Literature Review

Classification of Diabetic Retinopathy using Retinal Images

Prepared by: Mark Dhruba Sikder (26529548) & Asif Rana (27158632)

Course: Bachelor's in Computer Science, C2001

Date: 10 May 2018

Supervisor: Dr. Reza Zare

Table of Contents

Abstract

The global count of the people suffering from diabetes is increasing at an explosive rate. One of the main disease associated with diabetes is Diabetic Retinopathy(DR). This disease can be cured if it is detected at an early stage. If DR is not diagnosed it can cause permanent blindness. Even for the medical professionals it is challenging to detect DR in the initial stages. Application of machine learning can solve many classification problems. Recently, deep mining is playing a significant role in the medical domain. Classification of DR using deep neural network has been implemented by many researchers where they face several constraints includes lack of suitable training images. The accuracy of the previously approached classifier models for DR was not remarkable; this means there is a room for improvement. Building a classifier model which can gain the confidence of the professionals is the ultimate goal. In this literature review we investigate the proposed methodologies by the researchers in recent years and analyze the results. We note the key features of several modalities and the procedures used for preprocessing.

Keywords

CNN, AlexNet, Image Fusion, Exudates, Lesions, MIL, SVM, Fine-Tuning

Introduction

It is predicted that by the year 2035, the count of diabetic patient will rise to 562 million(Zhou, Zhao, Yang, Yu, & Xu, 2018) with Diabetic Retinopathy being one of the many diseases associated with diabetics(Pang, Luo, & Wang, 2018). Along with many complications a diabetic patient has a high chance to suffer from critical level vision loss and in worst case permanent blindness due to DR. 40-50% of the global population is in the treat of becoming a victim of DR(Bravo & Arbeláez, 2017). This catastrophic disease is claimed to be the major cause of blindness and with time the number of patients losing their eyesight is increasing rapidly(Bravo & Arbeláez, 2017). Rate of DR can be reduced significantly if the disease can be addressed in the early stages (Bravo & Arbeláez, 2017); detecting DR in the early stages is a challenge to modern science. Since, it has no visual indication of this disease in its preliminary stage(Bravo & Arbeláez, 2017). The necessity of detecting (DR) in early stage becomes an important task to accomplish in the health sector. In the past DR was classified as a disease which cannot be cured (Bravo & Arbeláez, 2017). Currently, there have been some proposed DR classifier models but there is a lot of room to improve in terms of efficiency and accuracy. Deep learning has shown some promising outcomes which has made it the to go choice for image feature extraction and classification in the medical domain(Quellec, Charrière, Boudi, Cochener, & Lamard, 2017). Despite having strong computational power, current deep learning algorithm is not able to gain the trust of the medical experts in classifying DR(Pang et al., 2018).

In the field of medical image analysis deep machine learning is playing a vital role (Quellec et al., 2017). Some research articles suggest that the detection of Diabetic Retinopathy can be made by applying the CNN (Convolutional Neural Network) and image fusion combined (Liu, Guo, Georgiou, & Lew, 2018). The use of the CNN model is recognized as a better alternative to the conventional methods used for visual learning (Girshick, Donahue, Darrell, & Malik, 2014).

In this literature review we will investigate the possibility of classifying DR using deep learning with CNN and image fusion as the classifying tool. First, we define some of the key features related to extraction and learning of exudates from the training images followed by the principals of the features used in classification. Then, we address the problem statements stated by the researchers. In the state of the art section, we discussed the proposed methodology of several DR classifying articles accompanied by the critical review of the research articles.

Problem Statements

Finding a dataset with a variety of images to test their model was difficult to find (Mansour, 2018). Most of the training data set in medical domain had skewed data set. Presence of skewed data set can cause the model to overfit the prediction level(Pratt, Coenen, Broadbent, Harding, & Zheng, 2016). Moreover, Retinal images can come in a various resolution in addition to variation in brightness and contrast level(Zhou et al., 2018). The variance in the image quality is accompanied by a missing part of the retina. (Mansour, 2018) faced difficulties with the dataset, in which the images were out of focus, over-exposed or under-exposed and noisy. (Quellec et al., 2017) stated that, due to the low resolution of the input images, the red lesions could not be detected easily.

The accuracy of handcrafted features can vary due to the resolution of the training image and the intensity of features in the image (al., 2016). Deep neural network is required for the classification of DR. Training a large neural network is considered as challenging to perform without the side effects of disappearing gradients and degradation of the images (Liu et al., 2018). Gathering datasets for training a classifier model in medical domain requires supervision from experienced clinicians to grade the retrieved images (Pratt et al., 2016) and if the training images are misclassified, it might affect the performance of the classifying model (Zhou et al., 2018).

Trained CNN can start to learn associations which are classified in a wrong way resulting in a wrong prediction of DR (Zhou et al., 2018). It is a challenge to achieve a satisfactory accuracy level using CNN to predict DR using the test images(Pratt et al., 2016).

Classification Using Deep Learning

A classifier model requires a training data set and a testing data set; in general it is suggested to have an equal fifty-fifty split of the entire data to obtain the taring and testing data sets(Bravo & Arbeláez, 2017). In general classification algorithms can be subdivided into two separate modules. First one being the feature extraction and the second one is the classification module(Qayyum, Anwar, Awais, & Majid, 2017).

Pre-processing techniques

It is suggested that to avoid the variance between the characteristics of the image, preprocessing methods should be applied before applying the classifying methods and feature extraction(Mansour, 2018). The retinal images of DR can come with different resolution, brightness and the level of contrast within the images. Further, the images might come in varied sizes. In order to avoid the disorientation caused due to the cropping and changing resolution of the images, it is advised not to crop the retina images(Zhou et al., 2018). Quality of the image can be determined by deploying many assessment strategies and the structural similarity (SSIM) and the root- mean-square error (RMSE) can be used as a scale for the quality of the image(Du, Li, Lu, & Xiao, 2016). Data set can be split into separate lesions individually and then training the model could be done on each individual lesions(Quellec et al., 2017). The training data set for DR can have different brightness level for the retinal image resulting to misclassification; in order to counterattack this constraint color normalization is being suggested(Pratt et al., 2016).

Feature Extraction

To retrieve the features from an image, the image needs to be transformed into a data format which can be applied to the CNN model. Choosing the shape of the region arbitrarily for feature extraction is considered as the simplest method out of the various approached procedures(Girshick et al., 2014). One proposed methodology is to wrap all pixels in the training image into a tight bounding box; despite the size and aspect-ratio of the training image (Girshick et al., 2014).

Deep Feature Extraction

Pre-trained classification models can be used for feature extraction. The implementation of the model can have a number of convolution layers and a variable amount of pooling layer which is followed by fully connected layer with different number of filters being applied(Zhang, Xia, Xie, Fulham, & Feng, 2017). By varying different parameters, the model can output a fine-tuned classification model.

Handcrafted Feature Extraction

LBP models are a common example of handcrafted feature extraction. LBP is used to describe the texture of the images which performs its task in three various stages. In the first stage, each pixel of the image is taken as a maximum value for its surrounding eight pixels which results in a eight binary number. Second stage, after getting the binary number are then merged together to transform into a 8-bit binary representing an integer. In the last stage, a histogram of frequency for the integer is created for the entire image resulting to a LBP descriptor(Zhang et al., 2017). Support Vector Machines (SVMs) classifier can retrieve the features from the training medical images using a high-order spectra method(Pratt et al., 2016). SVMs can be used to detect the variations in the shapes of the exudates and the dilated blood vessels in the retina(Pratt et al., 2016). Application of SVM in the extraction of features from images has one drawback which is that it cannot be applied in a real-time scenario unlike CNN(Pratt et al., 2016).

The architecture of deep learning

Convolution Neural Network

The objective of this literature review is to classify the images of Diabetic Retinopathy. Image classification does not require any form of localization for the objects present in the training images(Razavian, Azizpour, Sullivan, & Carlsson, 2014). We focused on two main methods for classification of the images which are CNN and Fusion techniques.

In general CNN consists of several convolution and pooling layers, it is then associated with multiple fully connected layers and the final layer consists of a regression layer to perform the classifier prediction(al., 2016). The beginning layers of CNN is responsible learn the lower level features of the training image and the end layers lean the high level features which are specific to the classification of the image(al., 2016).

One of the constraints for DR classification is the lack of labeled training data set. Since, CNNs have many parameters training the CNN with a small dataset might result to overfitting the model. One of the common way to deal with this problem is to use fine tuning which makes use of the pretrained networks which were built from a large data set and then continue to train it the relatively smaller current dataset(Yu, October 3, 2016). Using the pretrained model saves a lot of computational power which is accompanied with the save of time(Zhang et al., 2017) A research article recommends the use of large CNNs to compensate the lack of large data set. In addition, they stated that if the CNNs could be trained with fine-tuning it will give a better prediction result(Girshick et al., 2014). The best performance was achieved by a research team through fine tuning the layers of the pre-trained CNN. If the already tuned CNN is further tuned this might result into an even better model for classification(Zhou et al., 2018). It is claimed that CNN can achieve its maximum performance very easily with fine-tuning compared to the CNN which are made from scratch(al., 2016).

Fusion Technique

Image fusion is the merging of two or more images into a single image with a minimum loss of the features of the images(Chen & Huang, 2014).In medical research image fusion of multiple modalities plays a key role. Modalities from medical images can be extracted using image fusion(James & Dasarathy, 2014). Image fusion of multi modal training images can improve the quality of the features without any form of degradation in the features of the image(Du et al., 2016). It is recommended that applying image fusion technique can result into a classifying model with better accuracy(James & Dasarathy, 2014). Fusion technique can be used in the classifier model to fuse multiple focus of the same section of the image. The implementation can be divided into four methods- focus detection, initial segmentation, consistency verification and the last stage is fusion(Liu, Chen, Peng, & Wang, 2017). Application of image fusion can retrieve the features which cannot be detected by an average human eye(James & Dasarathy, 2014)

State of The Art

Detection of DR can be made by analyzing condition of the visual blood vessel dimensions of the eye. (Verma, Deep, & Ramakrishnan, 2011) were able to present a method for analyzing the dilation of the blood vessels and the physical damage done to the eye due to DR using a non-parametric method with a higher classification accuracy. Their implementation was generalized into training stage where they trained the model with medical images followed by the implementation of random forest for classification. In the training stage (Verma et al., 2011) decided to a mere training data set of 65 images. They used bagging to increase their training size but for such a complex and crucial task of detecting DR this small data set is not enough. Their choice for choosing a small dataset has water downed the achievement of having claimed accuracy of 90% and 87.5% on extreme DR cases. We have gone though many research articles; only this research paper introduces the use of random forest as the classifier model. Application of random forest might have better accuracy rate but it takes up a lot of memory thus training a large data set is not feasible.

In contrast to the implementation of using random forest as classifier, deep learning is becoming the leading methodology in building classifiers in the medical domain. (Quellec et al., 2017) implemented CNN to classifier model to detect the lesion in the retina for DR prediction. The researchers also proposed the use of heat maps to eliminate the artifacts which are being included during feature extraction. The artifacts could be filtered out by optimizing the CNN and the heat maps. (Quellec et al., 2017) preferred to used CNN which had a better efficiency compared to the random forest. (Quellec et al., 2017) claim that their implementation of tweaking the configuration of CNN with pixel level supervision produced a better heatmap compared to the general algorithms used for generating heat map. While testing the model in the Kaggle data set it received an impressive AUC of 0.954 after training the model with 108,000 images. However, the validity of the model is questioned when we consider the fact that the model was created with a testing size of 89 images from DIARETDB1 dataset.

Data set normalization was preferred by (Pratt et al., 2016) before moving forward with any form of classification. This method in preprocessing is quite common when classifying models in the medical domain is considered. Since in the medical domain the training images are more likely to have poor resolution and various color intensities. CNN has been proven as a best approach to classify model using deep learning (Pratt et al., 2016) has proposed to implement a CNN model along with fine tuning. Fine tuning relates to tweaking with the parameters of the CNN model to improve its efficiency. The first layer of the CNN is given the task to learn the edges of the training images; whereas, the last layer is being used to learn the features contained in the training images such as exudates. In the middle sub portion increased convolution layer learns the deeper features these layers are grouped together to form a convolution block. Basically, (Pratt et al., 2016) is trying to divide a complex CNN architecture into three sections and assigning each section an individual task of extracting features from the training image. This is an efficient approach as training 3 separate CNN to do the same task will require a lot of computational power and time. After the partition several of these blocks are grouped together to result into several convolution blocks. Batch normalization with max pooling is being applied to each of the convolution block. CNN with a smaller training data size have a high tendency overfit the model. In order, to prevent the model from overfitting, the convolution block is flattened to one-dimension. The final convolution layer is then associated with dropout dense layer and rectified linear unit in the end. The research article(Pratt et al., 2016) gave importance to pre-processing rather than carrying out building the classifier with the raw training images for DR. They addressed the shortcomings of raw data set and proposed normalization to be the solution to counter attack the drawback. Further, (Pratt et al., 2016) addressed the possibility of over fitting the model due to the presence of skewed data sets which is common in the medical domain training images. Implementing real-time class weights in the CNN with back propagation can be considered as the solution(Pratt et al., 2016). The results from the trained model states that the model achieved an accuracy of 75%, specificity-95% and sensitivity- 30% (Pratt et al., 2016). Achieving a sensitivity of 30% and specificity of 95% suggests that the model proposed is biased towards classifying the output as not having DR. This means even after considering the sewed test data set their applied procedure was not able to overcome the drawback. Moreover, 10% of images in their training datasets was not gradable by the professional.

Fine tuning was also used by (al., 2016) to build a CNN classifier model. In this paper the researchers went with a different approach to fine tuning rather than going with the common practices as in (Pratt et al., 2016). The common implementation of fine tuning CNN which only creates features from the training images and some fine tunes all of the layers of CNN(al., 2016). AlexNet (Gao, 2017)was used as their preferred of CNN. They generated a set of images by applying data augmentation which will be used for fine tuning the CNN classifier model. All the training images were given as an input to the trained CNN. The probabilistic output for the test exudates was averaged and FROC was as a measure of performance(al., 2016). The authors claimed that by incorporating the CNN model with handcrafted features will output an improved classifier model(al., 2016). The generated result shows that fine-tuned CNN had a slight better performance compared to the handcrafted features however fine tuning all the layers of CNN can improve the accuracy significantly(al., 2016). In this article, there was no specific method about the classification of DR. This article mainly focuses on the prediction difference between fine-tuning CNN compared to pretrained CNN and gave a brief idea about the level of accuracy using handcrafted features. Although the images for training the model was not used from a renowned data banks for medical images. They used a single video and used it to source the images for training the classifier this made us less confident in acknowledging their work as a better implementation to classify DR.

(Doshi, Shenoy, Sidhpura, & Gharpure, 2016) aims at automatic diagnosis of the disease into different stage of the disease using deep learning. The authors present a design and implementation of GPU accelerated deep convolutional networks to automatically diagnose and classify high-resolution images into 5 stages based on severity from a collection of RGB colored fundus retinal images. The model they propose automatically extracts and learns the features which are crucial in diagnosing the different stage of the disease without any explicit or manual feature extraction being required. In the pre-processing stage (Doshi et al., 2016) mentioned concern about CNN overfitting the training data which was previously mentioned in (Pratt et al., 2016). To overcome the chances of overfitting they proposed data augmentation which involves transformation with the training data set this includes shear, flop, transverse and last transpose. Among all of the reviewed articles it was only(Doshi et al., 2016) who mentioned the physical transformation of the images. In the medical domain the training images had to deal with skewed data sets. Instead of mentioning this drawback (Doshi et al., 2016) came up with a soothing solution. They suggested to drop some less major class for DR which might make the training data set less skewed to a certain class. Even though the proposed solution is ideal; but still following this procedure make working with the skewed data set more legible. After the stages of going through testing and training the model achieved a quadratic kappa score of 0.3066. After observation, (Doshi et al., 2016) pointed out that there was slight overfitting of the model. Model 2 was found out to be an improvement over Model 1 with a score of 0.35. Model 3 was even a better improvement and achieved a score of 0.38659. Finally, the authors performed an ensemble averaging which achieved a score of 0.3996. [Article 14] (Doshi et al., 2016) demonstrates an unique approach in improving their models accuracy which is noteworthy; however according to our shortlisted reviewed article none have mentioned about this procedure. Furthermore, they also gave an overview of their future work explicitly which is an improvement over the present model.

There is an approach which enhanced the interpretability through visualization based on results and furthermore. Thus, increasing the accessibility through web services and a message queue framework (Pang et al., 2018). In addition, the authors designed a report system for doctors and a mobile app for patients, by which, the accessibility can be improved. To obtain a good prediction, (Pang et al., 2018) planned in using two fundus images of each patient. Hence, this required them to implement feature fusion for two eyes. While creating the classifier model for DR (Pang et al., 2018) constructed a CNN model with two levels. Inception-V3 which is an implementation of a CNN architecture is used for feature extraction and VGG model. Final prediction was derived using both of these two procedures. Thus, the model was given two input images at a time for more accurate classification. The last convolution output was used for visualization. After building the model, the authors created a series of process which will guide the medical professionals for an easy access to the services provided by the model. (Pang et al., 2018) mainly focused on how the strong ability of CNN models in image processing can be transferred into real life applications and serve all kinds of industries. They focused more on promoting their application in actual environment. We agree on the view that only building a model and achieving a high accuracy is not enough to convince the medical professionals in using these architectures in practical implementation. Even though, models like CNN and deep MIL have been used in the model while building the classifier model for DR; (Pang et al., 2018) did not mention any specification of their model which we could take note of. In addition they did not mention the data set being used not did they reported any form of pre-processing steps befor implementing the model.

There are several approaches for preprocessing the training data set images in the medical domain. (Zhou et al., 2018) has proposed data set normalization same as (Pratt et al., 2016) to deal with the variance in resolution contrast and level of brightness. There is a concern to the tendency of the classifier to misclassify the DR result; this brings down the accuracy of the classifier model. Multiple Interface Learning (MIL) is nominated as the solution to solve the misclassified association in DR by (Zhou et al., 2018). AlexNet which is a common implementation of CNN architecture was chosen as the classifier architecture. The architecture consisted of 5 convolutional layers and 3 fully connected layer. Choosing Alexnet as the classifier model has an advantage over the other architectures; it can accelerate the processing time and overlap pooling to reduce the interconnected size of the network(Gao, 2017). The implementation of AlexNet was associated with fine tuning each layer in the CNN which is required due to the variance in depth of CNN from one application to another. This choice of tweaking the model using fine tuning contributed to a better model which will be discussed further in later part of the literature review. Multiple interface learning classifier model achieved an AUC value of 0.925 using the Kaggle-test due to its robust implementation of detecting the DR regions(Zhou et al., 2018). Even after achieving a high accuracy level, the model’s performance was not able to handle poor image quality datasets. Not being able to deal with noisy data is a major failure for a DR classifier model. Since, presence of noisy data is very common in the medical domain; thus, preventing the model from being applied in DR detection. The model proposed by (Zhou et al., 2018) did not go through rigorous testing to enforce its confidence of being a better DR classifier.

AlexNet can be used as the base for a Computer Aided Design (CAD) to solve intricate problems such as DR in the early stages (Mansour, 2018). Here, instead of using a MIL with conventional AlexNet architecture; they preferred to optimize the classifier in multiple levels of the architecture. This method can be used to focus on the specific region of interest of a training image. Features from the training image can be extracted by applying the AlexNet. Moreover, to reduce the computational overheads during the process, different feature selections and dimension schemes such as Principle component analysis and Linear discriminant analysis were applied. This was an ideal solution to make the classification process less computationally expensive. Furthermore, to examine the efficiency of the proposed CAD solution, a five-class classification was performed using an SVM classifier (Mansour, 2018). This five-classes were derived as normal, mild non-proliferative DR, severe proliferative DR, moderate non-proliferative and proliferative DR. These classifications played a vital role in making sure that the model does not get bias during the time of classification. Out of all the research articles that we have reviewed, only (Mansour, 2018) mentioned the use of 10 folds cross validation. Using 10 folds cross validation (Zacharski, 2018) which is very crucial to receive a much greater accuracy rate compared to splitting the data set into a certain percentage to receive a training data set and testing data set. The achieved accuracy of the proposed classifier model by (Mansour, 2018) achieved an accuracy of 97.93% which is quite impressive while using the Kaggle dataset of 35,126 fundus images. Even after having variations in the data set achieving such a high accuracy means the preprocessing steps taken before classifying the model made a positive impact to the classifier model. (Mansour, 2018) presented enough information to elaborate their work with a combination of self-explanatory images and diagrams which gave us a much clear view of their implementation. In the article they also mentioned that there is a scope for building an even better classifier by performing multiclass classification.

In a research article (Bravo & Arbeláez, 2017) approached on DR classification. Unlike (Zhou et al., 2018) and (Mansour, 2018) where both of them preferred to use AlexNet as their choice of CNN architecture (Bravo & Arbeláez, 2017) preferred to use VGG16 as the CNN architecture. The main goal of their proposed model was to detect DR in its early stage. Use of VGG16 as a classification architecture had an advantage over AlexNet which is VGG16 was designed for the implementation of large scale natural image classification. As a form of pre-processing the training data set they went with removing the background of the images which was not mentioned in any of the papers we reviewed and as a compromise they did not consider to apply and form of filtration to eliminate noise in the training data set. They did not stick with only a one form of preprocessing which, five different data sets were generated each set going through different form of preprocessing to identify which data set produced the best CNN classifier. None of the researchers gave that much of importance to the preprocessing of data set which (Bravo & Arbeláez, 2017) has considered. This was an impressive strategy since the performance of a classifier is highly dependent on the training dataset. The performance of the proposed model was evaluated using the confusion matrix. One advantage of using the confusion matrix over ROC and AUC is that with the confusion matrix we could determine the accuracy, precision, recall and some other performance measures to get to know how the model performs in different aspect which ROC or AUC ca not. (Bravo & Arbeláez, 2017) went with classifying the model into certain classes which required a larger data set. Rather than giving any excuses to justify their classifier is best by blaming the scarcity of "perfect" data set (Bravo & Arbeláez, 2017) did data augmentation. This taught us how can we deal with smaller data set and come up with a better classification model. The researchers went with four different CNN models rather than sticking with a singe model. Model best confusion matrix was elected as the proposed model. Then, the network was trained for 700 iterations and using the best pretrained weights which they retrained the network for the 5-class classification problem. In generating final classification method, they combined the scores of the best trained neural networks using the VGG architecture and chose the class with higher average score, for each image. However, after combining the scores of the best trained neural networks, they chose the class with the highest average score which was 0.505. Analyzing the score, we can comment that the applied methodology is not good enough with 50.5% accuracy. It can be stated that the model basically has only 50-50 chance of giving the correct prediction. Furthermore, the researchers are stating that the use of different networks trained with different parameters, made the final method more robust to outliers and capable of further generalization. That this statement does not solve the issue that they mainly aimed to achieve. Moreover, (Bravo & Arbeláez, 2017) did not give any valid information of which dataset they used leaving us quite anxious about the validation of the model.

In image fusion the combination of modalities used to classify a model depends on the presence of noise in the training data set. There is direct relationship between the performance of classifier model and the quality of the data set in the model. SVM (RAY, 2017) are modalities which can control both the data and the parameters in the data set can reject the outliers in the data sets which brings a much improvement in efficiency while using image fusion(James & Dasarathy, 2014). The performance of a modality in medical domain is dependent on the structure of the organ and the tissues(James & Dasarathy, 2014). The authors gave an overview of the application of image fusion in the medical domain and a brief idea about how image fusion can be applied in the medical domain. Since it was a review article there were no specific proposed methodology for applying image fusion. A preference for using SVM modality was made and they emphasized the requirement on the quality of the training images in the medical domain. They did not provide or referenced any form of accuracy of the applied models so that we can compare the accuracy of different modalities used with image fusion. They claimed that the quality of a classifier depends on the level of noise in the training image.

The application for using image fusion on the test images allows to improve the strength of the prediction of each individual layer and result into an improved prediction(Liu et al., 2018). As a preprocessing stage for the convolution fusion network they followed the conventional pre-processing strategy. In this research paper, (Liu et al., 2018) proposed a new fusion architecture from the ground up which combines the intermediate layers into changing weights. It is suggested that using early fusion accompanied by late prediction can be able to provide a high-resolution interpretation and decreases the number of parameters; thus, reducing the computational complexity. Locally connected layer can be implemented as a fusion module which can learn the weights of each features and produce a better representation. The first step of forming a CFN (Convolution Fusion Network) is to fuse the multi- level features in CNN which is then merged with several locally connected modules. The fusion results into a fusion module which can address image level classification; after that more training is carried out using the training images called the trained CFN module. There are three sub sections in the trained CFN module scene recognition, fine grained recognition and image retrieval. More development is being carried out for each of these dimensions until they reach a certain amount of accuracy. The final resulting model is called the CFN(Liu et al., 2018). This proposed methodology is claimed to be applicable for different visual recognition tasks (Liu et al., 2018). Using input specific weights can reduce the variance between images. This is the one of the first research paper to use locally connected layer as a fusion module. They proved that incorporating CNN with image fusion can output better result compared to CNN. CFN classifier model was tested in many different datasets and summarizing the result we found that CFN models can output a better prediction compared to the CNN and AlexNet models. In this article the researchers addressed about the shortcomings of a CNN model which was not addressed in any of the reviewed articles. This help us to consider the drawbacks of a CNN model before implementing it to classify DR.

Conclusion

In this literature review we tried to shed light to the devastating numbers of diabetic patients suffering from DR. There is no doubt that deep learning has the capability to classify DR. We reviewed a considerable number of articles where we found that the training data set played a vital role in the classification model. In the medical domain it is very common to get a skewed data set for train the classifier; this is accompanied with variable resolution, brightness and contrast level. Common filtering approaches were – normalization of the data set, weighted sampling, eliminating skewed classifier. After going through all the research paper and considering several modalities it can be concluded that using a CNN classifier with fine tuning and the incorporation of image fusion with the classifier is more like to output a better classifier model for DR. In the preprocessing stage, we would prefer to use batch normalization of the training images and while testing the accuracy of the model 10 folds validation process could be used.

References

al., N. T. e. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *EEE Transactions on Medical Imaging, vol. 35*(no. 5), pp. 1299-1312. doi:10.1109

Bravo, M. A., & Arbeláez, P. A. (2017). *Automatic diabetic retinopathy classification.* Paper presented at the 13th International Symposium on Medical Information Processing and Analysis.

Chen, H., & Huang, Z. (2014, 8-10 Nov. 2014). *Medical Image Feature Extraction and Fusion Algorithm Based on K-SVD.* Paper presented at the 2014 Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing.

Doshi, D., Shenoy, A., Sidhpura, D., & Gharpure, P. (2016, 19-21 Dec. 2016). *Diabetic retinopathy detection using deep convolutional neural networks.* Paper presented at the 2016 International Conference on Computing, Analytics and Security Trends (CAST).

Du, J., Li, W., Lu, K., & Xiao, B. (2016). An overview of multi-modal medical image fusion. *Neurocomputing, 215*, 3-20. doi:https://doi.org/10.1016/j.neucom.2015.07.160

Gao, H. (2017). A Walk-through of AlexNet. Retrieved from https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014, 23-28 June 2014). *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.* Paper presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition.

James, A. P., & Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. *Information Fusion, 19*, 4-19. doi:https://doi.org/10.1016/j.inffus.2013.12.002

Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion, 36*, 191-207. doi:https://doi.org/10.1016/j.inffus.2016.12.001

Liu, Y., Guo, Y., Georgiou, T., & Lew, M. S. (2018). Fusion that matters: convolutional fusion networks for visual recognition. *Multimedia Tools and Applications*. doi:10.1007/s11042-018-5691-4

Mansour, R. F. (2018). Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomedical Engineering Letters, 8*(1), 41-57. doi:10.1007/s13534-017-0047-y

Pang, H., Luo, C., & Wang, C. (2018). Improvement of the Application of Diabetic Retinopathy Detection Model. *Wireless Personal Communications*. doi:10.1007/s11277-018-5465-3

Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional Neural Networks for Diabetic Retinopathy. *Procedia Computer Science, 90*, 200-205. doi:https://doi.org/10.1016/j.procs.2016.07.014

Qayyum, A., Anwar, S. M., Awais, M., & Majid, M. (2017). Medical image retrieval using deep convolutional neural network. *Neurocomputing, 266*, 8-20. doi:https://doi.org/10.1016/j.neucom.2017.05.025

Quellec, G., Charrière, K., Boudi, Y., Cochener, B., & Lamard, M. (2017). Deep image mining for diabetic retinopathy screening. *Medical Image Analysis, 39*, 178-193. doi:https://doi.org/10.1016/j.media.2017.04.012

RAY, S. (2017). Understanding Support Vector Machine algorithm from examples Retrieved from https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014, 23-28 June 2014). *CNN Features Off-the-Shelf: An Astounding Baseline for Recognition.* Paper presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Verma, K., Deep, P., & Ramakrishnan, A. G. (2011, 16-18 Dec. 2011). *Detection and classification of diabetic retinopathy using retinal images.* Paper presented at the 2011 Annual IEEE India Conference.

Yu, F. (October 3, 2016). A Comprehensive guide to Fine-tuning Deep Learning Models in Keras (Part I). Retrieved from https://flyyufelix.github.io/2016/10/03/fine-tuning-in-keras-part1.html

Zacharski, R. (2018). Training Sets, Test Sets, and 10-fold Cross-validation. Retrieved from https://www.kdnuggets.com/2018/01/training-test-sets-cross-validation.html

Zhang, J., Xia, Y., Xie, Y., Fulham, M., & Feng, D. (2017). Classification of Medical Images in the Biomedical Literature by Jointly Using Deep and Handcrafted Visual Features. *IEEE Journal of Biomedical and Health Informatics*, 1-1. doi:10.1109/JBHI.2017.2775662

Zhou, L., Zhao, Y., Yang, J., Yu, Q., & Xu, X. (2018). Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images. *IET Image Processing, 12*(4), 563-571. Retrieved from http://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2017.0636