, pp. 247–254, Dec. 2018.
[43] M. Sushil, G. Suguna, R. Lavanya, and M. N. Devi, "Performance comparison of pre-trained deep neural networks for automated glaucoma detection," in *Proc. Int. Conf. ISMAC Comput. Vis. Bio-Eng. (ISMAC-CVB)*, D. Pandian, X. Fernando, Z. Baig, and F. Shi, Eds., 2019, pp. 631–637.
[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, pp. 1–15, Dec. 2014.
[45] N. S. Keskar and R. Socher, "Improving generalization performance by switching from adam to SGD," *CoRR*, vol. abs/1712.07628, pp. 1–10, Dec. 2017.
[46] N. S. Keskar and R. Socher, "Improving generalization performance by switching from adam to SGD," 2017, *arXiv:1712.07628*. [Online]. Available: http://arxiv.org/abs/1712.07628
[47] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 309–316.
[48] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Comput. Sci.*, vol. 90, pp. 200–205, Jan. 2016.
[49] F. Gao, Z. Yue, J. Wang, J. Sun, E. Yang, and H. Zhou, "A novel active semisupervised convolutional neural network algorithm for SAR image recognition," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–8, 2017.
[50] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn. (ICML)*, 2013, pp. 2–8.
[51] Y. Zhou, M. Kantarcioglu, and B. Thuraisingham, "Self-training with selection-by-rejection," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 795–803.
[52] A. Makhzani and B. J. Frey, "Winner-take-all autoencoders," in *Proc. Adv. Neural Inf. Process. Process. Syst. (NIPS)*, 2015, p. 2791–2799.
[53] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 241–246.
[54] D. Lee, S. Choi, and H.-J. Kim, "Performance evaluation of image denoising developed using convolutional denoising autoencoders in chest radiography," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 884, pp. 97–104, Mar. 2018.
[55] S. Zhai and Z. M. Zhang, "Semisupervised autoencoder for sentiment analysis," in *Proc. 13th AAAI Conf. Artif. Intell. (AAAI)*. Palo Alto, CA, USA: AAAI Press, 2016, pp. 1394–1400.
[56] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
[57] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Artificial Neural Networks and Machine Learning—ICANN*, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Berlin, Germany: Springer, 2011, pp. 52–59.
[58] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. 1506.02640, pp. 779–788, Jun. 2015.
[59] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*. [Online]. Available: http://arxiv.org/abs/1603.07285
[60] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation," in *Proc. 24th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2011, pp. 1–6.
[61] A. Almazroa, S. Alodhayb, E. Osman, E. Ramadan, M. Hummadi, M. Dlaim, M. Alkatee, K. Raahemifar, and V. Lakshminarayanan, "Retinal fundus images for glaucoma analysis: The RIGA dataset," *Proc. SPIE*, vol. 10579, Mar. 2018, Art. no. 105790B.
[62] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: The Messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, p. 231, Aug. 2014.
[63] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[64] F. Chollet. *Keras*. Accessed: 2015. [Online]. Available: https://keras.io
[65] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, and P. Tucker, "TensorFlow: A system for large-scale machine learning," in *Proc. Symp. Operating Syst. Design Implement.*, 2016, pp. 265–283.
[66] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," 2017, *arXiv:1711.00489*. [Online]. Available: http://arxiv.org/abs/1711.00489
[67] D. O. I. M. University of Nebraska and M. Center. *The Area Under an ROC Curve*. Accessed: Dec. 20, 2020. [Online]. Available: http://gim.unmc.edu/dxtests/roc3.htm
[68] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5353–5360.

MANAL ALGHAMDI (Member, IEEE) received the B.S. degree in computer science from King Abdulaziz University (KAU), Saudi Arabia, in 2008, and the M.S. degree in advanced computer science and the Ph.D. degree in computer vision from The University of Sheffield, U.K., in 2010 and 2015, respectively. Her study involved video representation and video similarity measurements. From 2016 to 2017, she was appointed as a Deputy Head of the Department of Computer Science, Umm Al-Qura University (UQU), Saudi Arabia. In 2019, she was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Miami, USA, where she developed interest in deep learning and its application in different areas of computer vision. She is currently an Assistant Professor with the Department of Computer Science, UQU. She focuses on developing and evaluating video and image processing techniques for various applications including video representation, similarity measurements and crowed analysis. Recently, she developed interest in computer security techniques including steganography and cryptography and her current interest in computer vision involves developing convolution deep learning techniques for healthcare applications. Her research interests include machine learning, computer vision, and security.

MOHAMED ABDEL-MOTTALEB (Fellow, IEEE) received the Ph.D. degree in computer science from the University of Maryland, College Park, in 1993. He was with Philips Research in Briarcliff Manor, New York, where he was a Principal Member of the Research Staff and a Project Leader. In 2001, he joined the University of Miami. Since 2011, he has been with the Institute of Electrical and Electronics Engineers. He currently chairs the Department of Electrical and Computer Engineering, where he is also a Professor. At Philips Research, he led several projects in image processing and content-based multimedia retrieval. He also represented Philips in the standardization activity of ISO for MPEG-7, where some of his work was included in the standard. He has published more than 150 journal and conference papers in the areas of image processing, computer vision, and content-based retrieval. He has coauthored the book *Human Activity Analysis from Video* (F. Niu, 2008). He holds 21 U.S. patents and more than 30 international patents. His research interests include 3-D face and ear recognition, dental biometrics, visual tracking, human-activity recognition, and medical image processing.