DIABETIC RETINOPATHY CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK AND STACKING ENSEMBLE OF CLASSIFIERS TECHNIQUES

BY

INUWA Rahmat
MTech/SICT/2018/8755

DEPARTEMENT OF COMPUTER SCIENCE
FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA

DECEMBER, 2021

DIABETIC RETINOPATHY CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK AND STACKING ENSEMBLE OF CLASSIFIERS TECHNIQUES

BY

INUWA Rahmat
MTech/SICT/2018/8755

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA, NIGERIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF TECHNOLOGY IN COMPUTER SCIENCE

DECEMBER, 2021

ABSTRACT

Early detection of diabetic retinopathy (DR) is critical, as prompt treatment can help reduce or even prevent visual loss. Most of the current state-of-the-art machine learning techniques for DR detection and classification make use of single classification for prediction. However this single classification models suffer from high variance, high bias, bottleneck in local optima, and the researcher also suffers the risk of choosing the wrong classifier. These issues can be solved by combining the predictions from multiple classifiers which produces predictions that are less sensitive to the specifics of the training data, the choice of training scheme and the serendipity of a single training run. Therefore, this research proposes an effective stacking ensemble technique for DR classification that will satisfy the drawbacks of using a single model, hence improve classification performance. The proposed stacking ensemble classifier was produced from a combination of four classifiers namely: Support Vector Machine (SVM), K-Nearest Neighour (KNN), Decision Tree (DT) and Naïve Bayes (NB). The proposed stacking ensemble technique was evaluated using the Messidor dataset. In comparison with the performance of the individual constituent models the proposed stacking ensemble technique achieved an acuracy of 99.17% which was better than the values achieved by the constituent models with accuracies of 98.33%, 96.67%, 93.33%, 95.00% for SVM, KNN, DT and NB, respectively. The proposed technique also produced better results than previous works based on Messidor dataset. These results suggest the robustness of the proposed model to DR detection and classification.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of study

Diabetes is the most common condition in the human body that causes many complications worldwide (Amin *et al.*, 2017). According to estimates from 2014, this disease's incidence rose from one hundred million patients in 1980 to four hundred and twenty-two million patients, with a global prevalence of 4.7% to 8.5% (Kirange *et al.*, 2019). Patients with a history of diabetes are more prone to diabetic retinopathy (Amin *et al.*, 2017). Diabetes is a condition in which the pancreas does not produce enough insulin or the body is unable to adequately process it (Verma *et al.*, 2011). Type 1 and Type 2 diabetes are the two forms of diabetes. A person with Type 1 diabetes has a pancreas that does not make insulin. Juvenile diabetes, often known as type 1 diabetes, is commonly diagnosed in children and teenagers. It can however, happen to adults. This kind of diabetes affects about 5% to 10% of diabetics (Wang and Lo, 2018).

Insulin resistance happens when the body does not create enough insulin or when the cells are unable to use it adequately, resulting in type 2 diabetes. Because type 2 diabetes is detected later in life, usually beyond the age of 45, it is referred to as "adult-onset diabetes." It is responsible for 90% to 95% of diabetics (Gulshan *et al.*, 2016). Type 2 diabetes has been identified more commonly in young adults, including children, in recent times than in the past (Gulshan *et al.*, 2016).

Diabetes has a negative impact on the kidneys, eyes, nerves, and heart (Tarr *et al.*, 2013). Diabetic nephropathy is the medical term for diabetic kidney illness (Selby and Taal, 2020). High blood glucose levels caused by diabetes can harm the component of the kidneys that purifies the blood in the case of diabetic kidney disease. When a filter is

broken, it becomes ‘leaky,' allowing protein into the urine (Sulaiman, 2019). Diabetes cause diabetic neuropathy, which is a type of nerve injury. High blood sugar levels can harm nerves all over the body (Nascimento *et al.*, 2016). The nerves in the legs and feet are the most commonly affected by diabetic neuropathy. Diabetic neuropathy symptoms can range from discomfort and stiffness in the legs and feet to difficulties with the digestive system, blood vessels, urinary tract, and heart, depending on which nerves are impacted (Gupta *et al.*, 2021). Diabetics may get Diabetic Foot Ulcers (Netten *et al.*, 2020). A diabetic foot ulcer is an open sore that affects about 15% of diabetic individuals and is usually found on the soles of the feet (Bus *et al.*, 2020).

Diabetic retinopathy (DR) is a vision-related consequence of diabetes. Diabetic retinopathy may cause no symptoms or just minor vision abnormalities at first. DR is a disease that tends to worsen and is one of the critical causes of blindness and vision loss (Arade and Patil, 2017). DR is a diabetes-related eye condition that arises when the retina's blood vessels swell and leak fluid, leading gradually to vision impairment (Khan, 2013). Anyone with type 1 or type 2 diabetes can have this DR disease. Diabetes causes excessive blood sugar levels to collect in blood arteries, obstructing or inhibiting blood flow to the vital organs, including the eyes, and affects up to 80% of all people with diabetes for 10 years or longer (Li *et al.*, 2019). This assumption facilitates the application of automated diagnostic screening methods to larger populations. DR symptoms include blurred vision, eyespots, and night vision difficulties (Rajalakshmi *et al.*, 2018).Early detection and regular eye exams could avoid vision loss and blindness, and they were critical in DR therapy (Ahmad *et al.*, 2011; Faust *et al.*, 2012; Ganesan *et al.*, 2014). Experienced optometrists or highly skilled eye technicians conduct retinal screening as a traditional and effective solution for the early detection of DR using retinal fundus photos acquired with a mydriatic or nonmydriatic camera (Rahimy, 2018;

Sandhu *et al.*, 2018). The modest differences between grades, as well as the availability of numerous minor fundamental qualities, make identification extremely challenging (Kumaran and Patil, 2018). Traditional manual DR screening, on the other hand, is difficult to perform and is subject to significant inter- and intra-observer variation, even among experienced ophthalmologists, which can lead to erroneous interpretation, a delay in appropriate diagnosis, and a strain on healthcare services (Arcadu *et al.*, 2019; ElTanboly *et al.*, 2017; Sarki *et al.*, 2020)

DR automated detection is required to address the issues of significant inter and intra-observer variation amongst ophthalmologists, which can lead to inconsistencies in interpretation, a delay in appropriate diagnosis, and a strain on health-care resources. Early-stage identification of DR, which can prevent blindness with appropriate care, is also crucial for diagnosis (Ahmad *et al.*, 2014). The creation of intelligent systems to assist ophthalmologists' decision-making has attracted the scientific community's attention in various works concerning incorrect diagnosis (Islam *et al.*, 2017; Pak *et al.*, 2020).

1.2 Statement of the Research

Deep learning models has been adapted by many previous works to perform diabetic retinopathy detection task (Dutta *et al.*, 2018; Sarki *et al.*, 2020; Zeng *et al.*, 2019). Deep learning models provide various advantages, such as features that are automatically determined and properly tailored for the desired conclusion. GPUs can also execute vast parallel computations and are scalable for big amounts of data (Benkaddour and Bounoua, 2017). Deep learning, on the other hand, suffers from the necessity for a vast amount of data to perform better than other techniques. There is no standard theory to aid one in choosing the correct deep learning architecture, since deep learning necessitates knowledge of topology, training method, and other factors. Deep

learning is also susceptible to high variance, a sort of error caused by a model's sensitivity to small perturbations in the training data. Because of the high variance, an algorithm would be forced to represent the noise in the training set (Nannia *et al.*, 2021).

Apart from deep learning models shallow learning models such as support vector machine, Naïve Bayes, Decision Trees and k-nearest neighbour have been used for DR detection and classification. However, each single mode is constrained to learn only some part of the structure of the data. Combining the strengths of different models can produce superior predictions. Single model prediction is also suffer from high variance or high bias (David and Suruliandi, 2019). When a model has a strong bias toward one of the outcomes of a problem it is trying to solve, it is said to have high bias. The model is most likely not learning enough from the training data when the bias is strong. The bagging ensemble method was used by Somasundaram and Alli (2017) for DR detection and classification. However the bagging ensemble method may still yield high variance estimator because trees are highly correlated, so may still over fit training data (Bühlmann, 2012).

Statistical, training data is mostly small relative to size of space needed to search, using stacking ensemble reduces risk of choosing the wrong classifier (Rahman and Tasnim, 2014). Computational, individual models get stuck in local optima; stacking ensemble can lead to better overall prediction. Training many models rather than a single model and combining the predictions from these models is an effective strategy to reduce the high variance and high bias of deep neural network and shallow learning models (Alabdulrahman, 2014).

In line with the identified problems a stacking ensemble method for DR classification is proposed which solves the problem of high variance, high bias, bottleneck in local

optima, and risk of choosing the wrong classifier which is identified with using a single classifier for prediction. The Alexnet convolutional neural network was utilized for feature extraction to take use of the benefits of deep learning models, such as features being automatically derived and optimally tailored for the intended conclusion.

1.3 Aim and Objectives

The aim of this study is to develop a model which will effectively classify Diabetic Retinopathy (DR) using convolutional neural network and stacking ensemble of classifiers techniques.

The Objectives of this study are to:

1. Extract deep features from diabetic retinopathy fundus images.
2. Develop an ensemble of classification models.
3. Evaluate the performance of the model in (ii) using accuracy, precision, recall, loss function, and f-score.

1.4 Scope of the Study

This research focuses on automatic diabetic retinopathy classification using the stacking ensemble. The stacking ensemble model was built using four classifiers namely Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT) and K-Nearest Neighbour (KNN). The bagging, boosting, averaging, and majority vote ensemble algorithms were not considered. Other single classifiers like discriminate analysis, Error-Correcting Output Codes, neural network and logistic regression as constituent classifiers for the ensemble were not explored. This thesis is also limited to just the Messidor dataset without consideration of other DR Dataset sources like e-ophtha, HRIS, DIARETDB1, DIARETDB0 and DRIVE.

1.5 Significance of the Study

This thesis makes contributions to the fields of medical image analysis and optometry. This study would be of benefit to clinicians on DR diagnosis and staging. Researchers would benefit

from this study as it gives insight on existing algorithms on DR disease detection and classification thus enhancing the decision making process towards selecting the appropriate detection and classification technique to implement or to modify towards diagnosis of DR.

1.6 Organization of Thesis

This thesis is divided into five chapters, beginning with Chapter 1 and ending with Chapter 5. The first chapter provides an overview of the research project. It also includes problem statements, goals and objectives, the scope of the research, and the significance of the research. Chapter 2 describes diabetic retinopathy in depth and analyses past studies on the subject undertaken by various researchers. The method utilized to analyze the topic is discussed in the third chapter. It supports the method for obtaining a solution to the problem. It goes over data collecting methods, feature extraction, and classifier ensemble. The details of the actual experimentation and the outcomes are presented in Chapter 4. The work was summarized, conclusions were reached, and recommendations for future work were offered in chapter five.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Preamble

The pathology and screening overview, as well as the strategies employed for DR detection in prior studies, are discussed in this chapter. The first portion covers the DR pathogen, which includes a fundamental explanation of the retina's structure, followed by a discussion of the disease's clinical symptoms.

2.2 Diabetes Retinopathy Pathology

Diabetes mellitus is a sugar metabolic illness caused by impaired insulin secretion and is characterized by elevated blood glucose levels (Ahmad *et al.*, 2011). Blood arteries that feed blood to essential organs can be damaged by hyperglycemia (high blood glucose levels). DR is a diabetic condition that affects the retinal vascular system, resulting in slow retinal degeneration and visual loss. In the working-age population, DR is considered as the primary cause of blindness (Ahmad *et al.*, 2014; Al-Hazaimeh *et al.*, 2018). Diabetes has been recognized as a major emerging global public health challenge (Banu *et al.*, 2016.); In the United Kingdom, 3 million people are estimated to have diabetes, with the number anticipated to double in the next 15-30 years. In the year 2000, 171 million individuals were predicted to have diabetes worldwide, with that number expected to climb to 366 million by 2030 (Dutta *et al.*, 2018) .If DR is caught early enough, laser therapy can help to prevent vision loss (Sopharak *et al.*, 2010). Interventions such as improved blood glucose control could also help mitigate future advancement of the DR when it is in its early stages (Anuradha and Velmurugan, 2015). Since there are often no signs or symptoms of DR, screening is critical for diagnosis. Patients can be handled appropriately once DR has been detected, with the goal of preventing vision loss (Kumudham and Shenbagavalli, 2013).

The retina is a light-sensitive tissue that lines the innermost surface of the eye. Millions of photo-receptors react to concentrated light and convert it to electrical signals in the retina. The optic nerve carries these signals to the brain, where they are converted into images. Retinal blood vessels, macula, optic disc, and fovea are all parts of the retina (Mangrulkar, 2017). The macula takes up a significant portion of the brain's visual capacity because of its large number of cone cells, which are responsible for fine visual acuity, colours and centre vision. The fovea, which is located in the centre of the macula, contains the greatest number of cone cells (Wang and Lo, 2018).

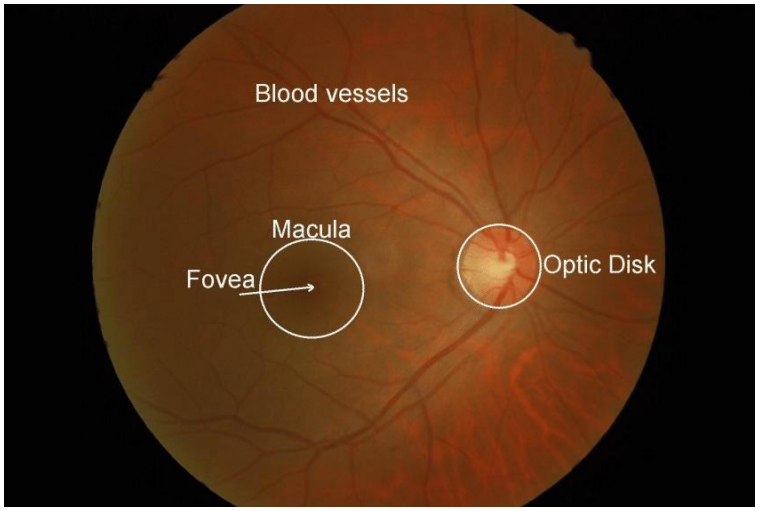

Figure 2.1 Retinal structures: Blood vessels, Optic Disk, and Fovea/Macula *(*Basit and Egerton, 2013)

Hyperglycemia causes diabetic retinopathy, a microangiopathy that affects the retinal vasculature. Blood and fluid escape from damaged retinal blood vessels, forming microaneurysms, exudates, haemorrhages, cotton wool patches, and venous loops (Tarr *et al.*, 2013). DR is a degenerative disorder, and areas of retinal ischaemia emerge as the blockages and damage of blood vessels deteriorate. The formation of new blood vessels is ignited in an attempt to revascularize the area. Due to the delicate nature of new vessels and the possibility of substantial bleeding, new vessels signify the later phases

of DR, posing a great risk of chronic vision loss. In the subsections below, the primary aspects of DR are detailed in further detail, with accompanying photos.

2.2.1 Microaneurysms

Microaneurysms are balloon-like entities that form on the edges of capillaries as the capillary walls deteriorates. Microaneurysms appear as single red dots unattached to any blood vessel on conventional fundus images because capillaries are not visible. They are frequently the first indicators of DR to be noticed (Somasundaram and Alli, 2017). Microaneurysms in the retina are depicted in Figure 2.2.

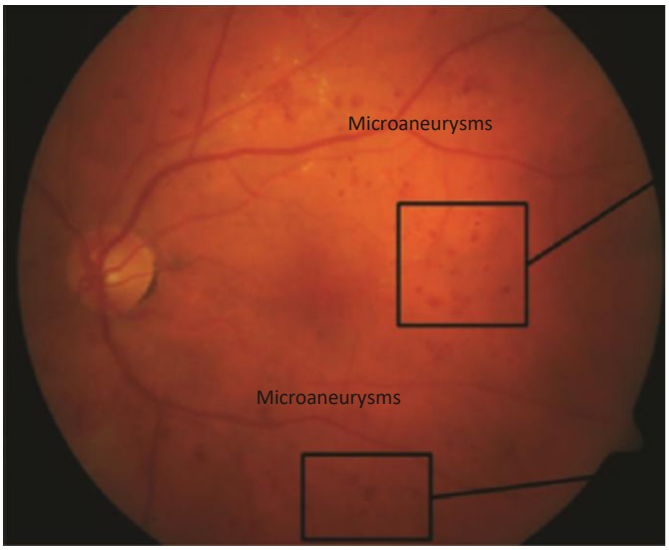

Figure 2.2 Microaneurysms in retina image *(Amin et al., 2016)*

2.2.2 Haemorrhages

Blood leaking comes from the shattering of capillary walls in this disorder, which can vary in size and form based on the retinal layer in which the capillaries are situated. Dot, blot, and flame haemorrhages are the three types of haemorrhages (Amin *et al.*, 2016). Figure 2.3 depicts a retinal picture with haemorrhages.

Figure 2.3 Haemorrhages in Retina Image *(Amin et al., 2016)*

2.2.3 Exudates

Oedema leaking is frequently caused by capillary collapse. Retinal thickness is caused by the accumulation of oedema. Macular oedema is the most common cause of vision loss in diabetics, and if it is medically severe, laser therapy will be required to mitigate the likelihood of vision loss (Dutta *et al.*, 2018). Oedema is a transparent fluid that cannot be seen with conventional 2D retinal imaging. Exudates are the lipid remnants from the oedema. Isolated patches, track lines, macular stars and circinates appear, as waxy yellow lesions with a variety of designs. Exudates are divided into two categories: hard and soft exudates.

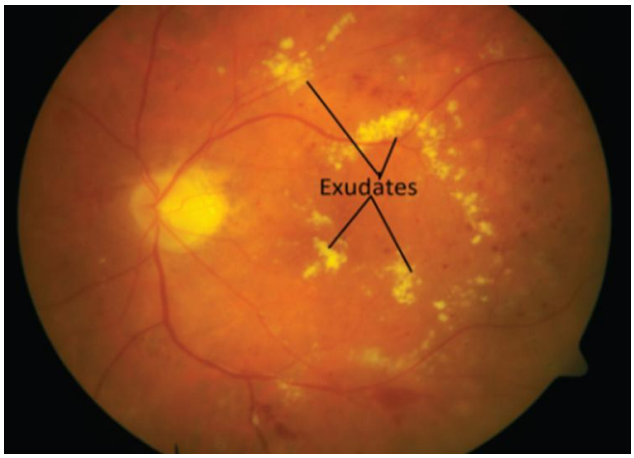

Figure 2.4 Exudates in Retina Image *(Sariera et al., 2020)*

1. ***Hard Exudates:*** These are a common symptom of DR and can range in dimension from microscopic dots to big spots with defined boundaries. In addition to blood, the eye contains liquid that is high in protein and fat, which leaks out to produce exudates. These can make it difficult to see because they block light from getting to the retina.
2. ***Soft exudates:*** These are sometimes known as "cotton wool patches," are more common in people with severe retinopathy.

Figure 2.5 Soft Exudate in Retina Image *(*Sariera *et al.*, 2020*)*

2.2.4 Neovascularisation

This is characterized by aberrant blood vessel proliferation in parts of the eye, particularly the retina, and is linked to blindness. This happens as a result of ischemia, or a lack of blood supply to the ocular tissues. Glaucoma can develop if these aberrant blood vessels develop around the pupil, raising the stress within the eye. The walls of these new blood arteries are thinner, making them more likely to rupture and haemorrhage, or to induce scar tissue to form, pulling the retina away from the back of the eye. When the retina peels away from the back of the eye, a retinal detachment happens, and if remain unattended, it can cause considerable blindness. Blood leaks can obstruct the vitreous and hinder light from reaching the retina through the pupil, resulting in distorted and hazy visions. Diabetic fibrous may grow on the retina in more advanced proliferate retinopathy (Tarr *et al.*, 2013).

2.3 Stages of Diabetic retinopathy

Table 2.1 summaries the stages of DR. Background DR, Proliferative DR, Pre-proliferative DR, and Maculopathy are the four primary types of DR. Background DR is the most mild kind of DR and poses no risk to vision. Pre-proliferative DR is a type of progressive retinal ischaemia that comes with an elevated risk of neovascularization. Proliferative DR is marked by neovascularization, and it is the most serious phase of the illness, with a high chance of vision loss. Table 2.1 demonstrates that proliferative DR encompasses characteristics other than new vasculature; however, keep in mind that these other characteristics are attributable to the existence of new vessels. Maculopathy can develop at any phase of DR, though it is more common as the disease develops. The presence of any DR characteristics at the macula is the formal definition of maculopathy, however it is typically reserved for vision-threatening macular oedema (Oloumi *et al.*, 2014).

Table 2. 1 Classification of DR

Stages of DR	Features
Background	Microaneurysms
	Dot and blot haemorrhages
	Exudates
Pre-proliferative	Multiple dot and blot haemorrhages.
	Cotton wool spots (CWS).
	Venous beading and loops.
	Intra-retinal microvascular abnormalities (IRMA).
Proliferative	New vessels elsewhere (NVE).
	New vessels at the disc (NVD).
	Pre-retinal/vitreous haemorrhage.
	Retinal detachment.
Maculopathy	Microaneurysms, haemorrhages, exudates at the macula.
	Macular oedema.

2.4 Diabetic Retinopathy Detection Techniques

In the last several decades, the topic of retinal photo analysis has sparked a lot of attention, with the automated identification of diabetic retinopathy receiving a significant portion of it. Landmark identification is another subject that has attracted a lot of attention. Optic disc, blood vessels, and the fovea are all landmarks. This section will begin with a quick overview of computerized blood vessel segmentation. Most DR detection strategies, especially new vessel detection methods, require it as a precondition before detecting diseased entities. There will be a brief discussion of the main approaches for detecting the key DR characteristics (microaneurysms, haemorrhages, and exudates). Following that, there will be a part that provides a detailed report of the detection of new vessels (proliferative DR). This chapter will conclude with a quick rundown of the machine learning algorithms described throughout this literature study. The majority of methods begin by preprocessing the photos. The primary preprocessing processes are used to compensate for reduced

lighting and contrast. Shade correction is a common method for dealing with poor lighting (Niemeijer *et al.*, 2005; Spencer *et al.*, 1996), whereby an image resembling the background is deducted from the initial image. The background image is created by using a median filter with a size that is much larger than the largest retinal feature. Contrast limited adaptive histogram equalization (CLAHE) is a common solution for poor contrast (Ramlugun *et al.*, 2012). This is a method for enhancing local contrast that is preferable to global contrast improvement. Preprocessing, on the other hand, can only repair to a certain extent; consequently, it is the photographer's obligation to ensure that an acceptable standard of photos is collected. Youssif *et al.* (2007) explains the preprocessing stages for retinal analysis in detail.

2.4.1 Vessel Segmentation

The rich red hue of vessels, their contrast with the background, and their gradient at vessel borders are all important characteristics that are used in segmentation algorithms (Cree *et al.*, 2005). Their cross section intensity profile resembles a Gaussian function and can be approximated as piecewise linear. Different methods have been published in the literature, and they can be divided into four categories based on mathematical morphology, matched filtering, vessel tracking, and machine learning. Fine, weak, and highly convoluted vessels might be difficult to find using vascular segmentation algorithms. The profile model can be complicated by vessel crossing and branching. Vessels can have high reflections along their midline, complicating the profile model even more. Pathologies can potentially reduce precision, resulting in false positives. On a pixel-by-pixel basis, vessel segmentation approaches are evaluated (Yin *et al.*, 2012).

2.4.2 Matched filtering

The matched filter was first introduced for vessel segmentation by Chaudhuri *et al.* (1989), and it has since become one of the most used approaches. It took advantage of the fact that a Gaussian function can estimate the cross section of the vessels, as well as the notion that vessels can be thought of as piecewise linear segments. As a result, a two-dimensional Gaussian filter was employed, which varied from a conventional isotropic Gaussian filter is that a single Gaussian function transverse profile was reproduced an amount of times and packed to make up the filter's length. The filter's length was chosen to be close to the distance at which vessel segments were considered to be linear. To "match" the blood vessel segments, this filter, which matched the geometry of vessel segments, was combined with the retinal picture. In order to identify vessels of various orientations, the filter was also rotated. The matching filter response (MFR) was created as a result of this, resulting in a significantly improved image. After that, a global threshold was used to create a binary vessel map.

Al-Rawi *et al.* (2007) increased the matched filter's performance by utilizing an optimization algorithm to identify the ideal filter settings automatically. Regrettably, the matched filtering approach responds to non-vessel edges as well as vessels. The most bothersome are the step borders generated between exudates and the background. A single global MFR threshold is insufficient and can lead to several false positives. As a result, a slew of modified matched filtering approaches have been presented.

Using vessel structural information, Hoover *et al.* (2000) proposed a piecewise threshold probing approach. The algorithm explored the MFR and employed a set of parameters to establish the threshold for each location in order to segment vessels during each probe. The point that the MFR highest point for a vessel is substantially superior to its surrounding points on either sides, while the MFR peak point for non-

vessel edges is not much superior than its neighbours on either sides was utilized by Lei Zhang *et al.* (2009). As a result, a two-sided thresholding method was proposed.

2.4.3 Mathematical Morphology

For the extraction of vessel midline, Mendonca and Campilho (2006) used variance of offset Gaussian filtering. Going back to the initial preprocessed photo, vessel improvement was conducted individually at diverse levels using a tweaked top hat operator with an expanding disc structural component to boost vessels of various widths. The double threshold operator was used to do morphological rebuilding at each scale, resulting in a binary vessel map. The final vessel segmentation was created by doing iterative region-growing utilizing the vessel mid-lines and various binary vessel maps. Fraz *et al.* (2012) proposed an improved Mendonca and Campilho model (2006). The vessel mid-lines were found using the Gaussian kernel's first order derivative. Bit plane slicing was utilized to construct a binary vessel map once the top hat vessel enhanced image was acquired; the sum of the higher order bit planes was used to generate a binary vessel map.

2.4.4 **Vessel** tracking

Cree *et al.* (2005) used a two-dimensional Gaussian framework to track vessels. A starting vessel point, as well as estimates of its breadth and direction, had to be manually picked. A tiny local area was removed around this point, and a Gaussian model with the same breadth and alignment was fitted by the non-linear least squares optimization approach. The fitted model was used to take precise measurements of vessel width and alignment. A tiny step was taken toward the vessel, and earlier dimensions were used as approximations to create a new fitting. Additional methods include a vessel tracking technique founded on utilizing a probabilistic design (Yin *et*

al., 2012) and a method based on multi-scale line tracking (Vlachos and Dermatas, 2010).

2.4.5 Machine Learning

In machine learning method image pixels are categorized into vessel or non-vessel. Supervised and unsupervised learning methods are used for this classification.

1. **Supervised Learning:** Sinthanayothin *et al.* (1999) used principal component analysis (PCA) to limit the photo to simply structural detail. To quantify edge strength, the canny edge detector was applied to the first principal component. The first principal component values and edge strong point were used as input data for a neural network classifier. The method utilized by Staal *et al.* (2004) was to extract ridges, which were then used to create photo primitives in the shape of line components. The image was then divided into patches using these line elements. A vector of 27 characteristics obtained from qualities of the spots and line components were used to categorize pixels using a KNN algorithm. Soares *et al.* (2006) produced a feature vector for each pixel based on the intensity of the pixel and the response of a two-dimensional Gabor wavelet applied at various measures and alignments. A Bayesian classifier was used to classify pixels, with class likelihoods given as a linear combination of Gaussian functions.
2. **Unsupervised:** For vessel tracking, Tolias and Panas (1998) suggested an unsupervised fuzzy method. A fuzzy C-means clustering approach was used to find the membership functions of the 2 linguistic values (vessel and non-vessel). No hypotheses regarding the shape of the vessels were established, and no edge information was necessary, because the suggested solution relied entirely on intensity data (usually corrupted by noise). By using matched filtering, Kande *et*

al. (2010) increased the contrast of blood arteries against the backdrop. To segment the vessels, a spatially weighted fuzzy C-means clustering technique was used to label the improved image. The spatial weighting component took into consideration the fact that grey level spatial distributions as photo intensity are not autonomous of one another.

2.5 Related Works

Computer vision field, the task of detecting DR early is a challenging issue. Diagnostic clarity criteria aim to identify clinical characteristics of Diabetic Retinopathy such as haemorrhages, microaneurysms, soft exudates, and hard exudates. It is an essential issue for a proper diagnosis to extract these signs as they help to determine the actual condition of DR.

Kirange *et al.*(2019) suggested a new technique for early-stage identification of DR by recognizing all microaneurysms, the first symptoms of DR, and correctly assigning labels to retinal fundus images grouped into five classes according to the seriousness of lesions. The five grading groups are: No DR, Mild DR, Medium DR, Severe DR, and Proliferative DR. Five standard classifiers were used in this proposed system to perform the classification task. These classifiers are SVM, KNN, Neural Networks (NN), NB, and Decision Tree (DT). The NB classifier was proposed to have surpassed the other four classifiers with an accuracy of 77.86%. Both the Gabor and the LBP descriptor were used for the extraction of features. However, the components extracted using the Gabor descriptor performed much better with an accuracy of 77.86% as compared to the LBP features that provided 41.84% accuracy. A drawback of this analysis is that it focused more on early-stage DR identification without considering the DR proliferation stage.

A graph-based approach to classifying retinal images was suggested by Mangrulkar (2017). The retinal images were pre-processed to eliminate noise and remove irrelevant

information. The Canny edge detector was then utilized to identify the edges of the items in the image. Using the kirsch template that defines the presence of an edge, the segmentation process was then performed. The Kirsch model is used for the retrieval of blood vessels from the retinal image. Together with the graph nodes extracted from the image, the Speed-Up Robust Features (SURF) features were extracted by finding the intersection points and the terminal ends. Using the graph-based method, classification was carried out, and the Artery Vein Ratio (AVR) was measured. The AVR ratio is a realistic measure to classify a diabetes-free or diabetes patient. The proposed process achieved an accuracy of 88%. Without considering a more advanced DR stage, this research only focused on the early phase identification of DR.

Sarwinda *et al.*(2017) provided a full model of Local Binary Pattern (LBP) as a texture feature descriptor technique for DR detection. In this study, the feature selection method was Expectation Maximization-Principal Component Analysis (EM-PCA), and the classification technique was KNN. The LBP feature descriptor approach was used to extract magnitude, sign, and mean values. The STARE diabetic retinopathy database was used in this investigation, which contains 66 DR photos and 44 normal images with a resolution of 700 x 600 pixels. A combination of the LBP sign and magnitude value showed a better sensitivity of 98.48% than a mixture of LBP mean value and sign, and the LBP magnitude and mean values with sensitivity of 97.5% and 97% respectively. A drawback of this study is that the obtained results were not benchmarked with related works that used the STARE database. Also, the extracted LBP features were only tested on the KNN classifier. More classifier can be used for testing the efficiency of these extracted features to improve the model robustness.

Costa *et al.*(2018) established a novel Multiple-Instance Learning-based weakly-supervised DR diagnosis system. The method used inherent local information to generate

predictions on new photos based on weak information about the absence or presence of the disease. The adoption of a joint-learning approach in which the encoding and classification phases are linked is the study's main contribution. SURF (Sped-Up Robust Features) were used to locate and define occurrences within retinal pictures in this study. The DR detection model was evaluated using the publicly available Messidor dataset. The proposed technique achieved an area under the curve of 90% on the Messidor dataset and 93% on the DR1 dataset. This study was limited to just the SURF feature; however, exploring more features such as texture, deep learning and image degrading features could help describe the DR disease more effectively which would in turn improve the proposed systems' performance.

A new approach to the diagnosis of Age-related Macular Degeneration (AMD) and DR, as proposed by Morales *et al.* (2017). The presentation of a new technique for the diagnosis of AMD and DR was the objective of this method. Five experiments were developed and tested using the suggested procedure: separating DR from normal, AMD from normal, pathological from normal, DR from AMD, and the three different classes (AMD, DR, and Normal): The LBP was used as the feature descriptor technique. The study's most noteworthy conclusion is that the new approach can distinguish between groups based on an analysis of the retina's spatial texture; thereby removing the retinal lesion's previous segmentation. The findings suggest that employing LBP as a texture descriptor for fundus photos provides useful characteristics for detecting retinal illness. This work, however, only investigated the LBP without further searching for more texture descriptors.

A multi-stage transfer learning system and an automated method for detecting the DR stages from a single human fundus image were proposed by Tymchenko *et al.* (2020). Three Convolutional Neural Network (CNN) architectures (EfficientNet-B4,

EfficientNet-B5, and SE-ResNeXt50) were ensemble. CNN was used as a function extractor and as a classifier. The CNNs pre-trained by Imagenet were used for encoder activation. The proposed technique was used for the early detection of DR and achieved a sensitivity and specificity of 0.99. The Shapley Addictive exPlanations (SHAP) were used to explain characteristics that lead to the disease process evaluation—using SHAP guarantees that the model learns beneficial features during preparation and uses correct characteristics at an inferential time. This approach's main advantage is that it increases generalization and eliminates uncertainty using a network ensemble, pre-maintained on a large dataset and precisely tuned to the target dataset. This analysis can be extended with SHAP calculation for the entire ensemble, not just for a particular network, which can provide a more precise optimization of hyper-parameters.

Al-Hazaimeh *et al.* (2018) suggested an effective image processing method for detecting DR illnesses from retinal fundus pictures. Preprocessing, blood vessel segmentation and removal, optic disc detection and removal, fovea elimination, feature selection, feature extraction, and classification were all part of the proposed automatic screening method for DR. The proposed method was benchmarked using the DIARETDB1 publicly available dataset. The Gray-level co-occurrence matrix (GLCM) was used to extract the microaneurysm, retinal hemorrhage and exudate features from the retina fundus images. Deep belief network was used to carry out feature selection. SVM model was adopted for data classification. The proposed technique was validated using sensitivity, specificity and accuracy performance metric. In this study an accuracy of 98.4%, sensitivity of 99% and specificity of 96% was achieved. The DIARETDB1 consist of 89 colour fundus images was used for training and testing the proposed method which makes the method limited to a small dataset size. Increasing the dataset size would produce a more robust model.

Sudha and Karthikeyan (2018) presented an analysis of DR using the Naïve Bayes classifier technique. There are 385 instances and 9 features in the collection. This data was obtained from the Sakarya University Educational and Research Hospital's Eye Clinic. Hemoglobin, URE, Glycated Hemoglobin, High-Density Lipoprotein, Diabetes Duration, Low-Density Lipoprotein, Creatine, Triglyceride, and Glucose are among the dataset's attributes. The classification accuracy of the Naïve Bayes classifier was 89%. To improve the clarity of the proposed system performance more performance metric like precision, loss function, recall, area under the curve and f-score should be used.

Li *et al.* (2019) introduced a novel method based on a Deep CNN (DCNN). In this paper, the regular DCNN max-pooling layers were replaced by a fractional max-pooling layer. Two DCNNs with differing numbers of layers were prepared for classification to achieve more discriminatory features. After integrating features from image metadata and DCNNs, the SVM classifier was trained to learn the inherent limits of dispersals of each category. The proposed DR method classifies DR phases into 5 groups, labeled with a number ranging from 0 to 4. The test results indicate that the suggested technique can reach a recognition rate of up to 86.17%. The dataset used for training in this study had an insufficient number of images of lesions 3 and 4, limiting the proposed method.

Islam *et al.* (2017) presented an automated DR detection technique based on bag of word approach. This method classifies diabetic retinopathy images into absence or presence DR based on five publicly available datasets: Messidor, STARE, DIARETDB1, DIARETDB0 and HRF. Speed up Robust Features (SURF) was used as the feature extraction technique. Using the K-means clustering algorithm, the extracted SURF feature was assigned to several clusters. Visual words are represented by the cluster centers, and these words together make up the lexicon, or bag of words. Each individual characteristic in the image is quantized to the nearest word, resulting in a histogram in

which each bin represents the frequency of a word inside that bag of words. SVM was used to classify the data. The proposed technique had a 94.4% accuracy, precision, recall, and f-score of 94%. In this study the DR images were just classified in normal or abnormal images without consideration of the severity level of the DR.

Zeng *et al.* (2019) suggested a binocular Siamese-like CNN for automatic DR detection. The suggested method takes as inputs binocular fundus images and discovers their correlation to aid in forecasting. The model basically takes two fundus photos belonging to the left and right eyes as inputs and sends them to the Siamese-like modules. In the fully-connected layer, the data from two eyes are integrated, and the model then produces the diagnosis result for each eye independently. In this study, the Kaggle DR Image dataset was employed. A total of 35126 high-resolution fundus pictures were obtained under a variety of imaging circumstances for the data set. The proposed binocular model performed well, with an area under the curve of 0.951, a recall of 82.2% and a specificity of 70.7%. The suggested method has the disadvantage that binocular models will have difficulty training or validating with datasets that do not feature paired fundus images.

Arade and Patil (2017) conducted a comparative study of DR using the K-NN and Bayesian classifier. An automated image processing system that detects DR gradation is presented in this paper. Blood vessel segmentation was done using the kirsch process, as it was found that retinal photos effectively differentiated the blood vessels. Differentiated vessels were extracted using moment invariants, grey level features. The DR severity was identified along with K-NN and Bayesian classifier using a feed-forward neural network. To validate the results obtained with an ophthalmologist, it was indicated that the Bayesian classifier generates results comparable to the expert opinion than the K-NN classifier. The accuracy of the Bayesian classifier obtained is 74%, while the precision

for K-NN is 66%. It is possible to expand this work and improve classification performance by training more classifiers or performing an ensemble.

2.6 Summary of Review

Several publications and individuals have published work on numerous machine learning models for DR discovery and classification in one form or another. Most of the techniques used are single models which suffer from high variance and high biasness that could lead to low model performance. To pick the best model for classification must of the publications trained several models and performed a comparison of these trained models to choose the best model which is time consuming. Some literature made use of the begging ensemble to prevent choosing the wrong classifier, however bagging ensemble method may still yield high variance estimator because trees are highly correlated, and so may still over-fit training data.

CHAPTER THREE

3.0 METHODOLOGY

3.1 Preamble

The phases involved in the automatic diagnosis of fundus images are covered in this chapter. It begins with a quick review of the steps involved in DR diagnosis as depicted in Figure 3.1. This is followed by a thorough examination of each of the processes involved in this research.

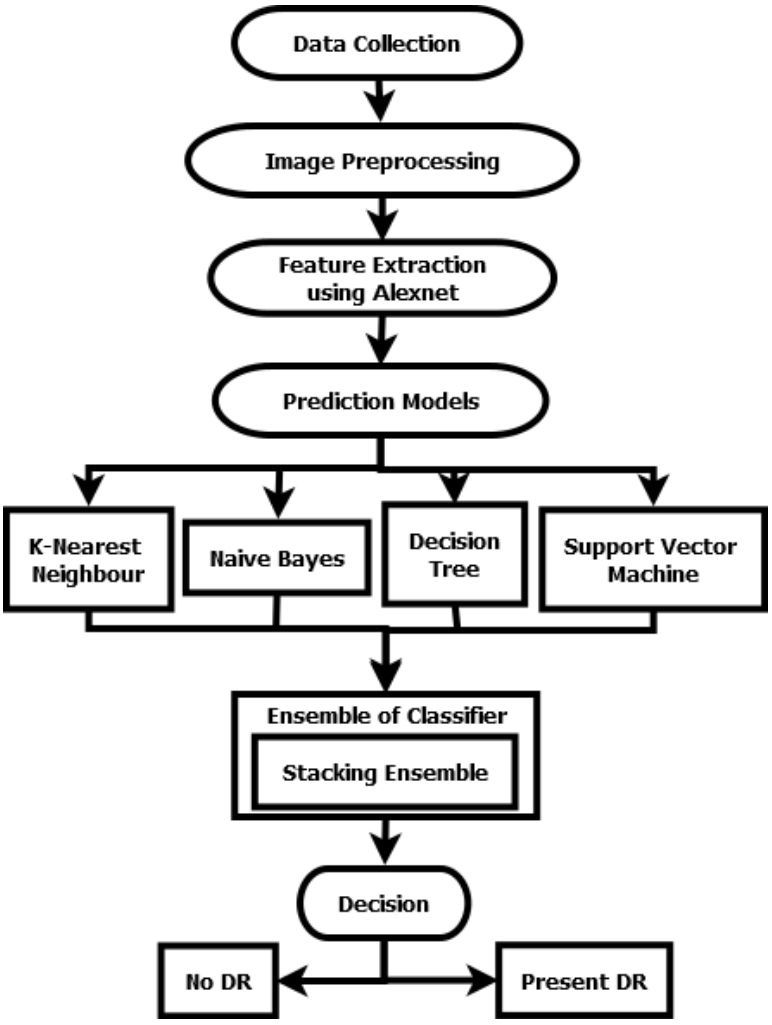

Figure 3. 1 Block Diagram of the Proposed System

3.2 Dataset

In this study the Messidor dataset (Messidor, 2021) was used for training and testing the proposed technique. The data set has 1200 RGB fundus images together with comments in an excel file (Kalyani *et al.*, 2021). Three ophthalmology units analyzed the photographs. Images with resolutions of 2240 by 1488, 1440 by 960, or 2304 1536 pixels were captured using 8 bits per color. Out of the 1200 images, 400 photos were taken without pupil dilation and 800 photographs were taken with pupil dilation (Zago *et al.*, 2020). The dataset is divided into four grades: 0, 1, 2, and 3, with 0 indicating no DR and the other three indicating DR. The severity levels are indicated by grades 1, 2, and 3, with 1 being the least severe and 3 being the most severe (Saxena *et al.*, 2020). In this study the images were categorized into two classes: 0 for normal and grades 1, 2, 3 were all changed to 1 to indicate diabetics' retinopathy. Figure 3.2 and 3.3 are examples of the Messidor images.

Figure 3. 2 Sample of Messidor normal image (Amin *et al*., 2016)

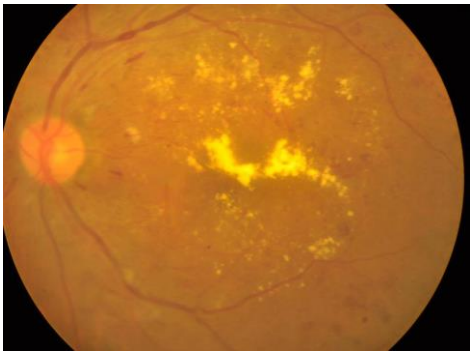

Figure 3. 3 Sample of Messidor diabetics retinopathy image (Sariera *et al*., 2020)

3.3 Image Preprocessing

The quality of the input used in model training determines the accuracy or efficiency of the categorization model. As a result, the training data in the datasets is preprocessed before being used as inputs. The retinal images in the databases come in a variety of

sizes (1440 by 960, 2304 by 1536 and 2240 by 1488). All photos are resized and reduced to the size 512 by 512 as the first stage in preprocessing.

3.4 Deep Feature Extraction

Feature Extraction is a dimensionality reduction method by which the initial raw data collection is reduced to more controllable classes. It also deals with creating variables to get around issues while describing the data with adequate accuracy. After the face pre-processing stage, the pre-processed image will be feed into the convolutional neural network for feature extraction. In this research the Convolutional Neural Network (CNN) Alexnet was used for deep feature extraction.

In this research 80% of the dataset was used for training the classification models and the remaining 20% was used to test the classification models.

3.4.1 Convolutional neural network (CNN)

Presently, in image processing and analysis field, convolutional neural network has become a vastly proficient method of feature extraction and identification (Benkaddour and Bounoua, 2017; Simonyan and Zisserman, 2015). CNN is the most representative model of deep learning (LeCun *et al.*, 2015). CNN is a multi-layer neural network; each layer consist of multiple 2D surfaces, and each plane consist of multiple independent neurons (Liu, 2018). CNNs have a large number of connections, and its design is made up of many layers, such as pooling, convolution, and fully-connected layers, that achieve some form of regularization (Ferreira and Giraldi, 2017). Deep architecture is used by CNN to learn complex features and functions that can be used to represent high-level abstractions. Deep architectures are made up of a high number of neurons and multi-level latent non-linearity calculations. Each level of CNN architecture signifies features at a particular degree of abstraction, which are defined as a set of lower-level features.(Albelwi and Mahmood, 2017). The standard model of CNN has a

structure composed of the input layer, alternating convolutional layers, pooling layers and non-linear layers (Vedaldi and Lenc, 2016). The convolutional layer and the pooling layers are responsible for feature extraction while the fully connected layers performs the classification on the features extracted by the convolutional and pooling layers (Benkaddour and Bounoua, 2017). The convolutional and pooling layers are discussed in the sub-sections below.

3.4.1.1 Convolutional Layers

CNN's basic building block is its convolutional layers. The main goal of convolution (Namatēvs, 2017). Is to extract different features from the input. Each kernel is used to generate a feature map, and these layers are made up of a succession of filters that try to extract local features from the input. Low-level significant features such as edges, corners, textures, and lines are retrieved in the first convolutional layer. The convolutional layer after that extracts higher-order features, but the highest-level features are extracted in the final convolution layer (Hossain and Alam Sajib, 2019).

3.4.1.2 Pooling Layer

The pooling layer, also known as a sub sampling layer, is used to decrease the resolution of prior feature maps by compressing features and lowering the network's computation cost (Sultana *et al.*, 2018). It frequently acts as a bridge between multiple convolutional layers (Liu, 2018). It fine-tunes the noise and disorder-resistant properties. In general, a pooling layer down samples the input map and reduces the dimensionality of the feature maps utilized by subsequent layers (Ferreira and Giraldi, 2017). Pooling divides the inputs into R x R-sized areas, with each region producing one output. The average of a rectangle neighbourhood, max pooling, and pooling via down-sampling are among pooling algorithms used by CNN.

To produce one output from each zone, pooling splits the inputs into regions of size *N x N*. (Padmanabhan, 2016). If an input of size *S x S* is fed to the pooling layer, the output size O may be calculated using equation 3.1:

$$O = floor(\frac{S}{N}) \quad (3.1)$$

Each output map can mix convolution with numerous input maps (Namatēvs, 2017).

Which can be expressed as follows in equation 3.2 (Namatēvs, 2017):

$$X_j^L = f(\sum_{i\epsilon M_j} X_j^{L-1} * K_{ij}^L + b_j^L) \quad (3.2)$$

Where
L – The convolutional layer;
L−1 – the downsampling layer;
X_j^{L-1} – input features of $L-1$ convolutional layer;
K_{ij}^L – Kernel maps of *L* convolutional layer;
b_j^L – Additive bias of *L* convolutional layer;
Mj – represents a selection of input maps;
i-th – input;
j-th –output.

In general, feature extraction with CNNs is made up of several similar processes, each of which is made up of three cascading layers: the convolution layer, the activation layer, and the pooling function (Liu, 2018).

3.4.2 Alexnet

The Alexnet was utilized to extract deep features from DR images in this research. In 2012, Alexnet took first place in the Imagenet large-scale visual recognition competition (Krizhevsky *et al.*, 2017). Alexnet is made up of eight layers, five of which are convolutional and three of which are completely connected. The first convolutional layer conducts convolution and maximum pooling with 11-by-11 filters. Max pooling is achieved using 3-by-3 filters with a stride size of 2. The second layer, which has a 5-by-5 filter layer, performs the same functions as the first. The max pooling procedures are carried out using 3-by-3 filters with a stride size of 2 pixels (Alom *et al.*, 2019). In the third, fourth, and fifth convolutional layers, the filter size is 3-by-3. At the fifth layer,

max pooling operations are done with 3-by-3 filters with a stride size of 2. Each of the sixth and seventh fully connected layers contains 4,096 neurons. The numbers of classes to be classified by ImageNet dataset consist of 1,000 classes. Therefore the final fully connected layer also contains 1,000 neurons (Alom *et al.*, 2019). The ReLU activation function is implements to the first seven layers respectively. A dropout ratio of 0.5 is applied to the sixth and seventh layer. The eighth layer output is finally supplied to a softmax function. Dropout is a regularization technique, being used to overcome the over fitting problem that remains a challenge in a deep neural network. Thus, it reduces the training time for each epoch (Zulkeflie *et al.*, 2019).

3.5 Ensemble Classification

Classification is a predictive modeling task in machine learning where a class label is predicted for a given instance of input data (Sultana *et al.*, 2018). The DR classification in this study was done using an ensemble classifier. A machine learning ensemble is a model that integrates the predictions of two or more models (Habib *et al.*, 2017).

A model's bias and variance must be minimized in order for it to attain acceptable classification performance. The significant variance of single classifier models can be mitigated by training numerous models and combining their predictions. The objective is to aggregate forecasts from a number of good but disparate models. When many neural networks' predictions are combined, a bias is introduced, which counteracts the variance of a single trained classifier model. The ultimate result is forecasts that are less sensitive to training data particular, training scheme selection, and the serendipity of a single training run.

The ensemble can produce better forecasts than any single best model, in addition to minimizing prediction variation. The suggested stacking model is part of the ensemble learning class, which offers methods for combining predictions from many models

prepared for the same task. Ensemble learning entails training many models on the same data set, then making predictions using each of the learned models before integrating the predictions in some fashion to provide a final outcome or prediction. Four classifiers were trained and integrated to reduce the high variance of a single classifier and maybe increase the accuracy of the final model. The subsections following go over these classifier models.

3.5.1 Support Vector Machine (SVM)

Statistical learning theory is used to develop the SVM algorithm (Cao *et al.*, 2019). The algorithm is based on the structural risk minimization principle, which allows it to compress an array of raw data into a support vector set and learn how to achieve a classification decision function (Ghosh *et al.*, 2019). The SVM model iterates over a collection of labeled training samples to locate a hyper-plane that produces an optimal path cap by finding data points. The use of support vectors improves class differentiation (Walsh, 2019). The decision function of a binary SVM in the input space is expressed in Equation 3.3.

$$\gamma = h(x) = sign\left(\sum_{j=1}^{n} u_j y_j \, K(x, x_j) + v\right) \tag{3.3}$$

Where x is the feature vector to be classified, j is the training instance index, n is the number of training example, and y_j is the training example label (1 or –1). u_j and v are fitted to the data to optimize the margin, and j, K (,) is the kernel function. Support vectors are training variables for which $u_j \neq 0$ (Sopharak *et al.*, 2010).

3.5.2 Naïve Bayes

This supervised learning method and statistical classification schemes are both demonstrated in the NB Model. It is based on an intrinsic probabilistic model and aids in measuring the results' probabilities to obtain principled uncertainty about the model (Sopharak *et al.*, 2010). The NB classifier is a probabilistic machine learning algorithm based on the Bayes theorem and the assumption of great feature independence. Learning involves numerous linear parameters in the number of problem functions, and NB classifiers are very scalable (Sudha and Karthikeyan, 2018). The Bayes theorem provides a way to compute the posterior probability $P(x|y)$from$P(x)$, $P(y)$ and $P(y|x)$ in NB. Equation (3.4) and (3.5) presented the equation for posterior probability$P(x|y)$.

$$P(x|y) = \frac{P(y|x) \times P(x)}{P(y)} \tag{3.4}$$

$$P(x|y) = \frac{P(y_1|x) \times P(y_2|x) \times \ldots \times P(y_n|x) \times P(x)}{P(y_1,\ldots,y_n)} \tag{3.5}$$

3.5.3 K-Nearest Neighbour

In KNN an item is classified based on its "distance" from its neighbours, and it is allocated to the most common class of its k closest neighbours (Bethanney*et al.*, 2015). If k = 1, the algorithm becomes the nearest neighbour algorithm, and the object is allocated to the nearest neighbour's class. This number K indicates how many neighbours an object has (Zhang, 2016).

The Euclidean distance is a linear distance between two points in Euclidean space (Cunningham and Delany, 2007; Zhang, 2016). If two vectors y_i and y_j are given where yi= (y_{i1}, y_{i2}, y_{i3}, …, y_{in}) and y_j= (y_{j1}, y_{j2}, y_{j3}, …, y_{jn}), Then the Euclidean distance between y_i and y_j is given in equation (3.6):

$$D(y_i, y_j = \sqrt{\sum_{k=1}^{n} (y_{ik} - y_{jk})^2} \quad (3.6)$$

The following is a description of the K-NN algorithm:

- Step 1: Assigns a positive integer k to each new sample.
- Step 2: In the database, select k entries that are closest to the new case.
- Step 3: The most common category is found for such entries.
- Step 4: We assign a category to the new sample.

3.5.4 Decision Tree (DT)

A Decision Tree (DT) is a simple predictive modeling tool that is widely utilized. DT is a type of supervised learning in which data is repeatedly separated based on a specific parameter (Patel and Singh, 2015). The decision tree employs a tree-like model to progress from observations about an item (represented by the branches) to inferences about the item's target value (defined in the leaves) (David *et al.*, 2015). Regression and classification problems can be solved using the DT algorithm. DT is easy to understand and view. Therefore, does not necessitate data standardization or preparation, and it requires less labour. The decision to do strategic splits has a significant effect on a tree's precision (Olaniyi *et al.*, 2017). Entropy, information gain and reduction invariance are techniques used in determining which attribute to the position at the root or the different levels of the tree.

The entropy of processed data is a measure of its randomness. The higher the entropy, the more difficult it is to make any conclusions from the information. A branch with an entropy of zero, for example, is picked as the root node, and a branch with an entropy greater than zero requires additional division (Olaniyi *et al.*, 2017). In equation 3.7, entropy for a single attribute is expressed.

$$E(S) = \sum_{i=1}^{n} -p_i log_2 p_i \quad (3.7)$$

Where S is the current state, p_i is the probability of an event i of state S.

Information Gain (IG) is a statistical feature that measures how effectively training data are classified according to a particular attribute's target classification. Information gain is mathematically described in equation 3.8.

$$IG = Entropy(before) - \sum_{j=1}^{N} Entropy\ (j, after) \quad (3.8)$$

Where "before" is the dataset before the split, N is the number of subsets generated by the division, and (j, after) is subset j after the division.

Reduction invariance is a technique for solving regression problems. To choose the optimal split, this algorithm uses the usual variance formula. The split with the lowest variance is selected as the criterion for dividing the population. The usual variance formula employed in this technique is stated in equation 3.9.

$$variance = \frac{\sum(X-\mu)^2}{n} \quad (3.9)$$

Where μ the mean of the values and X is the actual value and n is the number of values.

3.5.5 Stacking Ensemble

Ensemble modelling is a process in which many separate models are developed to predict an outcome, either via the use of many different modelling techniques or through the use of numerous training data sets (Nti *et al.*, 2020). Ensemble methods in machine learning combine many learning algorithms to achieve greater predicted performance than any of the component learning algorithms alone (Somasundaram and Alli, 2017). The goal of employing ensemble models is to lower the prediction's generalization error. When using the ensemble method, the prediction error lowers as long as the basis models are diverse and independent. As a result, the main benefit of

ensemble learning is that it reduces classification variance, which enhances prediction performance (Nannia *et al.*, 2021).

Stacking is a machine learning ensemble that combines predictions from numerous models to create a new model, which is then used to make predictions on the test data set (Lauría *et al.*, 2018). Firstly, all of the other models are trained with the available data, and then a combiner model is trained with all of the other algorithms' predictions as extra inputs to generate a final prediction (Nti *et al.*, 2020).The basic idea of stacking ensemble is to "stack" the predictions of models $(m_1, m_2, \dots, m_n)$ by a linear combination of weights $(w_1, w_2, \dots w_n)$as expressed in equation 3.10.

$$f_{stack}(x) = \sum_i^n w_i, f_{i\,(x)} \tag{3.10}$$

Where the weight vector "*w*" is learned by a meta-learner. The algorithm in algorithm 3.1 summarizes the stacking ensemble.

Algorithm 3.1: Stacking Ensemble Algorithm

Input: Dataset D = $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$;
Output: Ensemble Classifier H;
First-level: learn base-level classifiers $L_1, \dots, L_T$;
For $t = 1$ *to* T **do**
    Learn h_t based on D ($h_t = L_T(D)$)
End
Second-level: Construct new dataset of predictions
For $i = 1$ *to* n **do**
    $D_h = \{x_i', y_i\}$, *where* $x_i' = \{h_1(x_i), \dots, h_T(x_i)\}$
End
Third level: Learn a meta-classifier
Learn Hbased on D_h
Return H

Algorithm 3.1 consists of two classification level: the base-level classification and the meta-level classification. From algorithm 3.1 the required input is the dataset and the expected output is the stacked ensemble classifier. The first step is to learn and predict using the base (single) classifiers after which a new dataset is constructed at the second

level. Lastly a meta-classifier is used to learn the constructed new dataset which in turn returns the ensemble classifier.

3.6 Performance Metric

3.6.1 Accuracy

The rate of correct classifications is used to define accuracy. This is the number of correct guesses divided by the total number of right forecasts. The exact formula is given in equation 3.11:

$$\text{Accuracy}= \frac{\text{True positive + True negative}}{\text{True positive + True negative + False positive + False negative}} \quad (3.11)$$

3.6.2 Recall

Sensitivity is another term for recall. The amount of correct positive predictions that could have been made from all positive predictions is calculated by recall. The recall is calculated using the formula in equation 3.12.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}} \quad (3.12)$$

3.6.3 Precision

Precision is a metric used to calculate how many positive predictions are accurately made. The number of true positive elements is derived by dividing the total number of true positives by the total number of false positives. The formula in equation 3.13 is used to define precision.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \quad (3.13)$$

3.6.4 F-Score

The f-score of a model is defined as the harmonic average of precision and recall. F-Score is represented in equation 3.14.

$$\text{F-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3.14)$$

3.6.5 Binary Cross Entropy Loss Function

Loss functions are utilized to define the error between the output of an algorithms and the given target value (Ruby *et al.*, 2020). A loss function maps decisions to their associated costs. The lower the loss function the better the classification model. In this study the binary cross entropy was used as the loss function. Binary cross entropy is the negative average of the log of corrected predicted probabilities (Zhang and Sabuncu, 2018). The loss function formula is given in equation 3.15.

$$\text{Loss Function} = \frac{1}{N}\sum_{i=1}^{N} -(y_i * \log(P_i) + (1 - y_i) * \log(1 - P_i)) \quad (3.15)$$

Where N is the number of rows, P_i is the probability of class 1 and (1-P_i) is the probability of class 0.

MATLAB development environment was used to implement and evaluate the proposed technique.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1. Preamble

This chapter presents the results obtained by applying the stacking ensemble technique and the constituent individual classification techniques on the deep diabetic retinopathy features extracted using Alexnet CNN. The performance of the suggested technique is compared with existing workings using the Messidor dataset. After each presented result, a detailed explanation or discussion of each obtained result is given below the chart or table depicting the result.

4.2 Results and Discussion

In this work using MATLAB environment experiment was conducted on five models: SVM, KNN, Decision Tree, Naïve Bayes and the stacking ensemble. The classification data was split into two parts: a train dataset and a test dataset, with an 80:20 split. Out of 1200 instances, 960 were inputted as train data while 240 were used as test data. The results obtained for the single models and the ensemble method after testing is presented in Table 4.1.

Table 4. 1 Diabetic Retinopathy Classification Result

Algorithm	**Accuracy (%)**	**Loss Function**	**Precision (%)**	**Recall (%)**	**F-Score (%)**
Support Vector Machine (SVM)	98.33	0.0167	100	96.77	98.36
K-Nearest Neighbor (KNN)	96.67	0.0333	100	93.75	96.77
Decision Tree	93.30	0.0667	90.00	96.43	93.10
Naïve Bayes (NB)	95.00	0.0500	96.67	93.55	95.08
Stacking Ensemble (Proposed method)	**99.17**	**0.0083**	100	**98.36**	**99.17**

In Table 4.1, the stacking ensemble method achieved accuracy as high as 99.17% which is higher than the accuracy of any of the constituent models. Among the individual single models Support vector machine obtained the highest accuracy of 98.33% while decision tree achieved the least accuracy of 93.30%. Looking at the loss function performance measure the proposed stacking ensemble has the lowest loss function of 0.0083, followed by SVM with loss of 0.0167, then KNN with loss of 0.0333, Naïve Bayes with loss of 0.0500 and final decision tree with the largest loss function of 0.0667. This shows that the stacking ensemble method is more suitable for detection of diabetic retinopathy disease than the individual models. Using precision, recall and f-score to assess the performance of the stacking ensemble technique and the single constituent models, it can be seen from Table 4.1 that the stacking ensemble produced a superior recall and f-score than the individual models. However the proposed model achieved same precision of 1 with SVM and KNN models. The accuracy, recall, precision and f-score of all the five models are visualized in Figure 4.1.

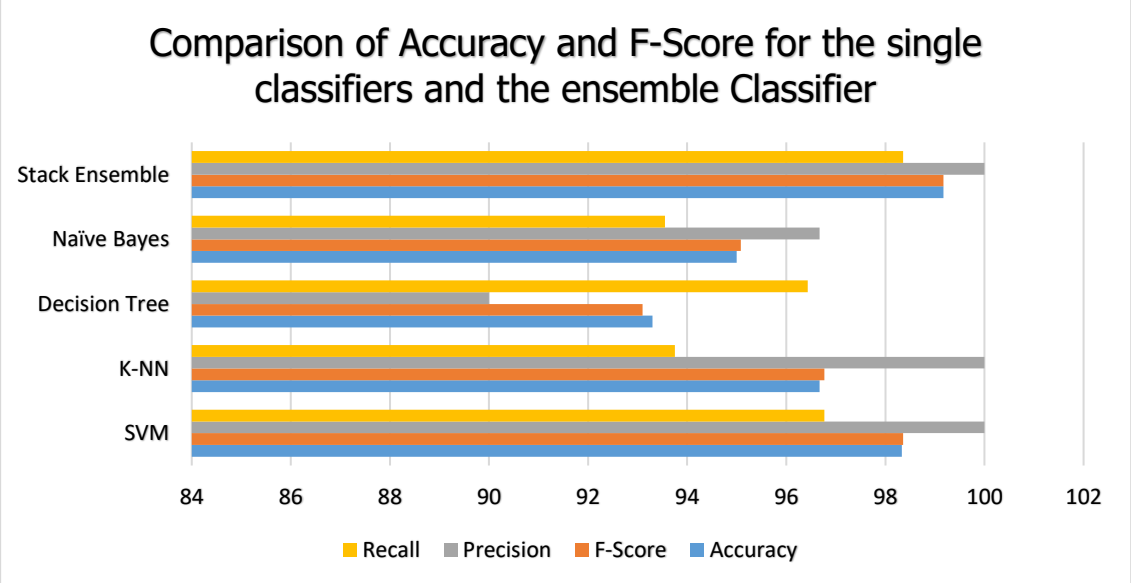

Figure 4. 1 Comparison of Accuracy and F-Score for the single classifiers and the ensemble Classifier

Figure 4.1 gives a clear visualization of the accuracy, recall, precision and f-score for both the single models and the ensemble model. For the chart it can be seen that decision tree has the least accuracy and f-score while the stack ensemble method has the highest accuracy and f-score. The confusion matrix for SVM, KNN, Decision Tree, NB and Stack ensemble is shown in figure 4.2, 4.3, 4.4, 4.5 and 4.6 respectively.

Actual: Positive (1)	120	0
Actual: Negative (0)	4	116
n = 240	Predicted: Positive (1)	Predicted: Negative (0)

Figure 4.2 SVM Confusion Matrix

Figure 4.2 shows the confusion matrix of SVM showing the True positive values, True negative values, false positive and false negative values. Out of the 240 examples used to test the model, 120 examples were correctly classified as diabetic retinopathy while 4 were wrongly classified as diabetic retinopathy. 116 examples were correctly classified as normal (no diabetic retinopathy) while no example was wrongly classified as normal images. Figure 4.3 is a confusion matrix of the KNN classification.

Actual: Positive (1)	120	0
Actual: Negative (0)	8	112
n = 240	Predicted: Positive (1)	Predicted: Negative (0)

Figure 4.3 KNN Confusion Matrix

Figure 4.3 displays the confusion matrix of KNN which resulted in an accuracy of 96.67%. 120 examples were correctly classified as diabetic retinopathy out of the 240 examples used to test the model, while 8 were wrongly classified as diabetic retinopathy. 112 examples were correctly classified as normal (no diabetic retinopathy) while no example was wrongly classified as normal images. Figure 4.4 is a confusion matrix of the Decision tree classification.

Actual: Positive (1)	108	12
Actual: Negative (0)	4	116
n = 240	Predicted: Positive (1)	Predicted: Negative (0)

Figure 4.4 Decision Tree Confusion Matrix

Figure 4.4 displays the confusion matrix of Decision Tree which resulted in an accuracy of 93.30%. 108 examples were correctly classified as diabetic retinopathy out of the 240

examples used to test the model, while 4 were wrongly classified as diabetic retinopathy. 116 examples were correctly classified as normal (no diabetic retinopathy) while 12 examples were wrongly classified as normal images. Figure 4.5 is a confusion matrix of the NB classification.

Actual: Positive (1)	116	4
Actual: Negative (0)	8	112
n = 240	Predicted: Positive (1)	Predicted: Negative (0)

Figure 4.5 NB Confusion Matrix

Figure 4.5 displays the confusion matrix of NB. After testing the model, 116 examples were correctly classified as diabetic retinopathy, while 8 were wrongly classified as diabetic retinopathy. 112 examples were correctly classified as normal (no diabetic retinopathy) while 4 examples were wrongly classified as normal images. Figure 4.6 is a confusion matrix of the Stack ensemble classification.

Actual: Positive (1)	120	0
Actual: Negative (0)	2	118
n = 240	Predicted: Positive (1)	Predicted: Negative (0)

Figure 4.6 Stack Ensemble Confusion Matrix

Figure 4.6 displays the confusion matrix of the stack ensemble. Out of the 240 examples used to test the model, 120 examples were correctly classified as diabetic retinopathy while 2 were wrongly classified as diabetic retinopathy. 118 examples were correctly classified as normal (no diabetic retinopathy) while no example was wrongly classified as normal images. Figure 4.7 is a Receiver Operating Characteristics (ROC) curve for SVM, KNN, Decision tree, NB and Stack ensemble classification

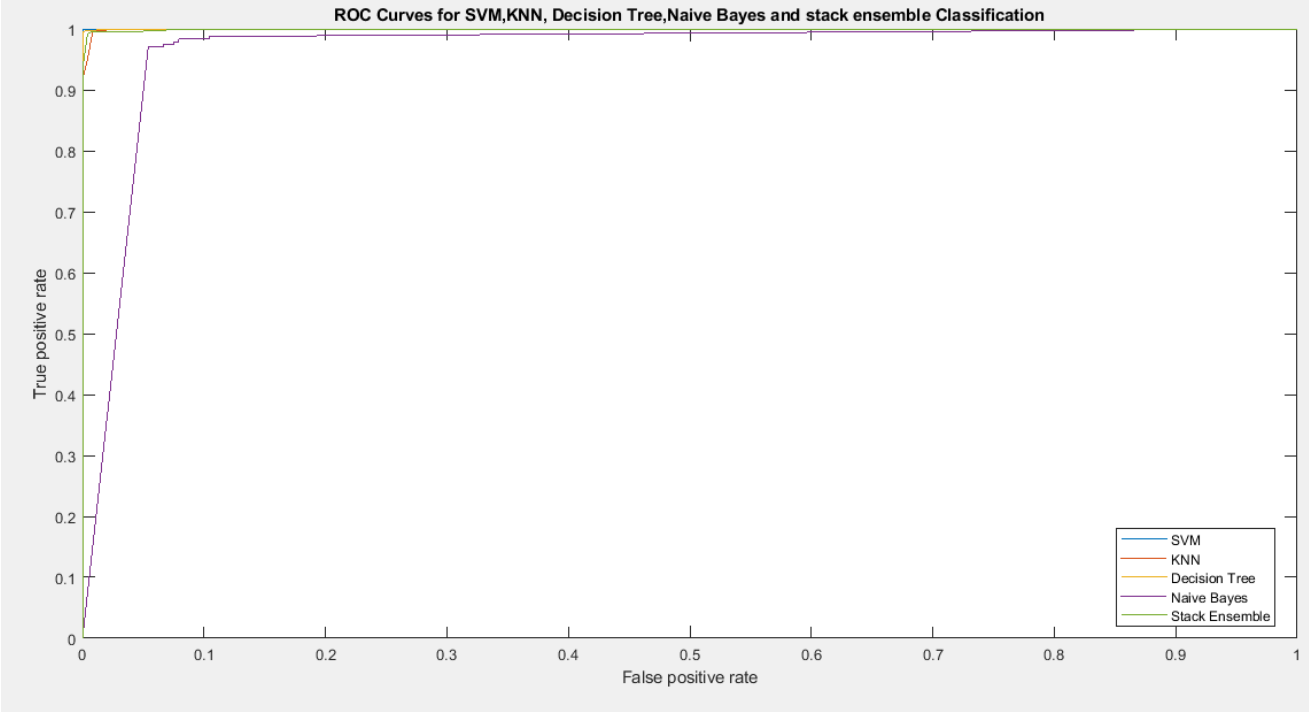

Figure 4.7 ROC curves showing a comparison of SVM, KNN, Decision tree, NB and Stack ensemble

The values in Table 4.2 represent the performance results of the proposed stack ensemble technique in comparison with related works that used the Messidor dataset for training and testing.

Table 4. 2 Comparison of Proposed Method with Related Works That Used the Messidor Dataset

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Saxena *et al.* (2020)	-	-	0.8880	-
Zhou *et al.* (2018)	-	-	-	-
Kalyani *et al.* (2021)	97.98	95.62	96.11	96.31
Zago *et al.* (2020)	-	-	95.00	-
Stacking ensemble (Proposed Method)	**99.17**	**100**	**98.36**	**99.17**

Looking at the values of the performance measures in Table 4.2 the proposed technique achieved the highest accuracy of 99.17, precision of 100%, recall of 98.36%, and f-score of 99.17% than the related works. The result of the proposed technique is highlighted in bold.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Summary

In this research Alexnet CNN was used to extract deep features from the diabetic retinopathy fundus images. These extracted deep features were feed as input to five classification models. For the classification process four classification techniques: SVM, KNN, NB, and Decision Tree predictions were ensemble using the stacking ensemble method to classify diabetic retinopathy disease using the Messidor fundus images. The single classifiers made predictions at first, which were then utilized to build a new model. The new model as employed for making predictions on the test dataset.

5.2 Conclusion

In conclusion, a method for diabetic retinopathy classification was proposed based on ensemble of classifiers using stacking. In this research work, prediction using the stacking ensemble model achieved a better performance than classification using any of the single constituent models. The first objective of identifying the challenges of existing diabetic retinopathy images detection and a classification technique was accomplished by reviewing several related works. The identified challenges includes: high bias and variance of the single classification models and difficulty in choosing the best classification model.

Four classification models were selected and trained using the extracted deep features. The predictions from these four classifiers were used to build a new model. This new model was built using the stacking ensemble method. This developed stack ensemble model is an accomplishment of the second objective. The stacking ensemble model was evaluated against the single constituent classification models and against related works on diabetic retinopathy classification which completed the third objective. The accuracy

precision, recall and f-score performance measure were used to execute this evaluation. The proposed stacking ensemble technique obtained an accuracy of 0.9917, precision of 1, recall of 0.9836, f-score of 0.9917 and loss function of 0.0083. These findings show that the proposed stacking ensemble technique is beneficial in improving DR categorization.

5.3 Contribution to Knowledge

This study contributed to knowledge by developing a stacking ensemble model which consists of four classification models for classification of diabetic retinopathy into No diabetic retinopathy (0) and Present diabetic retinopathy (1).

5.4 Recommendation

This study used the stacking ensemble technique to ensemble the single classifier predictions. For future work other ensemble techniques such as bagging, boosting, averaging, and major voting can be utilized. Also, in this study the deep classifiers such as Googlenet, Vgg19, ResNet and Lesnet were not considered in this ensemble. It is recommended that further research should explore deep classifiers techniques for ensemble. Only the Messidor dataset was used to train the suggested stacking ensemble model, which classifies DR into four stages ranging from 0 to 3. There are different datasets in which DR is classified into five categories. The system has not been trained for those datasets. The suggested stacking ensemble model will be trained for all conceivable classes of diabetic retinopathy as a future development of this research.

REFERENCES

Ahmad, A., Mansoor, A. B., Mumtaz, R., Khan, M., and Mirza, S. H. (2014). Image processing and classification in diabetic retinopathy: A review. *2014 5th European Workshop on Visual Information Processing (EUVIP)*, 1–6. https://doi.org/10.1109/EUVIP.2014.7018362

Ahmad F., M. H., Izhar, L. I., Nugroho, H., and Nugroho, H. A. (2011). Analysis of retinal fundus images for grading of diabetic retinopathy severity. *Medical and Biological Engineering & Computing*, *49*(6), 693–700. https://doi.org/10.1007/s11517-011-0734-2

Al Hazaimeh, O. M., Nahar, K. M. O., Al Naami, B., and Gharaibeh, N. (2018). An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images. *International Journal of Signal and Imaging Systems Engineering*, *11*(4), 206. https://doi.org/10.1504/IJSISE.2018.10015063

Alabdulrahman, R. (2014). *A Comparative Study of Ensemble Active Learning* [Masters degree]. ottawa.

Alom, Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Esesn, B. C. V., Awwal, A. A. S., and Asari, V. K. (2019). *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches*. 39.

Al-Rawi, M., Qutaishat, M., and Arrar, M. (2007). An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine*, *37*(2), 262–267. https://doi.org/10.1016/j.compbiomed.2006.03.003

Amin, J., Sharif, M., and Yasmin, M. (2016). A Review on Recent Developments for Detection of Diabetic Retinopathy. *Scientifica*, *2016*, 1–20. https://doi.org/10.1155/2016/6838976

Amin, J., Sharif, M., Yasmin, M., Ali, H., and Fernandes, S. L. (2017). A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions. *Journal of Computational Science*, *19*, 153–164. https://doi.org/10.1016/j.jocs.2017.01.002

Anuradha, C., and Velmurugan, T. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. *Indian Journal of Science and Technology*, *8*(15). https://doi.org/10.17485/ijst/2015/v8i15/74555

Arade, S. P., and Patil, J. K. (2017). COMPARATIVE STUDY OF DIABETIC RETINOPATHY USING K-NN AND BAYESIAN CLASSIFIER. *INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY*, *4*(5), 55–61.

Arcadu, F., Benmansour, F., Maunz, A., Willis, J., Haskova, Z., and Prunotto, M. (2019). Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *Npj Digital Medicine*, *2*(1), 92. https://doi.org/10.1038/s41746-019-0172-3

Banu, R., Arun, V., and Shankaraiah, N. (2016). Meta-cognitive Neural Network Method for Classification of Diabetic Retinal Images. *Second International Conference on Cognitive Computing and Information Processing (CCIP)*, 5.

Basit, A., and Egerton, S. J. (2013). Bio-medical imaging: Localization of main structures in retinal fundus images. *IOP Conference Series: Materials Science and Engineering*, *51*, 012009. https://doi.org/10.1088/1757-899X/51/1/012009

Benkaddour, M. K., and Bounoua, A. (2017). Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition. *Traitement Du Signal*, *34*(1–2), 77–91. https://doi.org/10.3166/ts.34.77-91

Bethanney, J., Divakaran, S., Abraham, S., Meera, G., and UmaShankar, G. (2015). Detection and classification of exudates in retinal image using image processing techniques. *Journal of Chemical and Pharmaceutical Sciences*, *8*(3). www.jchps.com

Bühlmann, P. (2012). Bagging, Boosting and Ensemble Methods. In J. E. Gentle, W. K. Härdle, and Y. Mori (Eds.), *Handbook of Computational Statistics* (pp. 985–1022). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21551-3_33

Bus, S. A., Lavery, L. A., Monteiro-Soares, M., Rasmussen, A., Raspovic, A., Sacco, I. C. N., Netten, J. J., and on behalf of the International Working Group on the Diabetic Foot. (2020). Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2019 update). *Diabetes/Metabolism Research and Reviews*, *36*(S1). https://doi.org/10.1002/dmrr.3269

Cao, J., Wang, M., Li, Y., and Zhang, Q. (2019). Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment. *PLOS ONE*, *14*(4), e0215136. https://doi.org/10.1371/journal.pone.0215136

Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., and Goldbaum, M. (1989). Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on Medical Imaging*, *8*(3), 263–269. https://doi.org/10.1109/42.34715

Costa, P., Galdran, A., Smailagic, A., and Campilho, A. (2018). A Weakly-Supervised Framework for Interpretable Diabetic Retinopathy Detection on Retinal Images. *IEEE Access*, *6*, 18747–18758. https://doi.org/10.1109/ACCESS.2018.2816003

Cree, M. J., Cornforth, D., and Jelinek, H. F. (2005). Vessel Segmentation and Tracking Using a Two-Dimensional Model. *Proceedings of the Image and Vision Computing Conference*, 6.

Cunningham, P., and Delany, S. J. (2007). *K-Nearest Neighbour Classifiers*. 18.

David, H. B. F., and Suruliandi, A. (2019). *Performance Evaluation of Ensemble Classifiers on Benchmark Datasets*. *6*(2), 6.

David K., K., A. Adepoju, S., and Kolo Alhassan, J. (2015). A Decision Tree Approach for Predicting Students Academic Performance. *International Journal of Education and Management Engineering*, *5*(5), 12–19. https://doi.org/10.5815/ijeme.2015.05.02

Dutta, S., Manideep, B. C., Basha, S. M., Caytiles, R. D., and Iyengar, N. Ch. S. N. (2018). Classification of Diabetic Retinopathy Images by Using Deep Learning Models. *International Journal of Grid and Distributed Computing*, *11*(1), 99–106. https://doi.org/10.14257/ijgdc.2018.11.1.09

ElTanboly, A., Ismail, M., Shalaby, A., Switala, A., El-Baz, A., Schaal, S., Gimel'farb, G., and El-Azab, M. (2017). A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Medical Physics*, *44*(3), 914–923. https://doi.org/10.1002/mp.12071

Faust, O., Acharya U., R., Ng, E. Y. K., Ng, K.-H., and Suri, J. S. (2012). Algorithms for the Automated Detection of Diabetic Retinopathy Using Digital Fundus Images: A Review. *Journal of Medical Systems*, *36*(1), 145–157. https://doi.org/10.1007/s10916-010-9454-7

Ferreira, A., and Giraldi, G. (2017). Convolutional Neural Network approaches to granite tiles classification. *Expert Systems with Applications*, *84*, 1–11. https://doi.org/10.1016/j.eswa.2017.04.053

Fraz, M. M., Barman, S. A., Remagnino, P., Hoppe, A., Basit, A., Uyyanonvara, B., Rudnicka, A. R., and Owen, C. G. (2012). An approach to localize the retinal blood vessels using bit planes and centerline detection. *Computer Methods and Programs in Biomedicine*, *108*(2), 600–616. https://doi.org/10.1016/j.cmpb.2011.08.009

Ganesan, K., Martis, R. J., Acharya, U. R., Chua, C. K., Min, L. C., Ng, E. Y. K., and Laude, A. (2014). Computer-aided diabetic retinopathy detection using trace transforms on digital fundus images. *Med Biol Eng Comput*, 10.

Ghosh, S., Dasgupta, A., and Swetapadma, A. (2019). A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, 24–28. https://doi.org/10.1109/ISS1.2019.8908018

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., and Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, *316*(22), 2402. https://doi.org/10.1001/jama.2016.17216

Gupta, M., Knezevic, N. N., Abd-Elsayed, A., Ray, M., Patel, K., and Chowdhury, B. (2021). Treatment of Painful Diabetic Neuropathy—A Narrative Review of Pharmacological and Interventional Approaches. *Biomedicines*, *9*(5), 573. https://doi.org/10.3390/biomedicines9050573

Habib, M. M., Welikala, R. A., Hoppe, A., Owen, C. G., Rudnicka, A. R., and Barman, S. A. (2017). Detection of microaneurysms in retinal images using an ensemble classifier. *Informatics in Medicine Unlocked*, *9*, 44–57. https://doi.org/10.1016/j.imu.2017.05.006

Hoover, A. D., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, *19*(3), 203–210. https://doi.org/10.1109/42.845178

Hossain, Md. A., and Alam Sajib, Md. S. (2019). Classification of Image using Convolutional Neural Network (CNN). *Global Journal of Computer Science and Technology*, 13–18. https://doi.org/10.34257/GJCSTDVOL19IS2PG13

Islam, M., Dinh, A. V., and Wahid, K. A. (2017). Automated Diabetic Retinopathy Detection Using Bag of Words Approach. *Journal of Biomedical Science and Engineering*, *10*(05), 86–96. https://doi.org/10.4236/jbise.2017.105B010

Kalyani, G., Janakiramaiah, B., Karuna, A., and Prasad, L. V. N. (2021). Diabetic retinopathy detection and classification using capsule networks. *Complex & Intelligent Systems*. https://doi.org/10.1007/s40747-021-00318-9

Kande, G. B., Subbaiah, P. V., and Savithri, T. S. (2010). Unsupervised Fuzzy Based Vessel Segmentation In Pathological Digital Fundus Images. *Journal of Medical Systems*, *34*(5), 849–858. https://doi.org/10.1007/s10916-009-9299-0

Khan, M. W. (2013). *Diabetic Retinopathy Detection using Image Processing: A Survey*. *1*(1), 6.

Kirange, D. K., Chaudhari, J. P., Rane, K. P., Bhagat, K. S., and Chaudhri, N. (2019). Diabetic Retinopathy Detection and Grading Using Machine Learning. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3570–3576. https://doi.org/10.30534/ijatcse/2019/139862019

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kumaran, Y., and Patil, C. M. (2018). *A Brief Review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques*. *7*(4), 11.

Kumudham, R., and Shenbagavalli, A. (2013). Enhancement and Classification for Colour Retinal Images. *International Conference on Engineering and Technology*, 5.

Lauría, E. J. M., Presutti, E., Kapogiannis, M., and Kamath, A. (2018). Stacking Classifiers for Early Detection of Students at Risk: *Proceedings of the 10th International Conference on Computer Supported Education*, 390–397. https://doi.org/10.5220/0006781203900397

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lei Zhang, Qin Li, You, J., and Zhang, D. (2009). A Modified Matched Filter With Double-Sided Thresholding for Screening Proliferative Diabetic Retinopathy. *IEEE Transactions on Information Technology in Biomedicine*, *13*(4), 528–534. https://doi.org/10.1109/TITB.2008.2007201

Li, Y.-H., Yeh, N.-N., Chen, S.-J., and Chung, Y.-C. (2019). Computer-Assisted Diagnosis for Diabetic Retinopathy Based on Fundus Images Using Deep Convolutional Neural Network. *Mobile Information Systems*, *2019*, 1–14. https://doi.org/10.1155/2019/6142839

Liu, Y. H. (2018). Feature Extraction and Image Recognition with Convolutional Neural Networks. *Journal of Physics: Conference Series*, *1087*, 062032. https://doi.org/10.1088/1742-6596/1087/6/062032

Mangrulkar, R. S. (2017). *Retinal Image Classification Technique For Diabetes Identification*. 6.

Mendonca, A. M., and Campilho, A. (2006). Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, *25*(9), 1200–1213. https://doi.org/10.1109/TMI.2006.879955

Messidor. (2021). Messidor. *Messidor*. https://www.adcis.net/en/third-party/messidor/

Morales, S., Engan, K., Naranjo, V., and Colomer, A. (2017). Retinal Disease Screening Through Local Binary Patterns. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 184–192. https://doi.org/10.1109/JBHI.2015.2490798

Namatēvs, I. (2017). Deep Convolutional Neural Networks: Structure, Feature Extraction and Training. *Information Technology and Management Science*, *20*(1). https://doi.org/10.1515/itms-2017-0007

Nannia, L., Ghidoni, S., and Brahnam, S. (2021). Ensemble of convolutional neural networks for bioimage classification. *Applied Computing and Informatics*, *17*(1), 19–35. https://doi.org/10.1016/j.aci.2018.06.002

Nascimento, O. J. M. do, Pupe, C. C. B., and Cavalcanti, E. B. U. (2016). Diabetic neuropathy. *Revista Dor*, *17*. https://doi.org/10.5935/1806-0013.20160047

Netten, J. J., Bus, S. A., Apelqvist, J., Lipsky, B. A., Hinchliffe, R. J., Game, F., Rayman, G., Lazzarini, P. A., Forsythe, R. O., Peters, E. J. G., Senneville, É., Vas, P., Monteiro-Soares, M., Schaper, N. C., and on behalf of the International Working Group on the Diabetic Foot. (2020). Definitions and criteria for diabetic foot disease. *Diabetes/Metabolism Research and Reviews*, *36*(S1). https://doi.org/10.1002/dmrr.3268

Niemeijer, M., van Ginneken, B., Staal, J., Suttorp-Schulten, M. S. A., and Abramoff, M. D. (2005). Automatic detection of red lesions in digital color fundus

photographs. *IEEE Transactions on Medical Imaging*, *24*(5), 584–592. https://doi.org/10.1109/TMI.2005.843738

Nti, I. K., Adekoya, A. F., and Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7*(1), 20. https://doi.org/10.1186/s40537-020-00299-5

Olaniyi, A. S., Kayode, S. Y., Abiola, H. M., Tosin, S.-I. T., and Babatunde, A. N. (2017). *STUDENT'S PERFORMANCE ANALYSIS USING DECISION TREE ALGORITHMS*. 8.

Oloumi, F., Rangayyan, R. M., Ells, A. L., Tranquillo, J. V., and Frize, M. (2014). *Digital Image Processing for Ophthalmology: Detection and Modeling of Retinal Vascular Architecture*. 177.

Padmanabhan, S. (2016). Convolutional Neural Networks for Image Classification and Captioning. *Stanford University's CS 231A (Computer Vision)*, 8.

Pak, A., Ziyaden, A., Tukeshev, K., Jaxylykova, A., and Abdullina, D. (2020). Comparative analysis of deep learning methods of detection of diabetic retinopathy. *Cogent Engineering*, *7*(1), 1805144. https://doi.org/10.1080/23311916.2020.1805144

Patel, N., and Singh, D. (2015). An Algorithm to Construct Decision Tree for Machine Learning based on Similarity Factor. *International Journal of Computer Applications*, *111*(10), 22–26. https://doi.org/10.5120/19575-1376

Rahimy, E. (2018). Deep learning applications in ophthalmology. *Current Opinion in Ophthalmology*, *29*(3), 254–260. https://doi.org/10.1097/ICU.0000000000000470

Rahman, A., and Tasnim, S. (2014). Ensemble Classifiers and Their Applications: A Review. *International Journal of Computer Trends and Technology*, *10*(1), 31–35. https://doi.org/10.14445/22312803/IJCTT-V10P107

Rajalakshmi, R., Subashini, R., Anjana, R. M., and Mohan, V. (2018). Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*, *32*(6), 1138–1144. https://doi.org/10.1038/s41433-018-0064-9

Ramlugun, G. S., Nagarajan, V. K., and Chakraborty, C. (2012). Small retinal vessels extraction towards proliferative diabetic retinopathy screening. *Expert Systems with Applications*, *39*(1), 1141–1146. https://doi.org/10.1016/j.eswa.2011.07.115

Ruby, .Usha A, Prasannavenkatesan, T., Jeena, J., and Vamsidhar. (2020). Binary cross entropy with deep learning technique for Image classification. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5393–5397. https://doi.org/10.30534/ijatcse/2020/175942020

Saleh Albelwi and Ausif Mahmood. (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy*, *19*(6), 242. https://doi.org/10.3390/e19060242

Sandhu, H. S., Eltanboly, A., Shalaby, A., Keynton, R. S., Schaal, S., and El-Baz, A. (2018). Automated Diagnosis and Grading of Diabetic Retinopathy Using Optical Coherence Tomography. *Investigative Opthalmology & Visual Science*, *59*(7), 3155. https://doi.org/10.1167/iovs.17-23677

Sariera, T. M. A., Rangarajan, L., and Amarnath, R. (2020). Detection and classification of hard exudates in retinal images. *Journal of Intelligent and Fuzzy Systems*, *38*. https://doi.org/10.3233/JIFS-190492

Sarki, R., Ahmed, K., Wang, H., and Zhang, Y. (2020). Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Information Science and Systems*, *8*(1), 32. https://doi.org/10.1007/s13755-020-00125-5

Sarwinda, D., Bustamam, A., and Wibisono, A. (2017). A Complete Modelling of Local Binary Pattern for Detection of Diabetic Retinopathy. *1st International Conference on Informatics and Computational Sciences (ICICoS)*, 4.

Saxena, G., Verma, D. K., Paraye, A., Rajan, A., and Rawat, A. (2020). Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets. *Intelligence-Based Medicine*, *3–4*, 100022. https://doi.org/10.1016/j.ibmed.2020.100022

Selby, N. M., and Taal, M. W. (2020). An updated overview of diabetic nephropathy: Diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes, Obesity and Metabolism*, *22*(S1), 3–15. https://doi.org/10.1111/dom.14007

Simonyan, K., and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. http://arxiv.org/abs/1409.1556

Sinthanayothin, C., Boyce, J. F., Cook, H. L., and Williamson, T. H. (1999). Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, *83*(8), 902–910. https://doi.org/10.1136/bjo.83.8.902

Soares, J. V. B., Leandro, J. J. G., Cesar, R. M., Jelinek, H. F., and Cree, M. J. (2006). Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, *25*(9), 1214–1222. https://doi.org/10.1109/TMI.2006.879967

Somasundaram, S. K., and Alli, P. (2017). A Machine Learning Ensemble Classifier for Early Prediction of Diabetic Retinopathy. *Journal of Medical Systems*, *41*(12), 201. https://doi.org/10.1007/s10916-017-0853-x

Sopharak, A., Dailey, M. N., Uyyanonvara, B., Barman, S., Williamson, T., Nwe, K. T., and Moe, Y. A. (2010). Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *Journal of Modern Optics*, *57*(2), 124–135. https://doi.org/10.1080/09500340903118517

Spencer, T., Olson, J. A., McHardy, K. C., Sharp, P. F., and Forrester, J. V. (1996). An Image-Processing Strategy for the Segmentation and Quantification of Microaneurysms in Fluorescein Angiograms of the Ocular Fundus. *Computers*

and Biomedical Research, *29*(4), 284–302. https://doi.org/10.1006/cbmr.1996.0021

Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., and van Ginneken, B. (2004). Ridge-Based Vessel Segmentation in Color Images of the Retina. *IEEE Transactions on Medical Imaging*, *23*(4), 501–509. https://doi.org/10.1109/TMI.2004.825627

Sudha, V., and Karthikeyan, C. (2018). Analysis of diabetic retinopathy using naive bayes classifier technique. *International Journal of Engineering and Technology*, *7*(2.21), 440. https://doi.org/10.14419/ijet.v7i2.21.12462

Sulaiman, M. K. (2019). Diabetic nephropathy: Recent advances in pathophysiology and challenges in dietary management. *Diabetology & Metabolic Syndrome*, *11*(1), 7. https://doi.org/10.1186/s13098-019-0403-4

Sultana, F., Sufian, A., and Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 122–129. https://doi.org/10.1109/ICRCICN.2018.8718718

Tarr, J. M., Kaul, K., Chopra, M., Kohner, E. M., and Chibber, R. (2013). Pathophysiology of Diabetic Retinopathy. *ISRN Ophthalmology*, *2013*, 1–13. https://doi.org/10.1155/2013/343560

Tolias, Y. A., and Panas, S. M. (1998). A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging*, *17*(2), 263–273. https://doi.org/10.1109/42.700738

Tymchenko, B., Marchenko, P., and Spodarets, D. (2020). Deep Learning Approach to Diabetic Retinopathy Detection. *ArXiv:2003.02261 [Cs, Stat]*. http://arxiv.org/abs/2003.02261

Vedaldi, A., and Lenc, K. (2016). MatConvNet—Convolutional Neural Networks for MATLAB. *ArXiv:1412.4564 [Cs]*. http://arxiv.org/abs/1412.4564

Verma, K., Deep, P., and Ramakrishnan, A. G. (2011). Detection and classification of diabetic retinopathy using retinal images. *2011 Annual IEEE India Conference*, 1–6. https://doi.org/10.1109/INDCON.2011.6139346

Vlachos, M., and Dermatas, E. (2010). Multi-scale retinal vessel segmentation using line tracking. *Computerized Medical Imaging and Graphics*, *34*(3), 213–227. https://doi.org/10.1016/j.compmedimag.2009.09.006

Walsh, P. (2019). Support Vector Machine Learning for ECG Classification. *Smart Healthcare and Safety Systems*, 10.

Wang, W., and Lo, A. (2018). Diabetic Retinopathy: Pathophysiology and Treatments. *International Journal of Molecular Sciences*, *19*(6), 1816. https://doi.org/10.3390/ijms19061816

Yin, Y., Adel, M., and Bourennane, S. (2012). Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognition*, *45*(4), 1235–1244. https://doi.org/10.1016/j.patcog.2011.09.019

Youssif, A. A. A., Ghalwash, A. Z., and Ghoneim, A. S. (2007). *A Comparative Evaluation of Preprocessing Methods for Automatic Detection of Retinal Anatomy*. 8.

Zago, G. T., Andreão, R. V., Dorizzi, B., and Teatini Salles, E. O. (2020). Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Computers in Biology and Medicine*, *116*, 103537. https://doi.org/10.1016/j.compbiomed.2019.103537

Zeng, X., Chen, H., Luo, Y., and Ye, W. (2019). Automated Diabetic Retinopathy Detection Based on Binocular Siamese-Like Convolutional Neural Network. *IEEE Access*, *7*, 30744–30753. https://doi.org/10.1109/ACCESS.2019.2903171

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218–218. https://doi.org/10.21037/atm.2016.03.37

Zhang, Z., and Sabuncu, M. (2018). Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *2nd Conference on Neural Information Processing Systems*, 11.

Zhou, L., Zhao, Y., Yang, J., Yu, Q., and Xu, X. (2018). Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images. *IET Image Processing*, *12*(4), 563–571. https://doi.org/10.1049/iet-ipr.2017.0636

Zulkeflie, S. A., Fammy, F. A., Ibrahim, Z., and Sabri, N. (2019). Evaluation of Basic Convolutional Neural Network, AlexNet and Bag of Features for Indoor Object Recognition. *International Journal of Machine Learning and Computing*, *9*(6), 801–806. https://doi.org/10.18178/ijmlc.2019.9.6.876

APPENDIX

```
%Load Images into datastore
imds =
imageDatastore('DBimages','IncludeSubfolders',true,'LabelSource','foldernames');
[imdsTrain,imdsTest] = splitEachLabel(imds,0.8,'randomized');

%Alexnet
net = alexnet;
net.Layers
inputSize       = net.Layers(1).InputSize
augimdsTrain  = augmentedImageDatastore(inputSize(1:2),imdsTrain);
augimdsTest   = augmentedImageDatastore(inputSize(1:2),imdsTest);

layer           = 'fc7';   %Alexnet
featuresTrain   = activations(net,augimdsTrain,layer,'OutputAs','rows');
featuresTest    = activations(net,augimdsTest,layer,'OutputAs','rows');

YTrain = imdsTrain.Labels;
YTest           = imdsTest.Labels;

% SVM with Gaussian kernel
rng('default') % For reproducibility
mdls{1} = fitcsvm(featuresTrain,YTrain,'KernelFunction','gaussian', ...
'Standardize',true,'KernelScale','auto');

% KNN classifer
rng('default')
mdls{2} = fitcknn(featuresTrain,YTrain, 'NumNeighbors',5);

% Decision tree
rng('default')
mdls{3} = fitctree(featuresTrain,YTrain);
```

```
% Naive Bayes
rng('default')
mdls{4} = fitcnb(featuresTrain,YTrain);

%Combine Models Using Stacking
rng('default') % For reproducibility
N       = numel(mdls);
Scores = zeros(size(featuresTrain,1),N);
cv      = cvpartition(YTrain,"KFold",5);
for ii = 1:N
   m = crossval(mdls{ii},'cvpartition',cv);
   [~,s] = kfoldPredict(m);
   Scores(:,ii) = s(:,m.ClassNames=='0');
end

%Fit ensemble
rng('default') % For reproducibility
t                = templateTree('Surrogate','on','MaxNumSplits',1);
stckdMdl      = fitcensemble(Scores,YTrain, 'OptimizeHyperparameters','auto', ...
'Learners',t,…'HyperparameterOptimizationOptions',struct('Verbose',0,'AcquisitionFunc
tionName','expected-improvement-plus'));

%Predict Labels and Scores on Test Data
label = [];
score = zeros(size(featuresTest,1),N);
mdlLoss = zeros(1,numel(mdls));
fori = 1:N
   [lbl,s] = predict(mdls{i},featuresTest);
   label = [label,lbl];
   score(:,i) = s(:,m.ClassNames=='0');
mdlLoss(i) = mdls{i}.loss(featuresTest, YTest);
end
```

```
[lbl,s] = predict(stckdMdl,score);
label = [label,lbl];
mdlLoss(end+1) = stckdMdl.loss(score,YTest);

score = [score,s(:,1)];

names = {'SVMGaussian',KNN,'DecisionTree',NB,'StackedEnsemble'};
array2table(mdlLoss,'VariableNames',names)


%Prediction and Confusion Matrix
SVM_te      = predict(mdls{1},featuresTest);
accuracy1   = mean(SVM_te == YTest);
SVM_Con     = confusionmat(YTest,SVM_te);
KNN_te      = predict(mdls{2},featuresTest);
accuracy2   = mean(KNN_te == YTest);
KNN_Con     = confusionmat(YTest,KNN_te);
tree_te     = predict(mdls{3},featuresTest);
accuracy3   = mean(tree_te == YTest);
tree_Con    = confusionmat(YTest,tree_te);
nb_te       = predict(mdls{4},featuresTest);
accuracy4   = mean(nb_te == YTest);
nb_Con      = confusionmat(YTest,nb_te);
accuracy5   = mean(lbl == YTest);
stacked_Con = confusionmat(YTest,lbl);

%Rows and Columns of confusion matrix for Support Vector Machine
SVM_con1 = SVM_Con(1:1,1:1);  SVM_con2 = SVM_Con(1:1,2:2);
SVM_con4 = SVM_Con(2:2,1:1);  SVM_con5 = SVM_Con(2:2,2:2);

%Rows and Columns of confusion matrix for Discriminate Analysis
KNN_con1 = KNN_Con(1:1,1:1);  KNN_con2 = KNN_Con(1:1,2:2);
KNN_con4 = KNN_Con(2:2,1:1);  KNN_con5 = KNN_Con(2:2,2:2);
```

```
%Rows and Columns of confusion matrix for Decision Tree
tree_con1 = tree_Con(1:1,1:1);  tree_con2 = tree_Con(1:1,2:2);
tree_con4 = tree_Con(2:2,1:1);  tree_con5 = tree_Con(2:2,2:2);

%Rows and Columns of confusion matrix for Naive Bayes
nb_con1 = nb_Con(1:1,1:1);  nb_con2 = nb_Con(1:1,2:2);
nb_con4 = nb_Con(2:2,1:1);  nb_con5 = nb_Con(2:2,2:2);

%Rows and Columns of confusion matrix for stacked ensemble
stacked_con1 = stacked_Con(1:1,1:1);  stacked_con2 = stacked_Con(1:1,2:2);
stacked_con4 = stacked_Con(2:2,1:1);  stacked_con5 = stacked_Con(2:2,2:2);

%Precision and Recall
SVM_precision       = SVM_con1/(SVM_con1+SVM_con2);
SVM_recall    = SVM_con1/(SVM_con1+SVM_con4);
KNN_precision       = KNN_con1/(KNN_con1+ KNN_con2);
KNN_recall          = KNN_con1/(KNN_con1+KNN_con4);
tree_precision = tree_con1/(tree_con1+ tree_con2);
tree_recall         = tree_con1/(tree_con1+tree_con4);
nb_precision        = nb_con1/(nb_con1+nb_con2);
nb_recall           = nb_con1/(nb_con1+nb_con4);
stacked_precision   = stacked_con1/(stacked_con1+stacked_con2);
stacked_recall      = stacked_con1/(stacked_con1+stacked_con4);



%F-Score
SVM_fscore    = 2*((SVM_precision*SVM_recall)/(SVM_precision+SVM_recall));
discr_fscore        =
2*((KNN_precision*KNN_recall)/(KNN_precision+KNN_recall));
tree_fscore         = 2*((tree_precision*tree_recall)/(tree_precision+tree_recall));
nb_fscore           = 2*((nb_precision*nb_recall)/(nb_precision+nb_recall));
stacked_fscore  =
2*((stacked_precision*stacked_recall)/(stacked_precision+stacked_recall));
```

```
%ROC Curve to get Area under the curve
[~,score_SVM]                = resubPredict(mdls{1});
[A,B,T_SVM,AUC_SVM]          = perfcurve(YTrain,score_SVM(:,2),1);
[~,score_KNN]                = resubPredict(mdls{2});
[C,D,T_discr,AUC_KNN]        = perfcurve(YTrain,score_discr(:,2),1);
[~,score_tree]               = resubPredict(mdls{3});
[E,F,T_tree,AUC_tree]        = perfcurve(YTrain,score_tree(:,2),1);
[~,score_nb]                 = resubPredict(mdls{4});
[G,H,T_nb,AUC_nb]            = perfcurve(YTrain,score_nb(:,2),1);
[~,score_stacked]            = resubPredict(stckdMdl);
[K,L,T_stacked,AUC_stacked]  = perfcurve(YTrain,score_stacked(:,2),1);
```