Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis

Article

Accepted Version

Multi-Modal Classifier Fusion with Feature Cooperation for Glaucoma Diagnosis

Nacer Eddine Benzebouchi*[1], Nabiha Azizi[2], Amira S. Ashour[3], Nilanjan Dey[4], R. Simon Sherratt[5]
[1, 2]*Computer Science Department, Labged Laboratory Badji Mokhtar Annaba University, PO BOX 12, 23000 Annaba, Algeria*
[3] *Department of Electronics Engineering and Communication Engineering, Tanta University, Egypt*
[4]*Department of Information Technology at Techno India College of Technology, India*
[5]*Department of Biomedical Engineering, University of Reading, UK*

[1]nasrobenz@hotmail.fr, [2]azizi@labged.net, [3]amirasashour@yahoo.com
[4]neelanjan.dey@gmail.com, [5]r.s.sherratt@reading.ac.uk

Abstract

Background: Glaucoma is a major public health problem that can lead to an optic nerve lesion, requiring systematic screening in the population over 45 years of age. The diagnosis and classification of this disease have had a marked and excellent development in recent years, particularly in the machine learning domain. Multimodal data have been shown to be a significant aid to the machine learning domain, especially by its contribution to improving data driven decision-making.
Method: Solving classification problems by combinations of classifiers has made it possible to increase the robustness as well as the classification reliability by using the complementarity that may exist between the classifiers. Complementarity is considered a key property of multimodality. A Convolutional Neural Network (CNN) works very well in pattern recognition and has been shown to exhibit superior performance, especially for image classification which can learn by themselves useful features from raw data. This article proposes a multimodal classification approach based on deep Convolutional Neural Network and Support Vector Machine (SVM) classifiers using multimodal data and multimodal feature for glaucoma diagnosis from retinal fundus images from RIM-ONE dataset. We make use of handcrafted feature descriptors such as the Gray Level Co-Occurrence Matrix, Central Moments and Hu Moments to co-operate with features automatically generated by the CNN in order to properly detect the optic nerve and consequently obtain a better classification rate, allowing a more reliable diagnosis of glaucoma.
Results: The experimental results confirm that the combination of classifiers using the BWWV technique is better than learning classifiers separately. The proposed method provides a computerized diagnosis system for glaucoma disease with impressive results comparing them to the main related studies that allow us to continue in this research path.

Keywords: Multimodal, Deep Learning, Convolutional Neural Networks, Image Classification, Ensemble Learning, Glaucoma diagnosis.

I. Introduction:

Glaucoma is considered the second reason for visual deterioration after age-related macular degeneration (AMD). Over 70 million people would be affected worldwide by 2020 [1]. Glaucoma can be managed, but can also cause blindness if not detected in time. Glaucoma is an eye disease that primarily affects people over 45 years of age. This illness can cause optic nerve lesion; the nerve begins with the retina in the back of the eye and carries the images to the brain. When this nerve is

damaged, the visual field is reduced, the vision is then modified and this can lead to long-term blindness. In most cases, glaucoma is linked to an increase in the pressure inside the eye, also termed intraocular hypertension or intraocular pressure (IOP).

Intraocular pressure is usually measured with a tonometry test, which is an elementary test, as an elevated IOP is an important risk factor of the appearance of glaucoma. However, an elevated IOP is not always synonymous with glaucoma, and normal IOP does not necessarily mean a patient will never have glaucoma.

Glaucoma is a tricky disease, so it very difficult for a person to notice any glaucoma-related visual impairment because of the complete absence of symptoms. When glaucoma is diagnosed in time it can be treated and vision can be stabilized. Therefore, if it is not detected and is not taken care of early, glaucoma can evolve and can result in complete blindness. On the other hand, the reduction of the catch rate of this silent and severe disease is one of the main interests of public health, in order to take charge at the first appearance and control its progression, thus providing improved diagnosis of glaucoma.

In order to help ophthalmologists to detect glaucoma at an early stage, several research studies have been directed towards automated decision making of glaucoma. The main objective of automated systems is to improve diagnostic accuracy. In fact, they are used as a second opinion by doctors to get the final diagnosis [2], which can decrease human errors, in order to provide a uniform large-scale screening at a better price.

Once trained, computers can obtain much faster classifications, so this helps doctors in their real-time classification. Glaucoma classification has undergone excellent development in recent years, particularly in machine learning domain. Generally, the issues in classification reduce to finding a better decision area that separates objects into categories or classes. In order to define the best separation, we introduce the concept of margin or plane between two categories. To simplify the idea, we show an example in a two-dimensional space to explain how each classifier finds its margin separating two classes. The red points represent the first class samples and the blue points represent the second class. This idea can be generalized to a space of high dimensions. The idea is illustrated in figure 1.

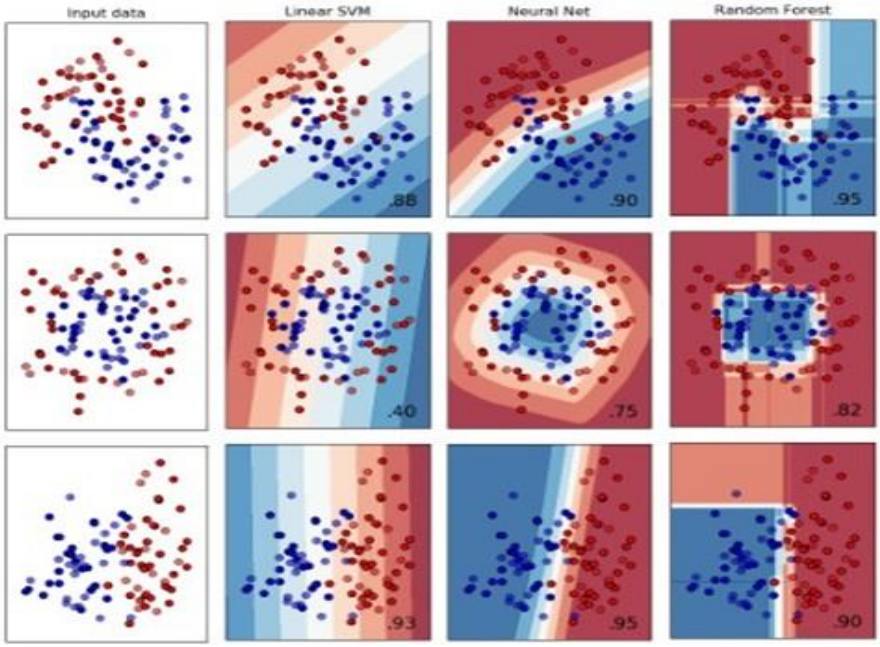

Figure 1 : Different separation surfaces generated by different classifiers[1].

[1] https://scikit-learn.org/

We notice that each classifier has its own way of generating the margin that separates the classes and the resulting model differs from one classifier of another; generally, it is not possible to build a perfect partition of space, so the role of the classifier will often be to give a probability of belonging to an object to a class. Therefore the choice of the classifier is not obvious because we do not have a single classifier that is considered the best to solve all the problems; this choice is influenced by the database content and the dispersion of data.

Recently, a combination of classifiers has been proposed as a research path to enable more reliable recognition by using the complementarity that can exist between classifiers. While the first experiments in classifiers combination date back to the 1980s [3], this technique has become an increasingly used way to improve the quality of recognition systems in several applications, namely medical image recognition [4], handwritten digits recognition [5], face recognition [6] and speech recognition [7]; these systems differ by output type of the classifiers combined and by the nature of the classifiers used.

The main reason why the pattern recognition community has a real interest in the combination of classifiers is its ability to take into account a large number of features used by different classifiers through exploiting the marginal performance and behavior of each of these classifiers. In the machine learning domain and disease diagnosis, features are considered the most important information in pattern recognition. In recent years, classification systems have used feature extraction techniques such as shape and texture primitives. Such results are encouraging [8, 9, 10], but remain insufficient because we are not sure to have the optimum representation according to the base used, and we do not know if the handcrafted characteristics are optimal in their performance.

Classical methods of classification from shapes use feature extraction techniques to represent the shape -while testing/ analyzing different families according to the processed basis-. The choice of the latter is not justifiable and in no way guarantees the ability of the characteristics chosen to represent new images; indeed, the modification of the initial base or its enrichment calls into question the already adapted characteristics and imperatively involves redoing the feature extraction phase.

Recently, the Convolutional Neural Network (CNN) has become widespread and represents a deep learning architecture that generates features in an automatic way [11]. In other words, CNN learns and extracts the most discriminative characteristics from the forming data. It has been shown to give statistically impressive results in image recognition applications [12-14]. The main benefit of using the Deep Convolutional Neural Network (CNN) is to take the entire image instead of the defective part, which avoids the sophisticated design of handcrafted features, which is a tedious step; this saves a lot of time and memory. Thus, as a second benefit of using CNN, it avoids segmentation and provides powerful features to properly classify sick patients and patients who are not sick [12-14]. Another benefit of CNN is that it does not require any pretreatment step that may affect performance.

However, with the major problem of the data limit and the choice of hyper-parameters, such as the filter number, shape, and the max-pooling shape...etc., for this purpose, it is not really known whether the features considered by a CNN are the most representative, Therefore, an improvement would be to merge them with other families of characteristics to give improved results.

In pattern recognition field, and more particularly that of medical diagnosis by the content of images, benefit can be gained from considering several modalities into decision-making. Indeed, multimodality can provide comprehensive information about image content, by increasing interpretation capabilities, the characteristics improvement for analysis, and by producing more reliable results. However, most existing research has focused on a single modality to diagnose a disease such as glaucoma [15], although recent studies have shown that learning using multimodal

data can provide complementary information [16] in order to obtain better performance [17] of extracting features and classification.

In this paper, we propose a multimodal classification approach to glaucoma disease based on two types of commonly used classifiers, namely the deep Convolutional Neural Network (CNN) and the Support Vector Machine (SVM). We investigated the use of two types of modalities, on the one hand considering resources and on the other hand to consider features in order to obtain the most representative features of the classification phase and subsequently increase diagnostic performance.

In this work, we are the first to propose two levels of multimodality for glaucoma classification:

- We rely on two modalities for the input images, the original RGB images and other binary modality using the Otsu technique; where each modality brings a certain types of information to be added to the system that cannot be inferred or obtained from other modalities. In mathematical terms, this added information is known as diversity.
- The use of two image representation modalities (RGB and binary), which are the characteristics automatically generated by CNN and the texture and shape characteristics used with SVM, we want to benefit from these representations by using the ensemble learning techniques. Indeed, each representation generates its own vision of the image and the combination of several points of view certainly increases the performance of the diagnosis.
- We have also added a fifth system (combining the features extracted by the convolution layers of the two image modalities using the SVM classifier).
- We suggest a new fusion approach called hybrid fusion: the features concatenation of different modalities (early fusion) and the multimodality classification (late fusion) using the best-worst weighted vote (BWWV) technique in order to generate the final decision of our multimodal system.

This paper is structured as follows: an overview of the related work presented in Section 2. In section 3, the paper presents the background in fusion techniques and proposes a hybrid fusion approach. Basic concepts on the Deep CNN are described in Section 4. Section 5 represents our CNN architecture and explains the different steps of our method to classify retinal fundus images into glaucoma or normal. Section 6 illustrates the experimental results from this work. Finally, a conclusion of this study presented in section 7.

II. Related studies:

Several studies have been conducted to develop tools to diagnose glaucoma. The automated detection of glaucoma was obtained using different machine learning (ML) techniques.

1) Glaucoma Diagnosis Using Classical Computer Vision Techniques

In recent years, many glaucoma diagnostic systems have been proposed using various traditional computer vision strategies to extract characteristics that will represent fundus images in classification. Texture and Shape features are the most used such as Gray Level Co-Occurrence Matrix (GLCM) features [8, 9], fractal dimension (FD) features [18], higher order spectra (HOS) features [19, 20], Wavelet-Based Features [20, 21, 22], local configuration pattern (LCP) features [23], correntropy features [22], fast Fourier transform (FFT) features [24] and GIST feature descriptor [25]. Most of these works [8, 18, 20, 24, 25] used support vector machine (SVM) as a classifier technique. Maheshwari et al. [22] used a variant of the SVM classifier which is the Least Squares SVM (LS-SVM) classifier with Radial Basis Function (RBF), Morlet wavelet and Mexican-hat wavelet kernels. The accuracy of the proposed approach is 98.33% using 3-fold cross-validation. Acharya et al. [23] tested several classifiers for glaucoma diagnosis, namely probabilistic neural network (PNN), decision tree (DT), k-nearest neighbor (kNN), support vector machine (SVM) and discriminant classifiers; the KNN classifier gave a better accuracy of 95.7%. Noronha et al. [19]

employed the SVM and Naïve Bayesian (NB) classifiers based on HOS cumulant characteristics to classify digital fundus images into three classes: normal, mild glaucoma and moderate/severe glaucoma. Dua et al. [21] formed several classifiers such as LibSVM, SMO, random forest, and naïve Bayes based on DWT and texture characteristics, which were computed from diverse wavelet filters. An accuracy of 93.00% is obtained using an SMO classifier with the ten-fold cross-validation method.

In 2003, Zheng and Essock [26] proposed a computer aided diagnosis system for glaucoma, using Fisher's Linear Discriminant Function (LDF) in the classification phase, and a novel Wavelet-Fourier Analysis (WFA) technique for the features extraction phase in order to obtain a better classification result. The results showed a sensitivity of 77.5%, a specificity of 96.5% and a Receiver Operating Characteristics (ROC) area of 94.1%.

In 2007, Meier et al. [27] proposed a two-stage classification system for the diagnosis of glaucoma without the segmentation stage. They used four feature extraction techniques, namely Principal Component Analysis (PCA) to select the relevant features directly from the image, Gabor filters, Fast Fourier Transform (FFT) and the Histogram features, and used the SVM-classifier (nu-SVM) to make the classification. The results showed an accuracy of 86%.

In 2009, Nayak et al. [28] proposed an automatic system for glaucoma diagnosis. They passed by the classical features extraction step using digital fundus images, then through the classification step that used the neural network classifier; they segment optical disk and cup by morphological operations in order to compute the cup-to-disc ratio. The results showed an accuracy of 90%.

In 2011, Acharya et al. [29] presented a glaucoma diagnostic system based on a fusion of higher order spectra (HOS) and texture features from digital fundus images using a random forest classifier, merged with z-score normalization and feature selection techniques. The results showed an accuracy of 91.7%.

In 2012, Mookiah et al. [20] used two techniques for features extraction, the Discrete Wavelet Transform (DWT) and Higher Order Spectra (HOS) using digital fundus images in order to automatically classify glaucoma and the normal class using SVM classifier with kernel function of polynomial order 2; the classification rate is 95%.

In 2014, Kumbhare et al. [30] used Naïve Bayes (NB) and minimum distance classifiers for automatic diagnosis of glaucoma using Higher Order Spectrum (HOS) and texture (Gray Level Co-Occurrence Matrix (GLCM) and Run Length Matrix (RLM)) features; the classification rate is 91%.

In 2016, Singh et al. [31] proposed a glaucoma diagnostic system using five of the most used automatic learning algorithms, Random Forest, NB, *k*-Nearest Neighbor (KNN), SVM and Artificial Neural Network (ANN) in order to choose the classifier that gave best results. The proposed approach used evolutionary attribute selection for feature selection and PCA as the feature reduction technique using wavelet features from the segmented optic disk. The results showed an accuracy of 94.7%.

In 2017, Maheshwari et al. [32] used the LS-SVM classifier for automatic diagnosis of glaucoma disease. The Binary Relief algorithm was used to extract relevant features. The recognition rate was 95.19% using the 3-fold cross-validation method.

In 2018, Kausu et al. [33] suggested an automatic glaucoma detection method using wavelet and morphological characteristics from the fundus images. This study proposed a system for glaucoma classification based on the segmentation of the region of interest (ROI). An accuracy of 97.67% is obtained using an MLP classifier with the 10-fold cross-validation technique.

All the above-mentioned approaches have shown better results for automatic glaucoma classification that require a traditional prior step of extracting features prior to the main recognition step, which costs a lot of energy and resource; therefore they are not applicable in real time.

2) Glaucoma Diagnosis Using Deep Learning Models

Recently, Deep Convolutional Neural Networks (CNNs, or ConvNets) are widespread because of finding themselves the best representation of the image. In fact, CNNs automatically extract the feature map using its Convolutional and Pooling layers. CNNs representing Deep Learning *(DL)* architectures have encouraging results for image recognition applications, including medical imaging [12, 13]. In addition, *DL* models have also demonstrated impressive results for difficult applications such as handwritten character recognition [34], object detection [35], natural language processing [36], and speech recognition [37].

Numerous studies have shown superior performance using deep convolutional neural networks for glaucoma diagnosis. Chen et al. [38] used convolutional neural network (CNN) for glaucoma detection using six layers. CNN is formed using segmented images of the region of interest (ROI). This work employs 1,676 images of SCES dataset and 650 images of ORIGA dataset in order to validate the efficiency of CNN. The results of this approach are based on area under curve (AUC) values, which respectively represent 83.1% and 88.7% of the ORIGA and SCES datasets. Chai et al. [39] proposed a glaucoma classification system using a Two-Branch Convolutional Neuron Network (CNN) in order to analyze both the entire image and the optic disc region automatically extracted using the Faster-RCNN model. The authors add a fusion layer to combine the characteristics extracted from two branches and a fully connected layer for the classification phase. The best classification rate is 81.69% using five convolutional layers. Zilly et al. [40] used ensemble learning to segment optic cup and disc from retinal images using CNN architectures and computed the cup-to-disc ratio for automatic classification of glaucoma. In order to reduce computational complexity and provide better performance, an entropy sampling technique is used. Orlando et al. [41] suggested a CNN model for automatic classification of glaucoma using two different architectures namely OverFeat and VGG-S from fundus images. The segmentation of the optic nerve head (ONH) area and the technique of vessel inpainting were applied to fundus images to improve the quality of the image and thus assess the improvement of the characteristic discrimination. The performance of this method is evaluated according to the area under ROC curve (AUC). The AUC values of the two CNN architectures (OverFeat and VGG-S) used are 76.3% and 71.8%, respectively. Raghavendra et al. [42] proposed a computer-aided diagnosis (CAD) system for automatic classification of glaucoma using CNN and the linear discriminant analysis (LDA) classifier. In this study, the authors used 1,426 fundus images to form CNN using eighteen layers and showed good diagnostic performance. For more details about Computer-Aided Diagnosis (CAD) systems of Glaucoma, authors can be referred to the recent review [43]. In our previous study [13], we trained two CNNs using different modalities, namely original color fundus images and binary images converted by the Otsu method in order to distinguish well between glaucomatous cases and non-glaucomatous cases. The cooperation of two fully automated CNNs has led to better results.

This article is a continuation of the work already done [8, 9, 13] using ensemble learning techniques by benefiting handcrafted feature descriptors and features automatically extracted by CNN in order to obtain a better representation of the image and therefore optimal performance. Based our knowledge, this is the first study to suggest a multimodal classification method for automated diagnosis of glaucoma using multimodal data and multimodal features from retinal fundus images.

III. Basic Concepts of Deep Convolutional Neural Network (CNN):

Deep Learning (DL) refers to a particular type of Artificial Intelligence (AI) using especially the Neural Network (NN) and certain models of particular algorithms such as the CNN model in order to generate intelligent models through learning. Just like a NN, the DL algorithm will take X inputs to return Y results. The input value will be processed and analyzed through many neuron successions that take as input the outputs of the previous neuron layers. DL models are built on the same model as Multilayer Perceptron (MP), however it should be noted that there are more numerous intermediate layers. Each of the intermediate layers maybe subdivided into sub-part, treating a sub-problem, simpler and providing the result to the next layer, and so on.

CNNs are multi-layered NNs that specialize in pattern recognition tasks. They are renowned for their robustness at low input variations, the low pretreatment rate necessary for their operation, and do not require any choice of a specific feature extractor. The proposed architecture is based on several deep NNs alternating between the convolution and the pooling layers. The architecture consists of a succession of convolutional layers, and aggregation is dedicated to the automatic extraction of features. While the second part is made up of layers of completely connected neurons, it is specifically dedicated to classification.

A convolutional neural network architecture is formed by a stack of processing layers:

- The Convolutional layer (Conv), which processes the data of a receiver field.
- The Pooling layer (Pool), which compresses the information by reducing the size of the intermediate image (often by sub-sampling).
- The correction layer (ReLU), refers to the activation function (linear rectification unit).
- The Fully Connected (FC) layer, which is a perceptron type layer.
- The loss layer (LOSS), specifies how network training penalizes the difference between the expected and actual signal, normally it is the last layer in the network.

1) Convolutional layer:

A convolutional layer L^i (layer i of the network) is characterized by its number N of convolution maps $M_j^i (j \in \{1, \dots N\})$, the size of convolution Kernel $K_x \times K_y$ (often square) and the connection schema to the previous layer L^{i-1}. Each M_j^i convolution card is the result of a convolution sum of cards from the previous layer M_j^{i-1} by its respective convolution core. A b_j^i bias is then added and the result is passed to a non-linear transfer function [44].

$$\phi(x) = 1.7159\, tanh(\tfrac{2}{3}x) \tag{1}$$

In the case of a map completely connected to the previous layer maps, the result is then calculated by:

$$M_j^i = \phi(b_j^i + \sum_{n=1}^{N} M_n^{i-1} * K_n^i) \tag{2}$$

$*$ is the convolution operator.

Figure 2 presents an example of convolution. The sky blue grid represents the input feature map and in this example, just one map is represented to make the drawing easy to understand. A value (dark blue zone) kernel glides over the map with a kernel (k) size 3×3 applied to a 5×5 input (i) using 1×1 strides (s). The final results of this operation are called output feature maps.

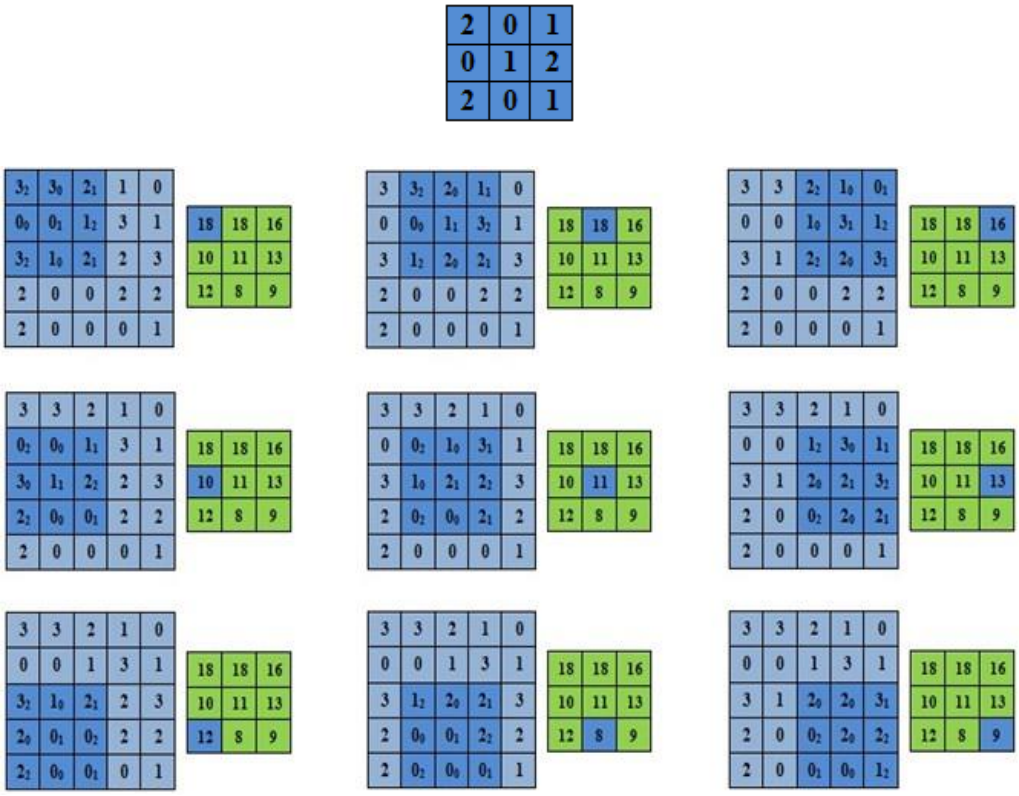

Figure 2: An example of calculating the output values of a convolution.

The schema in figure 3 shows an example of the first stage of convolution.

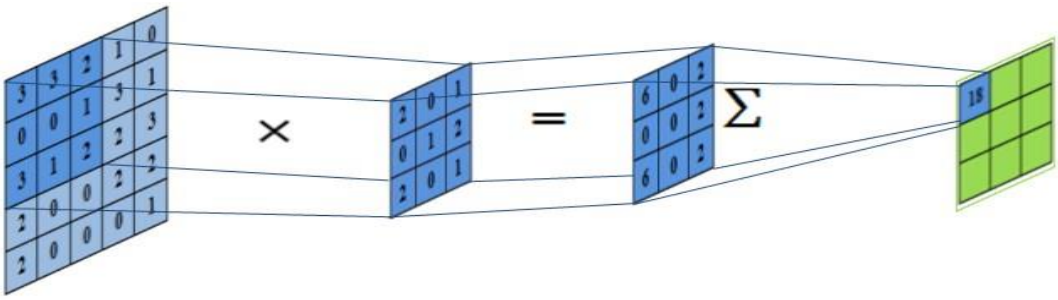

Figure 3 : A stage of the convolution calculation.

Stride defines how a filter slips around the input volume and zero-padding fills it with zeros around the border. So these stride and zero-padding are used to control the spatial dimension of the output volume. Figure4 illustrates an example of using padding and strides with clearer vision.

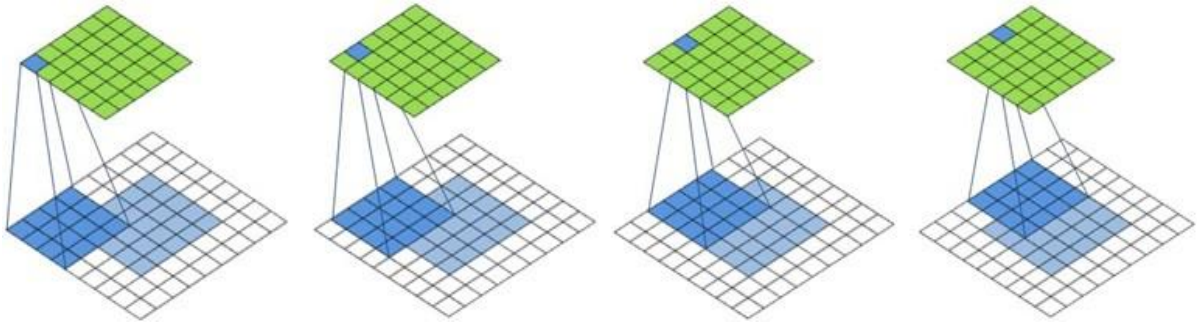

Figure 4: Convolution with p=2, k=4, i=5 and s=1

2) Pooling layer

Another important concept of CNNs is pooling, which is a form of sub-sampling of the image. In the CNN literature, convolutional layers are generally followed by sub-sampling layers in order to reduce over-learning. A sub-sampling layer reduces the size of the maps, reducing the number of parameters and calculation in the network. The pooling operation also created a form of invariance to (weak) rotations and translations that can appear as input.

Max-pooling is a variant of this layer that has shown satisfactory results. *The max-pooling layer* output is given by the maximum activation value within the input layer for different regions of size $K_x \times K_y$ does not overlap. In a similar way to a convolutional layer, a bias is added and the result is passed to the transfer function $\phi()$ defined above, an example of calculating the output values of max-pooling is shown in figure 5.

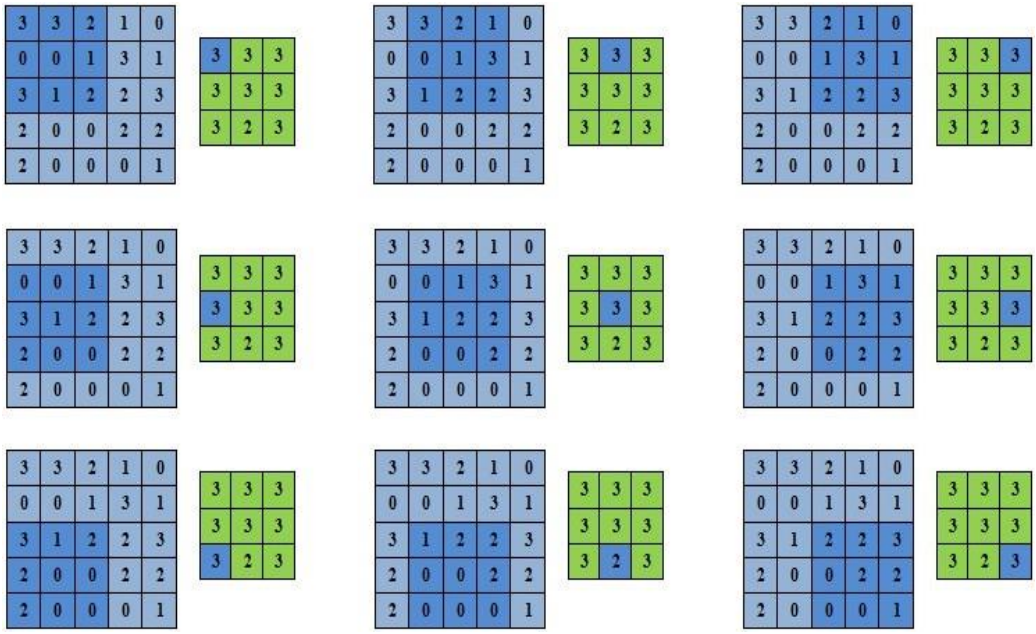

Figure 5: *Example of a 3×3 max pooling procedure over a 5×5 input using 1×1 strides*

3) Correction layer

Frequently, it is preferable to interpose between the processing layers, a layer that will operate a mathematical function (termed an activation function) on the output signals in order to obtain major efficiency of the processing. In particular:

- ReLU (Rectified Linear Unit) Correction: "non saturating activation function"

$$f(x) = \max(0, x) \tag{3}$$

- Hyperbolic Tangent Correction:

$$f(x) = \tanh(x) \tag{4}$$

- The correction by the Sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

Generally, the correction *Relu* is desirable; it allows a training of deep NNs several times faster and efficient on large databases. The graphical representations of these functions defined above are illustrated in figure 6.

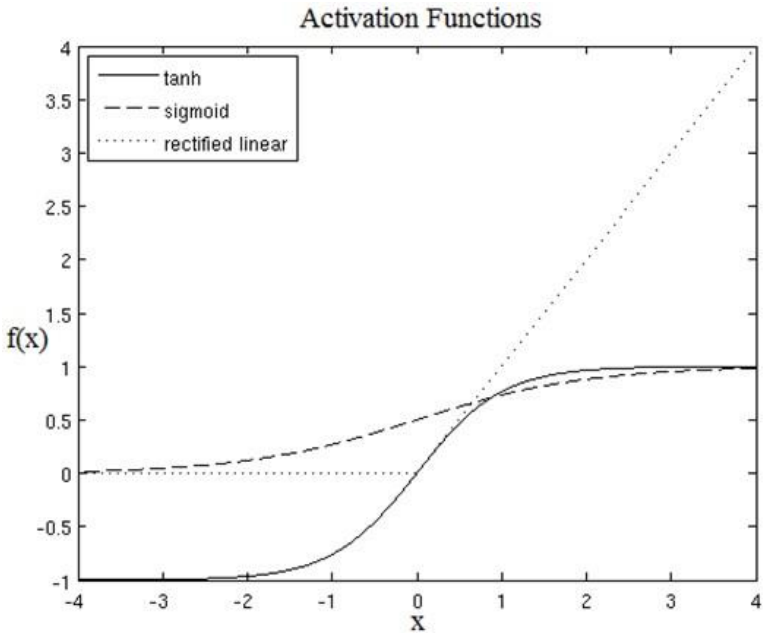

Figure 6 : Graphical representations of tanh, sigmoid and the rectified linear unit function.

4) Fully connected layer

At each ending of a convolution and pooling process, fully connected layers are added in order to realize the classification. The neurons that make up the fully connected layer have connections to all the outputs of the previous layer so that their activation functions can be calculated with a matrix product followed by a polarization shift.

5) Loss layer

This layer is considered as the last layer in the network. It defines how the formation of the network penalizes the difference between the expected and actual signal, one will find many functions adapted to specific tasks. In particular, the Softmax loss is widely used for classification which assigns a class to each object to be classified (in this work we have two classes, namely glaucoma and normal), Euclidean loss is used for regression, and Sigmoid cross-entropy loss is used to classify K probability values in the range [0,1].

IV. Multimodal fusion

Today, many real-world applications require multimodal data processing. More and more information collected from the real world is composed by nature consisting of data with different modalities. The term multimodality refers to the use of several modalities for carrying out the same task. A modality is a particular concrete form of a communication mode (visual mode, sound mode, gestural mode, etc.). For example, noise, music, speech are modalities of the sound mode. Theoretically, a multimodal computer system is a system capable of integrating several modalities (even if it integrates only one mode). The use of a multimodal approach can give not only better performance but also more robustness when one of these modalities is acquired in a noisy environment. Because of the rich characteristics of natural phenomena, it is rare that a single modality provides a complete knowledge of interest phenomenon.

The primary concept behind multimodality is complementarity, where each modality brings a certain types of information to be added to the system that cannot be deduced or obtained from other modalities. Mathematically, this added information is termed diversity [45]. Diversity allows supplying circumstances to a system to improve the uniqueness, the interpretability, the robustness, the performance, the decision-making and to obtain a global vision of the system. Multimodal data classification has become a very active area of research in recent years and has been used in many applications of practical interest [46], including the use of ensemble learning techniques [13].

Ensemble Learning is a machine learning technique that is considered a difficult task in the pattern recognition community [47, 48]. EL consists of training multiple basic models (multiple classifier systems (MCS)) as ensemble members and then combining their results into a single output in order to obtain an optimal predictive model with more accurate and reliable decisions; numerous studies have demonstrated the exceptional performance of EL for classification tasks [49, 50, 51, 52, 53], that can outperform the ensemble's individual members and several works have shown that multiple classifier systems generally generalize better than a unique classifier [54, 55, 56]. In 1965 [57], the notion of ensembles appeared in the classification literature and has then been studied in various ways, such as bagging [58], boosting [59], model averaging [60], stacking [61] and the mixture of experts [62]. MCS, another name for the mixture of experts or ensembles of classifiers, are especially helpful when different classifiers are formed on various parts of the characteristic space or when heterogeneous sets of characteristics are available for use in a multimodal classification problem.

There are two main types of commonly used ensemble methods:

-Heterogeneous ensemble methods: Combine a set of hypotheses $h_1 \dots h_T$, produced by different algorithms $L_1 \dots L_T$ on the same learning set A.

-Homogeneous ensemble methods: Combine a set of hypotheses $h_1 \dots h_T$, produced by the same algorithm $L_1 \dots L_T$ on a different learning set A. They use adaptive (boosting) or random (bagging) strategies.

By providing complementary information, multimodal data is generally used for achieving a good performance in classification tasks. However, the different information combination comes from various modalities (multimodal data) is a complex task, especially when one is interested in heterogeneous data. The purpose of the fusion is to correlate the elements of each modality and improve the quality of what is displayed by choosing to display the best of each modality. There are two main types of commonly used architecture to combine multimodal information to distinguish: early fusion (feature fusion) and late fusion (decision fusion).

1) *Early fusion:*

Early fusion consists of directly using the features extracted from the modalities to be combined in order to make the final decision of the system. In this case, there is no intermediate decision phase of each modality. For example, in our case, this would mean that no decision is made on the classification of glaucoma disease once the features are extracted from multimodal and fused images. This process is illustrated in figure 7:

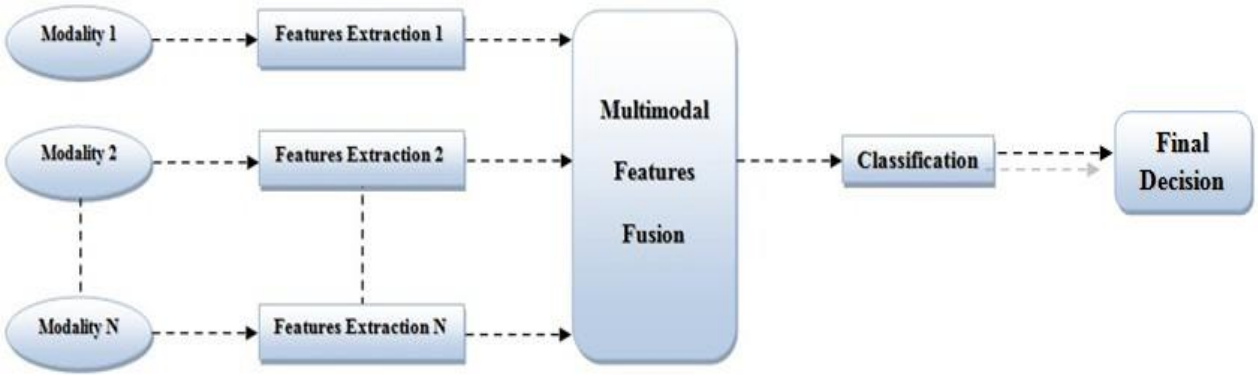

Figure 7: Illustration of the general early fusion process: fusion is applied directly to characteristics extracted from different modalities.

2) *Late fusion:*

Late fusion consists of merging decisions made on those modalities, rather than the features extracted directly from the modalities. It enables semantic concepts to be learnt directly at the unimodal level. For example, in our study, we first apply separate classifiers for each modality by considering that the modalities are independent, and then merging their outputs by applying a fusion method such as the voting method, figure 8 illustrates this principle.

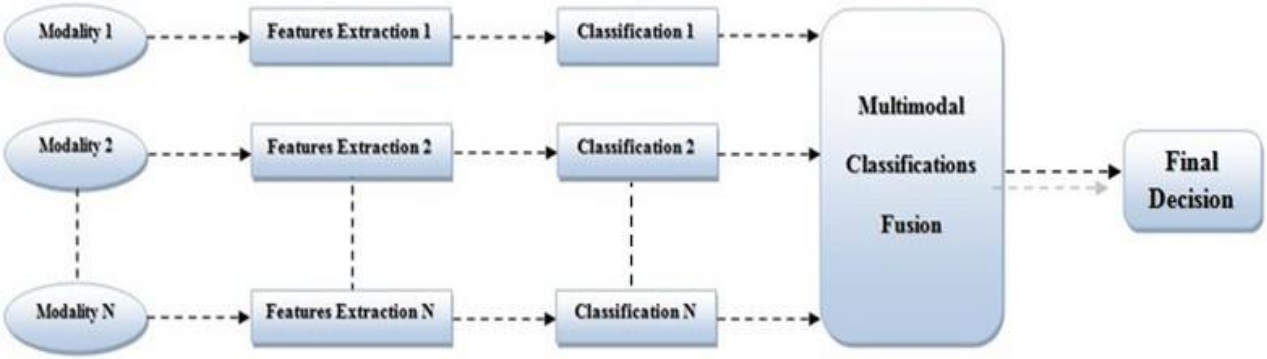

Figure 6: Illustration of the general late fusion process: fusion is applied to a decision set taken at the unimodal level.

In this study, we propose applying a new fusion approach, called the hybrid approach in order to benefit from two early and late fusion techniques, on the one hand by combining the different modalities before learning, and on the other hand we use separate classifiers for each modalities combination to really ensure the concept of multimodality (EL). We then merge the results using a voting method to obtain the final decision of our system. This principle is illustrated in figure 9.

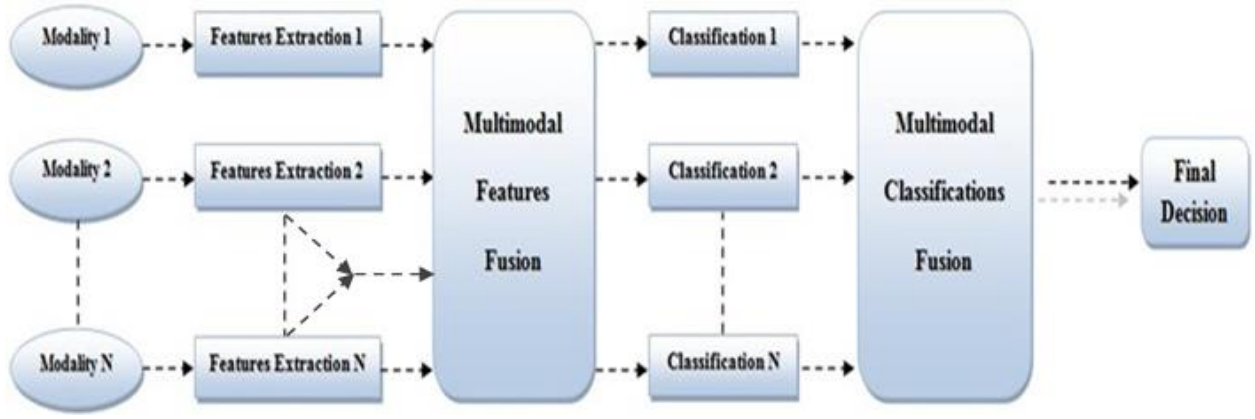

Figure 7 : A general illustration of our hybrid fusion approach.

V. Proposed Method:

The creation of our network by the proposed approach, illustrated in Figure 10, was obtained after several experiments and after a deep study of the literature for other tasks of pattern recognition and information fusion.

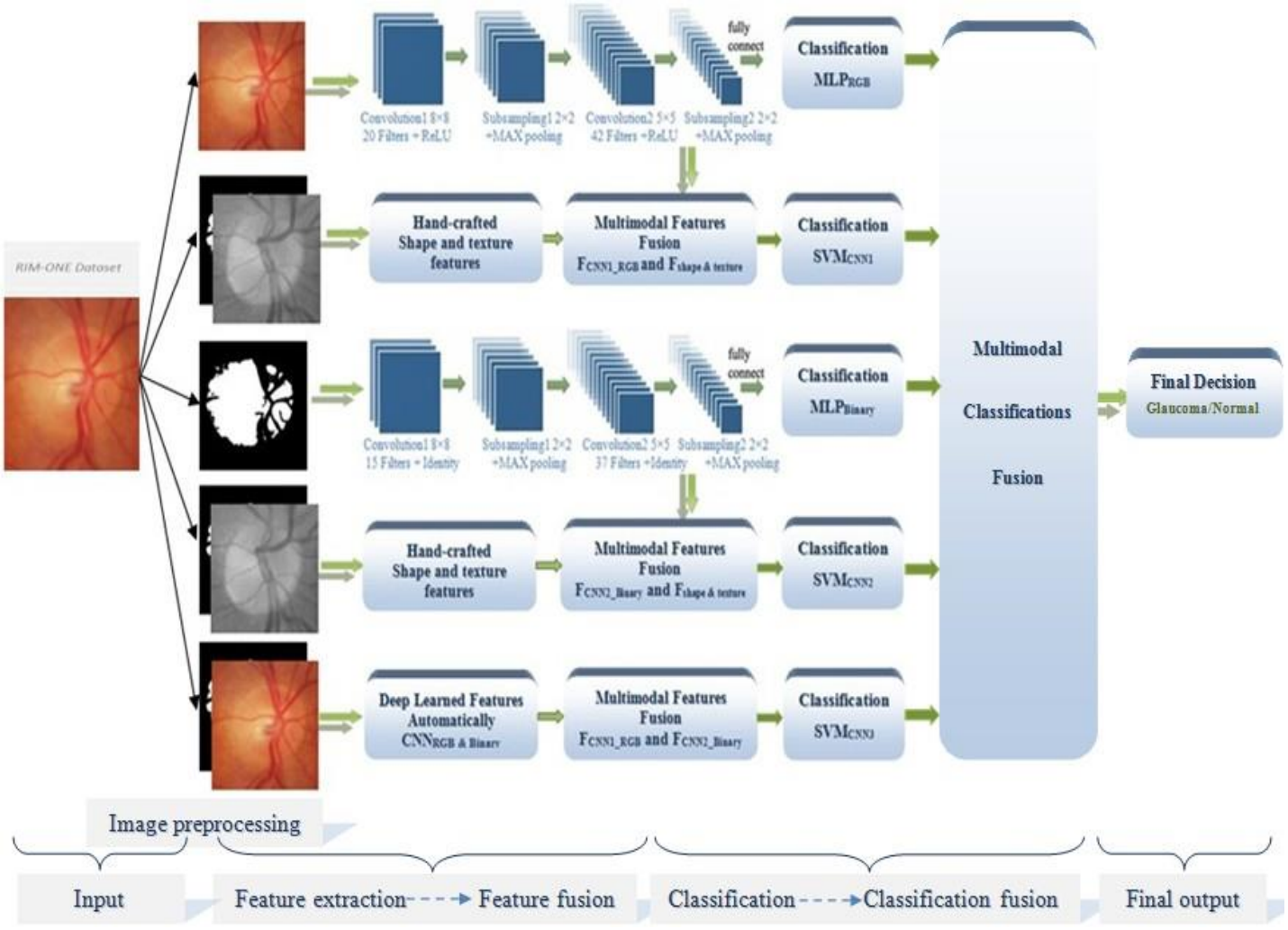

Figure 8 : Proposed Network Architecture for Glaucoma diagnosis

Our proposed multimodal classification system is based on two multimodalities; on the one hand, of the resource type and on the other hand of the characteristic type, in order to cooperate to obtain a very performing system. Firstly, we train two CNNs using multimodal data namely the RBV color images modality and the binary images modality from retinal fundus images. Secondly, we apply the principle of early fusion, which consists in extracting relevant characteristics from the different modalities and merging them- *bag of features*-, to form only a single vector on which the learning is realized; on the one hand, we combine the characteristics automatically generated by the CNN from the two modalities (RGB images and binary images) and the characteristics manually produced by traditional techniques such as GLCM, Hu Moments and Central Moments in a single multimodal representation in order to obtain the most representative characteristics, while on the other hand we merge the characteristics generated directly by the CNN from two modalities in a single multimodal vector for better classification of images. Finally, after obtaining the multimodal representation of the unimodal characteristics, we then apply separate classifiers for each multimodal feature vector by merging the five multimodal outputs using the commonly used voting method to ensure the principle of late fusion, and thus obtain the final diagnosis. The use of the five different models allows the system to avoid the incompatibility problem in the final decision of our system. We adopt an SVM classifier in addition to a CNN classifier for the classification phase because it is considered the best binary separator in the medical field.

1 Image preprocessing

The preliminary pretreatment stage allows obtaining the binary image modality which represents another point of view of the database by using the commonly used Otsu method (see Figure 11) because of its better performance in many areas of image processing and which reveals a better contrast between retinal structures.

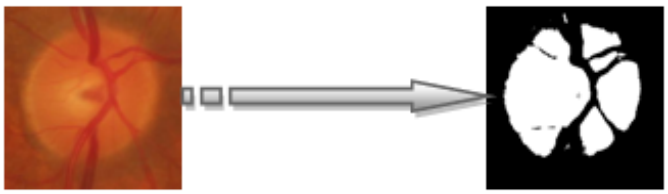

Figure 11: An example of a Binarize image from the RIM-ONE dataset using the Otsu method.

Binarization is an important step in any process of image processing and analysis. Binarization is an operation that produces two classes of pixels; in general, they are represented by black pixels and white pixels. The binarization of an image can be done using a threshold: pixels whose gray level is below the threshold become black, and those above become white. The binarisée image of good quality can produce more precision in pattern recognition by comparing with the source image because the binarisée image does not contain noise [63].

Otsu method is one of the automatic threshold calculation methods for the unavoidable binarization with that of Kittler & Illingworth. The principle of Otsu's thresholding method is to find the threshold that minimizes the intra-class weighted variance, as well as maximizes the interclass variance [64]. In another way, the Otsu method tries to find the threshold, t, which separates the histogram optimally into two segments. The Otsu method steps are presented below:

1- The weighted intra-class variance to find the threshold is defined as follows:

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t) \tag{6}$$

- The class probabilities are calculated according to:

$$q_1(t) = \sum_{i=1}^{t} p(i) \qquad and \qquad q2(t) = \sum_{i=t+1}^{L} p(i) \tag{7}$$

Where Lrepresents the bins of the histogram,$t$is the threshold which separates these two classes and $\sigma_1^2\ and\ \sigma_2^2$ are the variances of these latter.

2- The class means are defined as follows:

$$\mu_1(t) = \sum_{i=1}^{t} \frac{ip(i)}{q_1(t)} \qquad and \qquad \mu_2(t) = \sum_{i=t+1}^{L} \frac{ip(i)}{q_2(t)} \tag{8}$$

3- The total mean level of the original image is defined as follows:

$$\mu_T = \mu(L) = \sum_{i=1}^{L} ip(i) \tag{9}$$

4- The class variances are computed as follows:

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{p(i)}{q_1(t)} \tag{10}$$

and

$$\sigma_2^2(t) = \sum_{i=t+1}^{L} [i - \mu_2(t)]^2 \frac{p(i)}{q_2(t)} \tag{11}$$

5- The total variance of levels :

$$\sigma_T^2(t) = \sum_{i=1}^{L} (i - \mu_T)^2 p(i) \tag{12}$$

6- The between-class variance is given by:

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = q_1(\mu_1 - \mu_T)^2 + q_2(\mu_2 - \mu_T)^2 = q_1(t)q_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{13}$$

7- Finally, the$\sigma_b^2(t)$maximum represents the desired threshold.

2 CNN architecture and conception

We used 6 layers in both CNN Architectures: convolutional layer C1, subsampling layer S1, convolutional layer C2, subsampling layer S2, dense layer D and output layer O (see figure 10).

We developed a network with a CNN architecture that avoided the phase of extracting traditional handcrafted features by processing the extraction of features and classification at one time within the same network of neurons and therefore provide an automatic diagnosis.

CNNs are currently the most powerful models for classifying images [12, 13]. They have two distinct parts. At the input, an image is provided in the form of a matrix of pixels. It has 2 dimensions (width and height) to a grayscale image and 3 dimensions for color image RGB (with 3 depth units, the third of which corresponds to the stacking of 3 images according to each color, red, green and blue). The first part of a CNN is the conventional part itself. It functions as an extractor of image characteristics. An image is passed through a succession of filters, or convoluted nuclei, creating new images called convolution maps. Some intermediate filters reduce the resolution of the image by a local maximum operation. In the end, the convolution maps are combined in a feature vector, called a CNN code. This code CNN got out of it from the convolutive party is then connected in the entry of a second part, constituted by completely connected layers (multilayer perceptron). The role of this part is to combine the characteristics of the code CNN to classify the image. The output is a last layer with one neuron per category.

3 Training

Since the images are generally too voluminous to be used directly in a CNN, consequently each image in our database is resized to 100 x 100 in order to reduce the computational complexity and ensure a standard scale for all images used in the training.

Our neural network is performed after several performance tests. We start through the convolution blocks creation; a batch normalization step is applied after each convolutional layer to decrease the number of feature maps. A *stochastic gradient descent* is used with a *momentum* value of 0.9. *L2 regularization* method is also applied to weight and biases with a threshold equal to 0.0005. Finally, a low *learning rate* is fixed at 0.0001 to train our neural network. We used two convolution layers and two sub-sampling layers, are structured one after the other with a *ReLU* and *Identity* activation function, for the convolution layers we perform a *stride (1, 1)*, and for sub-sampling layers we use size *2x2* for *kernel Size* and *stride* with the Max-pooling function. For the dense layer we used the *ReLU* and *Cube* functions, also the *Square Mean Error (MSE)* function has been used to optimize the loss function. Finally, for the classification we use the function *Softmax* widely used. Figure 12 shows the training process of our CNN.

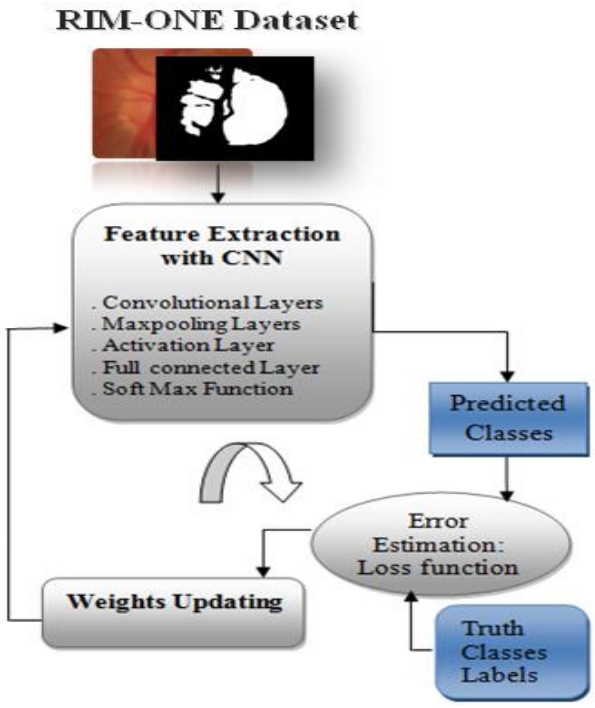

Figure 12: Supervised training process of CNN Classifier.

4 Computational complexity

Recently, there has been a growing interest in speeding up the execution of CNNs. the actual execution time may be responsive to implementations and equipment. Our models in this study have an affordable and low-cost computational cost, which takes only one day of forming on a single processor, but once our models are pre-forming (off-line), the concept of computational cost does not intervene in the real-time classification phase (on-line). The following figure shows that, from a number of iterations, our CNN classifiers converge; this figure presents model score versus iteration of the two CNNs, this is the value of the loss function on the current mini-batch.

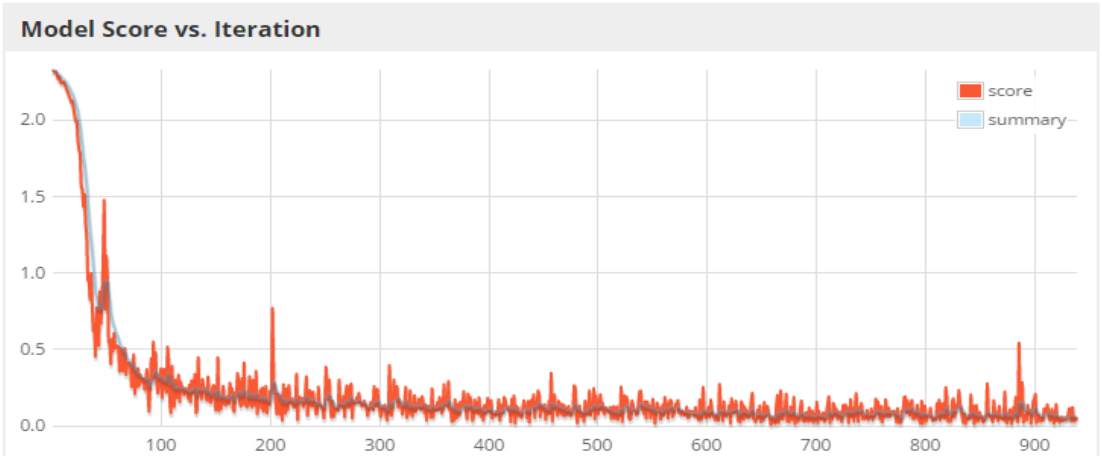

Figure 13: Model score versus iteration

The main model parameters and training information of proposed CNN architecture are described in the tables below.

Table 1: Model and Training Information of $CNN1_{RGB}$.

Model and Training Information

Model Type	MultiLayerNetwork
Layers	6
Total Parameters	431080
Start Time	
Total Runtime	
Last Update	2018-05-12 23:57:39
Total Parameter Updates	938
Updates/sec	3,47
Examples/sec	111,11

Table 2: Model and Training Information of $CNN2_{Binary}$.

Model and Training Information

Model Type	MultiLayerNetwork
Layers	6
Total Parameters	431080
Start Time	
Total Runtime	
Last Update	2018-05-12 23:39:38
Total Parameter Updates	938
Updates/sec	3,36
Examples/sec	107,38

Limiting network complexity is one way to understand the impacts of factors such as depth (number of layers), width (numbers of filters), filter sizes, and strides of the architectures in network conception. CNN's accuracy is evaluated using these factors and their increase increases the computational cost.

The computational cost of the convolutional layer is $O(F_I MNmnF_O)$, where $M \times N$ represents the feature map size of each input, $m \times n$ is the convolution kernel size and F_I, F_O are the input and output channels in a layer, respectively. The computational cost of the pooling layer is frequently highly low compared to that of the convolutional layer, which is defined as follows: $O(F_I MN)$ [65].

Finally, our method offers accurate recognition with limited computational complexity and affordable time cost.

5 Handcrafted feature extraction

In pattern recognition, characteristics are the measurable properties of an observed physical phenomenon. The discriminant characteristics extraction is a fundamental step in the recognition process prior to classification. The feature extraction phase consists of calculating a set of measures to represent each class as unique as possible and also to reduce dimensionality. The performance of recognition systems depends largely on the choice of descriptors used and the techniques associated with their extraction. Many descriptors are used in image search systems by content and more particularly in diagnostic aided systems to describe forms such as descriptors of colors, textures, and shape. Wen-Jie and Woo [66] proposed an automatic system for diagnosing breast tumors, using the characteristics of the auto-covariance texture and the morphological characteristics of solidity. These features are extracted and merged to introduce them into the SVM classifier to identify abnormalities in breast images. Bob et al. [67] used two-dimensional Gabor filter with different scales and orientations for extracting the texture features. The facial color gamut was used for extracting the color features; the color and texture characteristics extracted are combined to achieve a tool for facial diagnosis. The results showed an accuracy of 99.83%. Chang et al. [68] suggested combining the shape and color features to obtain an automatic tongue diagnosis system. The results showed that this combination can give better diagnostic accuracy.

In this study, as regards the traditional feature extraction methods, we rely on two families of heterogeneous features commonly used namely the texture and shape in order to concatenate them with the features automatically generate by the CNN to ensure a better vector representation of the medical image features. In terms of texture features, we used the GLCM method, and we rely on both methods of shape features extraction: Central Moments and Hu Moments. These three adopted methods are widely used and have given better results in our related work compared to the literature [8, 69, 70].

1) *Gray Level Co-Occurrence Matrix (GLCM):*

Because of their richness in texture information, Co-Occurrence Matrix has become the best known and most used to extract these textures features [71, 72]. The co-occurrences matrices consist in measuring the probability of appearance of pixel values pairs located at a certain distance in the image. It is based on the calculation of the probability p (i, j, d, θ) which represents the number of times a color level pixel i appears at a relative distance d of a color level pixel j and according to a given orientation θ. The angular directions θ conventionally used are 0, 45, 90 and 135 degrees. This matrix characterizes the identifiable patterns in gray levels of a pixels region. As co-occurrences matrices count a very large amount of information difficult to exploit directly to characterize the textures. Haralick et al. [73] proposed the first fourteen parameters, characterizing the textures, resulting from these matrices. Recently, in medical diagnosis, only the first thirteen most appropriate characteristics are commonly used [74, 75] and which are considered in our study. The Haralick features considered in this study are presented as follows:

Angular Second Moment:

$$P_1 = \sum_i \sum_j (p(i,j))^2 \tag{14}$$

Where p (i, j) are the elements of the matrix P.

Contrast:

$$P_2 = \sum_{n=0}^{N_g^{-1}} n^2 \left(\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right) \tag{15}$$

The variable N_g is the gray levels number used in the image.

Correlation:

$$P_3 = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{16}$$

The variables $\mu_x \sigma_x$ represent the mean and standard deviation respectively measured for the vector$p_x(i)$

Sum of squares (Variance):

$$P_4 = \sum_i \sum_j (i - \mu) p(i,j) \tag{17}$$

Inverse Difference Moment:

$$P_5 = \sum_i \sum_j \frac{1}{1 + (i-j)} p(i,j) \tag{18}$$

Sum Average:

$$P_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i) \tag{19}$$

Sum Variance:

$$P_7 = \sum_{i=2}^{2N_g} \left(i - P_g\right)^2 p_{x+y}(i) \tag{20}$$

Sum Entropy:

$$P_8 = \sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i) + \epsilon) \tag{21}$$

Entropy:

$$P_9 = \sum_i \sum_j p(i,j) \log(p(i,j) + \epsilon) \tag{22}$$

Difference Variance:

$P_{10} =$ variance (p_{x-y}) (23)

p_{x-y} represents the difference of two vectors $p_x(i)$ and $p_y(i)$.

Difference Entropy:

$$P_{11} = -\sum_{i=2}^{2N_g} p_{x-y}(i) \log(p_{x-y}(i) + \varepsilon) \tag{24}$$

Information measures of Correlation (2): Expresses the measure of the linear dependence of gray level between pixels in specific positions.

$$P_{12} = \frac{P_9 - HXY1}{\max(HX, HY)} \tag{25}$$

$$P_{13} = [1 - \exp(-2.0(HXY2 \; - \; P_9))]^2 \tag{26}$$

The variables$HX$and HYrepresent the measured entropy on$p_x(i)$ and $p_y(i)$ respectively.

The variables HXY1 and HXY2 are computed as follows:

$$HXY1 = -\sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\} \tag{27}$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\} \tag{28}$$

2) Central Moments:

Moments are scalar quantities used to describe a function and to capture its important characteristics. The moments notion in mathematics, are projections of a function on a polynomial basis; different systems of moments can be recognized according to the polynomial base used. The Central Moments have become one of the most used shape descriptors in many fields [76-78], which have shown superior performance. A central moment is a moment of a probability distribution of a random variable on the mean of the random variable. Geometric moment (raw moment) of order (p+q) for a 2-dimensional discrete function is calculated as follows:

$$M_{pq} = \sum_x \sum_y x^p \, y^q \, f(x,y) \tag{29}$$

Where $p, q = 0,1,2, ...,$ and (x, y) the pixel position.

The central moments are generally used to substitute the raw moment which are defined as follows [79]:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) \, dx \, dy \tag{30}$$

Where $\bar{x} = \frac{M_{10}}{M_{00}}$and $\bar{y} = \frac{M_{01}}{M_{00}}$ are the components of the centroid, $f(x, y)$ is the image function andM_{00} is the area for binary images and for grey-tone images it is the sum of grey level.If f(x, y) is a numerical image, then the precedent equation becomes:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y) \tag{40}$$

The central moments used of order up to 3 are:

$$\mu_{00} = M_{00} \tag{41}$$

$$\mu_{01} = \; \mu_{10} = 0 \tag{42}$$

$$\mu_{11} = M_{11} - \bar{x}M_{01} = M_{11} - \bar{y}M_{10} \tag{43}$$

$$\mu_{20} = M_{20} - \bar{x}M_{10} \tag{44}$$

$$\mu_{02} = M_{02} - \bar{y}M_{01} \tag{45}$$

$$\mu_{21} = M_{21} - 2\bar{x}M_{11} - \bar{y}M_{20} + 2\bar{x}^2 M_{01} \tag{46}$$

$$\mu_{12} = M_{12} - 2\bar{y}M_{11} - \bar{x}M_{02} + 2\bar{y}^2 M_{10} \tag{47}$$

$$\mu_{30} = M_{30} - 3\bar{x}M_{20} + 2\bar{x}^2 M_{10} \quad (48)$$

$$\mu_{03} = M_{03} - 3\bar{y}M_{02} + 2\bar{y}^2 M_{01} \quad (49)$$

3) Hu Moments:

A reformulation of the previously represented moments is necessary to allow the rotation invariance, the rotation Hu-moment invariants (HMI) was proposed by Hu [79] in 1962. Hu obtained these expressions thanks to algebraic invariants carried out at the generating function of the moment under a rotation transformation. An ensemble of nonlinear centralized moment expressions constitutes the HMIs, which are completely orthogonal (i.e. rotation) invariant. HMIs are frequently used in image processing and play a very important role in pattern recognition; they are calculated from normalized centralized moments up to order 3 and are described as follows:

$$I_1 = \eta_{20} + \eta_{02} \quad \textbf{(50)}$$

$$I_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \quad \textbf{(51)}$$

$$I_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2 \quad \textbf{(52)}$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad \textbf{(53)}$$

$$I_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\,[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad \textbf{(54)}$$

$$I_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad \textbf{(55)}$$

$$I_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad \textbf{(56)}$$

In order to calculate the central moments and the Hu moments to extract the shape features, we need to convert the color images to binary images; for this purpose, we adopted the Otsu method as presented previously. The co-occurrence matrix is computed from grayscale images using the ImageJ tool[2].

6 Fusion and classification

As there is no unique model for all pattern recognition problems and no unique method is applicable to all problems; in other words, there is no "best" classifier able to process (learn) any distribution of learning data. For this reason, it has not been possible to highlight the undeniable superiority of one classification method over another or of a feature extractor over another, which leads us to take an interest in Ensemble Learning (EL).

As already mentioned in Section 4, we applied our approach to hybrid fusion: the features concatenation of different modalities (early fusion) and the multimodality classification (late fusion). In order to benefit from the complementarity that can exist between classifiers on the one hand, and of the set of features multimodal generated according to the adopted methods (CNN, GLCM, Central Moments and Hu Moments) on the other hand, five models have been used which are $MLP_{RGB,}$ SVM_{CNN1}, MLP_{Binary}, SVM_{CNN2} and SVM_{CNN3} (see figure 10).

[2] http://imagej.net/Welcome

In this study, we adopt the parallel ensemble method where the five basic learners are generated in parallel. The basic motivation that we use the parallel method is to exploit the independence between the five basic classifiers because the merger/averaging can greatly reduce the error. An appropriate combination method should be used after the ensemble's formation algorithms (the multiple base learners) have been formed in order to combine their training outputs into a unique form as the final classifier. Despite a large number of combination methods existing in the literature, only three of them have been widely used and have shown considerable potential for amelioration in many applications of EL, which are linear combiner, product combiner, and vote combiner.

To generate the final decision of our multimodal system, a vote-based merge technique is applied. Majority voting is considered to be one of the effective methods of fusion [80, 81]. In the process of majority voting method, a decision to choose the label of an input sample is produced by each classifier. The class that has the highest number of votes is determined as the representative class of all the classifiers in the set [82]. Let $C = [c_1, c_2, \dots c_L]$ be a set of L classifiers, x is the input sample and $c_{i,j}$ is the output of the i^{th} classifier for thej^{th} class. The final decision using the majority voting can be defined as follows:

$$MV(x) = max_{j\in\Omega} \sum_{i=1}^{L} c_{i,j} \tag{57}$$

However, the specific accuracy of each classifier is not considered in the final decision, which is considered to be the main flaw in the majority voting methods. In general, the chosen classifiers do not have a similar skill. Consequently, the weighted voting method is used to combine the decision of the chosen classifiers [83]. In this aggregated method, the result of each classifier is weighted by a coefficient that affects the combination process. Note that w_i is the weight of the i^{th} classifier**,** the weighted majority voting is defined as:

$$WMV(x) = max_{j\in\Omega} \sum_{i=1}^{L} w_i c_{i,j} \tag{58}$$

and $\sum_{i=1}^{L} w_i = 1$

Many schemes have been suggested to estimate the weights of classifiers [84]. Usually, these weights are estimated using the specific accuracy of each classifier. Let a_i and a_j be the accuracies of the i^{th} and j^{th}classifiers on the validation set. The weight w_i is calculated by:

$$w_i = \frac{a_i}{\sum_{\boldsymbol{j}} a_j} \tag{59}$$

In this study, the best-worst weighted Vote (BWWV) scheme [83] is used as a measure to quantify the weights. The main idea behind this scheme is to identify the best and worst members of the set using their estimated error on the validation set; the a_i values are determined using the following expression:

$$a_i = \mathbf{1} - \frac{e_k - e_b}{e_w - e_b} \tag{60}$$

Where e_w indicates the maximum error among the classifiers, and e_b is the minimum error.
The a_i value varies in [0,1], where 0 indicates the worst classifier and 1 corresponds to the best classifier.

VI. Experimental results

In order to evaluate our multimodal approach for the classification of glaucoma, we use the k-fold cross-validation method with k=10. The implementation of the proposed work was done with Deeplearning4j[3], the first commercial-grade, open-source, distributed deep-learning library written for Java and Scala. In the sub-sections that follow, the details of the experiments and their results are represented.

1 Used dataset

The proposed method for glaucoma diagnosis uses a set of 455 images from the open Retinal Image Database for Optic Nerve Evaluation (RIM-ONE) database, of which 200 images represent glaucoma disease and 255 images are normal; they are centered on the ONH with a field-of-view (FOV) of 34°. Figure 14 shows examples of retinal fundus images.

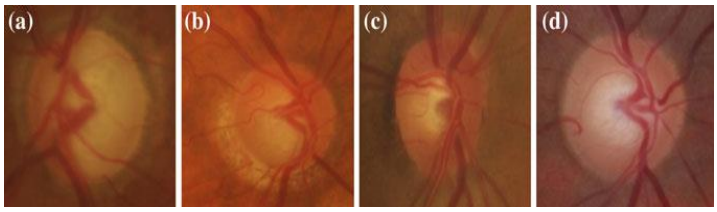

Figure 14: Examples of fundus images from RIM-One database: (a) and (b) glaucoma, (c) and (d) normal.

2 Evaluation criteria

In order to evaluate the performance of our system, we use evaluation criteria commonly used in the medical field which are described as follows:

$$Sensitivity\ (SEN) = \frac{TP}{TP + FN} \tag{61}$$

$$Specificity\ (SPE) = \frac{TN}{TN + FP} \tag{62}$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP + FP} \tag{63}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN + FN} \tag{64}$$

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + TN + FP + FN} \tag{65}$$

$$Matthews\ Correlation\ Coefficient\ (MCC) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{66}$$

Where,

True Positives (TP) describe the samples of sick people with a positive test.

True Negatives (TN) describe the samples of non-sick people with a negative test.

False Positives (FP) represent the samples of non-sick people with a positive test.

False negatives (FN) represent the samples of sick people with a negative test.

[3] https://deeplearning4j.org/

3 Results

In this work, five models are used: $CNN1_{RGB}$ based on the color fundus images (original images from Rim-One dataset), SVM_{CNN1} using all features (GLCM, Central Moments, Hu Moments and features map generated by $CNN1_{RGB}$), $CNN2_{Binary}$ based on the binary images (original images from the Rim-One dataset converted by the Otsu algorithm), SVM_{CNN2} using all features (GLCM, Central Moments, Hu Moments and features map generated by $CNN2_{Binary}$) and SVM_{CNN3} based on the $CNN1_{RGB}$ and $CNN2_{Binary}$ features. The following tables summarize the individual performance of each classifier used.

Table3 (MLP_{RGB}) illustrates that 197 images of glaucoma were formally recognized as glaucomatous images by the CNN_{RGB} model, 252 normal images are properly classified as non-glaucomatous images. In summary, 449 images are accurately labeled, resulting in 98.68% Accuracy with Sensitivity 98.5%, Specificity 98.82%, Positive Predictive Value (PPV) 98.5%, Negative Predictive Value (NPV) 98.82% and Matthews Correlation Coefficient (MCC) 97.32%.

Table 3: Confusion Matrix of $CNN1_{RGB}$ results

	Glaucoma	Healthy
Glaucoma	197	3
Healthy	3	252

Table4 (MLP_{Binary}) shows that 199 images of glaucoma were correctly recognized as glaucomatous images by the $CNN2_{Binary}$ model, 253 normal images are properly classified as non-glaucomatous images. In summary, 452 images are accurately labeled, resulting in 99.34% Accuracy with Sensitivity 99.5%, Specificity 99.22%, Positive Predictive Value (PPV) 99%, Negative Predictive Value (NPV) 99.61% and Matthews Correlation Coefficient (MCC) 98.66%.

Table 4: Confusion Matrix of $CNN2_{Binary}$ results

	Glaucoma	Healthy
Glaucoma	199	1
Healthy	2	253

In table 5 (SVM_{CNN1}), we demonstrate that 198 images of glaucoma were correctly detected as glaucomatous images by the SVM_{RGB} model, 253 normal images are properly classified as non-glaucomatous images. In summary, 451 images are accurately labeled, resulting in 99.12% Accuracy with Sensitivity 98.50%, Specificity 99.61%, Positive Predictive Value (PPV) 99.49%, Negative Predictive Value (NPV) 98.83% and Matthews Correlation Coefficient (MCC) 98.22%.

Table 5: Confusion Matrix of SVM_{RGB} results

	Glaucoma	Healthy
Glaucoma	197	3
Healthy	1	254

As shown in table 6 (SVM_{CNN2}), 198 images of glaucoma were correctly detected as glaucomatous images by the SVM_{Binary} model, 255 normal images are properly classified as non-glaucomatous images. In summary, 453 images are accurately labeled, resulting in 99.56% Accuracy with Sensitivity 99%, Specificity 100%, Positive Predictive Value (PPV) 100%, Negative Predictive Value (NPV) 99.22% and Matthews Correlation Coefficient (MCC) 99.11%.

Table 6: Confusion Matrix of SVM_{Binary} results

	Glaucoma	Healthy
Glaucoma	198	2
Healthy	0	255

Table 7 (SVM_{CNN3}) illustrates that 200 images of glaucoma were formally recognized as glaucomatous images by the $SVM_{RGB\&Binary}$ model, 252 normal images are properly classified as non-glaucomatous images. In summary, 452 images are accurately labeled, resulting in 99.34% Accuracy with Sensitivity 100%, Specificity 98.82%, Positive Predictive Value (PPV) 98.52%, Negative Predictive Value (NPV) 100% and Matthews Correlation Coefficient (MCC) 98.67%.

Table 7: Confusion Matrix of $SVM_{RGB\&Binary}$ results

	Glaucoma	Healthy
Glaucoma	200	0
Healthy	3	252

A comparative study on the SVM classifier performance using several feature families combinations has been achieved with figure 15 illustrating the obtained results. Comparing the obtained results, we find that the combination of the three techniques of handcrafted characteristics extraction (GLCM, Central Moments and Hu Moments) with the characteristics automatically produced by CNN (from both modalities) is much better than using the separate combination of these last three techniques with CNN's characteristics, which leads to the improvement of the diagnostic performance.

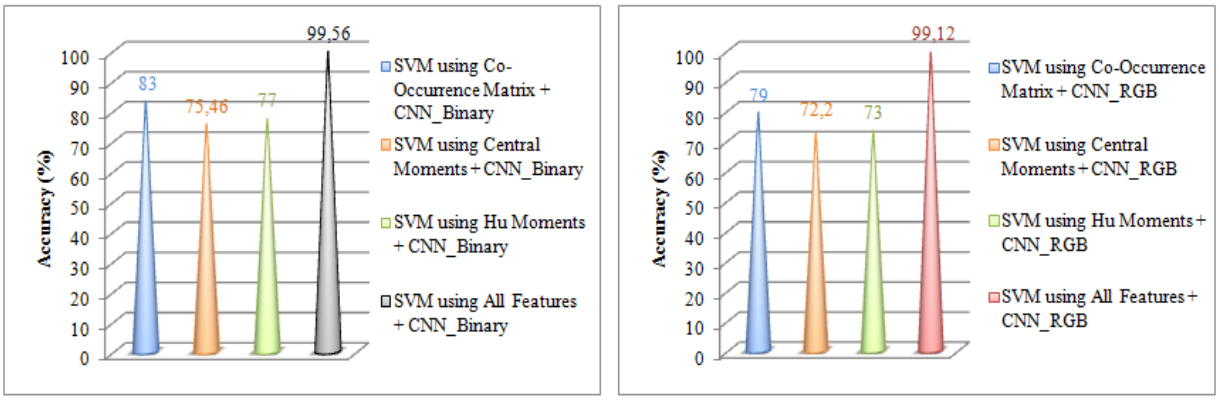

Figure 15: Comparison of obtained results using combinations different from characteristics

The important parameters to tune for the SVM classifier are defined as follows: Radial Basis Function (RBF) kernel with a gamma value of 0.5 and the c parameter equal to 4.

The proposed approach was also evaluated according to the ROC curves [85]. The ROC curves of our models are plotted in figure 16.

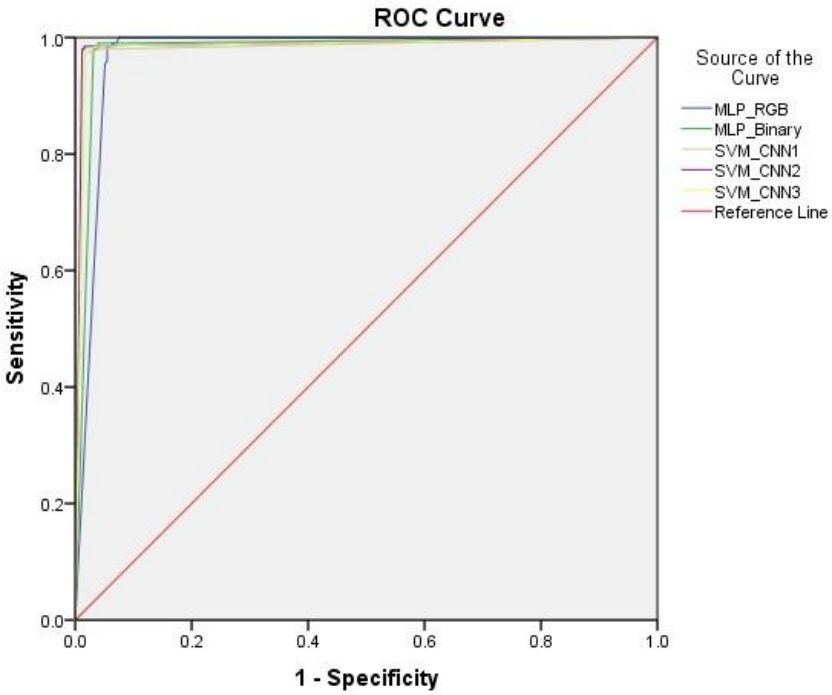

Figure 16: ROC Curve of five models used

The areas under the curve (AUC) values for five models are given in table 8.

Table 8: AUC values of five models used

Area Under the Curve

Test Result Variable(s)	Area
MLP_RGB	.973
MLP_Binary	.979
SVM_CNN1	.982
SVM_CNN2	.987
SVM_CNN3	.984

The next table summarizes the obtained results for each model.

Table 9: Obtained results from five models used

Model	ACC(%)	SEN(%)	SPE(%)	PPV(%)	NPV(%)	MCC(%)	AUC(%)
MLP_{RGB}	98.68	98.50	98.82	98.50	98.82	97.32	97.30
MLP_{Binary}	99.34	99.50	99.22	99.00	99.61	98.66	97.90
SVM_{CNN1}	99.12	98.50	99.61	99.49	98.83	98.22	98.20
SVM_{CNN2}	99.56	99.00	100	100	99.22	99.11	98.70
SVM_{CNN3}	99.34	100	98.82	98.52	100	98.67	98.40

4 Discussion

In this study, several experiments are performed about the CNN's performance, notably on the number of epochs and the activation function. Figure 17 demonstrates the impact of the number of epochs on the accuracy of two $CNN1_{RGB}$ and $CNN2_{Binary}$ classifiers according to the activation function used.

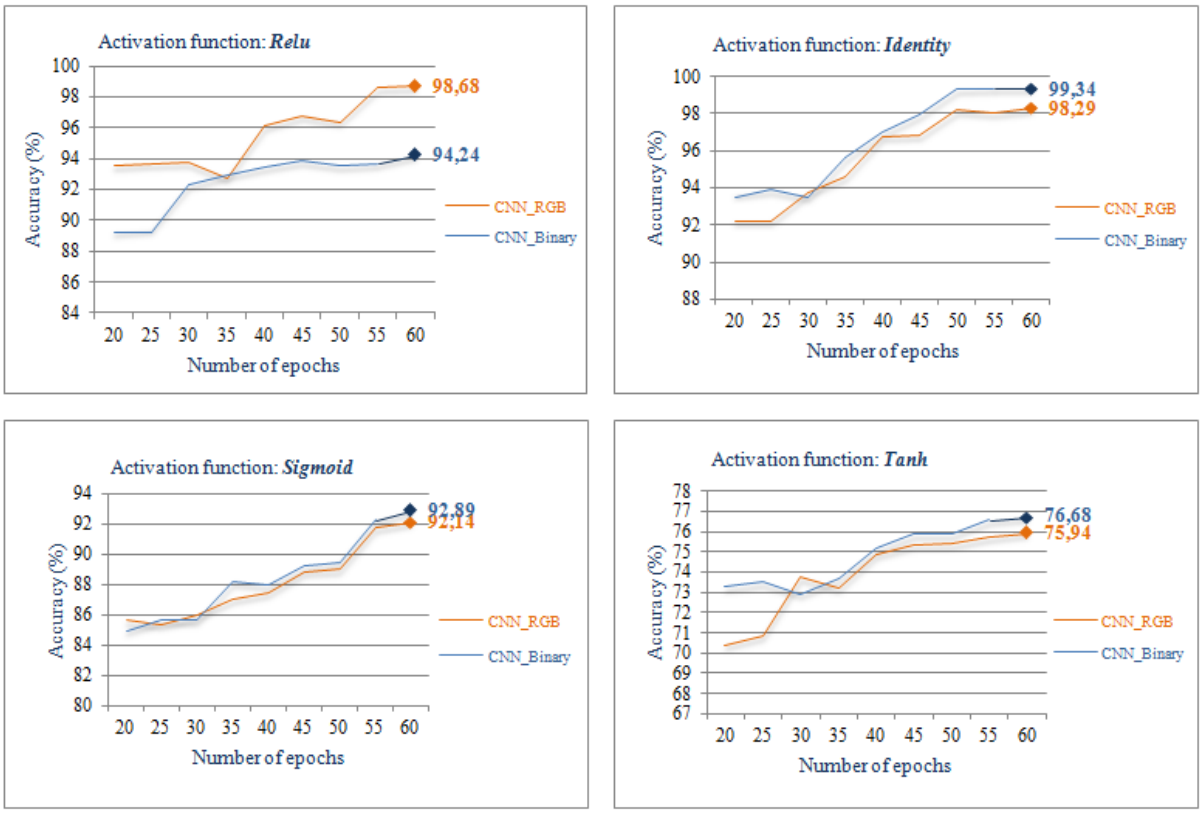

Figure 17: Plot of accuracy of two $CNN1_{RGB}$ and $CNN2_{Binary}$ models compared with the different epochs according to the *Relu*, *Identity*, *Sigmoid* and *Tanh* activation functions.

In general, we observed that the increase in the number of epochs resulted in an increase in the accuracy of both CNNs ($CNN1_{RGB}$ and $CNN2_{Binary}$) for the four activation functions considered. According to this figure, we found that the $CNN1_{RGB}$ and $CNN2_{Binary}$ classifiers achieved maximum performance when the number of epochs equals 60 for the four activation functions used: *Relu*, *Identity*, *Sigmoid* and *Tanh*; and that the accuracy is almost stable after 55 epochs. We also noted that the $CNN1_{RGB}$ model obtained better accuracy by using the *Relu* function with a 98.68% rate while $CNN2_{Binary}$ model by using the *Identity* function outperform all other activation functions with a 99.38% accuracy.

In order to improve the sensitivity and specificity of our multimodal classification and strengthen overall performance of diagnostic on the one hand, and to generate the final decision of our approach, on the other hand, three aggregation methods have been applied: Majority voting (MV), Weighted Majority Voting (WMV) and Best-Worst Weighted Vote (BWWV). Table 10 shows the total results obtained.

Table 10: Obtained results from three aggregation methods used

Model	ACC(%)	SEN(%)	SPE(%)	PPV(%)	NPV(%)	MCC(%)	AUC(%)
MV	99.12	99.00	99.22	99.00	99.22	98.22	98.60
WMV	99.56	99.50	99.61	99.50	99.61	99.11	98.80
BWWV	99.78	99.50	100	100	99.61	99.55	99.20

Figure 18 illustrates the ROC curves [85]. The AUC values of three aggregation methods considered are given in table 11. According to the results presented above, we find that the Best-Worst Weighted Vote (BWWV) method gave the best result and with high accuracy, comparing with state-of-the-art results using the same database (RIM-ONE).

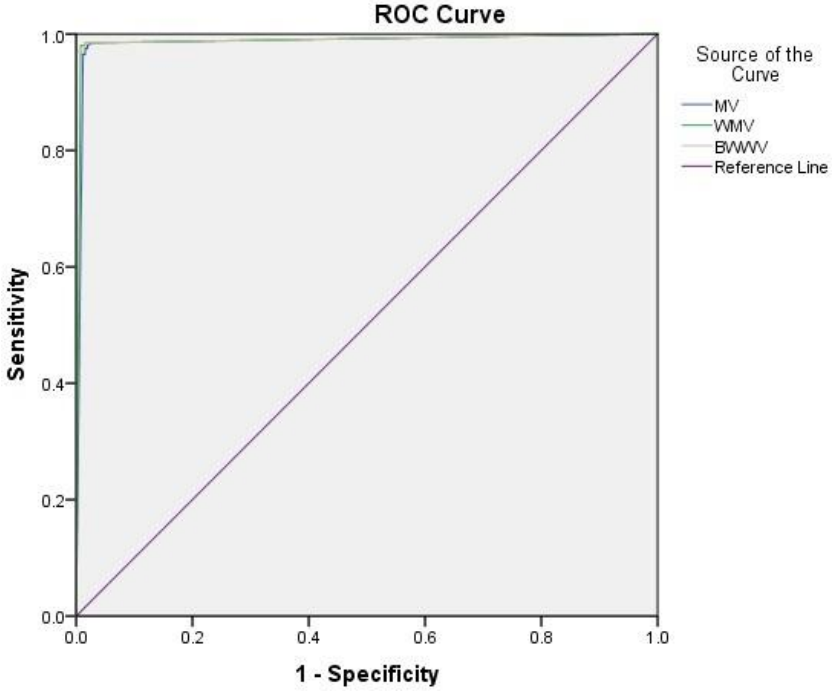

Figure 18: ROC Curve of three aggregation methods used

Table 11: AUC values of three aggregation methods used

Area Under the Curve

Test Result Variable(s)	Area
MV	.986
WMV	.988
BWWV	.992

The analysis of the experimental results led to the conclusion that the combination of classifiers using the BWWV method is preferable to the separate learning of classifiers, and that the combination of multimodal features has endorsed its effectiveness in classifying glaucoma.
The main advantage of this work is access to the design of a CAD system of glaucoma that outperforms other systems of literature; this thanks to the fusion of two types of characteristics (handcrafted characteristics represented by GLCM, Central Moments and Hu Moments, and the characteristics automatically generated by CNN) in a multimodal architecture.

Table 12 illustrates the main features suggested in the literature and presents a comparative study of the proposed method performance with the leading existing methods of glaucoma diagnosis in the literature.

Table 12: Comparative study on the performance of the proposed method compared to other work for the glaucoma diagnosis.

Authors	N° of Classes	Methods/Features	Classification method	Number of images	ACC(%)	SEN(%)	SPE(%)
Kolar et al. [18]	Tow (Normal/ Glaucoma)	Fractal Dimension (FD)	SVM	30	93.80	-	-
Noronha et al. [19]	Three (N, Mild G, severe G)	HOS cumulant features and LDA	Naïve Bayesian (NB)	272	92.65	100	92.00
Mookiah et al. [20]	Tow (Normal/ Glaucoma)	High Order Spectral (HOS) and wavelet features	SVM	60	95.00	93.33	96.67
Dua et al. [21]	Tow (Normal/ Glaucoma)	DWT and texture features	SMO	60	93.00	-	-
Bock et al. [24]	Tow (Normal/ Glaucoma)	Raw pixel intensities, FFT, B-spline and PCA	SVM	575	80.00	73.00	85.00
Raghavendra et al. [25]	Tow (Normal/ Glaucoma)	RT, MCT and GIST descriptor	SVM	1000	97.00	97.80	95.80
Nayak et al. [28]	Tow (Normal/ Glaucoma)	Morphological Discussed above	ANN	61	90.00	100	80.00
Acharya et al. [29]	Tow (Normal/ Glaucoma)	HOS and texture features	Random Forest	60	91.70	-	-
Singh et al. [31]	Tow (Normal/ Glaucoma)	wavelet features	Random Forest and ANN, SVM and k-NN	63	94.70	-	-

Maheshwari et al. [32]	Tow (Normal/ Glaucoma)	VMD, entropy and FD	LS-SVM	488	95.19	93.62	96.71
Kausu et al. [33]	Tow (Normal/ Glaucoma)	Morphological and wavele features	MLP	86	97.67	98.00	97.10
Chen et al. [38]	Tow (Normal/ Glaucoma)	Six layers CNN	-	1,676	-	-	-
Chai et al. [39]	Tow (Normal/ Glaucoma)	Two-branch CNN (five-conv layers)	-	3,554	81.69	-	-
Zilly et al. [40]	Tow (Normal/ Glaucoma)	CNN, Hough transform and entropy sampling	-	155	94.10	92.30	95.60
Orlando et al. [41]	Tow (Normal/ Glaucoma)	Eight layers CNN	-	101	-	-	-
Raghavendra et al. [42]	Tow (Normal/ Glaucoma)	Eighteen layers CNN	LDA	1,426	98.13	98.00	98.30
Deepti et al. [86]	Tow (Normal/ Glaucoma)	GLCM	ART/ANN	100	90.00	-	-
Kevin et al. [87]	Tow (Normal/ Glaucoma)	HOS	NB/SVM	272	92.60	100	92.00
Rajendra et al. [88]	Tow (Normal/ Glaucoma)	Gabor Filters	SVM	510	93.10	89.73	96.20
Nagarajan et al. [89]	Tow (normal/ Glaucoma	MVEP	ANN	399	94.00	95.00	94.00
Ashish et al. [90]	Tow (Normal/ Glaucoma)	CDR, NRR and blood vessels in ISNT quadrants	SVM and ANN	67	94.00	100	-
Simonthomas [91]	Tow (Normal/ Glaucoma)	Haralick texture	KNN	60	98.00	-	-
Issac et al. [92]	Tow (Normal/ Glaucoma)	morphological features (CDR, NRR, BV and IQ)	SVM and ANN	67	94.11	100	90.00
Acharya et al. [93]	Tow (Normal/ Glaucoma)	Gabor transform and entropy	SVM	510	93.10	89.75	96.20
Proposed Method	Tow (Normal/ Glaucoma)	Multimodal features fusion: GLCM, Central Moments, Hu Moments and CNN features	Multimodal Classifications fusion: BWWV CNN_{MLP}+SVM	455	99.78	99.50	100

Conclusion

Glaucoma is a tricky disease; it is the leading cause of visual impairment after age-related macular degeneration. Several studies have been conducted to develop tools to diagnose this disease.

In this study, we presented a multimodal classification method for diagnosing glaucoma using multimodal data and multimodal features from retinal fundus images, as a first work using the multimodality concept. This approach uses the hybrid fusion technique that we proposed, based on the deep convolutional neural networks (CNN) and SVM classifier.

One of the key benefits of using CNN is the elimination of traditional steps such as extraction and selection of features, by automatically generating the most informative features. The merge of the latter with the artisanal characteristics gave a highly robust and accurate system.

In the future, in order to durably exploit our system, it would be interesting that each well classified and approved image by the doctor will be integrated/added to the original learning base. In this case, the proposed system will be able to recognize new cases (unknown cases). We also want to invest in the representation of the feature map generated by CNN by looking for the best structuring of the new convolutional layers. Another interesting future direction would be the possibility of further improving our system by including other types of modalities, mainly topographic map images, which is a new representation of the retina, as well as textual information (concerning the patient: age, with or without: diabetes, myopia...).

In conclusion, the proposed method provides a more robust computer-aided diagnosis (CAD) system that enables ophthalmologists to quickly diagnose patients with glaucoma and is considered a second opinion by doctors with high accuracy.

References

1. Quigley, H.A., Broman, A.T., "The number of people with glaucoma worldwide in 2010 and 2020",Br. J. Ophthalmol. 90(3), pp. 262–267,2006
2. Doi, K., "Computer-aided diagnosis in medical imaging: historical review, current status and future potential",Comput Med Imaging Graph, 31,pp. 198–211,2007
3. Duerr, B., Haettich, W., Tropf, H., Winkler, G., "A combination of statistical and syntactical pattern recognition applied to classification of unconstrained handwritten numerals", Pattern Recognition,12(3), pp.189-199, 1980
4. Chou, Y.Y., Shapiro, L.G., "A hierarchical multiple classifier learning algorithm", Pattern Analysis and Applications, 6(2), pp 150–168, 2003
5. Azizi, N. ,Farah, N., Sellami, M., "Ensemble classifier construction for Arabic handwritten recongnition", 7th International Workshop on Systems, Signal Processing and their Applications, WoSSPA 2011, Tipâza, Algeria pp. 271-274
6. "A method of combining multiple experts for the recognition of unconstrained handwritten numerals", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17(1), pp.90-94, 1995
7. Brunelli, R., Falavigna, D., "Person identification using multiple cues",IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17(10), pp.955-966, 1995
8. Chibelushi, C.C., Mason, J., Deravi, F., "Integration of acoustic and visual speech for speaker recognition", Third European Conference on Speech Communication and Technology, Berlin, Germany, September 22-25, pp.175-160, 1993
9. Zemmal, N., Azizi, N., Dey, N., Sellami, M., "Adaptative Semi Supervised Support Vector Machine Semi Supervised Learning with Features Cooperation for Breast Cancer Classification",Journal of Medical Imaging and Health Informatics, 6(1): 53-62, 2016
10. Zemmal, N., Azizi, N., Sellami, M., Zenakhra, D., Cheriguene, S., Dey, N., Ashour, A.S.: "Robust feature selection algorithm based on transductive SVM wrapper and genetic algorithm: application on computer-aided glaucoma classification", IJISTA 17(3): 310-346 (2018)
11. Anushikha, S., Malay, K.D., Partha, S.M., Vaclav, U., Radim, B., "Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image",Computer Methods and Programs in Biomedicine 124: 108-120, 2016
12. Dey, N., Ashour, A. S., & Nguyen, G. N. Recent Advancement in Multimedia Content using Deep Learning
13. Benzebouchi, N.E., Azizi, N., Ayadi, K., "A Computer-Aided Diagnosis System for Breast Cancer Using Deep Convolutional Neural Networks," Proceedings of the 4thInternational Conference on Computational Intelligence in Data Mining, Advances in Intelligent Systems and Computing, Springer, In book: Computational Intelligence in Data Mining, No. 52, pp.583-593, 2019, DOI. 10.1007/978-981-10-8055-5_52
14. Benzebouchi, N.E., Azizi, N., Bouziane, S.E, "Glaucoma Diagnosis Using Cooperative Convolutional Neural Networks",International Journal of Advances in Electronics and Computer Science, vol.5(1), pp. 31-36, 2018
15. Li, Z., Dey, N., Ashour, A. S., Cao, L., Wang, Y., Wang, D., ... & Shi, F. (2017). Convolutional Neural Network Based Clustering and Manifold Learning Method for Diabetic Plantar Pressure Imaging Dataset. Journal of Medical Imaging and Health Informatics, 7(3), 639-652
16. Thakur, N., Juneja, M., "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma", Biomedical Signal Processing and Control, Vol.42, pp.162-189, 2018

17. Liu, M., Zhang, D., Chen, S., “Attribute relation learning for zero-shot classification”, Neurocomputing, vol.139,pp. 34-46, 2014
18. Zhang, D., Shen, D., “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease”, Neuroimage,vol.59, pp. 895-907, 2012
19. Kolar, R., Jan, J.: "Detection of Glaucomatous Eye via Color Fundus Images Using Fractal Dimensions", Radioengineering, 17(3), pp.109-114, 2008
20. Noronha, K.P., Acharya, U.R., Nayak, K. P., Martis, R.J., Bhandary, S.V., “Automated classification of glaucoma stages using higher order cumulant features,” Biomedical Signal Processing and Control, vol. 10, pp. 174–183, 2014
21. Mookiah, M.R.K., Acharya, U.R., Lim, C.M., Petznick, A., Suri, J.S., “Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features”, Knowledge-Based Systems, vol.33, pp. 73–82, 2012
22. Dua, S., Acharya, U.R., Chowriappa, P., Sree, S.V., ”Wavelet-based energy features for glaucomatous image classification”, IEEE Transactions on Information Technology in Biomedicine 16 (1) (2012) 80–87
23. Maheshwari, S., Pachori, R.B., Acharya, U.R., “Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images”, IEEE journal of biomedical and health informatics 21 (3) (2017) 803–813
24. Acharya, U.R., Bhat, S., Koh, J.E.W., Bhandary, S.V., Adeli, H., "A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images", Computers in Biology and Medicine, vol.88, pp. 72-8, 2017
25. Bock, R., Meier, J., Ny´ul, L.G., Hornegger, J., Michelson, G., “Glaucoma risk index: automated glaucoma detection from color fundus images”, Medical image analysis 14 (3) (2010) 471–481
26. Raghavendra, U., Bhandary, S.V., Gudigar, A., Acharya, U.R., “Novel expert system for glaucoma identification using non-parametric spatial envelope energy spectrum with fundus images”, Biocybernetics and Biomedical Engineering, 38(1), pp. 170-180, 2018
27. Zheng, Y., Essock, E.A., “A Novel Feature Extraction Method – Wavelet-Fourier Analysisand Its Application to Glaucoma Classification”,Proceedings of 7th Joint Conference on Information Sciences, Cary, North Carolina, pp. 672-675, 2003
28. Meier, J., Bock, R.,Nyúl, J., Michelson, G., “Eye Fundus Image Processing System for Automated Glaucoma Classification”, Proceedings of52nd International Scientific Colloquium. TechnischeUniversitätIlmenau, pp.81-85, 2007
29. Nayak, J., Acharya, U.R., Bhat, P.S., Shetty, N., Lim, T.C., “Automated Diagnosis of Glaucoma Using Digital Fundus Images”, Journal of Medical Systems, Springer, vol.33, pp. 337-346, 2009
30. Acharya, U.R., Dua, S., Du, X., Sree, V.S., Chua, K.C., “Automated diagnosis of glaucoma using texture and higher order spectra features”, IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 3, pp. 449-455, 2011
31. Kumbhare, P., Turkar, M., Kularkar, R., “Computer Aided Automatic Glaucoma Diagnosis”,International Journal of Electrical, Electronics and Data Communication, vol.2, pp. 28-32, 2014
32. Singh, A., Dutta, M.K., ParthaSarathi, M., Uher, V., Burge, R., “Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image”, Computer Methods and Programs in Biomedicine, vol.124, pp. 108-120, 2016
33. Maheshwari, S., Pachori, R.B., Kanhangad, V., Bhandary, S.V., Acharya, U.R., “Iterative variational mode decomposition based automated detection of glaucoma using fundus images”, Computers in Biology and Medicine, vol.88, pp. 142-149, 2017
34. Kausu, T. R., Gopi, V. P., Wahid, K. A., Doma, W., Niwas, S. I., "Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images", Biocybernetics and Biomedical Engineering, 38(2), pp. 329–341, 2018
35. Cun, Y.L., Boser, B., Denker, J.S., Howard, R.E., Habbard, W., Jackel, L.D., and al. : Advances in neural information processing systems 2. Citeseer. ISBN 1-55860-100-; (1990) 396–404
36. Ren, S., He, K., Girshick, R., Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks", in: Advances in neural information processing systems, pp. 91–99, 2015
37. Benzebouchi, N.E., Azizi, N., Aldwairi, M., Farah, N., “Multi-classifier system for authorship verification task using word embeddings”, 2018 2nd IEEE International Conference on Natural Language and Speech Processing (ICNLSP), IEEE, Algiers, Algeria, pp.1-6, 2018
38. Graves, A., Mohamed, A.-r., Hinton, G., “Speech recognition with deep recurrent neural networks”, in: Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, IEEE, 2013, pp. 6645–6649
39. Chen, X., Xu, Y., Wong, D.W.K., Wong, T.Y., Liu, J.,”Glaucoma Detection based on Deep Convolutional Neural Network”, 37th Annual International conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, pp.715-718, 2015
40. Chai, Y., He, L., Mei, Q., Liu, H., Xu. L., “Deep learning through two-branch convolutional neural network for glaucoma diagnosis”, International Conference on Smart Health (ICSH), pp. 191-201, 2017
41. Zilly, J., Buhmann, J. M., Mahapatra, D., “Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation”, Computerized Medical Imaging and Graphics 55 (2017) 28-41
42. Orlando, J.I., Prokofyeva, E., del Fresno, M., Blaschko, M.B., “Convolutional neural network transfer for automated glaucoma identification”, 12th International Symposium on Medical Information Processing and Analysis, Tandil, Argentina, vol. 10160, 2017
43. Raghavendra, U., Fujita, H., Bhandary, S.V., Gudigar, A., Tan, J.H., Acharya, U.R., “Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images”, Information Sciences 441 (2018) 41-49
44. Hagiwara, Y., Koh, J.E.W., Tan, J.H., Bhandary, S.V., Laude, A., Ciaccio, E.J., Louis, T., Acharya, U. R., “Computer-aided diagnosis of glaucoma using fundus images: A review”, Computer Methods and Programs in Biomedicine, 165, 1–12, 2018

45. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, vol. 88(11), pp. 2278–2324, 1998
46. Lahat, D., Adali, T., Jutten, C., "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects", Proceedings of the IEEE, 103(9), 2015
47. Zhang D., Wang Y., Zhou L., Yuan H., Shen D., “Multimodal classification of Alzheimer's disease and mild cognitive impairment”, Neuroimage, vol.55, pp. 856-867, 2011
48. Han, M., Liu, B., “Ensemble of extreme learning machine for remote sensing image classification”, Neurocomputing 149 (2015) 65–70
49. Mashhoori , A., “Block-wise two-directional 2DPCA with ensemble learning for face recognition”, Neurocomputing 108 (2013) 111–117
50. Tumer, K., Oza, N.C., "Input decimated ensembles", Pattern Analysis & Applications, 6(1), pp.65–77, 2003
51. Oza, N.C., Tumer, K., "Classifier ensembles: Select real-world applications", Information Fusion, 9(1), pp.4–20, 2008
52. Barandela, R., Sanchez , J. S., Valdovinos, R.M., "New Applications of Ensembles of Classifiers", Pattern Analysis & Applications, 6(3), pp.245–256, 2003
53. Tumer, K., Ghosh, J., "Classifier Combining: Analytical Results and Implications",13th National Conference on Artificial Intelligence, Portland, Oregon, August 1996
54. Porwik, P., Doroz, R., Wrobel, K., "An ensemble learning approach to lip-based biometric verification, with a dynamic selection of classifiers", Expert Systems With Applications (2018), https://doi.org/10.1016/j.eswa.2018.08.037
55. Rokach , L., “Pattern Classification Using Ensemble Methods”, first ed., World Scientific Publishing Company, Singapore, 2010
56. Zhou, Z.-H., “Ensemble Methods: Foundations and Algorithms,” first ed., Taylor & Francis, Boca Raton, FL, 2012
57. Martínez-Muñoz, G., Hernández-Lobato, D., Suárez, A., “An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation”, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 245–259
58. Nilsson, N.J., “Learning Machines: Foundations of Trainable Pattern-Classifying Systems”, McGraw-Hill, NY, 1965
59. Breiman, L., “Bagging predictors”, Machine Learning 24 (2) (1996) 123–140
60. Drucker, H., Cortes, C., Jackel, L.D., LeCun, Y., Vapnik, V., “Boosting and other ensemble methods”, Neural Computation 6 (6) (1994) 1289–1301
61. Perrone, M.P., Cooper, L.N., “when networks disagree: ensemble methods for hybrid neural networks”, in: Mammone, R.J. (Ed.), Neural Networks for Speech and Image Processing, Chapman-Hall, 1993, Chapter 10
62. Wolpert, D.H., “Stacked generalization”, Neural Networks 5 (1992) 241–259
63. Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., “Adaptive mixtures of local Experts”, Neural Computation, 3(1991) 79-87
64. Sezgin, M., Sankur, B., “Survey over image thresholding techniques and quantitative performance evaluation”, Journal of Electronic Imaging, vol.13(1), pp.146–168, 2004
65. Otsu, N.,“A Threshold Selection Method from Gray-Level Histograms”, IEEE Transactions on Systems, Man and Cybernetics, vol.9(1), pp. 62-66, 1979
66. Maji, P., Mullins, R., “On the Reduction of Computational Complexity of Deep Convolutional Neural Networks”, Entropy 2018, 20(4), 305, https://doi.org/10.3390/e20040305
67. Wen-Jie,W., Woo, K.M.,“Ultrasound Breast Tumor Image Computer-Aided Diagnosis With Texture and Morphological Features”, Academic Radiology,vol.15(7), pp. 873–880, 2008
68. Bob, Z, Xingzheng, W., Fakhri, K., Zhimin, Y.,David, Z., “Computerized facial diagnosis using both color and texture features”,Information Sciences, vol.221, pp. 49–59, 2013
69. Chang, J.J., Keun, H.K., Young,J.J., Jinsung, K., “Improving color and shape repeatability of tongue images for diagnosis by using feedback gridlines”,European Journal of Integrative Medicine, vol.6(3), pp. 328–336, 2014
70. Azizi, N., Tlili-Guiassa, Y., Zemmal, N., “A Computer-Aided Diagnosis System for Breast Cancer Combining Features Complementarily and New Scheme of SVM Classifiers Fusion”, International Journal of Multimedia and Ubiquitous Engineering, vol.8(4), pp. 45–58, 2013
71. Azizi, N., Zemmal, N., Sellami, M., Farah, N., “A new Hybrid Method Combining Genetic Algorithm and Support Vector Machine Classifier: Application to CAD system for mammogram images”,International Conference of Multimedia Computing and Systems, pp. 415-420, 2014
72. Shradhananda. B., Banshidhar, M., Ratnakar, D.,“Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer”,Neurocomputing, vol.154, pp. 1–14, 2015
73. Loris, N., Sheryl, B., Stefano, G., Emanuele, M., Tonya, B., “A comparison of methods for extracting information from the co-occurrence matrix for sub cellular classification”, Expert Systems with Applications, vol.40(18), pp. 7457–7467, 2013
74. Haralick, R.M., “Statistical and structural approachs to texture”, Proceedings of the IEEE, vol.67(5), pp.786–804, 1979
75. Manisha, V., Balasubramanian, R., “Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval”,Journal of Visual Communication and Image Representation, vol.32, pp. 224-236, 2015
76. Mellisa, P.A., Jeklin, H., Sakka, N.,“Mammograms Classification Using Gray-level Co-occurrenceMatrix and Radial Basis Function Neural Network”,Procedia Computer Science,vol.59, pp. 83-91, 2015
77. Luxin, Y., Mingzhi, J., Houzhang, F., Hai, L., Tianxu, Z.,“Atmospheric-Turbulence-Degraded Astronomical Image Restoration by Minimizing Second-Order Central Moment”, IEEE Geoscience and Remote Sensing Letters, vol. 9(4), pp. 672-676, 2012

78. Angshuman, P., Nilotpal, B., Ananda, S.C.,"Digit Recognition from Pressure Sensor Data using Euler Number and Central Moments", International Conference on Communications, Devices and Intelligent Systems, pp. 93-96, 2012
79. Saeed, K.P., Karim, F., Farshid, H.,"Face Detection Based on Central Geometrical Moments of Face Components",IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan,2006
80. Hu,M., "Visual Pattern Recognition by Moment Invariants", IRE Transactions on Information Theory, vol.8(2), pp. 179-187, 1962
81. Krawczyk, B., "Untrained weighted classifier combination with embedded ensemble pruning", Neurocomputing, vol.196, pp. 14–22, 2016
82. Cheriguene, S.,Azizi, N.,Farah, N., Ziani, A., "A two stage Classifier Selection Ensemble based on mRMR Algorithm and Diversity Measures", Computing Systems and ApplicationsConference, (2016)
83. Azizi, N., Farah, N., Sellami, M., Ennaji, A., "Using Diversity in Classifier Set Selection for Arabic Handwritten Recognition, Multiple Classifier Systems", International Workshop on Multiple Classifier Systems, pp. 235-244, 2010
84. Moreno–Seco, F., Iñesta, M.J., León, P.J., Micó, L., "Comparison of classifier fusion methods for classification in pattern recognition tasks", Lecture Notes in Computer Science, vol.4109, pp. 705–713, 2006
85. Onan, A., Korukoğlu, S., Bulut, H.,"A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification", Expert Systems with Applications, vol.62, pp. 1-16, 2016
86. Bradley, A.P., "The use of the area under the roc curve in the evaluation of machine learning algorithms",Pattern Recognition, vol.30(7), pp. 1145–1159, 1997
87. Deepti, Y., Partha, M.S., Malay, K.D., "Classification of Glaucoma Based on Texture Features Using Neural Networks",Seventh International Conference onContemporary Computing, 109–112, 2014
88. Kevin, P., Noronhaa, U., Rajendra, A.K., Prabhakar, N., Roshan, J., "Automated classification of glaucoma stages using higher order cumulant features", Biomedical Signal Processing and Control, vol.10, pp. 174–183, 2014
89. Rajendra, A., Lim, W.J., Kevin, P.N, Lim, C.M, Sulatha, V.B., "Decision support system for the glaucoma using Gabor transformation", Biomedical Signal Processing and Control, vol.15, pp. 18–26, 2015
90. Nagarajan, R., Balachandran, C., Gunaratnam, D., Klistorner, D., "Neural network model for early detection of glaucoma using multi-focal visual evoked potential (M-Vep)", Investigative Ophthalmology and Visual Science,vol.42, 2002
91. Ashish, I., Partha, M.S., Malay, K.D., "An adaptive threshold based image processing technique for improved glaucoma detection and classification", Computer Methods and Programs in Biomedicine, vol.122(2), pp. 229-249, 2015
92. Simonthomas, S., Thulasi, N., Asharaf, P., "Automated diagnosis of glaucoma using Haralicktexture features", IEEEInternationalConference onInformation Communication and Embedded Systems, pp. 1-6, 2014
93. Issac, A., Partha Sarathi, M., & Dutta, M. K., "An adaptive threshold based image processing technique for improved glaucoma detection and classification", Computer Methods and Programs in Biomedicine, 122(2), 229–244, 2015
94. Acharya, U.R., Ng, E.Y.K., Lim, W.J.E., Noronha, K.P., Lim, C. M., Nayak, K.P., Bhandary, S.V., "Decision support system for the glaucoma using Gabor transformation", Biomedical Signal Processing and Control, 15, pp. 18-26, 2015