SFFT-CapsNet: Stacked Fast Fourier Transform for Retina Optical Coherence Tomography Image Classification using Capsule Network

Michael Opoku[1], Benjamin Asubam Weyori[2], Adebayo Felix Adekoya[3], Kwabena Adu[4]
Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana[1, 4]
Department of Electrical and Computer Engineering, University of Energy and Natural Resources, Sunyani, Ghana[2]
Faculty of Computing, Engineering, and Mathematical Sciences, Catholic University of Ghana, Sunyani[3]

Abstract*—The work of the Ophthalmologist in manually detecting specific eye related disease is challenging especially screening through large volume of dataset. Deep learning models can leverage on medical imaging like the retina Optical Coherence Tomography (OCT) image dataset to help with the classification task. As a result, many solutions have been proposed based on deep learning-based convolutional neural networks (CNNs). However, the limitations such as inability to recognize pose, the pooling operations which affect resolution of the featured maps have affected its performance in achieving the best accuracies. The study proposes a Capsule network (CapsNet) with contrast limited adaptive histogram equalization (CLAHE) and Fast Fourier transform (FFT), a method we called Stacked Fast Fourier Transform-CapsNet (SFFT-CapsNet). The SFFT was used as an enhancement layer to reduce noise in the retina OCT image. A two-block framework of three-layer convolutional capsule network each was designed. The dataset used for this study was presented by University of California San Diego (UCSD). The dataset consists of 84,495 X-Ray images categorized into four classes (NORMAL, CNV, DME, and DRUSEN). Experiment was conducted on the SFFT-CapsNet model and results were compared with baseline models for performance evaluation using accuracy, sensitivity, precision, specificity, and AUC as evaluation metrics. The evaluation results indicate that the proposed model outperformed the baseline model and state-of-the-arts models by achieving the best accuracies of 99.0%, 100%, and 99.8% on overall accuracy (OA), overall sensitivity (OS), and overall precision (OP), respectively. The result shows that the proposed method can be adopted to aid Ophthalmologist in retina disease diagnosis.

Keywords—Capsule network; convolution neural network; medical imaging; optical coherence tomography

I. INTRODUCTION

The concept of identifying specific medical conditions in the human body through the analysis of medical images to establish basis for the existence and growth rate of a particular disease can be very tedious and stressful. As a radiologist, one is confronted with the burden of finding and interpreting extracted features from medical images to diagnose and monitor different kinds of diseases associated with human body [1-3]. Human beings in our nature are not only slow in processing medical images but are prone to errors especially when stressed out. Considering a sensitive area like medical field, a wrong diagnose can be quite expensive as its implications can lead to unexpected consequence [4]. As a result, the attention of many researchers has been diverted into finding a substantive solution to assist the radiologists to deal with the complications confronted on daily bases as part of their work [1]. Computer aided models can help but the major challenge has been selecting the right method and acquiring good performance such as high prediction accuracy, achieving low runtime and low computational cost [5-12].

The introduction of deep learning (DL) brought a promising breakthrough especially in the field of medical science. The Artificial Neural Networks (ANN) [13-14] and Convolutional Neural Networks (CNN) [15] which are DL techniques became the most predominantly employed methods in identifying and diagnosing anomalies using radiological imaging technologies. However, both the ANN and the CNN have their benefits and limitations which usually affect performance of the model. They both require huge dataset for efficient training of models and increasing the image resolution adds up to number of trainable parameters which affects runtime and computational cost of the model [16]. According to Noord, and Postma, [17] ANN is not flexible and does not allow for easy customization of the model. Moreover, ANN also has deficiency with diminishing and exploding of gradient [17]. The CNNs framework gained more attention as a result of the high performance it can offer and its flexibility to use. The enormous computer-aided algorithms providing state-of-the-art early disease detection results for the medical imaging diagnosing tasks have depended on the building blocks of CNNs for most of the models produced [18].

However, the million unanswered yet important question is whether CNN models truly generalize. According to Gu, [19], a well-trained CNN model is still prone to adversarial attack as the network can easily be fooled by images that are carefully designed with imperceptible perturbations. According to Xi et al., [20], the CNNs face two major challenges which are lack of rotational invariance and failure to consider the spatial orientation hierarchies between features of the image. CNNs therefore, require huge dataset with different poses for same images, if one wants to achieve high performance for such classification model.

In an attempt to address these challenges, Hinton et al., [21] introduced novel type of neural network known as the capsule network (CapsNet). Sabour et al., [22] enhanced the CapsNet architecture by introducing the dynamic routing by agreement between capsules. The CapsNet used the routing by agreement to resolve the problem resulting from the pooling operations of the CNNs which affects the resolution of the feature maps. Again, the CapsNet employed what is known as the reconstruction regularization to recognize the spatial hierarchical relationships among the entity parts. Moreover, since the CapsNet is equivariant in nature, it does not require huge dataset or data augmentation like rotation and scaling during training and testing also, adversaries' attacks such as the imperceptible pixel perturbation of the CNNs is also addressed by the CapsNet.

The study therefore adopts the CapsNet to classify retina optical coherence tomography (OCT) images due to the many advantages it offers over the other DL models. The study reconstructs the CapsNet architecture to include other controlled parameters to enhance its performance significantly. To ensure effective distribution of coupling coefficient which can enhance performance accuracy and convergence, the study performs normalization using the sigmoid function [23] instead of the SoftMax [22]. The study makes the following contributions:

- Proposes new capsule networks architecture named Stacked Fast Fourier Transform capsule network (SFFT-CapsNet).
- Evaluate the proposed model on retina OCT dataset.
- Compares results of proposed model with original capsule networks and state-of-the-art deep learning models performance.
- Provides visualization of internal processing results to establish better explainability of the proposed architecture.

The rest of the study is organized in the following manner;

The Section II of the study provides insight on related works. The Section III provides the methods of the study. The Section IV presents the experimental setup and results discussion. Finally, the Section V presents conclusion and recommendation for future expansion.

II. Related Works

Every computer vision has a simple task of performing classification on given images or objects. Deep learning models have been applied to different domains like medical field for classification task. Wang et al., [24] implemented a deep learning model for automatic detection of metastatic breast cancer. The method employed enhanced the localization task. Nithya et al., [25] implemented ANN to detect kidney disease like kidney stones. Arunkumar et al. [26] implemented another algorithm based on ANN to classify the abnormal magnetic resonance (MRI) image which was used to identify brain tumor.

According to Karri et al., [27], transfer learning was easily included in CNN architecture to train on small dataset after which the results were successfully implemented on OCT image for classification of DME and dry AMD. However, due to the limitation of the CNNs mentioned earlier, the CapsNet [22] was introduced to provide better classification which also addressed most of the failures of the CNNs.

The performance of CapsNet has always been compared to CNNs [28] in many researches to help enhance the algorithm and architecture of the CapsNet. The results of the comparison also indicate that CapsNet also has its own challenges [29]. A major challenge identified by Sabour et al., [22] indicated that the CapsNet architecture attempted to extract features on every entity part found of the image which might not be necessary for classification task. Extracted information such as background information makes CapsNet implementation more vulnerable to misclassification [21]. The explanation given to this was that the shallow structure of the CapsNet implemented with single convolutional layer is not sufficient enough to extract only required features. Liu et al., [30] demonstrated in their article that the original CapsNet performed very poor on complex data due to insufficient convolutional layers required to extract better features to enhance performance. This conclusion was made when the original CapsNet was compared with their proposed model called the DDRM-CapsNet for performance efficiency. In their study, they modified the original CapsNet architecture to include more convolutional layers to ensure better feature extraction. The study also enhanced the dynamic routing mechanism to include two stages and changed the output vector to 24 dimensions. So, in all, the network had three standard convolutional layers, a primary capsule layer, two-digit capsule layers and three fully connected layers.

To improve the architecture of the CapsNet, many researchers attempted improving the algorithm to include new features. In the CapsNet architecture, information of each capsule is sent to the next available layers at the full magnitude of its activation value but still lacks an appropriate control mechanism for selecting discriminant features from the outputs of each layer [29]. This means in the routing mechanism, encoded information from one capsule to the next capsule layer is not filtered even if it contains background information or unwanted information that is not required for classification decision. Also, CapsNet introduced an initial logit b_{ij} in the routing Softmax function to represent the log prior probability of how tight the coupling of the initial capsule i is with capsule j which depended on location and type of two capsules. Nonetheless, the function instead transformed the logits of the coupling coefficients to what is known as concentrative values. This means background information can mistakenly be sent to the next capsule layer with very high coefficient that can impact large values for summation of the predicting vectors [23]. Hence, Zhao et al., [31] concluded in their article that the SoftMax function implemented for normalization process to ensure uniform assignment of probability values between capsules prevented fair distribution of coupling coefficients which affected the performance. In another research, Yang & Wang, [29] also proposed two different methods to address the issue of capsules obtaining high activation values. Their study also introduced Cubic-Increase Squash (CI – squash) and Powered Activation (PA) which was modification to the original version of the squash function. The study demonstrated information sensitiveness was a major reason why CapsNet was not achieving high performance with color background images. Again, the study concluded there was the need to restrict the capsules from achieving unreasonably high distribution of activation values as it also affected the performance. As a result, Mensah et al, [32] implemented the power squash function as the original squash is susceptible to generating high activation values. The activation values of the original capsules experienced faster growth at the initial stage leading to very high generated activation values. It was therefore important to include a sparsity which constrains the capsules from achieving such high activation values so that the capsules would be able to identify distinguishing features that are of greater interest to the classification process.

Many researchers have also tried to find different means to improve the performance of the CapsNet models to resolve the issue of the information sensitiveness and the high activation values. Some focused on reengineering the CapsNet architecture while others tried to improve the algorithm [32-35]. Nguyen and Ribeiro [36] reconstructed the vector CapsNet architecture to include more convolution layers and also varied the fully connected layers to leverage better filter input images for best feature extraction and image restoration. In another study, Xiang et al. [37] developed a multi-scale capsule network for classifying images which sorted to address the shortfalls of the original CapsNet architecture. The proposed multi-stage CapsNet included structural and semantic information on the first layer. Phaye et al., [38] enhanced extraction of the discriminative feature maps by introducing dense and diverse CapsNet. Their study replaced the convolutional layer with densely connected convolutional layer to improve the performance. In another studies by Huang et al., [39], to address the issue of vanishing gradient problem, the study introduced a dense connection between every layer. The results also indicated a significant improvement of the model. Other studies also tried to introduce residual connections in their attempt to address the vanishing gradient problem [40-41]. Bhamid & El-Sharkawy, [42] also implemented a CapsNet model which allowed the primary capsule to carry information at three different image scales. Their study dealt with complex data such as CIFAR10 which the model showed a significant improvement in the classification accuracy.

The capsule architecture has been applied to many different areas where image classification was a problem. Li et al. [43] implemented a capsule network model to recognize and monitor the growth rate of the rice crops through the images captured with unmanned aerial device. According to another article by Paoletti et al. [44], a CNN-Based-CapsNet was implemented in their study to classify remotely taken hyperspectral images. A similar study was also conducted by on hyperspectral image classification [45-46]. The overall performance also indicated a promising result than using the convolutional network. The implementation of CapsNet in the area of medical imaging has also shown very promising results. Adu et al, [47], implemented Dilated CapsNet in their research on Brain Tumor Classification Also, Afshar et al, [9] implement another CapsNet model for brain tumor classification. Koresh and Chacko, [48] also implemented CapsNet noiseless image classification algorithm which was used to classify corneal optical coherence tomography (OCT) dataset.

In our studies, we look at introducing two different enhancement layers to increase the information sensitivity to enhance the performance of the model. We also replace SoftMax activation function (AF) with sigmoid AF and reconstruct the original architecture to include six convolutional layers which were as a result of best performance from different modifications.

A. *Optical Coherence Tomography Images*

The Optical Coherence Tomography (OCT) is an accepted standard clinical practice diagnostic imaging technique for diagnosing retinal diseases. The results of this method provide an OCT image with possible features enough to provide sufficient visualization for detecting if there is a change in the retinal vessels from the imprint of the OCT film. The evaluation parameter is usually to identify if there is an increase or decrease in the retina layer [26]. The OCT is usually employed to acquire very high-resolution cross-sectional images from the retina. It uses low-coherence light to capture two and three-dimensional images from optical scattering media like biological tissue. The OCT can be used to capture large number of images through which the level of deterioration of the optic nerves can be determined with time.

The method is very easy to implement and does not involve any ionizing radiation [49]. This makes it possible to differentiate between the various layers of the retina so that specific diagnose can be established based on the measurement of the retina thickness to detect a retinal disease. The OCT now serve as a baseline retinal assessment and popular choice capturing retinal image for clinical practice before therapy session is initiated [50-51]. Ophthalmologist might depend on the results of the OCT image to select a choice of treatment for their patient. Therefore, emphasis on the increasing essential role of the OCT imaging cannot be overlooked. There are so many diseases that can be diagnosed using OCT imaging. Disease like diabetic retinopathy which is as a result of damage to blood vessels of retina, Macular pucker, Glaucoma, Macular hole, Age-related macular degeneration, Drusen, Central serous retinopathy, Macular edema, Vitreous traction, and Optic nerve abnormalities other macular and other related diseases are visible to OCT images. There are many eye diseases resulting from deteriorating of these retina cells. The focus of this study is centered on classification of three major Macular diseases. According to research age is a major risk factor associated with macular disease. The common diseases that may affect the healthiness of the macula are age-related macular degeneration [AMD], choroidal neovascularization (CNV) and diabetic macular edema (DME) [50] [52-53].

B. *The CapsNet Architecture*

The concept of CapsNet was first introduced in the Transforming Auto-Encoders in 2011 by Hinton et al., [21]. In their article, it was concluded that the CNN loose valuable information such as pose and spatial relationships among features maps due to the Max pool operation. As a result, the concept of the Transforming Auto-Encoders which used capsules to encode entities instead of neuron was introduced to keep the meaningful information that could influence the output of the classification task. Nonetheless, the concept did not receive any attention until the dynamic routing by agreement using CapsNet [22] was introduced. The CapsNet uses capsules to encode entities such that each capsule is represented as activity vector to indicate an instantiation parameter (e.g. Texture, color, angle etc.) of a specific entity. The activity vector elements encode the properties that represent the entity and their activation vector indicates a probability of pose that an instantiation feature of an entity exist within its limited domain. This means the direction of the capsules indicates detailed characteristics of the features and the length of the capsule indicates its probability of existence of different features. The CapsNet architecture can be represented as an inverse computer graphic [22]. The implementation of the CapsNet in many research have always resulted in producing high accuracy [54-55].

In its operation, the CapsNet takes input vector from the lower capsule layer and multiply them by the weight matrices and a coupling coefficient. The CapsNet is seen as an equivariant in nature and therefore is able to recognize pose such as size, position and direction as well as varieties of features that have been randomly placed. In an article presented by Lenssen et al., [56], it was indicated that the CapsNet does not only represent equivariance of the characteristics of the entity type but can also signify invariance of the existence probability. The design of the CapsNet is basically meant for classification of images through feature extraction like that of CNN. This brands the CapsNet as an efficient platform when it comes to dealing with establishing spatial relationship and dealing with hierarchical data. The meaningful information is stored at different levels of capsules and the higher the levels the greater the information they can accumulate. The various levels can be seen as lower level which represent the primary capsule and the higher level which represent the digitCapsule. The capsules become activated when certain conditions are satisfied.

Through the dynamic routing process which implements routing by agreement mechanism, each active capsule must select another capsule from the next layer that the features captured can be passed to. The process ensures that the lower-level capsules agree on a feature before it is sent to the higher-level (digit Capsule). Through the training process, each capsule is able to capture certain features or characteristics of the image and the assumption is that the capsule is activated if there are properties of the image required by the higher layer for which the capsule must respond. The routing process is a major feature in the CapsNet. Fig. 1 shows the Dynamic routing process. The dynamic routing process is implemented to update weights between capsules from one layer to next which allows characteristics or properties captured by lower node capsules to be propagated to the next suitable capsule at the upper layer. If the prediction matches the higher-level capsule's output, then the coupling coefficient for these two capsules is increased. Let ui be the output of capsule i and its prediction from parent capsule j is expressed as:

$$S_j = \sum_i c_{ij}\hat{u}_{j|i} \tag{1}$$

A nonlinear function is used to shrink long and short vectors to 1 and 0 respectively. Eq. (2) shows the non-linear squash function.

$$v_j = \frac{\|s_j\|^2}{1+\|S_j\|^2}\frac{s_j}{\|S_j\|} \tag{2}$$

where *sj* in Eq. (6) is the input vector to the *jth* capsule and *vj* is the output vector. CapsNet adopts non-linearity squashing function on output vectors (*vj*) in each iteration [11]. This shows the likelihood of the vector between 0 and 1, which means that it squashes small vectors and maintains long vectors in the unit length.

$$\{v_j \approx \|s_j\|s_j = 0 \;\; v_j \approx \frac{\|s_j\|}{\|s_j\|} \tag{3}$$

The log probabilities are updated in the routing process based on the agreement between *vj* for the fact that the agreement between two vectors will be increased and have a large inner product. Therefore, agreement *aij* for updating the log probability and coupling coefficient is defined as:

$$a_{ij} = v_j\hat{u}_{j|i} \tag{4}$$

Capsule *k* in the last layer is connected with a loss *lk*. This puts a big loss value on capsules with long output instantiation parameters when the entity does not exist. The loss function *lk* is expressed as follows:

$$l_k = T_k max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k)max(0, \|v_k\| - m^-)^2 \tag{5}$$

where *Tk* is 1 when class *k* is present, and is 0 otherwise. The *m*+, *m*−, and *λ* are hyperparameters that are set before the learning process.

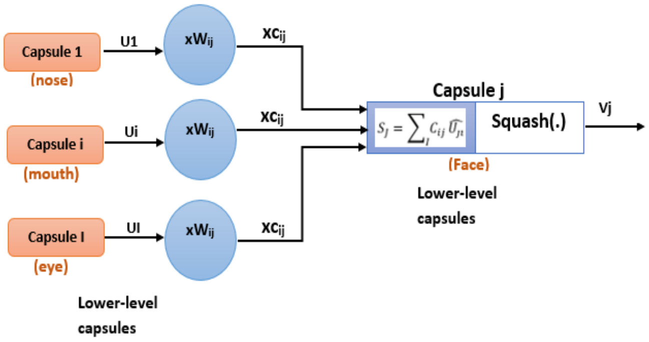


Fig. 1. Dynamic routing process.

III. Proposed Method

The main objective of the study is to design a capsule network model that include image processing tools to ensure textural enhancement for better feature extraction to improve performance. Increasing the depth of the model too much can lead to overfitting, so the study carefully introduced two blocks of three convolutional layers through several model modifications to address the insufficiency nature of the convolutional layer in the original architecture. This was implemented while ensuring the model does not become too complex which can lead to overfitting. The study employed two different enhancement techniques. The network after reconstruction makes the model more sensitive to input image and still suppresses overfitting as we introduce a dropout technique by setting the early-stopping hyperparameter (patience) to 10 epochs if validation loss does not improve during training to save only best models [59]. After exploring several model modifications, we arrived at the following conclusions:

- Implementation of contrast limited adaptive histogram equalization enhancement (CLAHE) technique and Fast Fourier Transform (FFT) enhancement to reduce noise in the various input images for better textural feature extraction whiles reducing number of trainable parameters.
- Power Squash: The study adopted the power version of the original squash function $||Vj||^n \frac{Vj}{||vj||}$ based on [30] [32]. The power squash is able to suppress smaller activation values (see Fig. 2).
- Sigmoid Activation: The sigmoid [23] activation function improved the distribution of the coupling coefficients leading to the overall improvement of the model performance. Even though the original CapsNet used the SoftMax function which was believed to constrain the Cij within a smaller interval [31]. The goal is to acquire a coefficient that is sufficient enough to produce large values for better distribution. This way, it will be able to establish relevant features that are required by prediction vectors. Based on experimental results obtained from the different model modifications, there is a clear indicated that sigmoid activation function improved the convergence and overall accuracy of the model.
- Loss Function: The study introduced a loss function called E-swish to enhance the performance of the model instead of using existing activation function. This was an enhanced version of the original swish function. We compare the performance of our activation function with RELU and from the experimental results our function outperformed the existing baseline activation functions.

Power Squash: The study adopted the power version of the original squash function based on [29] [32]. The power squash has capability to compress short vectors to almost zero length whiles extending the long vectors to a value slightly below one. However, the original squash function generated high activation values for smaller $||sj||$ which intend generated very high activation values that are not sufficient to maintain high information sensitivity for the CapsNet model. Therefore, sparsity is required to constrain the capsules from achieving high activation values. Sparsity as explained by [32] is employed to differentiate and consider highly discriminant capsules that can extract required information from complex images especially ones with varied backgrounds. Fig. 2 shows the original squash function verses the suppressed small values of the power squash function. The power squash however, introduced sparsity to the model by controlling how the initial activation values are computed by the primary capsule. Therefore, for high values of $n,$ the function experience very slow values at the initial stage and appreciated as the $||sj||$ increased. This was not the case of the original squash function which experienced a sharp increase at the initial stage as indicated in the Fig. 2.

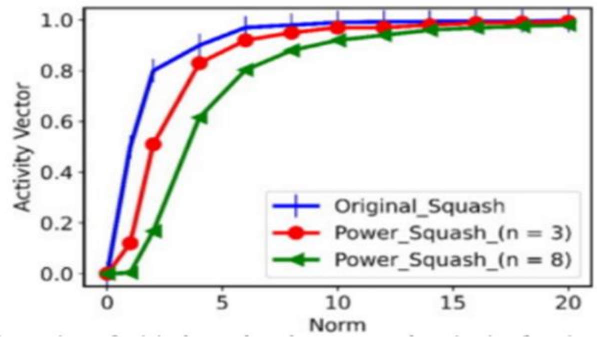


Fig. 2. The power squash verses the original squash.

A. *Contrast limited Histogram Equalization and Fourier Transform*

The study employed Contrast Limited Adaptive Histogram Equalization and Fourier Transform techniques to enhance the contrast of textural features and spatial patterns of the various input images. The conclusion for employing the two enhancements strategy was due to the results from the best performance of many model modifications. The method ensures visibility of the features for better feature extraction. The two enhancement methods have been deployed separately in many different researches due to their flexibility in implementation with low computation work load. The contrast limited adaptive histogram equalization (CLAHE) has mostly been deployed in research to reduce noise and improve the color of the x-ray images. It works best for all biomedical images by removing noise that can lead to miss classification. Histogram equalization produces over brightness of the input image which makes it difficult for the model to identify most required patterns and hidden objects.

Therefore, the CLAHE is employed to amplify the contrast of the image and limit neighboring pixel's procedure to reduce noise on the image. On the other hand, the Fourier Transform allow for images to be represented as a sum of complex exponentials of broad range frequencies. It has been implemented successfully in image processing applications such as enhancement, restoration or compression. It's usually employed as a processing tool to decompose images into its sine and cosine components. The output can then be represented in a Fourier or frequency domain when their input images are represented in a spatial domain equivalent. The Fourier domain allows each point to represent specific frequency in the spatial domain. It is best known for proving geometric characteristics of the spatial domain image as the images are decomposed into a sinusoidal component which makes it more flexible to analyze or process.

B. *The Dataflow Analysis of the Model*

Data augmentation was done by resizing the dataset images due to varied sizes of 1024x1050, 784x950, and 800x1020. Though the amount of retina OCT dataset used in this study was small however, CapsNet does not require huge dataset for training a model compared to CNNs. Fig. 3 shows summary diagram depicting the dataflow of the model.

C. Model Architecture

The paper proposes a capsule network model named Stacked Fast Fourier Transform CapsNet (SFFT-CapsNet). Fig. 4 illustrates the proposed Stacked Fast Fourier Transform CapsNet (SFFT-CapsNet) architecture. The SFFT-CapsNet consists of two blocks of convolutional layers. The convolutional layer block 1 consists of CLAHE layers, three convolutional layers, and batch normalization whereas the convolutional block 2 consists of Fourier transform layers, three convolutional layers, and batch normalization. The model consists of a Primary Capsule layer and a classification layer (retinaCaps). The process begins with passing all the input images with the dimension of 48x48x3 through CLAHE layer and the Fourier transform layer in the two blocks. The Fourier transform and CLAHE layer are image enhancement layers and therefore does not contribute additional parameters to the model. The output feature maps are forwarded to convolutional layers followed by batch normalization layers. After the input image goes through the enhancement layers, the output feature maps from the enhancement layer will still have the same dimension of 48 × 48 × 3 as the input image.

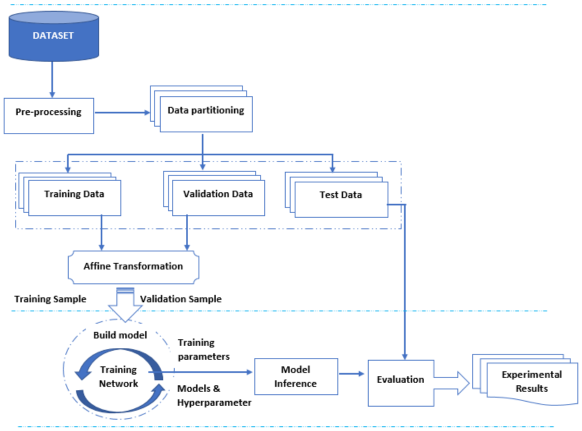


Fig. 3. Data flow diagram of the proposed model.

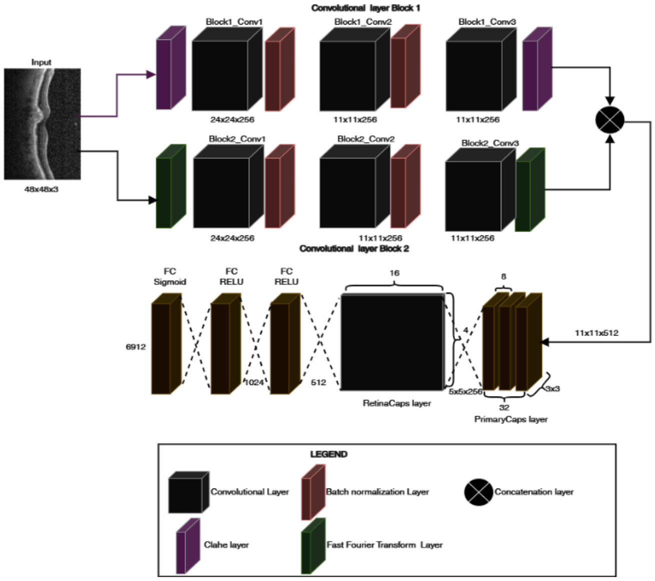


Fig. 4. Proposed Stacked Fast Fourier Transform CapsNet architecture.

The output 48x48x3 feature maps are sent to Block1_Conv1 and Block2_Conv1 with filter of 256, kernel size of 3x3, and stride of 2 which gives output feature map of 24x24x256. These output feature maps are forward to the next convolutional layers which are Block1_Conv2 and Block2_Conv2 with the filters of 256, kernel sizes of 3x3, and strides of 2. This will convolve to feature maps of 11x11x256. Again, these feature maps of 11x11x256 are sent to Block1_Conv3 and Block2_Conv3 with filters of 256, kernel sizes of 1x1, and strides of 1 to produce the feature maps of 11x11x256. These feature maps from Block1_conv3 and Block2_Conv3 are concatenate to produce the feature map of 11x11x512 for the next layer. The feature maps from the feature extractions layers are forward to a PrimaryCaps with filter 256, kernel size of 3x3, and stride of 2 which will obtain output feature map of 5x5x256. At the PC layer, a tensor product between u and the weights (W) produces uˆj|i made up of 576 (i.e., 4 × 4 × 16), 8-dimensional vectors. At the Digit Caps layer, the Recognition Caps will form k, 16D vectors, where k = number of classes. There are three fully connected (FC) layers in the decoder network consisting of 512, 1024, and 6912 neurons in the first, second, and third layers respectively.

D. Experimental Settings

This practical aspect of the study was deployed using a Windows system with NVIDIA 394 GeForce GTX 1650 6GB GPU. The codes which used TensorFlow as the backend was implemented via Keras Libraries and python (Anaconda). Both the proposed CapsNet model (CLAHE-FT) and the original CapsNet were trained for 100 epochs respectively in order to compare their performance. The batch size of the input images was set to 32 while the learning rate was maintained at 0.0001. Through the deployment process, the study made use of the Adams algorithm with momentum as the gradient optimizer. The momentum was adjusted to 0.9 whiles the descent rate was set 10-6. To avoid overfitting as part of the reconstruction process, the early stopping hyperparameter thus patience was set to 10 during the training so the algorithm can only save the best model. To complement the reconstruction layer (FC) in avoiding overfitting, we set the early stopping hyperparameter; patience, to 10 during training and saved only the best model. The extracted code implemented is a modified code which is available at https://github.com/XifengGuo/CapsNet-Keras.

E. Dataset

The dataset was made up of 84,495 x-ray images (jpeg) when it was downloaded from Kaggle.com. The folder contained subfolders which each contained specific category of images. In all, there were four categories of images which have been sampled into directories as Normal, CVN, DME and DRUSEN. However, the dataset had imbalance classes and since CapsNet could work with small size dataset, we decided to make it balanced for fare distribution by reducing the size of the various classes with large dataset. The images have been labeled as follows; (disease type)-(patient ID)-(image number of the patient). Again, these were OCT images that have been selected from retrospective cohorts of adult patients from institutions such as Shiley Eye Institute at University of California San Diego, California Retinal Research Foundation, Medical Center Ophthalmology Associates, Shanghai First

People's Hospital, and Beijing Tongren Eye Center. It took barely four years to make such selections which were between the year 2013 and 2017. The inclusion criteria for eligible images were performed by different levels of trained and expect graders with enough experience to verify and establish correct data labels of the images into their respective classes. The first groups of graders were made up of undergraduate and medical students who had successfully passed an OCT interpretation course review.

These first graders were able to initiated quality control and excluded OCT images containing critical artifacts or significant image resolution reductions. Four ophthalmologists with a lot of experience were the second graders who independently graded the image that had passed the first grading. These second graders had a primary task of recording the present or absent of specific disease on each OCT scan. CNV, macular edema, drusen, and other pathologies which are present or absent on the OCT scan were recorded. The third group of graders consisted of two senior independent retinal specialists. Each specialist has over 20 years of clinical retinal experience, who varied the true label of the images. The Images that were imported into the database started with a label matching the recent diagnosis of the patient. The sample dataset selection is illustrated in a CONSORT-style as indicated in Fig. 5

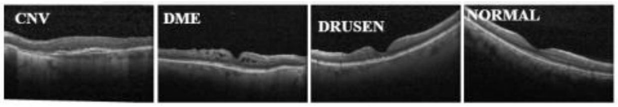


Fig. 5. Sample retina OCT image representing different classes of images.

IV. Experimental Results

This section presents the results of the proposed model used on the retina OCT image dataset. The SFFT-CapsNet model was established based on different modifications. We then compare the results to the original CapsNet by Sabour et al., [22] which has been serving as the baseline for most models in CapsNet. We evaluated the model by conducting comparative analysis on performance-accuracy with the current state-of-the-art models which have compelling efficient results on the same retina OCT dataset. This comparison is conducted to establish the best model for classification of the retina OCT images.

The study also establishes the efficiency of proposed model and its ability to generalize by comparing and visualizing the clusters formed through the routing process. An evaluation matrix such as accuracy (ACC), sensitivity (SE), precision (PR), specificity (SP), confusion matrix, receiver operating characteristic-area under ROC curve (ROC-AUC) are used to examine the performance of the models. The Precision and Recall were used to evaluate the model in order to achieve the Receiver Operating Characteristics (ROC) as well as the Precision Recall curves. The overall accuracy (OA), overall sensitivity (OS) and overall precision (OP) are also calculated. The computation of the OA for instance is by finding the average of the total accuracy scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq.(6).

Thus,

$$OA = \frac{correct\ classified\ classes}{total\ number\ of\ classes} \quad (6)$$

The computation of the OS for instance is by finding the average of the total sensitivity scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq. (7).

$$OS = \frac{total\ sensitivity\ for\ the\ four\ classes}{\text{total number of classes}} \quad (7)$$

The computation of the OP for instance is by finding the average of the total Precision-recall scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq. (8).

$$OP = \frac{total\ precision-recall\ for\ the\ four\ classes}{\text{total number of classes}} \quad (8)$$

The results of the comparison between the original CapsNet and the SFFT-CapsNet is presented in Table I. The results indicated that the SFFT-CapsNet obtained overall accuracy of 99.0% which establishes a very high performance of the model over the original CapsNet which had an overall accuracy of 94.2% when applied to the retina OCT images. The performance of the model also indicated an overall sensitivity (OS) of 100% and overall precision of 99.8% for the SFFT as against the original CapsNet which had 94.5% and 97.0% respectively. Fig. 6 shows a histogram representing the comparison of accuracies based on the overall accuracies of OA, OS, and OP. Fig. 7 shows the validation and loss accuracy. Also, Fig 8 and 9 shows the Confusion Matrix and the ROC-AUC, respectively. Fig. 10 shows the Precision and Recall curves as applied on the dataset.

TABLE I. Comparison of Results of the SFFT-CapsNet Model and Original CapsNet

Method	Classes	ACC (%)	SE (%)	PR (%)	SP (%)	AUC (%)	OA (%)	OS (%)	OP (%)
Original CapsNet [22]	CNV	95.5	97.5	99.0	100	100	94.2	94.5	97.0
	DME	95.0	94.3	97.0	95.8	99			
	DRUSEN	98.8	88.2	96.0	97.6	99			
	NORMAL	87.6	98.2	96.0	98.2	97			
SFFT CapsNet [Ours]	CNV	100	100	100	100	100	99.0	100	99.8
	DME	98.8	100	100	100	100			
	DRUSEN	100	97.2	100	100	100			
	NORMAL	97.1	98.7	99.0	100	100			

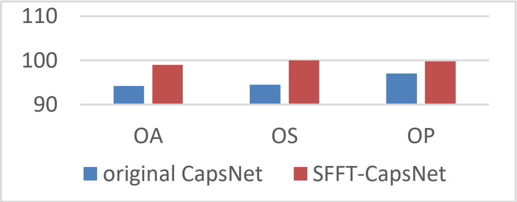
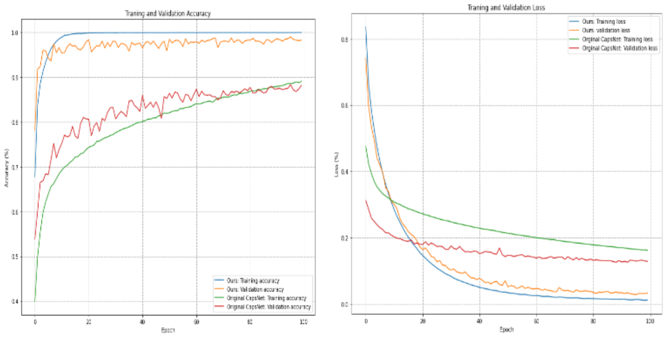
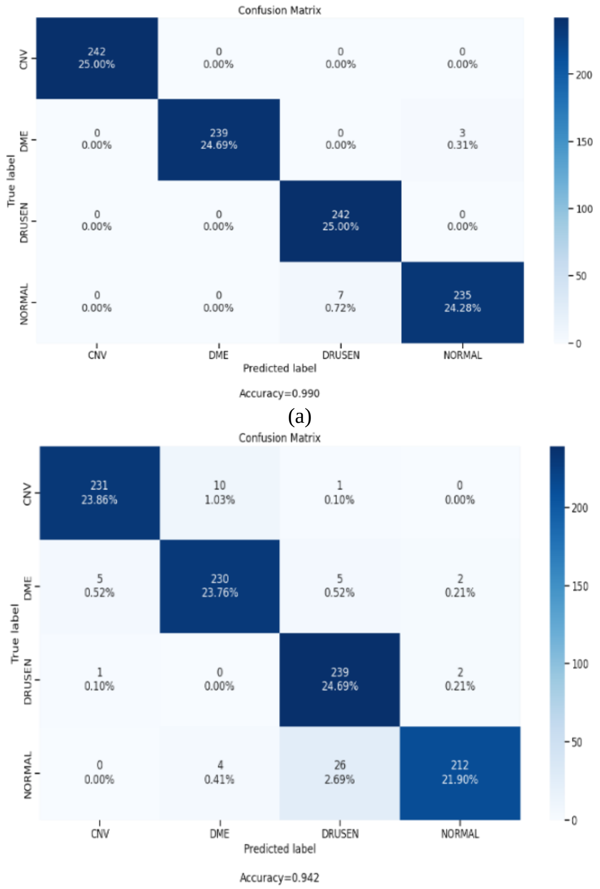


Fig. 6. Histogram comparing accuracies based on the overall accuracies of OA, OS, and OP.



(a) (b)

Fig. 7. Training, validation accuracy and loss curve on retina OCT images. (a) Training and validation accuracy curves of SFFT-CapsNet and original CapsNet., and (b) Training and validation loss curves of SFFT-CapsNet and original CapsNet.

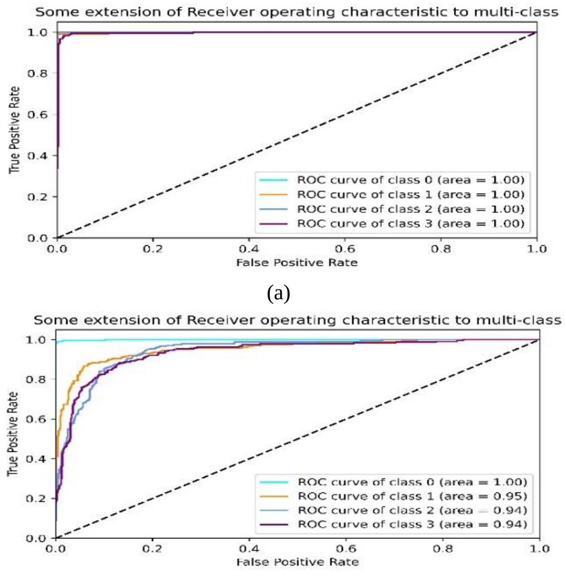


(b)

Fig. 8. Comparison of the proposed SFFT-CapsNet model and original CapsNet based on confusion matrix. (a) SFFT-CapsNet and (b) Original CapsNet.

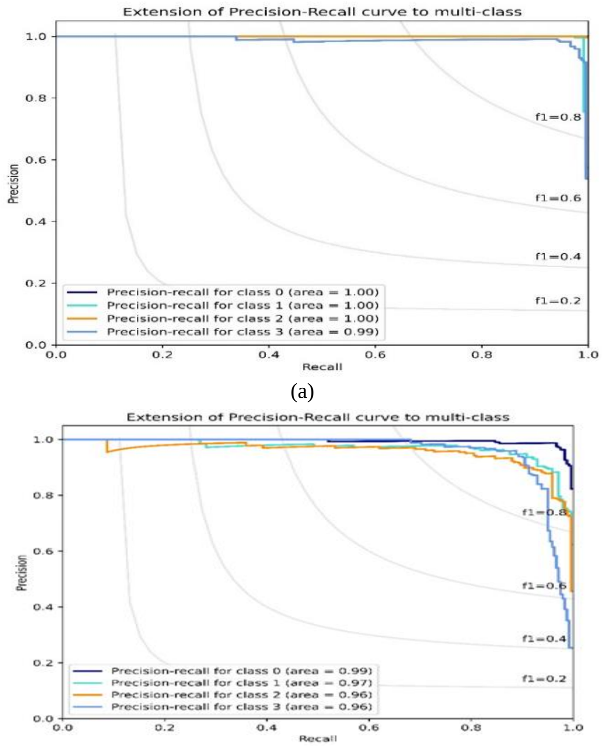
A. *Ablation Study*

For every model, it is imperative to identify the various components which impacted significantly in the performance of the model. This can help to establish the robustness of our method and the various layers that contributed to improve the performance. To determine these components which impacted on the validation accuracy, an adjustment is made to the proposed architecture and its hyperparameters through several experimental modifications to find the level of impact each component makes to the model. The results are then recorded and presented for further analysis. Table II shows the results of the ablation on retina OCT dataset. From the Table II, combination of the block-1 and the block-2 layers gave the best accuracy of 99.0% on the retina OCT dataset.



(b)

Fig. 9. Comparison of ROC-AUC on the proposed model and original CapsNet. (a) SFFT-CapsNet ROC (b) Original CapsNet ROC curve.



(b)

Fig. 10. Precision-recall curves comparison on the SFFT-CapsNet model and original CapsNet. (a) represents the SFFT-CapsNet Precision-Recall curve, and (b) Original CapsNet Precision-Recall curve.

TABLE II. RESULTS OF ABLATION STUDY ON SFFT-CAPSNET

No.	Layers	Number of Layers	Normalizer	Validation accuracy
1	Block-1(Conv layer)	1	SoftMax	94.22%
2	Block-1(Conv layer)	2	SoftMax	94.83%
3	Block-1(Conv layer)	3	SoftMax	95.02%
4	Block-1(Conv layer)	3	Sigmoid	95.91%
5	Block-1 (CLAHE layer, Conv layer)	1, 3	SoftMax	96.03%
6	Block-1(CLAHE layer, Conv layer)	2, 3	SoftMax	96.85%
7	Block-1(CLAHE layer, Conv layer)	2, 3	Sigmoid	97.70%
8	Block-2(FT,Conv layer)	1, 3	SoftMax	96.35%
9	Block-2(FT,Conv layer)	2, 3	SoftMax	97.52%
10	Block-2(FT, Conv layer)	2, 3	Sigmoid	98.01%
11	Block-1(CLAHE layer, Conv layer), Block-2(FT, Conv layer)	2,3, 2,3	SoftMax	98.71%
12	Block-1(CLAHE layer, Conv layer), Block-2(FT,Conv layer)	2,3, 2,3	Sigmoid	99.0%

A further analysis and evaluation were conducted by comparing our proposed SFFT-CapsNet with state-of-the-art results from other models that made use of the same retina OCT dataset. The evaluation matrix was based on accuracy, sensitivity, precision, specificity, overall accuracy, overall sensitivity, and overall precision. Table III presents the results of comparison of SFFT-CapsNet and previous works. The comparison was done considering the performance of the models on the individual classes of the retina OCT dataset. The letter “x” used in the table shows areas where the research paper failed to report the expected results for a particular evaluation metrics.

From the Table III, the best results have been highlighted in bold. The results from the Table III shows our proposed model achieved the best performance in all instances for all the classes using the ACC, SE, PR, SP, and AUC evaluation metrics. It can be observed from Table III that SFFT-CapsNet obtained OA, OS, and OP results of 99.0%, 100%, and 99.8%, respectively. This means the results from Table III concludes that the proposed model outperformed all the other state-of-the-art works compared. The second-best emanated from another work presented by Rajagopalan et al., [57] using CNN which obtained 97.0%, and 93.4%, on OA and OS. Though in their paper, the study failed to report the result for OP. The third best model with accuracies of 90.1%, 86.8% and 86.3% for OA, OS, and OP, respectively was also presented in a study known as the Lesion Attention Convolutional Neural Network (LACNN) which was proposed by Leyuan et. al. [53]. In all, the HOG-SVM model achieved the least performance with the accuracies of 78.1%, 65.3%, 460 and 71.8% on OA, OS, and OP, respectively.

Fig. 11 shows Histogram representing of the overall accuracies of OA, OS, and OP for HOG-SVM, Transfer Learning, VGG16, LACNN, IFCNN, LGCNN, CNN by Rajagopalan, and SFFT-CapsNet. Fig. 11 indicates that the proposed model outperformed the current-state-of-art models in all the metrics tested.

TABLE III. COMPARISON OF RESULTS OF THE SFFT-CAPSNET AND THE STATE-OF-THE-ART WORKS

Method	Classes	ACC (%)	SE (%)	PR (%)	SP (%)	AUC	OA	OS	OP
HOG-SVM [62]	CNV	85.7	87.6	82.0	84.0	92.2	78.1	65.3	71.8
	DME	91.4	53.8	74.6	97.2	87.3			
	DRUSEN	90.2	29.5	52.6	97.0	81.3			
	NORMAL	89.1	90.4	78.1	88.4	94.6			
Transfer Learning [58]	CNV	86.9	76.2	93.9	95.9	96.1	79.5	7.9	73.1
	DME	91.6	75.5	66.9	94.1	93.8			
	DRUSEN	87.2	70.7	42.0	89.1	89.2			
	NORMAL	93.3	88.9	89.5	95.2	98.1			
VGG16 [59]	CVN	91.0	86.6	93.2	94.7	97.2	83.2	6.2	76.4
	DME	92.8	70.9	74.6	96.2	93.6			
	DRUSEN	90.7	54.7	54.5	94.7	88.7			
	NORMAL	91.8	92.6	83.3	91.5	97.2			
LACNN [53]	CNV	92.7	89.8	93.5	95.1	97.7	90.1	6.8	86.3
	DME	96.6	87.5	86.4	98.0	97.4			
	DRUSEN	93.6	72.5	70.0	95.6	93.4			
	NORMAL	**97.4**	97.3	94.8	97.4	99.2			
IFCNN [60]	CNV	92.4	94.8	87.9	90.9	X	87.3	82.5	84.7
	DME	94.4	79.2	81.9	97.2	X			
	DRUSEN	93.0	64.4	76.8	97.3	X			
	NORMAL	98.4	91.5	92.2	96.4	X			
LGCNN [61]	CNV	93.3	93.3	91.5	93.3	X	88.4	84.6	82.9
	DME	93.6	85.7	79.4	96.8	X			
	DRUSEN	95.4	71.0	65.2	96.0	X			
	NORMAL	94.6	88.5	95.5	97.9	X			
Rajagopalan et. al., [57]	CNV	X	X	X	X	X	97.0	93.4	X
	DME	X	X	X	X	X			
	DRUSEN	X	X	X	X	X			
	NORMAL	X	X	X	X	X			
SFFT-CapsNet **[ours]**	CNV	100	**100**	**100**	**100**	**100**	99.0	100	99.8
	DME	98.8	**100**	**100**	**100**	**100**			
	DRUSEN	100	**97.2**	**100**	**100**	**100**			
	NORMAL	97.1	**98.7**	**0.99**	**100**	**100**			

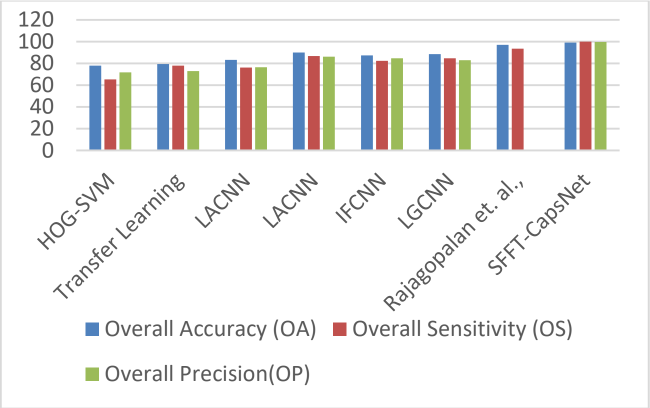


Fig. 11. Histogram representing the overall accuracies of OA, OS, and OP for HOG-SVM, Transfer Learning, VGG16, LACNN, and SFFT-CapsNet.

Fig. 12 shows an evaluation of clusters generated from the raw data and the routing process. Fig. 12(a) represents the raw clustering generated from the input dataset which has no visible clusters. It can be seen that the content is scattered on one cannot observe any possible clusters. Fig. 12(b) also shows the result of the clusters acquired from using the original CapsNet and finally Fig. 12(c) also shows the clusters formed from using the SFFT-CapsNet. From the Fig. 12, it can be visualized that the Fig. 12(c) which was acquired from the proposed model produced better clustering after the routing process than the original CapsNet.

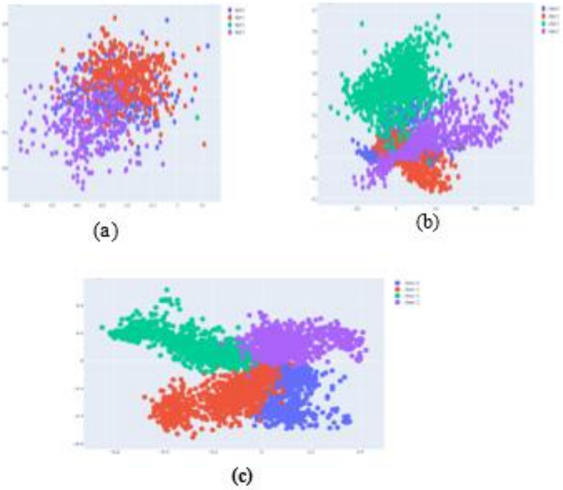


Fig. 12. Visualization of the clusters formed at the retina caps layer (a) raw dataset before routing (b) clusters formed after the routing process of the original CapsNet (c) clusters formed after the routing process of the proposed model.

This can be attributed to the fact that during the routing process, the primary capsules combine with class capsules to establish high agreement which forms better clusters at the retinaCaps layer in the SFFT-CapsNet. The various separations created by the cluster which have been indicated using different colors can be used to determine how efficient the routing process could be and hence determine the efficiency of the model.

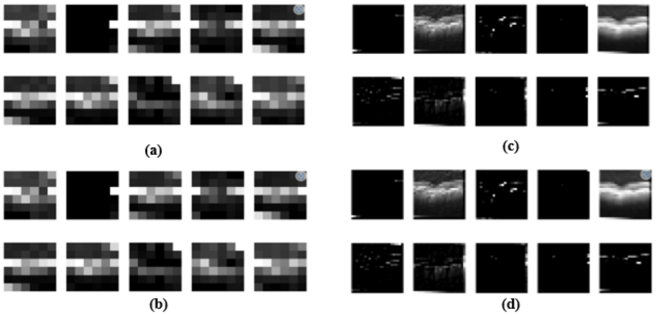


Fig. 13. Activation maps from digit capsule and evaluation capsule of the original CapsNet and SFFT-CapsNet. (a) digit capsule activation from original CapsNet. (b) evaluation capsule activation map from original CapsNet, (c) digit capsule activation from SFFT-CapsNet, and (d) evaluation capsule activation map from SFFT-CapsNet.

Fig. 13 visualizes activation maps performance of the various layers in the original CapsNet and the proposed SFFT-CapsNet as they receive input images. From the visualization, it is observed that the activation from the SFFT-CapsNet provides enough details on how a layer extracts features from the input image to influence the final output of the model. It provides better insight on the actual operations of the CapsNet.

B. Discussion

The study proposed SFFT-CapsNet for classification of retina OCT images. The result of the model is compared to the original CapsNet which is serving as a baseline for the study. From Table I, the results indicated that the accuracy of proposed SFFT-CapsNet outperformed the baseline model by a difference of 4.8% upon achieving accuracy of 99%. This is a very significant improvement in the field of computer vision. The outstanding performance is due to the novelty employed in the method. Increasing the convolutional layers to six was a good strategy for improving the feature extraction by the model. Secondly, the two enhancement layers introduced in the model improved visibility of hidden patterns and controlled over brightening of the images. Replacing the SoftMax with sigmoid and introducing the power squash prevented the model from generating high activating values that can prevent effective learning of features of the model. From the results of the ablation study in Table II, it can be concluded that such significant improvement was achieved because every layer we introduced strategically had significant impact on the model.

Again, the validation and loss accuracy in Fig. 7 shows that our model achieved higher performance compared to the baseline. Also, Fig. 8(a) and 8(b) show the confusion matrix of the proposed evaluated SFFT-CapsNet models against the original CapsNet. The diagonal outputs indicated in blue illustrated the correct prediction from the models (True positives and true negatives). The misclassifications generated from the model are indicated in white colors as output at the upper and bottom part of the correct predictions. Fig. 8(a) represents the confusion matrix of the SFFT-CapsNet. Fig. 8(b) on the other hand represents the confusion matrix of the original CapsNet. From the confusion matrix, the four instances of the dataset have 242 images each as test samples. The results from the confusion matrix can be concluded that the proposed model obtained the highest correct predictions thus 242, 239, 242, and 235 on CNV, DME, DRUSEN, and

NORMAL, respectively whereas the original CapsNet obtained 231, 230, 239, and 212 on CNV, DME, DRUSEN, and NORMAL, respectively. Fig. 9 presents the ROC-AUC curves and Fig. 10 illustrates the result of Precision-Recall curve. From the results, it can be deduced that the proposed SFFT-CapsNet controls misclassification better than the original CapsNet. From the output of the confusion matrix, our proposed model recorded the least misclassifications when compared to that of the baseline CapsNet. The results of the visualization in Fig. 12 are strong indication that our model is able to generalize well by learning the required features for better classification performance.

The performance of the proposed model can be associated to fact that the introduction of the two layers thus the CLAHE and Fourier transform could sufficiently reduce the noise in the input images. Also, increasing the convolution layer also enhanced the chances of the model extracting required features that could impact on the results of the model.

However, the high misclassification of the original CapsNet can be attributed to the insufficient convolutional layers to extract the required features to support the prediction and classification task. This is in line with the findings proposed by Cao et al., [63] which concluded that CapsNet is unable to perform well on complex images because the convolutional layer is not sufficient able to extract the required features which ends up including features that may not be required and can lead to misclassification.

Also, the performance of our model could compete and outperform the state-of-the-art models that have been implemented on the retina OCT dataset as indicated in Table III. The results indicate a strong confirmation that the proposed SFFT-CapsNet achieved the best performance in all scenarios of the evaluation Metrix. The performance of the CapsNet was not surprising as compared to the other methods which were implemented using CNN. This is because the CapsNet with the dynamic routing algorithm was able to recognize the pose, texture and spatial relationship which contributed to the performance of the classification model.

V. Conclusion

The study proposed an efficient CapsNet architecture for classifying retina OCT images. The study reconstructed the original CapsNet to include two extra layer components thus the contrast limited adaptive histogram equalization layer and Fourier transform layer. The two layers are all image enhancement layers which presented the model with more visibility of the input images. Again, the study increased the convolutional layers to six to ensure better feature extraction and replaced the SoftMax activation function with sigmoid.

Four-class (CNV, DME, DRUSEN, and NORMAL) retina OCT image dataset presented by UCSD were used for training and testing the proposed capsule framework. Evaluations of models were conducted using evaluation metrics such ACC, SE, PR, SP, AUC on the individual Class while OA, OS, and OP to measure the overall performance of the models. The results of the proposed model were compared with that of the original CapsNet in terms of performance accuracies. A further comparative analysis was conducted using the results of the propose model and that of the state-of-the-arts deep learning standard models that have been applied to the retina OCT image dataset.

The summary of the results has been presented in Table III. The results indicated that the proposed SFFT-CapsNet obtained OA, OS, and OP results of 99.0%, 100%, and 99.8%, respectively which were the best results in all instances compared with the other existing works. This performance indicates that the proposed technique is better in detecting eye diseases from retina OCT images. The method can be adopted to help ophthalmologists in detecting eye disease from retina OCT images. Although the proposed SFFT-CapsNet model achieved high performance compared to the state-of-the-art models, however, it was found that the model still needs improvement. As part of the future works, the study aims to propose an effective activation function that can help the convolutional layers implemented to extract better features required for the model performance. Also, the study seeks to test the final version of the model on complex images to establish the robustness of the model.

Acknowledgment

The authors would like to thank the editor and the reviewers for their helpful suggestions and valuable comments.

References

[1] G. Litjens, T. Kooi et al., “A survey on deep learning in medical image analysis”. CoRR abs/1702.05747 (2017). http://arxiv.org/abs/1702.05747.

[2] P. Afshar P., A. Oikonomou F. Naderkhani, N. P. Tyrrell, N. Konstantinos Plataniotis, K Farahani and A. Mohammadi, “3D-MCN: A 3D Multi-scale Capsule Network for Lung Nodule Malignancy Prediction” Scientific Reports (2020) 10:7948 | https://doi.org/10.1038/s41598-020-64824-5,.

[3] S. Shen., S. X. Han., D. R. Aberle, A. A. Bui and W. Hsu, “An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification”, Expert Systems With Applications 128 (2019) 84–9. https://doi.org/10.1016/j.eswa.2019.01.048 0957-4174/ 2019.

[4] , M. Avendi, A. Kheradvar and H. Jafarkhani, “A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI”, Med. Image Anal. 30, 2016. 108–119.

[5] A. P. Sunija, S. Kar, S. Gayathri, P. Varun Gopi and P. Palanisamy, OctNET: A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images, Computer Methods and Programs in Biomedicine (2020), doi: https://doi.org/10.1016/j.cmpb.2020.105877.

[6] O. J. P Charry and González OFA., "A Systematic Review of Deep Learning Methods Applied to Ocular Images", Cien.Ing.Neogranadina, vol. 30, no. 1, pp. 9-26, Nov. 2019.

[7] N. Gurudath, M. Celenk, and H. B. Riley, "Machine Learning Identification of Diabetic Retinopathy from Fundus Images," In 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2014, pp. 1-7. doi:10.1109/SPMB.2014.7002949 [Links].

[8] R Priyadarshini, N. Dash, and R. Mishra, "A Novel Approach to Predict Diabetes Mellitus Using Modified Extreme Learning Machine," In 2014 International Conference on Electronics and Communication Systems (ICECS), 2014, pp. 1-5. doi:10.1109/ECS.2014.6892740 [Links].

[9] P Afshar, A Mohammadi, K N Plataniotis, “Brain Tumor Type Classification via Capsule Networks”, [C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

[10] R. A. Welikalaa et al ., "Automated Detection of Proli-ferative Diabetic Retinopathy Using a Modified Line Operator and Dual Classification," Computer Methods and Programs in Biomedicine, vol. 114, no. 3, pp. 247-261, 2014. doi:10.1016/j.cmpb.2014.02.010 [Links].

[11] R LaLonde, "U Bagci. Capsules for Object Segmentation", arXiv:1804.04241, 2018.

[12] O. Zhicheng Jia, and P . Tae-Euei Kam, "Dynamic Routing Capsule Networks for Mild Cognitive Impairment Diagnosis" 2019. DOI: 10.1007/978- 3-030-32251-9 68.

[13] M. Mehdy, P. Ng, E. F. Shair, N. Saleh, and C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer", Comput. Math. Methods Med. 2017, 2610628.

[14] P. M. Kwabena, A. Felix Adekoya, A. Abra Mighty et al., "Capsule Networks – A survey", Journal of King Saud University – Computer and Information Sciences, 2019, https://doi.org/10.1016/j.jksuci.2019.09.01.

[15] J. Rathod, V. Waghmode, A Sodha and P. Bhavathankar, "Diagnosis of skin diseases using Convolutional Neural Networks" In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 1048–1051.

[16] J. Naranjo-Torres, M, Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, A. Valenzuela, "A Review of Convolutional Neural Network Applied to Fruit Image Processing", Appl. Sci. 2020, 10, 3443.

[17] N. Noord and E. Postma, "Learning scale-variant and scale-invariant features for deep image classification. Pattern Recognit", 61, 583–592, 2017,.

[18] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation". arXiv preprint arXiv:1505.04597, 2015.

[19] J. Gu, "Interpretable Graph Capsule Networks for Object Recognition. The Thirty-Fifth AAAI Conference on Artificial Intelligence", Association for the Advancement of Artificial Intelligence (AAAI-21), 2021.

[20] E Xi, S. Bing, Y. Jin, "Capsule Network Performance on Complex Data" (2017).

[21] G. E. Hinton, A. Krizhevsky and S. D. Wang, "Transforming Auto-Encoders", In Artificial Neural Networks and Machine Learning—ICANN 2011 Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6791, pp. 44–51. ISBN 978-3-642-21734-0. https://doi.org/10.1007/978-3-642-21735-7_6.

[22] S. Sabour, N. Frosst, and G. E. Hinton. "Dynamic Routing Between Capsules." 2017. arXiv preprint arXiv:1710.09829.

[23] B. Jia, and Q. Huang, "DE-CapsNet: A diverse enhanced capsule network with disperse dynamic routing", Appl. Sci. 10 (884), pp 1–13. 2020.

[24] D. Wang, A. Khosla, R, Gargeya, H. Irshad, A. H. Beck, "Deep learning for identifying metastatic breast cancer" 2016, arXiv preprint arXiv:1606.05718.

[25] A. Nithya, A. Appathurai, N. Venkatadri, D. R. Ramji and C. A. Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images". Measurement 149:106952, 2020.

[26] N Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahimb, J. Rodrigues, V. H. C. de Albuquerque, "Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks", Concurr Comput Pract Exp 32(1):4962, 2020.

[27] S. P. K., Karri, D. Chakraborty and J. Chatterjee, "Transfer learning-based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration" Biomed. Opt. Express 8, 579–592, 2017. doi: 10.1364/BOE.8.000579.

[28] H. Ren, J. Su and H. Lu, "Evaluating generalization ability of convolutional neural networks and capsule networks for image classification via top-2 classification" 2019. ArXiv:1901.10112v2 Cs.CV.

[29] Z. Yang and X. Wang, "Reducing the dilution: An analysis of the information sensitiveness of capsule network with a practical solution". arXiv, 2019, arXiv:1903.10588.

[30] J. Liu , F. Gao , R. Lu , Y. Lian , D. Wang, X. Luo , and C, "Wang. DDRM-CapsNet: Capsule Network Based on Deep Dynamic Routing Mechanism for Complex Data". ICANN 2019, LNCS 11727, pp. 178–189, 2019. https://doi.org/10.1007/978-3-030-30487-4_15.

[31] Z. Zhao, A. Kleinhans, G. Sandhu, I. Patel and K. P. Unnikrishnan,. "Capsule networks with max-min normalization", ArXiv:1903.09662v1 [Cs.CV], 1–15. 2019.

[32] P. K Mensah, B. W Asubam and A. A Mighty, "Exploring the performance of LBP-capsule networks with KMeans routing on complex images", Journal of King Saud University – Computer and Information Sciences,2020. https://doi.org/10.1016/j.jksuci.2020.10.006.

[33] E. Xi, S. Bing, and Y. Jin. "Capsule network performance on complex" data. 2017. arXiv preprint arXiv:1712.03480.

[34] Z. Zhang , S. Ye, P. Liao , Y. Liu, G. Su and Y Sun, "Enhanced Capsule Network for Medical image classification", 978-1-7281-1990-8/20/$31.00 IEEE, 2020.

[35] F. Shaukat, G. Raja, R. Ashraf, S. Khalid, M Ahmad and A Ali, "Artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features". J Ambient Intell Hum Comput 10: pp 4135–4149, 2019.

[36] H. P. Nguyen and B. Ribeiro, "Advanced Capsule Networks via Context Awareness", ICANN 2019, LNCS 11727, pp. 166–177, 2019. https://doi.org/10.1007/978-3-030-30487-4_14.

[37] C. Xiang, L. Zhang, Y. Tang, W. Zou and X. Chen, "MS-CapsNet: a novel multi-scale capsule network", IEEE Signal Process Lett 25(12):1850–1854, 2018.

[38] S. S. R. Phaye, A. Sikka, A. Dhall and D. Bathula, "Dense and diverse capsule networks: Making the capsules learn better", ArXiv : 1805 . 04001v1 [Cs . CV] pp 1–11. 10 May 2018.

[39] G. Huang, Z. Liu, L. Van der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

[40] G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-deep neural networks without residuals", ArXiv:1605.07648v4 [Cs.CV], pp1–11. 2017. Retrieved from http://arxiv.org/abs/1605.07648.

[41] G. Deborshi and R. Sun, "Application of capsule networks for image classification on complex datasets". 2019.

[42] S. B. S Bhamidi and M. El-Sharkawy, "Residual capsule network. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference", UEMCON 2019, 0557–0560. https://doi.org/10.1109/ UEMCON47517.2019.8993019.

[43] Y. Li, M. Qian, P. Liu, Q. Cai, X. Li X, J. Guo, H. Yan et al "T recognition of rice images by UAV based on capsule network". Clust Comput 22(4):9515–9524, 2019.

[44] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza and J. Li and F. Pla, "Capsule networks for hyperspectral image classification", IEEE Trans Geosci Remote Sens 57(4):2145–2160. 2018.

[45] F Deng, P. Shengliang, X. Chen, Y. Shi, T. Yuan and P. Shengyan, "Hyperspectral image classification with capsule network using limited training samples". Sensors 18(9):3153, 2018.

[46] W. Y Wang, H. C Li, L. Pan, G. Yang and Q. Du, "Hyperspectral image classification based on capsule network". In: IGARSS 2018 IEEE international geoscience and remote sensing symposium. IEEE, pp 3571–3574, 2018.

[47] K. Adu, Y. Yu, J. Cai and N. Tashi, "Dilated Capsule Network for Brain Tumor Type Classification Via MRI Segmented Tumor Region". 2019 IEEE International Conference on Robotics and Biomimetics, 2020. DOI: 10.1109/ROBIO49542.2019.8961610.

[48] H. J. D. Koresh and S. "Chacko. Classification of noiseless corneal image using capsule networks. Soft Computing, 2020" https://doi.org/10.1007/s00500-020-04933-5.

[49] J. Wang, G. Deng, W. Li, C. Yiwei, G. Feng, H. Liu, Y. He, G. Shi, "Deep learning for quality assessment of retinal OCT images", Biomedical Optics Express, 2019, https://doi.org/10.1364/BOE.10.006057

[50] P. A. Keane, P. J. Patel, S. Liakopoulos, F. M. Heussen S. R, Sadda, A. Tufail. "Evaluation of age-related macular degeneration with optical coherence tomography" Surv Ophthalmol. 2012; 57: 389e414. 5.

[51] T. Ilginis, J. Clarke and P. J. Patel. "Ophthalmic imaging". Br Med Bull. 2014;111(1):77-88. doi:10.1093/bmb/ldu022.

[52] C. Neely, K. J. Bray, C. E. Huisingh, M. Clark, G. J. McGwin, and C. Owsley, "Prevalence of undiagnosed age-related macular degeneration in primary eye care," JAMA Ophthalmol., vol. 135, no. 6, pp. 570-575, 2017

[53] F. Leyuan, C. Wang, S. Li, H. Rabbani, X. Chen, and Z Liu, "Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification", 2019. DOI 10.1109/TMI.2019.2898414

[54] T. Hahn, M. Pyeon and G Kim, "Self-routing capsule networks". In: Wallach HM, Larochelle H, Beygelzimer A, d'Alch´e-Buc F, Fox EB, Garnett R, editors. Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada. 2019. p. 7656–65. URL: http: //papers.nips.cc/paper/8982-self-routing-capsule-networks.

[55] Malmgren C. A, "Comparative Study of Routing Methods in Capsule Networks. Master's thesis" Link¨oping University, Computer Vision; 2019.

[56] J. E. Lenssen, M. Fey, P. Libuschewski, "Group equivariant capsule networks", In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnet R, editors. Advances in Neural Information Processing Systems 31. Curran Associates, Inc.; 2018. p. 8844–53. URL: http://papers.nips.cc/paper/8100-group-equivariant-capsule-networks.pdf.

[57] N. N. V. Rajagopalan, and A. N. Josephraj, "Diagnosis of retinal disorders from optical coherence tomography images using cnn", PLOS ONE 16 (7) (2021),1–17.doi:10.1371/journal.pone.0254180.

[58] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122–1131, 2018.

[59] K. Simonyan, A. "Zisserman, Very deep convolutional networks for large-scale image recognition", In: ICLR 2015 Conference Proceedings, pp. 1–14. ArXiv:1409.1556v6 [Cs.CV] 2015.

[60] F. Leyuan, Y. Jin, L. Huang, S. Guo, G. Zhao and X. Chen, "Iterative fusion convolutional neural networks for classifiication of optical coherence tomography images," J. Vis. Commun. Image Represent., vol. 59, pp. 327– 333, 2019.

[61] L. Huang,, X. He,, L. Fang, H. Rabbani, and X. Chen, "Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network". IEEE Signal Process. 2019. Lett. 26, 1026–1030. doi: 10.1109/LSP.2019.2917779

[62] P. P. Srinivasan et al ., "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. doi:10.1364/BOE.5.003568

[63] S. Cao, Y. Yao, G An, E2-capsule neural networks for facial expression recognition using AU-aware. IET Image Process. Electron. Lett., 1–2 2019. https://doi. org/10.1049/iet-ipr.2020.0063