Retina Disorders Classification via OCT Scan: A Comparative Study between Self-Supervised Learning and Transfer Learning

Saeed Shurrab
Department of Computer Information Systems, Jordan University of Science and Technology, Jordan
sashurrab18@cit.just.edu.jo

Yazan Shannak
Department of Computer Information Systems, Jordan University of Science and Technology, Jordan
ywshannak19@cit.just.edu.jo

Rehab Duwairi
Department of Computer Information Systems, Jordan University of Science and Technology, Jordan
rehab@just.edu.jo

Abstract: *Retina disorders are among the common types of eye disease that occur due to several reasons such as aging, diabetes and premature born. Besides, Optical Coherence Tomography (OCT) is a medical imaging method that serves as a vehicle for capturing volumetric scans of the human eye retina for diagnoses purposes. This research compared two pretraining approaches including Self-Supervised Learning (SSL) and Transfer Learning (TL) to train ResNet34 neural architecture aiming at building computer aided diagnoses tool for retina disorders recognition. In addition, the research methodology employs convolutional auto-encoder model as a generative SSL pretraining method. The research efforts are implemented on a dataset that contains 109,309 retina OCT images with three medical conditions including Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), DRUSEN as well as NORMAL condition. The research outcomes showed better performance in terms of overall accuracy, sensitivity and specificity, namely, 95.2%, 95.2% and 98.4% respectively for SSL ResNet34 in comparison to scores of 90.7%, 90.7% and 96.9% respectively for TL ResNet34. In addition, SSL pretraining approach showed significant reduction in the number of epochs required for training in comparison to both TL pretraining as well as the previous research performed on the same dataset with comparable performance.*

Keywords: *OCT images, retina disorders, self-supervised learning, transfer learning, resnet34.*

Received September 20, 2021; accepted December 1, 2022
https://doi.org/10.34028/iajit/20/3/8

1. Introduction

Retina impairment is the most common sight-threatening disorder that may leads to blindness [37]. Retinal damage can result from various conditions such as diabetes, premature born and aging. Choroidal Neovascularization (CNV) is part of the spectrum of retinal abnormalities that leads to retinal swelling and disruption. CNV is a typical sign of late-stage Age-Related Macular Degeneration (AMD) and is characterized by abnormal growth of blood vessels on the choroid, a connective tissue that provides blood supply to the retina [10]. AMD affects mainly patients older than 60 years and accounts for about 8% of the blindness cases worldwide [45].

Age-related ocular signs also include DRUSEN, a small deposit of extracellular waste that accumulate under the retina, which can seriously impair vision [39]. These deposits are composed of lipids, carbohydrates and proteins that aggregate between the basal lamina and the Bruch membrane of the retina leading to visual defects that affect up to 24 per 1000 population [5].

Diabetic Macular Edema (DME) is another retinal complication caused by diabetes and affects approximately 75,000 patients per year in the United States [6]. DME results from fluid accumulation in the center of the retina or the so-called macula as a result of the disrupted blood-retinal barrier [44]. All these conditions should be rapidly and precisely diagnosed to accelerate the appropriate intervention and prevent visual loss. Retinal Optical Coherence Tomography (OCT) is a non-invasive diagnostic technique that is routinely used to provide cross-sectional images for the internal structure of the eye. Retinal OCT enables ophthalmologists to easily visualize the retinal compartments [9]. However, to interpret and classify the retinal abnormality using retinal OCT images, the ophthalmologist conducts multiple eye exams that are time-consuming and subjective to unreliability [22]. These restrictions can be skipped by the computer-aided diagnosis that reduces the consumed time and inter-observer variability in ocular image interpretation.

Artificial Intelligence (AI) played an important role in revolutionizing many fields of science including the healthcare sector. Analyzing the relationship between the medical approaches and the patient outcomes in order to achieve better medical service and disease management in terms of time and cost in an efficient and effective manner is the primary goal of AI in health care [11].

Deep learning is a subdivision of AI that enables computers to employ a set of mathematical models with multiple layers that are able to learn the latent patterns in certain types of data in a form of hierarchical representation automatically without human intervention [14, 23]. One of those deep learning models is the Convolutional Neural Network (CNN) which is a neural model that deals with data of grid representation such as time series data which is a single dimension (1D) grid data and images which are two dimensional (2D) grid data [14].

Medical image processing has benefited from CNNs, this benefit can be viewed as shifting the classical medical image processing operations such as classification, segmentation, localization, registration and detection from the manual mode into automated mode by developing dedicated models that are able to handle these tasks [19]. Among this spectrum of medical image processing tasks is the classification which aims at determining whether a medical image contains a certain medical condition or not. CNN can act as Computer-Aided Diagnosis (CAD) tool that makes image interpretation by radiologist easier and robust by supporting it with a second computerized opinion [38].

In the same vein, Transfer Learning (TL) is another important aspect that comes with CNN models and their applications. Transfer learning is the process conveying and generalizing the knowledge that has been learned in a certain task to another task from the same domain in a supervised fashion in order to improve the learning scheme in the latter task [14, 15, 33]. Technically, this means using a pretrained weights for a certain model to initialize the training of another model. As an example, ImageNet dataset [8] which contains approximately 14 million images and 22,000 visual categories is mainly used as a reference dataset for building pretrained models for various computer vision tasks such as image classification, objects detection, image reconstruction and semantic segmentation.

Typically, there are four options when using transfer learning. These are determined by the dataset size and its similarity with the source data (on which the model was originally trained):

1. Retrain only the output layer in the case the dataset is small and very similar to the source data.
2. Retrain all layers with initial weights taken from pretrained model when the dataset is large and very similar to the source data.
3. Retrain the last layers and freeze early layers when dataset is small and less similar to the source data.
4. Train the model from scratch if the dataset is large and very different from the source data [18].

By comparison, Self-Supervised Learning (SSL) is another pretraining approach which aims at learning the representation features for a certain task in an unsupervised manner by relaying only on the input data [7]. More clearly, the main intuition behind SSL approach is to learn the representative features in the target task by initially developing a pretext predictive task using the input data of the target task. The pretext task applies certain transformations to the input data, such as rotation, to get pseudo labels automatically. Then, a classifier is trained on this input data with pseudo labels. Consequently, the learned features are transferred from the pretext task to the target task to accomplish it with respect to the true labels. Comparatively, SSL approach is susceptible to have less discrimination power than TL approach as it does not rely on the actual labels, however; it has lower probability to be biased toward the class labels in the original task at which the pretrained model was trained on such as in the case of TL approach [46].

In regard to the pretext tasks, a variety of methods have been developed which can be mainly divided into three categories including generative learning whose main purpose is to (learn to reconstruct), contrastive learning whose main purpose is to (learn to compare) and adversarial learning which is a combination of generative and contrastive approaches [29].

This research aims at building a computer aided diagnosis tool for retina abnormalities classification by comparing the performance of two pre-training strategies including SSL and TL. ResNet34 architecture is employed for experimenting the research questions in two training schemes. The first scheme employs convolutional auto-encoder as a self-supervised pre-training approach for ResNet34 architecture followed by supervised training for classification purposes, while the second scheme employs ResNet34 model pretrained on ImageNet Dataset according to the third transfer learning scenario. The research efforts are accomplished using dataset reported in [27] which contains 109,309 OCT images with three medical condition CNV, DME, DRUSEN in addition to the NORMAL condition.

To the best of our knowledge, this is the first attempt to compare the effect of both self-supervised learning and transfer learning approaches on the performance of ResNet34 model for classifying retina disorders using the same dataset. Comparatively, SSL pretraining approach achieved better performance than TL training approach in terms of accuracy, sensitivity and specificity. Furthermore, employing SSL showed considerable reduction in the computational efforts required to achieve the optimum performance in comparison to both TL pretraining as well as the previous research works on the same dataset. The research code and all programming implementations are available on the following GitHub repository: https://github.com/SaeedShurrab/OCT-Scans-Classification.

The remainder of the article is structured as follows: section two discusses the related works, while

section three presents the research methodology, results and findings are presented in section four and whereas results discussion and research limitations are discussed in section five and finally section six concludes the article and suggests future research directions.

2. Related Works

Classification of retinal disorders using AI tools had been well studied in the previous literature with a variety of models that range from using primitive models such as statistical machine learning approaches along with classical medical image processing techniques to more sophisticated models using convolutional neural networks which can be briefly summarized as follow:

With respect to statistical machine learning approaches, Liu *et al*. [28] proposed an OCT-based image classifier to distinguish three retinal disorders including Macular Edema (ME), Macular Hole (MH) and AMD from normal retina. Local Binary Pattern method was employed as a local feature extractor while Support Vector Machine (SVM) was trained in binary fashion as normal versus each of the three disorders. Another similar study performed by Albarrak *et al*. [1] that integrated oriented gradient histogram with local binary pattern histogram to extract features from 3D retinal OCT scans to be classified as AMD versus Normal retina using Bayesian classification network. Similarly, Srinivasan *et al*. [40] combined oriented gradient histogram and SVM to discriminate between DME, AMD and normal retina conditions using OCT images.

Venhuizen *et al*. [42] proposed a methodology for discriminating between normal and AMD in retina OCT images by combining K-mean clustering with Random Forest Classifier (RF). K-means algorithm is used for features extraction which outcomes are used to develop patch occurrence histogram (Bag of words) which in turn is fed into RF classifier. Another work that adopted the same previous methodology was performed by Lemaitre *et al*. [25] which aimed to discriminate DME OCT images from normal ones. Local binary pattern was used as feature extraction instead of K-means clustering. They reported better performance in term of sensitivity and specificity in comparison to [43]. This work had been further extended in 2016 by examining different input images preprocessing techniques, feature extraction techniques as well as classification algorithms [26]. In 2017, Venhuizen *et al*. [43] extended their earlier work transforming it from a binary classification problem to multi-class classification problem with gradual severity levels of AMD disease including no AMD, early AMD, intermediate AMD, advanced AMD Geographic Atrophy and advanced AMD CNV. The results showed relatively comparative performance to the human performance in terms of sensitivity and specificity.

Several similar works had been performed by other researchers such as the works accomplished in [3, 17, 41]. It can be observed that the main theme of these researches is oriented toward manually generating representative features to be fed into a certain machine learning algorithm.

On the other side, utilizing CNNs in retina disorders classification provided a vehicle for better performance and less preprocessing. A summary of the most recent retina disorders classification using CNN are briefly presented as follow:

One of the earliest works that employed CNNs in retina disorder classification is performed by Apostolopoulos *et al*. [4]. They proposed a binary classification CNN that distinguishes between AMD and normal OCT scans that is called RetiNet. The proposed architecture consists of two-phase learning, where in the first phase, 2D scans are used to learn features representation in the input image using CNN (RetiNet-B). While in the second phase, learned features are used to train a second network (RetiNet-C) to classify 3D retinal scans. Their model showed better performance in terms of AUC score in comparison to VGG16, DenseNet, ResNet and 2DSeg architectures. Another study performed by Lee et al. [24] developed a CNN model to classify OCT images as AMD versus normal. VGG16 architecture with Xavier weights initialization was employed.

Kermany *et al*. [20] released a relatively large retina OCT images dataset [21]. The dataset contains (109, 312) OCT images distributed among four classes namely CNV, DME, DRUSEN and Normal condition. Inception V3 architecture pretrained on ImageNet dataset [A8] was evaluated on the same dataset and achieved accuracy score of (96.6%). Following that, other researches had been performed that examine the performance of different CNN architectures on the same dataset. Among those researches is the work that accomplished by Li *et al*. [27] who evaluated the performance of VGG16 architecture pretrained on ImageNet dataset on Kermany *et al*. [21] dataset. Their implementation was able to outperform the previous results by achieving accuracy score of (98.6%). Rastogi *et al*. [35] compared the performance of four variants of DenseNet architecture namely vanilla DenseNet, DenseNet-B, DenseNet-C and DenseNet-BC on the same dataset. The best result was achieved by DenseNet-BC model with accuracy score of (97.65%) on test set of 5415 OCT images created by the authors which outperformed [20] results but not [27]. Fang *et al*. [12] developed an Iterative Fusion Neural Networks (IFCNN) for classifying retinal disorders. The role of iterative fusion strategy is to join the feature maps from all previous convolutional layers with the present convolutional layer to improve the classification accuracy. The performance achieved by IFCNN in terms of accuracy is (87.3%) which is the least performance achieved on the same dataset.

Alqudah [2] combined Kermany dataset [27] with Farsiu *et al*. [13] dataset which contains OCT images for both AMD disease and normal condition to increase the number of diseases included in the dataset. A special CNN architecture, called AOCT-NET, was developed which was able to achieve overall accuracy of (94.4%).

Nugroho [32] compared two manual feature extraction methods, namely, Linear Binary Pattern (LBP) and Histogram Oriented Gradient (HOG) against two convolutional feature extraction DenseNet169 and ResNet50 on the same dataset. The extracted features from each method were fed into a single perceptron classifier. The results showed better performance achieved under neural based features extraction methods against manual methods. Table 1 summarizes the performance measures in terms of accuracy, sensitivity, specificity and Receiver Operating Curve (ROC). Other similar works that adopted CNNs in classifying retinal disorders using different datasets had been performed by Schlegl *et al*. [36], Perdomo et al. [34] and Sun and Sun [41].

To sum up, it can be observed from the previous literature that the main theme of the performed researches on Kermany's dataset is either using pretrained models or specific convolutional architecture designed specifically for this dataset. This research distinguishes itself by comparing the performance of ReNet34 model using both self-supervised learning and transfer learning which is to the best of our knowledge the first attempt to approach retina OCT classification via the two strategies.

Table 1. Performance measures summary of all works performed on Kermany et al. dataset.

Author	Model	Acc. (%)	Sens. (%)	Spec. (%)	AUC (%)
Li *et al*. [27]	VGG16	98.6	97.8	99.4	100
Rastogi *et al*. [35]	DenseNet	97.65	95.57	99.15	97.75
Kermany *et al*. [20]	Inception V3	96.6	97.8	97.4	99.9
Alqudah [2]	AOCT-NET	94.4	99.91	98.52	N/A
Fang *et al*. [12]	IFCNN	87.3	N/A	N/A	99.85

3. Materials and Methods

3.1. Research Dataset

The research dataset is publicly available for the research community from Mendeley data website since January 2018 [27]. The dataset contains (108, 309) OCT images from 5319 patients as a training set divided into three medical conditions including CNV, DME, DRUSEN in addition to the Normal scans. All images in the data set comes with (jpeg) format with varying dimensions between (496) and (1536) pixels as well as gray scale coloring system. In addition, the distribution of the images over the four classes is imbalanced distribution with Normal having largest number of images (51, 140) followed by CNV condition (37, 205), while DME and DRUSEN have (11, 348) and (8, 616) images respectively. Further, additional (1,000) images from 633 patients are available as a test set divided as 250 image for each label [20]. Figure 1 shows four sample images from the research dataset while Tables 2 and 3 summarize the dataset properties.

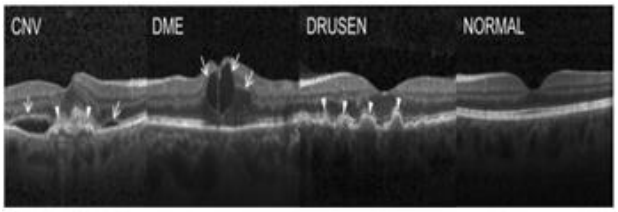


Figure 1. A sample OCT image from the four labels CNV, DME, DRUSEN and Normal [20].

Table 2. Properties of research dataset.

Property	Total images	Extension	Resolution	Coloring sys.
Value	109309	Jpeg	496-1536	Gray-scale

Table 3. Training and test data distribution.

Class	Normal	CNV	DME	DRUSEN	Total
Train data	51,140	37,205	11,348	8,616	108,309
Test data	250	250	250	250	1,000

3.2. Dataset Preprocessing

In order to make the data ready for modeling, the research dataset has undergone the following operations:

1. All images have been resized into (224x224).
2. All images have been converted from gray-scale into RGB to suit the ResNet requirements and normalized using the global mean and standard deviation of the training datasets with values of (0.2003) and (0.2042) respectively.
3. Further for pretrained models, all images have been normalized according to the standard values of the mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225) for the ImageNet dataset which are the preferred values for the pre-trained models.

3.3. Research Models

3.3.1. Residual Neural Networks (ResNets)

Residual neural networks are convolutional architecture that had been proposed by He *et al*. [16] in 2015. Standard residual networks have (5) architectures with different number of layers including (18, 34, 50, 101, 152). ResNet34 is selected to investigate the research assumptions for the sake of simplicity, however, any other architecture is applicable. Further, the performance of this architecture has not been investigated on the same dataset. ResNet34 architecture consists of initial single

convolutional layer, four residual blocks and single fully connected layer. The first convolutional layer has zero padding and stride values of (3) and (2) respectively and followed by a max pooling layer with zero padding and stride values of (3) and (1) respectively. With respect to residual blocks, each block consists of a predefined number of convolutional layers with skip connection between each two convolutional layers and stride value of (2) for the second, third and fourth residual blocks. Ultimately, the fully connected layer represents the number of classes in the research dataset and preceded by an adaptive average pooling layers that convert the input image into (1x1) vector. Each convolutional layer is followed by batch normalization layer. Further the input dimensions are halved after each block while the number of channels is doubled.

3.3.2. ResNet34 Auto-Encoder

Auto-encoders are neural models that consist of two sub-models, namely, encoder and decoder, the encoder compresses the input data into a latent dimensional space; while the decoder reproduces the input data from the compressed latent space [30]. Convolutional Auto-Encoders (CAE) are special case of the auto-encoders that use the convolution layers. Features extraction is one of the most significant applications of CAEs where an auto-encoder is trained until capturing the features in the input data, then the decoder is discarded while the encoder is kept to perform its intended function. For this research purposes, an auto-encoder is implemented to capture the features in the research dataset as a self-supervised learning technique. The convolutional blocks of ResNet34 architecture were employed as the backbone (encoder model) of the auto-encoder; while the fully connected block was discarded. On the decoder side, five convolutional upsampling layers were employed each with stride value of (2), batch normalization layer and ReLU activation except for the decoder output layer where sigmoid activation was employed to match the input format. Table 4 depicts the detailed structure of the ResNet34 auto-encoder architecture.

3.4. Self-Supervised Pre-Training

Two stages of training are implemented in a consecutive manner. In the first stage, the auto-encoder is trained to capture properties of training data until it is able to produce images that are relatively similar to the input images with no further improvement on the validation set. Consequently, as the auto-encoder backbone constituted of the convolutional blocks in ResNet34 architecture, the encoder weights are transferred into the classification model as initial weights to start the second stage of training. Eventually, the obtained results, via SSL approach, are compared with pre-trained ResNet34 on ImageNet dataset where the initial layers are frozen while the remaining layers are fine-tuned.

3.5. Experimental Setup

Initially, the training dataset was divided into two divisions for training and validation with percentages of (97477 images, 90% of data) and (10832 images, 10% of data), respectively. In addition, the original test set as indicated in Table (3) is used for testing purposes. In regard to the pretrained ResNet34 on ImageNet dataset, the parameters of the first four blocks including Conv1, Res1, Res2 and Res3 were set as frozen (non-learnable) parameters as they account for approximately (8) million parameters out of (21.28) million parameters while the remaining layers have been set unfrozen. Ultimately, each of the trained models has its own set of hyperparameter values, part of these hyperparameter values are shared between the three models while some models may have different values. In addition, early stopping regularization techniques was employed to avoid overfitting. Table 5 summarizes the experimental settings for each model.

Table 4. The architecture of ResNet34 and the Auto-encoder.

Block	Number of layers	Kernel Size	in channels	out channels	activation
Encoder					
Conv1	1	7x7	3	64	ReLU
Max Pool	1	3x3	64	64	N/A
Res1	6	3x3	64	64	ReLU
Res2	1	3x3	64	128	ReLU
Res2	7	3x3	128	128	ReLU
Res3	1	3x3	128	256	ReLU
Res3	11	3x3	256	256	ReLU
Res4	1	3x3	256	512	ReLU
Res4	5	3x3	512	512	ReLU
AVG Pool	1	7x7	512	512	N/A
FC	1	N/A	512	4	SoftMax
Decoder					
Block1	1	2x2	512	256	ReLU
Block2	1	2x2	256	128	ReLU
Block3	1	2x2	128	64	ReLU
Block4	1	2x2	64	64	ReLU
Block5	1	2x2	64	3	Sigmoid

All coding efforts have been accomplished using Pytorch3 deep learning framework using NVIDIA RTX 2080 TI GPU with (11 GB) RAM and AMD Ryzen (9 3950x) CPU with (16 GB) RAM. Further, Automatic Mixed Precision (AMP) was activated during the training phase to accelerate the training process by reducing the training time as well as memory requirements as dictated in [31]. Overall accuracy, sensitivity and specificity were employed as performance measures.

4. Results

4.1. Auto-Encoder Performance

Figure 2 depicts the auto-encoder performance in terms of learning curve and images reconstruction. As

mentioned previously, the main judgment criterion on the auto-encoder performance is its performance on the validation set with respect to the development loss and the quality of the reconstructed images. It can be clearly seen from Figure 2-a) that the development loss began to stabilize after the third epoch while the lowest development loss was achieved at the fourth epoch and the remaining epochs showed no significant improvement. On the other side, the quality of the reconstructed image as shown in Figure 2-c) is similar to the input image as shown in Figure 2-b) to a large extent. And thus, having these two criteria satisfied, the weights can be transferred from the encoder to the classification model.

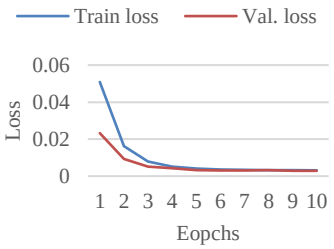
Table 5. Hyperparameters of research models.

Setting	Auto-encoder	SSL ResNet34 model	TL ResNet34 model
Batch size	16	16	16
Learning rate	1e-4	5e-5	5e-6
Weight decay	1e-5	1e-5	1e-5
LR scheduler	StepLR	StepLR	StepLR
LR Scheduling factor	0.5	0.5	0.5
LR Scheduling step	2	1	1
Loss function	MSE	CrossEntropy	CrossEntropy
Optimizer	Adam	Adam	Adam
Max epochs	10	20	20

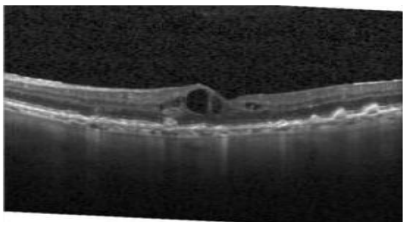
4.2. Overall Performance Evaluation

Figure 3 summarizes the training and validation history in terms of losses and accuracy for both models. With respect to the SSL pre-trained ResNet34 model, it can be observed from Figure 3-a) that the model achieved faster convergence on the validation set in which it nearly reached the optima after three epochs; while the best validation loss was achieved on the fifth epoch with approximately (0.0883) cross-entropy loss before beginning to stabilize with tiny steps on both training and validation sets. In addition, the same behavior is nearly held in terms of accuracy measure for both training and validation sets as shown in Figure 3-b) with highest accuracy score of (97.1%) achieved on the validation set. On the other side as shown in Figure 3-c) and Figure 3-d), the TL pre-trained ResNet34 model began to converge on the validation set right after the seventh epoch with a slightly fluctuated behavior before achieving the best validation loss at the eleventh epoch with nearly (0.176) cross-entropy loss. In regard to the accuracy, the maximum accuracy score achieved on the validation set is (94%) achieved at the fifteenth epoch.

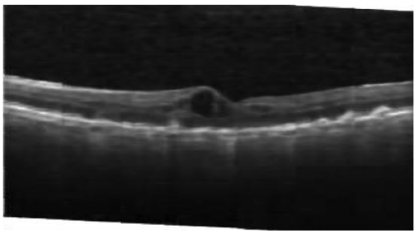
With respect to the research models' performance in terms of accuracy, sensitivity and specificity on the test set, both models were able to achieve performance evaluation metrics greater than (90%) as shown in Table 6. More clearly, SSL ResNet34 model achieved accuracy, sensitivity and specificity scores of (95.2%), (95.2%) and (98.4%), while TL ResNet34 model achieved scores of (90.7%), (90.7%) and (96.9%) respectively.



a) Learning curve.



b) Sample input image.



c) The reconstructed image.

Figure 2. Auto-encoder performance in terms learning curve and reconstruction.

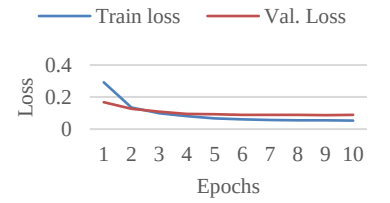
Table 6. Research models' performance summary on the test set.

Model	Acc. (%)	Sens. (%)	Spec. (%)
SSL ResNet34 model	95.2	95.2	98.4
TL ResNet34 model	90.7	90.7	96.9

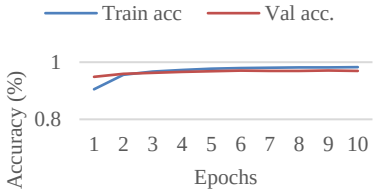
4.3. Per Class Performance Evaluation

Table 7 shows the sensitivity and specificity of both models at the class level. It can be clearly observed from Table 7 that both models achieve better performance in terms of specificity in comparison to the sensitivity measure on the level of the four labels, namely, CNV, DME, DRUSEN, and NORMAL with specificity scores of (94.4.2%), (99.5%), (99.7%) and (100%) respectively for the SSL ResNet34 model as well as specificity scores of (91.5%), (98.7%), (100%) and (97.5%), respectively, for the TL ResNet34 model.

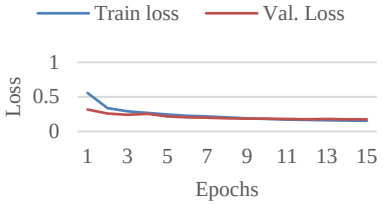
On the sensitivity side, SSL ResNet34 model was able to achieve sensitivity scores of (99.2%), (98%), (84.4%) and (98.8%) for CNV, DME, DRUSEN and NORMAL respectively. On the other side, the TL ResNet34 model was able to achieve sensitivity scores of (91.5%), (98.7%) (100.0%) and (97.5%) following the same previous order.



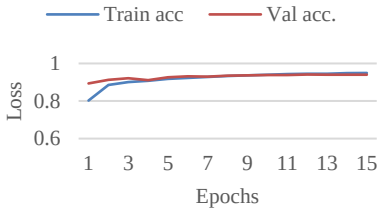
a) SSL learning curve.



b) SSL accuracy curve.



c) TL learning curve.



d) TL accuracy curve.

Figure 3. Research models performance history in terms of loss and accuracy.

Table 7. Class-specific performance measures evaluation on the test set.

Measure	CNV	DME	DRUSEN	NORMAL
SSL ResNet34 model				
Sensitivity (%)	99.2	98	84.8	98.8
Specificity (%)	94.4	99.5	99.7	100.0
TL ResNet34 model				
Sensitivity (%)	99.6	98	66.4	98.8
Specificity (%)	91.5	98.7	100	97.5

5. Discussion and Limitations

This research provided a deep learning based framework for classification purposes of four human eye retina medical conditions including CNV, DME, DRUSEN and NORMAL situation using retina OCT scans. The research methodology employs convolutional auto-encoder as a self-supervised pre-training method that is used to initialize the classification model, while residual neural network (ResNet34) was employed as the research model. Further, the obtained results by applying SSL approach was compared with ResNet34 model pre-trained on ImageNet dataset with partially frozen weights from the initial layers.

Indeed, the nature of the medical images differs significantly from any other images category. And thus, using pretrained models on dataset such as ImageNet may not be a convenient solution in the medical sector. This is because working with medical problems require achieving higher performance to ensure the robustness of the decisions made upon the predictive model outcomes. In addition, training a convolutional model from scratch requires longer training time, large amount of data and more tuning efforts until achieving considerable results which is a relatively a difficult issue in the medical sector. Hence, employing SSL approach as a pre-training method played a considerable role in capturing the visual features in the retina OCT images as well as improving the performance of the predictive model in comparison to the transfer learning based model.

To gain a clear vision, SSL ResNet34 model was able to achieve better performance in terms of the overall accuracy, sensitivity and specificity with values of (95.2%), (95.2%), and (98.4%) respectively in comparison to the TL ResNet34 model which achieved performance scores of (90.7%), (90.7%), and (96.9%) respectively.

Additionally, from a computational perspective, employing SSL pretraining approach helped significantly in reducing the total number of training epochs in which it required only (20) epochs to achieve its optimum performance for training both auto-encoder as well as the classification model. Whereas the pretrained model on ImageNet dataset which is trained for (90) epochs on the ImageNet dataset in addition to additional (15) epochs on the retina OCT dataset which in total accounts for (105) epochs. The main point to mention here is that SSL approach enabled training relatively large model from scratch with limited amount of data with better performance than transfer learning approach as well as less computations.

With respect to the model performance on the level of medical condition in terms of sensitivity and specificity, SSL ResNet34 model achieved higher scores than the TL ResNet34 model, thus, it will be considered for further interpretation. In regard to the specificity scores, NORMAL and DME DRUSEN conditions achieved the highest specificity scores with percentage of (100%) and (99.7%) respectively followed by DRUSEN medical condition with percentage of (99.5%) and finally CNV medical condition with percentage of (94.4%). Technically, Specificity score with respect to a certain class refers to the percentage of cases not belonging to the same class and were correctly identified. Equally important, having a high specificity score with respect to a certain

class indicates less chance of false positive occurrence. As an example on CNV medical condition, out of all cases which do not actually have CNV disorder only (5.6%) were identified as having CNV whereas in reality those cases belong to DME (0.67%) and DRUSEN (4.93%).

On the other side, sensitivity score is an interesting measure in the medical field that represents the percentage of cases that belong to a certain medical condition which has been correctly identified. More clearly, less false negative occurrence is associated with higher sensitivity score. Pertaining the SSL ResNet34 model performance in terms of sensitivity scores, CNV condition achieved the highest sensitivity of (99.2%) followed by NORMAL and DME with percentages of (98.8%) and (98.0%) respectively, while the least sensitivity score was attributed to DRUSEN medical condition with percentage of (84.8%). To elaborate more, (0.8%) out of all CNV cases were incorrectly classified as DME. Similarly, (2%) out of all DME cases were incorrectly classified as CNV (0.69%). In addition, (15.2%) out of all DRUSEN cases were incorrectly classified as CNV (14.8%) and DME (0.4%). Lastly, (1.2%) out of all NORMAL cases were incorrectly classified as DME (0.2%) and DRUSEN (0.8%).

6. Conclusions

This research has compared the performance of SSL and TL approaches to train ResNet34 architecture. The proposed approach was tested on an OCT dataset that contains (109, 309) images with three medical conditions including CNV, DME, DRUSEN and NORMAL condition. Further, the research methodology employs convolutional auto-encoder as an SSL pretext task. The main research outcomes showed that:

1. SSL pretraining approach proved its effectiveness in training relatively a large model such as ResNet34 from scratch.
2. The performance of the SSL ResNet34 model outperformed TL ResNet34 model pretrained on ImageNet dataset.
3. The overall accuracy, sensitivity and specificity on the test set achieved by SSL ResNet34 model are (95.2%), (95.2%) and (98.4%) respectively, while the TL ResNet34 model was able to achieve scores of (90.7%), (90.7%) and (96.9%) respectively.
4. SSL pretraining approach achieved the best performance from a computational perspective in which it required only (20) epochs to reach the optimum performance which is not achieved by the other methodologies on the same dataset.

As a future research direction, this research can be further improved by employing the same pre-training approach to different convolutional architectures such as residual neural networks variants (50, 101, 152), DenseNet Variants and InceptionNet variants to ensure its capability for generalization with respect to the model.

References

[1] Albarrak A., Coenen F., and Zheng Y., "Age-related Macular Degeneration Identification in Volumetric Optical Coherence Tomography Using Decomposition and Local Feature Extraction," *in Proceedings of the International Conference on Medical Image, Understanding and Analysis*, Aberdeen, pp. 59-64, 2013.

[2] Alqudah A., "A Oct-Net: a Convolutional Network Automated Classification of Multiclass Retinal Diseases Using Spectral-Domain Optical Coherence Tomography Images," *Medical and Biological Engineering and Computing*, vol. 58, no. 1, pp. 41-53, 2020.

[3] Alsaih K., Lemaitre G., Rastgoo M., Massich J., Sidibé D., and Meriaudeau F., "Machine Learning Techniques for Diabetic Macular Edema (DME) Classification on SD-OCT Images," *Biomedical Engineering Online*, vol. 16, no. 1, pp. 68, 2017.

[4] Apostolopoulos S., Ciller C., De Zanet S., Wolf S., and Sznitman R., "Retinet: Automatic Age-related Macular Degeneration Identification in OCT Volumetric Data," *Investigative Ophthalmology and Visual Science*, vol. 58, no. 8, pp. 387, 2017.

[5] Auw-Haedrich C., Staubach F., and Witschel H., "Optic Disk Drusen," *Survey of Ophthalmology*, vol. 47, no. 6, pp. 515-532, 2002.

[6] Bhagat N., Grigorian R., Tutela A., and Zarbin M., "Diabetic Macular Edema: Pathogenesis and Treatment," *Survey of Ophthalmology*, vol. 54, no. 1, pp. 1-32, 2009.

[7] Chen L., Bentley P., Mori K., Misawa K., Fujiwara M., and Rueckert D., "Self-supervised Learning for Medical Image Analysis Using Image Context Restoration," *Medical Image Analysis*, vol. 58, pp. 101539, 2019.

[8] Deng J., Dong W., Socher R., Li L., Li K., and Fei-Fei L., "ImageNet: A Large-Scale Hierarchical Image Database," *in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, pp. 248-255, 2009.

[9] Drexler W. and Fujimoto J., "State-of-the-Art Retinal Optical Coherence Tomography," *Progress in Retinal and Eye Research*, vol. 27, no. 1, pp. 45-88, 2008.

[10] Dumar A. and Arias J., "Choroidal Neovascularization: OCT Angiography Findings," https://eyewiki.aao.org/Choroidal_Neovasculariz

ation:_OCT_Angiography_Findings, Last Visited, 2023.

[11] Ellahham S., Ellahham N., and Simsekler M., "Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges," *American Journal of Medical Quality*, vol. 35, no. 4, pp. 341-348, 2020.

[12] Fang L., Jin Y., Huang L., Guo S., Zhao G., and Chen X., "Iterative Fusion Convolutional Neural Networks for Classification of Optical Coherence Tomography Images," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 327-333, 2019.

[13] Farsiu S., Chiu S., O'Connell R., Folgar F., Yuan E., Izatt J., and Toth C., "Quantitative Classification of Eyes with and without Intermediate Macular Degeneration Using Optical Coherence Tomography," *Ophthalmology*, vol. 121, no. 1, pp. 162-172, 2014.

[14] Goodfellow I., Bengio Y., and Courville Y., *Deep Learning*, The MIT Press, 2016.

[15] Hameed S., Ashraf M., and Ya-nan Q., "Multi-lingual Language Variety Identification Using Conventional Deep Learning and Transfer Learning Approaches," *The International Arab Journal on Information Technology*, vol. 19, no. 5, pp. 705-712, 2022.

[16] He K., Zhang X., Ren S., and Sun J., "Deep Residual Learning for Image Recognition," *in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 770-778, 2016.

[17] Hussain M., Bhuiyan A., Luu C., Smith R., Guymer R., Ishikawa H., Schuman J., and Ramamohanarao K., "Classification of Healthy and Diseased Retina Using SD-OCT Imaging and Random Forest Algorithm," *PloS One*, vol. 13, no. 6, pp. 0198281, 2018.

[18] Karpathy A., "Cs231n: Convolutional Neural Networks for Visual Recognition," http://cs231n.stanford.edu/2016/, 2016, Last Visited, 2023.

[19] Ker J., Wang L., Rao J., and Lim T., "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375-9389, 2017.

[20] Kermany D., Goldbaum M., Cai W., Valentim C., Liang H., Baxter S., McKeown A., Yang G., and et al.,"Identifying Medical Diagnoses and Treatable Diseases by Image-based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122-1131, 2018.

[21] Kermany D., Zhang K., and Goldbaum M., "Large Dataset of Labeled Optical Coherence Tomography and Chest X-Ray Images," *Mendeley Data*, vol. 172, no. 5, pp. 1122-1131, 2018.

[22] Koh J., Acharya U., Hagiwara Y., Raghavendra U., Tan J., Sree S., Bhandary S., Rao A., Sivaprasad S., Chua K., Laude A., and Tong L., "Diagnosis of Retinal Health in Digital Fundus Images Using Continuous Wavelet Transform and Entropies," *Computers in Biology and Medicine*, vol. 84, pp. 89-97, 2017.

[23] LeCun Y., Bengio Y., and Hinton G., "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[24] Lee C., Baughman D., and Lee A., "Deep Learning is Effective for Classifying Normal Versus Age-related Macular Degeneration OCT Images," *Ophthalmology Retina*, vol. 1, no. 4, pp. 322-327, 2017.

[25] Lemaitre G., Rastgoo M., Massich J., Sankar S., Mériaudeau F., and Sidibé D., "Classification of SD-OCT volumes with LBP: application to DME Detection," *in Proceedings of the Ophthalmic Medical Image Analysis International Workshop*, Munich, pp. 9-16, 2015.

[26] Lemaître G., Rastgoo M., Massich J., Cheung C., Wong T., Lamoureux E., Milea D., Mériaudeau F., and Sidibé D., "Classification of SD-OCT Volumes Using Local Binary Patterns: Experimental Validation for DME Detection," *Journal of Ophthalmology*, vol. 2016, 2016.

[27] Li F., Chen H., Liu Z., Zhang X., and Wu Z., "Fully Automated Detection of Retinal Disorders by Image-based Deep Learning," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 257, no. 3, pp. 495-505, 2019.

[28] Liu Y., Chen M., Ishikawa H., Wollstein G., Schuman J., and Rehg J., "Automated Macular Pathology Diagnosis in Retinal OCT Images Using Multi-Scale Spatial Pyramid and Local Binary Patterns in Texture and Shape Encoding," *Medical Image Analysis*, vol. 15, no. 5, pp. 748-759, 2011.

[29] Liu X., Zhang F., Hou Z., Wang Z., Mian L., Zhang J., and Tang J., "Self-supervised Learning: Generative or Contrastive," *arXiv preprint arXiv:2006.08218*, 2020.

[30] Maggipinto M., Masiero C., Beghi A., and Susto G., "A Convolutional Autoencoder Approach for Feature Extraction in Virtual Metrology," *Procedia Manufacturing*, vol. 17, pp. 126-133, 2018.

[31] Micikevicius P., Narang S., Alben J., Diamos G., Elsen E., Garcia D., Ginsburg B., Houston M., Kuchaiev O., Venkatesh G., and Wu H., "Mixed Precision Training," *in Proceedings of the 6th International Conference on Learning Representations*, Vancouver, pp. 1-12, 2018.

[32] Nugroho K., "A Comparison of Handcrafted and Deep Neural Network Feature Extraction for Classifying Optical Coherence Tomography

(OCT) Images," *in Proceedings of the 2nd International Conference on Informatics and Computational Sciences*, Semarang, pp. 1-6, 2018.

[33] Olivas E., Guerrero J., Martinez-Sober M., Magdalena-Benedito J., Serrano L., Benedito J., and Lopez A., *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Technique*, IGI Global, 2009.

[34] Perdomo O., Rios H., Rodríguez F., Otálora S., Meriaudeau F., Müller H., and González F., "Classification of Diabetes-Related Retinal Diseases Using a Deep Learning Approach in Optical Coherence Tomography," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 181-189, 2019.

[35] Rastogi D., Padhy R., and Sa P., "Detection of Retinal Disorders in Optical Coherence Tomography Using Deep Learning," *in Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies*, Kharagpur, pp. 1-7, 2019.

[36] Schlegl T., Waldstein S., Bogunovic H., Endstraßer F., Sadeghipour A., Philip A., Podkowinski D., Gerendas B., Langs G., and Schmidt-Erfurth U., "Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning," *Ophthalmology*, vol. 125, no. 4, pp. 549-558, 2018.

[37] Shah S., "Blindness and Visual Impairment due to Retinal Diseases," *Community Eye Health*, vol. 22, no. 69, pp. 8, 2009.

[38] Shiraishi J., Li Q., Appelbaum D., and Doi K., "Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging," *Seminars in Nuclear Medicine*, vol. 41, n. 6, pp. 449-462, 2011.

[39] Silvestri G., Williams M., McAuley C., Oakes K., Sillery E., Henderson D., Ferguson S., Silvestri V., and Muldrew K., "Drusen Prevalence and Pigmentary Changes in Caucasians Aged 18-54 Years," *Eye*, vol. 26, no. 10, pp. 1357-1362, 2012.

[40] Srinivasan P., Kim L., Mettu P., Cousins S., Comer G., Izatt J., and Farsiu S., "Fully Automated Detection of Diabetic Macular Edema and Dry Age-related Macular Degeneration from Optical Coherence Tomography Images," *Biomedical Optics Express*, vol. 5, no. 10, pp. 3568-3577, 2014.

[41] Sun Z., and Sun Y., "Automatic Detection of Retinal Regions Using Fully Convolutional Networks for Diagnosis of Abnormal Maculae in Optical Coherence Tomography Images," *Journal of Biomedical Optics*, vol. 24, no. 5, pp. 056003, 2019.

[42] Venhuizen F., Ginneken B., Bloemen B., Grinsven M., Philipsen R., Hoyng C., Theelen T., and Sánchez C., "Automated Age-related Macular Degeneration Classification in OCT Using Unsupervised Feature Learning," *Medical Imaging: Computer-Aided Diagnosis, International Society for Optics and Photonics*, vol. 9414, pp. 391-397, 2015.

[43] Venhuizen F., Ginneken B., Asten F., Grinsven M., Fauser S., Hoyng C., Theelen T., and Sánchez C., "Automated Staging of Age-related Macular Degeneration Using Optical Coherence Tomography," *Investigative Ophthalmology and Visual Science*, vol. 58, no. 4, pp. 2318-2328, 2017.

[44] Wilkinson C., Ferris F., Klein R., Lee P., Agardh C., Davis M., Dills D., Kampik A., Pararajasegaram R., and Verdaguer J., "Proposed International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677-1682, 2003.

[45] Xiang D., Tian H., Yang X., Shi F., Zhu W., Chen H., and Chen X., "Automatic Segmentation of Retinal Layer in OCT Images with Choroidal Neovascularization," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5880-5891, 2018.

[46] Yang X., He X., Liang Y., Yang Y., Zhang S., and Xie P., "Transfer Learning or Self-supervised Learning? A Tale of Two Pre-training Paradigms," *arXiv preprint arXiv:2007.04234*, 2020.

Saeed Shurrab completed an MSc in Data Science from Jordan University of Science and Technology with the German Academic Exchange Scholarship (DAAD). Prior to that, he completed a BSc in Industrial and Systems Engineering from the Islamic University of Gaza - Palestine in 2014. Currently, Saeed focuses on developing deep neural networks and their applications to healthcare problems.

Yazan Shannak is a machine learning engineer at Tarjama and a Masters student at Jordan University of Science and Technology. His research interest is deep learning for natural language processing and computer vision.

Rehab Duwairi is professor of computer science currently working for Jordan University of Science and Technology. Her research interests include: data science and machine learning, Arabic Language Technologies, and AI in health.