Fully automated diagnosis of papilledema through robust extraction of vascular patterns and ocular pathology from fundus photographs

Khush Naseeb Fatima,1 Taimur Hassan,2 M. Usman Akram,1,* Mahmood Akhtar,1 and Wasi Haider Butt1

1Department of Computer Engineering, National University of Sciences and Technology, Pakistan
2Department of Electrical Engineering, Bahria University, Islamabad Pakistan
*usmakram@gmail.com

Abstract: Rapid development in the field of ophthalmology has increased the demand of computer aided diagnosis of various eye diseases. Papilledema is an eye disease in which the optic disc of the eye is swelled due to an increase in intracranial pressure. This increased pressure can cause severe encephalic complications like abscess, tumors, meningitis or encephalitis, which may lead to a patient’s death. Although there have been several papilledema case studies reported from a medical point of view, only a few researchers have presented automated algorithms for this problem. This paper presents a novel computer aided system which aims to automatically detect papilledema from fundus images. Firstly, the fundus images are preprocessed by going through optic disc detection and vessel segmentation. After preprocessing, a total of 26 different features are extracted to capture possible changes in the optic disc due to papilledema. These features are further divided into four categories based upon their color, textural, vascular and disc margin obscuration properties. The best features are then selected and combined to form a feature matrix that is used to distinguish between normal images and images with papilledema using the supervised support vector machine (SVM) classifier. The proposed method is tested on 160 fundus images obtained from two different data sets i.e. structured analysis of retina (STARE), which is a publicly available data set, and our local data set that has been acquired from the Armed Forces Institute of Ophthalmology (AFIO). The STARE data set contained 90 and our local data set contained 70 fundus images respectively. These annotations have been performed with the help of two ophthalmologists. We report detection accuracies of 95.6% for STARE, 87.4% for the local data set, and 85.9% for the combined STARE and local data sets. The proposed system is fast and robust in detecting papilledema from fundus images with promising results. This will aid physicians in clinical assessment of fundus images. It will not take away the role of physicians, but will rather help them in the time consuming process of screening fundus images.

© 2017 Optical Society of America

OCIS codes: (110.4500) Optical coherence tomography; (100.2960) Image analysis; (100.0100) Image processing; (170.4470) Ophthalmology; (100.5010) Pattern recognition.

References and links
1. J. R., Basic Ophthalmology, 4th ed., (New Delhi, 2009).
2. H. S. Nguyen, K. M. Haider, and L. L. Ackerman, “Unusual causes of papilledema: Two illustrative cases,” Surg. Neurol. Int. 4(60), 60 (2013).
3. C. L. Fraser, M. A. Ridha, V. Biousse, and N. J. Newman, “Vitreous Hemorrhage Secondary to Optociliary Shunt Vessels from Papilledema,” J. Neuroophthalmol. 32(4), 332–334 (2012).
4. L. Tang, R. H. Kardon, J.-K. Wang, M. K. Garvin, K. Lee, and M. D. Abràmoff, “Quantitative Evaluation of Papilledema from Stereoscopic Color Fundus Photographs,” Invest. Ophthalmol. Vis. Sci. 53(8), 4490–4497 (2012).
5. E. Z. Karam and T. R. Hedges, “Optical coherence tomography of the retinal nerve fibre layer in mild papilloedema and pseudopapilloedema,” Br. J. Ophthalmol. 89, 294–298 (2005).
1. Introduction

Rapid development in the field of ophthalmology has increased the demand of computer aided classifiers for various eye diseases. Papilledema is an eye disease in which the optic disc of the eye is swelled due to increase in intracranial pressure. It causes blurred vision, headaches and nausea in its initial stages. If it is left untreated, it can result in permanent loss of vision and it might lead to death in some cases. Papilledema occurs due to increase in intracranial pressure which happens due to increased blood pressure i.e. hypertension. Any patient showing papilledema from hypertension has malignant hypertension and should be considered as a medical emergency. Papilledema is not just caused by hypertension; there are several other causes too. It is well known that papilledema itself is not just a disease, but a symptom of several other diseases [1]. If papilledema is found in a patient, its underlying cause should be found as soon as possible, because higher stages of papilledema indicate that some serious disease (such as brain tumor or malignant hypertension) might be progressing in the patient. Therefore papilledema should be detected on time before it is too late to save the patient’s life [1].

Different signs of papilledema as seen on a fundus image include blurring of the optic disc boundary, appearance of a circumferential ring around the optic disc, dilation of blood vessels, appearance of hemorrhages or cotton wool spots, loss of major vessels as they leave the optic disc and blurring of the vessels on the optic disc [1]. Figure 1 shows a fundus image from STARE data set with papilledema. In the image, it can be seen that the boundary of optic disc is blurred and not quite distinguishable. The blood vessels are also dilated while some vessels are blurred.
Fig. 1. Fundus image with papilledema: It can be noted that the optic disc boundary is blurred and blood vessels are dilated.

In literature, there have been several papilledema case studies carried out from medical perspective, where papilledema was found along with presence of some life threatening diseases, thus helping doctors in their diagnosis. It is recommended that every patient of papilledema should be treated very carefully. Generally, a computed tomography (CT) scan, magnetic resonance imaging (MRI) scan or lumbar puncture are used to detect the real cause of papilledema but these are not very viable sources. Neurosurgical operations also provides a convenient way to cure papilledema however it also fails in majority of the cases, as described in [2]. The best way to detect papilledema is through thorough physical examination and maintaining the patient checkup history. Therefore, it is of grave importance to do a complete physical examination and get a thorough history of the patient to diagnose the underlying cause of papilledema and not just undergo neurosurgical operations for treating papilledema. Papilledema can be easily diagnosed through regular checkups [3].

Some of researchers have worked on the automated detection of papilledema using fundus and OCT images. The major sign of papilledema is optic disc swelling as shown in Fig. 1 and automated detection of papilledema mainly focuses on the analysis of optic disc. L. Tang et al. used digital stereoscopic color fundus images to evaluate volume of the optic disc in [4]. By measuring optic disc volume, papilledema can be diagnosed and managed because increase in volume would indicate presence of papilledema. The authors have used a total of 29 sets of fundus photographs along with optic nerve head centered spectral domain optic coherence domain tomography scans (SD-OCT scans) obtained from fifteen patients having papilledema. In order to estimate the 3-dimensional shape of the optic nerve head, an automated multi-scale stereo correspondence algorithm is used. This algorithm automatically finds dense association between the stereo images giving a stereo derived depth map which is further interpolated to smooth all the noise present using a thin plate spine (TPS). For determining the change in volume of the optic nerve head using OCT, an automated 3-D segmentation technique was applied which segments the OCT volume. Finally the correlation between SD-OCT volume and stereo volume measurements was assessed through a reference plane which also helps in grading papilledema using Friesen scale. The volumetric measurements of retinal surface elevation found from both stereo fundus photographs and OCT scans are positively correlated (correlation coefficient r2 = 0.60; P < 0.001). Therefore, it is concluded that the elevation of the retinal surface in patients suffering from papilledema are extracted from stereoscopic fundus photographs and it correlates well with the results that are obtained from OCT scans. Stereoscopic color imaging of the optic nerve head together with some method of automated shape reconstruction can prove to be a low-cost substitute of SD-OCT scans that has potential for a more cost-effective diagnosis and management of papilledema.

Another method to determine the degree to which OCT can discriminate variation in the retinal nerve fiber layer thickness (RNFL) between healthy eyes, eyes with mild papilledema
and eyes with pseudo-papilledema has been reported in [5]. The data set contained a total of
41 subjects out of which 17 were normal (healthy) people, 11 people had congenitally
crowded optic nerves and 13 people had mild papilledema. All of these subjects went through
fundus photography, complete neuro-ophthalmic examination and testing of visual fields.
Among these patients, spinal fluid pressure measurements were calculated in 5 patients of
pseudopapilledema and eleven patients of mild papilledema. The OCT scans were made on
both eyes of all the subjects. These scans were circular in shape and had 3.38mm diameter
surrounding the optic disc. The swelling in the fundus images (papilledema) was graded by 2
experts after analyzing the fundus photographs. Authors reported that the RNFL thickness in
superior and inferior parts was greater showing more correlation between all groups of
subjects. The mean RNFL thickness in normal subjects and patients with papilledema was
significantly different however it showed no significant difference between patients with
congenitally crowded optic nerves and those having papilledema. This study concluded that
although OCT gave statistical difference in RNFL thickness between patients of papilledema,
pseudo-papilledema and normal people, it did not distinguish patients with papilledema from
those having congenitally crowded optic nerves.

Apart from this, an automated system to monitor different stages of papilledema has been
reported in [6]. Authors extracted 24 features based on different signs of papilledema. The
algorithm was applied on a local data set consisting of images with papilledema of grades 0 to
4 taken from 39 patients showing progression of papilledema for over 2 years. For
classification, the decision tree methodology was used. However, no work was done to
classify grade 5 papilledema cases. The algorithm was evaluated by calculating Cohen’s
weighted k coefficient which tells if the result from algorithm is in agreement with ground
truth. The value of k varies from −1 to 1 where −1 means the results are in complete
disagreement and 1 means the results are in complete agreement. Out of 100 images, 64
images were correctly classified into their respective grades of papilledema. Another method
for classifying papilledema from fundus images employs 13 different features [7]. This
method gives an accuracy of 96.67%, when applied to 30 images from the STARE data set.
One of the biggest limitations of this algorithm is that it does not use any vascular features or
other features specifically related to blurring of optic disc boundary. Previously we have
presented an automated method to detect papilledema from fundus images by extracting 10
different vascular and textural features in [8] but here, we propose a fully automated decision
support system by incorporating different combinations of color, texture, vessel and disc
margin based features for automated diagnosis of papilledema from fundus images. We have
also validated our proposed system on more images from STARE as compared to [8],
particularly the proposed system was tested on 160 fundus images in which 90 of them are
from STARE data set and 70 of them are from our local data set that is acquired from AFIO.
The rest of this paper is structured as follows. Section 2 gives an in-depth description of our
proposed methodology. Section 3 demonstrates our results, followed by Section 4 which
discusses the evaluation extracted feature set. Section 5 depicts the overall results followed by
Section 6, which gives a discussion on our proposed implementation and outlines conclusions.

1.1 Data set description

There are two data sets that have been used in this research. The first one is the online STARE
data set and the second one is our custom data set acquired from AFIO. STARE data set has
annotations for optic nerve (ON) swelling [17]. These images are further re-annotated with
the help of two ophthalmologists. These new annotations were made by the ophthalmologists
without any prior information about annotations from STARE database. Apart from this, a
local data set was annotated by two ophthalmologists working in AFIO. We considered a
subset of 90 and 70 images from STARE and local data set respectively for which both
ophthalmologists gave same annotations. The annotations have been carried out using
Modified Friesen Scale [18]. Modified Friesen Scale defined 6 grades for papilledema,
however, in this research we are only focusing on two i.e. normal and papilledema. The images are not further graded into sub grades of papilledema. Table 1 shows original Modified Friesen Scale [18].

| Papilledema Grade | Normal optic disc | Minimal degree of edema | Low degree of edema | Moderate degree of edema | Marked degree of edema | Severe degree of edema |
|-------------------|-------------------|-------------------------|--------------------|-------------------------|------------------------|------------------------|
| 0                 | Normal optic disc | Prominence of the retinal nerve fiber layer at the nasal, superior and inferior poles in inverse proportion to disc diameter. Radial nerve fiber layer striations, without tortuosity | C-shaped halo that is subtle and grayish with a temporal gap; obscures underlying retinal details * | Disruption of normal radial nerve fiber layer arrangement striations | Temporal disc margin normal | |
| 1                 | Minimal degree of edema | | Elevation (nasal border) | No major vessel obscuration | | |
| 2                 | Low degree of edema | | Circumferential halo | Elevation (nasal border) | | |
| 3                 | Moderate degree of edema | | Obscuration of ≥1 segment of major blood vessels leaving disc | Circumferential halo | Elevation (all borders) | Halo (irregular outer fringe with finger-like extensions) |
| 4                 | Marked degree of edema | | Total obscuration on the disc of segment of a major blood vessel on the disc | Elevation (whole nerve head, including the cup) | Border obscuration (complete) | Halo (complete) |
| 5                 | Severe degree of edema | | Obscuration of all vessels on the disc and leaving the disc | | | |

* Key features (major findings) for each grade.

Table 2 gives a summary of different data sets used in this study.

| Data set | Total images | Normal images | Abnormal images |
|----------|--------------|---------------|-----------------|
| AFIO     | 70           | 40            | 30              |
| STARE    | 90           | 50            | 40              |
| Combined | 160          | 90            | 70              |

2. Proposed methodology

We propose a fully automated system to detect papilledema from fundus images. First of all, the input fundus images are preprocessed. After preprocessing, several features are extracted from each candidate image. Out of those features, the optimal features are selected and combined to form a feature matrix to be used for classification between normal and papilledema images. Figure 2 shows a block diagram of our proposed system.
2.1 Preprocessing

The preprocessing is the first step in our proposed system for the localization of optic disc region and for the extraction of blood vessels from input fundus image. As we know, papilledema causes swellings within the optic disc region, also shown in Fig. 3, so rather than processing the whole fundus image, the proposed system only extracts the specific region of interest (ROI) that depicts swelled optic disc and blood vessels.

Since optic disc is a bright and circular region from which all blood vessels emerge, these properties (intensity and vessel density) can be used to determine its position [9]. First of all, vessels are enhanced for which 2-D Gabor wavelet is used due to its ability to enhance directional structures. After vessel enhancement, vessel segmentation is performed through a multilayered thresholding technique [10]. The reason for using segmentation method
proposed in [10] is because it is quite robust to noise and it can even detect small vascular patterns with an average accuracy of 94.85% [10]. The images are then cropped such that the resulting image has optic disc and a small region surrounding the optic disc (2-disc diameter). Apart from this, we have also extracted the optic disc boundary by removing blood vessels around the optic disc region using the method which we proposed in [11] and the boundary of OD is smoothened using ellipse fitting. The extracted disc boundary plays a vital role in extracting color features as described in section 2.2.1 below.

2.2 Feature extraction

The preprocessing step results into our required ROI which includes optic disc and a small region containing optic disc as well as vascular segmentation of this region. Figure 4 shows the ROI that can be used to extract different representative features.

![Fig. 4. The ROIs obtained because of preprocessing step (a) original fundus image (b) ROI after optic disc localization (c) ROI after vessel segmentation.](image)

The feature extraction process is divided into four sub-steps, which include:

- **Color** features extraction, such as maximum or minimum pixel values.
- **Texture** features extraction, such as contrast, correlation and standard deviation values.
- **Vascular** features extracted from the vascular structured ROI including maximum area and minimum area values.
- **Disc margin obscuration** features extraction, like nasal to temporal ratio and maximum nasal to temporal ratio values.

2.2.1 Color features

In case of normal images, the optic disc boundary is sharp, bright and clearly visible while for the images with papilledema the optic disc boundary becomes dull, less bright and blurry. Therefore, as papilledema advances to higher grades, the RGB values of the pixels on the
optic disc starts to vary. The color subset in feature extraction calculates these varied values. A total of 3 features are extracted from the color images. For these 3 features, the ROI obtained after optic disc localization is directly used whereas for rest of the subsets, the image is first converted into gray-scale image. The features that have been used from this color image include maximum, minimum and mean pixel value. In order to calculate these values, 8 pixels, at an equal distance on the optic disc boundary are automatically selected. Figure 5 shows selected pixel values on a normal image and on an image with papilledema. The quantitative values for colored features against these two images are also given in Fig. 5.

![Selected pixels and their respective values on disc boundary of (a) normal image, and (b) image with papilledema.](image)

Note that the selected pixel values are all lying on the optic disc boundary. This is done to exploit the fact that the images with papilledema would have blurred disc margins. As one can see that the boundary of image with papilledema is less sharp as compared to that of a normal image without papilledema. Therefore, calculating pixel values on the boundary gives good information for detecting papilledema. To automatically select the 8 pixels, we first extracted the center of the localized optic disc region. Then with the radial distance from the center, we have calculated 8 pixels with equal interval between 0 to $2\pi$. The radial distance is determined by finding the difference between the boundary of optic disc from its center as shown in Fig. 5. Since the optic disc is not entirely circular so that’s why some of the boundary pixels do not honor the length of the radial line. In that case, we picked those boundary pixels which lies on the same orientation as radial line and are closest to the radial lines as shown in Fig. 6.
From these selected pixels, we calculated the following three features.

1. **Maximum pixel value**: Out of all selected eight pixels, three maximum pixel values are selected which gives the maximum intensity information of R, G and B channel respectively. The maximum value out of those three R, G and B values is then selected to be used as our first color based feature.

2. **Minimum pixel value**: From the manually selected 8 pixels, minimum pixel value is selected and once again minimum out of the 3 R, G and B values is chosen as the second color based feature.

3. **Mean pixel value**: The mean of all manually selected 8 pixel values is calculated and then the mean of the obtained R, G and B values is calculated to be the third color based feature.

### 2.2.2 Texture features

Texture is one of the most important characteristics which can be used to identify different objects or region of interests. In any image, the texture gives a great deal of information such as contrast, regularity or uniformity, etc. present in that image. Texture of an image is associated with the spatial distribution of its intensity values. Different methodologies can be used to extract a significant number of quantitative texture features from images which can then be used to obtain further classification results [12]. In a fundus image, as the swelling of the optic disc increases due to papilledema, its peripapillary texture properties changes [6]. These textural variations are due to abnormal optic disc pathology resulted from the spread of papilledema. Based on this motivation, we have extracted some of the useful textural features that contributes positively in discriminating the diseased and healthy pathology.

The first four textural features subset have been extracted using gray-level co-occurrence matrix (GLCM) of the image. The GLCM deals with statistical properties of spatial distribution of image gray level values. In a gray-scale image, each pixel has some intensity value called the gray level. The GLCM tabulates how frequently different gray-level combinations co-occur in the image or a particular section of the image [13]. In our proposed implementation, we have extraction following texture based features:

1. **Contrast**: It is a measure of intensity contrast between a pixel and its neighbor over the entire image. The contrast of the image would be very low if the neighboring
pixels are too similar in their grey level. If the image is constant, its contrast will be equal to 0. As we know that papilledema causes swelling in the optic disc region which leads to the blurriness. This blurriness causes the notable decrease in the contrast as compared to the healthy fundus scan.

2. **Correlation:** It measures how much a pixel is correlated to its neighbor pixel over the full image. Since papilledema causes a decrease in the inter-pixel variability (contrast) so it will produce a strong correlation between a pixel and its neighbors especially near the optic disc region.

3. **Energy:** It measures local homogeneity in an image and tells us about the uniformity of the texture. As the homogeneity of the texture increases, its energy value also increases. A constant image has energy value equal to 1. In the case of papilledema, the uniformity within the local neighborhood will be greater so we will have more energy as compared to the healthy one.

4. **Homogeneity:** It determines the uniformity of non-zero entries in the GLCM. If the GLCM concentrates along the diagonal then the homogeneity of the texture is high which indicates that there are several pixels having same or quite similar gray level values. However, if there are bigger changes in the value of gray levels then the homogeneity will be lower and the GLCM contrast will be higher. The value of homogeneity ranges from 0 to 1. In case of no variation in the image, the GLCM homogeneity will be equal to 1. Therefore, high homogeneity refers to textures that contain ideal repetitive structures, while low homogeneity refers to big variation in both texture elements and their spatial arrangements. An inhomogeneous texture would refer to an image that has almost no repetition of texture elements indicating an absence of the spatial similarity [14]. So, the GLCM of papilledema affected scans will consist of pixel entries near the diagonal as we have less contrast (inter-pixel variability) and high homogeneity. But for healthy images, the homogeneity will be low and GLCM will have sparse entries.

5. **Entropy:** It is a measure of randomness in the image. For papilledema affected images, as we have fine textural representation so the entropy would be low whereas healthy scans will have coarse texture that would produce a high entropy rate.

6. **Standard Deviation:** The square root of variance is called standard deviation. It shows how much the data varies from its mean value. If the value of standard deviation is low, it would imply that the data is very close to its mean, whereas if the value of standard deviation is high it would imply that the data is wide spread. As we have low contrast scans in the case of papilledema so we will also have low value of standard deviation whereas for the healthy scan, the value of standard deviation would be high.

7. **Maximum Value of Histogram:** It shows the pixel's distribution of the image in the gray level scale. It may be visualized as if each pixel is placed in a bin matching the intensity of that pixel. Subsequently, all pixels in every bin are added up and showed in the form of a bar chart called histogram. For a gray-scale image, the total number of bins is 256. We have used the maximum value of this gray-scale histogram as one of our feature. The value for this feature will be low for the scans that have papilledema syndromes because the histogram spread will be uniformly distribution with almost similar probabilities for each pixel whereas for the healthy case, this feature will have a higher value due to non-uniformity.

8. **Cluster Shade:** It measures the skewness i.e. how much the image lacks in symmetry. A high value of cluster shade indicates the image is less symmetric and a low value
indicates that there is little variation in the gray levels of the image [15]. To calculate this feature, the image is divided into two clusters. The mean gray level value of both clusters is then calculated and finally their absolute difference is taken. Since papilledema scans have symmetric histogram due to low variability so the value of cluster shade would be much lower as compared to the healthy scans which will have skewed histogram profile.

9. **Profile:** It calculates the intensity values along a particular path. It selects equally spaced points along the specified path, and then uses interpolation to find the intensity value for each point. As in papilledema, the boundary of the optic disc is blurred; the profile along a path cutting the optic disc gives different values as compared to the healthy profile as shown in Fig. 7(c), 7(d). We have used a single path for each image by drawing a line which cuts the optic disc. Figure 7 shows the path selected to calculate profile on a healthy fundus image and on an image with papilledema. It also shows a plot of the intensity versus the distance values along the specified path. Profile based feature is the range of intensity values which is the difference of maximum and minimum intensity values present in the profile. In papilledema case, the intensity difference between the optic disc and its surrounding region is low due to blurring and swelling as compared to a normal case.

![Fig. 7. The selected profile paths in case of (a) healthy fundus image, and (b) image with papilledema, and their computed profiles shown respectively in (c) and (d).](image)

### 2.2.3 Vascular features

When papilledema occurs, it affects the vascular structure of the fundus image. As the disease progresses, the vascular properties vary substantially. Vessels become obscured, dilated and in some cases tortuosity of vessels occurs. With the increase in the grade of papilledema, the obscuration of major vessels also increases making them less defined. Keeping in mind these properties, several vessel features have been extracted from fundus images which are [6]:

1. **Maximum vascular area:** The image obtained after vessel segmentation has been used to compute area for each connected region. The maximum area would be considered as our first vessel feature.
2. Minimum vascular area: The second feature is the minimum area in the vascular segmented image.

3. Mean vascular area: The mean of areas of all connected regions in the vascular segmented image is used as our third feature. It is calculated by adding area of all the regions and dividing it by total number of regions in the image.

4. Standard deviation of vascular areas: The standard deviation of all areas on the regions in the vascular segmented image is our fourth feature.

5. Kurtosis of vascular areas: It measures the flatness of a given distribution. The value of kurtosis is 3 for a normal distribution and > 3 for a distribution with outliers.

6. Vessel discontinuity index: In papilledema, the visual permanence along the length of major vessels is decreased as the blood vessels become obscured. The blood vessels then start appearing as discontinuous segments rather than their original shape. Vessel discontinuity index (VDI) is the feature that can quantify the connectivity alongside the extent of veins and arteries in the fundus image [6]. To calculate VDI, the number of disjoint regions in the vascular segmented image is counted. With the increase in vessel obscurity, more gaps can be seen in the vessel mask thus indicating that the value of VDI is directly proportional to the swelling of optic disc. If the swelling is more advanced, the value of VDI will be higher. Figure 8 shows fundus images from our local data set, vessel segmentations of these fundus images and their corresponding VDI. It can be noted that as papilledema progresses, the vessels become more disconnected and thus their VDI also increases.

2.2.4 Disc margin obscuration features

The optic disc can be divided into four equal poles of 90°. They are called superior, nasal, inferior and temporal poles. The nasal and temporal poles are further subdivided into three equal segments of 30° each which are called superior, central and inferior segments of nasal or temporal poles. Figure 9 shows the division of optic disc into poles and the subdivision of nasal pole. When the optic disc is swollen due to papilledema, optic disc (OD) margin gets
blurry. As the swelling progresses, so does the obscuration of OD margin. According to modern Friesen scale [16], the nasal pole is first to start becoming blurred. In a higher degree of papilledema, the temporal pole also becomes obscured along with the nasal pole. In case of extreme papilledema, the obscuration progresses all around the optic disc. The superior and inferior poles have large variability in the level of obscuration of disc margin due to which they are not used. Based on these properties, the disc margin obscuration features are extracted. Although Echegaray et al. [6], have proposed disc margin obscuration features but we have defined the disc margin obscuration features in a different way. Authors in [6] measure the blurriness in different sectors of optic disc by calculating standard deviation of the radial distance from the center of the optic disc to each point on its margin (where the change in pixel intensity was greatest every 2°). In our algorithm, we have used mean intensity value of gradient magnitude in different sectors of optic disc for calculating the disc margin obscuration features which are:

1. Superior nasal to temporal ratio: It gives the ratio of mean intensity of the magnitude of gradient in the superior nasal segment of the optic disc to the mean intensity of the magnitude of gradient in the temporal pole of the optic disc.

2. Central nasal to temporal ratio: It gives the ratio of mean intensity of the magnitude of gradient in the central segment of nasal pole to the mean intensity of the magnitude of gradient in the temporal pole of the optic disc.

3. Inferior nasal to temporal ratio: It gives the ratio of mean intensity of the magnitude of gradient in the inferior nasal segment of the optic disc to the mean intensity of the magnitude of gradient in the temporal pole of the optic disc.

4. Nasal to superior temporal ratio: It gives the ratio of the mean intensity of the magnitude of gradient in the nasal pole of optic disc to that of superior segment of temporal pole of the optic disc.
5. **Nasal to central temporal ratio:** It gives the ratio of the mean intensity of the magnitude of gradient in the nasal pole of the optic disc to that of the central temporal pole of the optic disc.

6. **Nasal to inferior temporal ratio:** It gives the ratio of the mean intensity of the magnitude of gradient in the nasal pole to that of inferior temporal pole of the optic disc.

7. **Maximum nasal to temporal ratio:** It gives the ratio of the maximum intensity of the magnitude of gradient in the nasal pole to the mean intensity of the magnitude of gradient in the temporal pole of the optic disc.

8. **Nasal to temporal ratio:** It gives the ratio of mean intensity of the magnitude of gradient in the nasal pole to the mean intensity of the magnitude of gradient in the temporal pole of the optic disc. Table 3 gives summary of all extracted features.

| Number | Feature Name                              | Feature Group               |
|--------|-------------------------------------------|----------------------------|
| F1     | Maximum pixel value                       | Color                      |
| F2     | Minimum pixel value                       | Color                      |
| F3     | Mean pixel value                          | Color                      |
| F4     | Contrast                                  | Texture                    |
| F5     | Correlation                               | Texture                    |
| F6     | Energy                                    | Texture                    |
| F7     | Homogeneity                               | Texture                    |
| F8     | Entropy                                   | Texture                    |
| F9     | Standard deviation                        | Texture                    |
| F10    | Maximum value of histogram                | Texture                    |
| F11    | Cluster shade                             | Texture                    |
| F12    | Profile                                   | Texture                    |
| F13    | Maximum vascular area                     | Vascular                   |
| F14    | Minimum vascular area                     | Vascular                   |
| F15    | Mean vascular area                        | Vascular                   |
| F16    | Standard deviation of vascular areas      | Vascular                   |
| F17    | Kurtosis of vascular areas                | Vascular                   |
| F18    | Vessel discontinuity index                | Vascular                   |
| F19    | Superior nasal to temporal ratio           | Disc margin obscuration    |
| F20    | Central nasal to temporal ratio            | Disc margin obscuration    |
| F21    | Inferior nasal to temporal ratio           | Disc margin obscuration    |
| F22    | Nasal to superior temporal ratio           | Disc margin obscuration    |
| F23    | Nasal to central temporal ratio            | Disc margin obscuration    |
| F24    | Nasal to inferior temporal ratio           | Disc margin obscuration    |
| F25    | Maximum nasal to temporal ratio            | Disc margin obscuration    |
| F26    | Nasal to temporal ratio                   | Disc margin obscuration    |
2.3. Feature selection

All extracted 26 features were combined to form a feature vector. However, all of these features were not used directly for classification. The optimum features were selected through Wilcoxon rank sum test so that redundant, irrelevant or noisy features responsible in reduction of performance of proposed system can be removed. The Wilcoxon rank sum test analyzes whether the normal or abnormal class median values present in the data set differ significantly or not. If the features do not have different values of median for both classes then those features are considered less effective for classification. Table 4 shows the results obtained after applying Wilcoxon rank sum test to our original set of 26 features. The features are written in descending order of their absolute score. The features become less significant as their absolute score decreases and the p-value increases. The top ten features obtained according to this test are features F15, F16, F7, F25, F12, F13, F4, F11, F2 and F18 (i.e., representative mean area of all regions in vascular map, STD of areas in vascular map regions, homogeneity, maximum nasal to temporal ratio of OD boundary obscuration, profile value, maximum area in the vascular map regions, contrast, cluster shade, minimum pixel value of the colored fundus images and VDI features).

Table 4. Performance of features calculated using Wilcoxon rank sum test

| Feature | Absolute Score | p-value |
|---------|----------------|---------|
| F15     | 4.373          | <10^{-5}|
| F16     | 4.361          | <10^{-5}|
| F7      | 4.266          | <10^{-5}|
| F25     | 4.243          | <10^{-5}|
| F12     | 4.160          | <10^{-5}|
| F13     | 4.148          | <10^{-5}|
| F4      | 4.136          | <10^{-5}|
| F11     | 3.139          | 0.0017  |
| F2      | 3.128          | 0.0018  |
| F18     | 2.830          | 0.0047  |
| F3      | 2.285          | 0.0223  |
| F9      | 2.047          | 0.0406  |
| F8      | 1.881          | 0.0599  |
| F6      | 1.727          | 0.0842  |
| F20     | 1.596          | 0.1104  |
| F17     | 1.335          | 0.1818  |
| F26     | 1.335          | 0.1818  |
| F21     | 1.276          | 0.2020  |
| F23     | 1.252          | 0.2106  |
| F22     | 1.192          | 0.2329  |
| F1      | 1.016          | 0.3097  |
| F24     | 0.943          | 0.3454  |
| F5      | 0.884          | 0.3766  |
| F14     | 0.840          | 0.4007  |
| F6      | 1.727          | 0.0842  |
| F19     | 0.813          | 0.4163  |
| F10     | 0.552          | 0.5811  |
2.4. Classification

After feature selection, the selected features were combined to form a feature set which is then used for classification between normal and papilledema cases using the SVM classifier. We chose SVM as a classification algorithm in our proposed implementation because it is fast and requires less time for training as compared to other supervised classifiers. Also, the major benefit of SVM is that it requires only its support vectors to hold a decision boundary instead of storing all of its training samples, which makes it memory efficient as well. Also it can easily classify non-linearly separable samples through the use of its non-linear hyper-plane kernels. In order to perform classification, the data set was first divided into training and testing samples such that the training samples were used to train the classifier while the testing samples were used to test the output of the classifier. It is essential that the training samples must not have any information about the testing samples. If the training and testing samples are not divided properly then it is possible that same samples can be used for training as well as testing thus giving overly optimistic or rather invalid results. We used 10-folds cross validation i.e. whole data set was divided into 10 equal folds or subsets having samples of both classes. Out of 10 folds, 9 were used for training the classifier while the remaining one fold was used for testing. This process was repeated 10 times so that each time a different fold was used for testing. Thus after obtaining results for 10 times, their average was taken and used as the final classification result.

3. Results

The proposed algorithm was tested on two different data sets including a local data set and an online data set. First data set that we have used contains images which are locally collected from AFIO, Rawalpindi, Pakistan. This data set contains a total of 70 RGB fundus images including 40 normal images and 30 abnormal images (images with papilledema). All these images are of 1504 × 1000 resolution. The second data set that we used for our evaluation is a publically available data set called STARE [17]. It contains approximately 400 RGB fundus images of 700 × 605 resolution. However, out of those 400 images, only 10 images with optic disc swelling (papilledema) are present in the data set. Since we wanted to have a bigger data set to test our proposed methodology, we needed more images with papilledema. To compensate this shortage of fundus images with papilledema, we obtained another 30 papilledemic images from online and these images were annotated by multiple expert ophthalmologists working in AFIO.

3.1 Group-wise features evaluation

We divided the extracted features into 4 groups or categories namely color, texture, vascular and disc margin obscuration features. In order to evaluate these groups, we performed group-wise classification using all possible combination of different groups. Therefore, for 4 groups, total 14 possible group-wise combinations were used. The evaluation results on the combined data set are shown in Table 5.
It can be observed from the results table that the group which performs well individually as compared to other groups was ‘vascular features’ giving an accuracy of 75.9%, followed by ‘color features’ with accuracy of 72.3%. One can also note that the worst classification results in case of an individual feature set were obtained for the ‘disc margin obscuration features’ with an accuracy of 59.2%. We report that the best results were obtained in case of combined color, texture and disc margin obscuration features, giving an accuracy of 87.8%, specificity of 90.6% and sensitivity 84.1%.

It can be further noted that the color features when combined with other groups increase the accuracy of that group significantly. For example, the accuracy in case of texture features alone was 65.6% which was increased to 79.8% when combined with color features. Similarly the accuracies in case of vascular features and disc margin obscuration features were increased respectively from 75.9% to 84.3% and from 59.2% to 72.9% after combining them with color features. Similar to color features, the vascular features also seem to increase the accuracy of groups. The accuracy for texture features was increased from 65.62% to 80.9% by combining them with vascular features. Similarly the disc margin obscuration had an increase in accuracy from 59.2% to 78%. The color and vascular features when combined together gave an accuracy of as high as 84.34% (which is the second best out of all group-wise combinations and the best of all 2 group combinations). It can be concluded that the most significant groups are color features and vascular features.

### 4. Individual features evaluation

All features were individually evaluated using Wilcoxon rank sum (Table 4) and top features were used one by one to study the effect of number of features on accuracy, sensitivity and specificity. The results were summarized in the form of a graph indicating this effect, as shown in Fig. 10.
The best percentage accuracy, sensitivity, and specificity results (respectively 87.4%, 88.7%, and 85.72%) were obtained using top 12 features. An abrupt change in all 3 performance measures can be noticed at top 4 features; otherwise the graph remains rather smooth from top 5 features to top 10 features.

5. Overall results

In the end, the proposed method was evaluated on all databases. Table 6 gives summary of our evaluation results.

| Data set | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|----------|--------------|-----------------|-----------------|
| AFIO     | 87.41        | 88.69           | 85.72           |
| STARE    | 95.62        | 93.74           | 97.96           |
| Combined | 85.89        | 83.94           | 88.39           |

It can be seen that although same methodology was applied to AFIO and STARE data set and the ratio of images with and without papilledema were same for both cases, but the performance measures for STARE are more accurate than AFIO. It is so because the AFIO data set contained images that were mostly of mild papilledema cases (grade 1, 2, 3), whereas in case of STARE the images were mostly of severe papilledema cases (grade 4, 5). Apart from this, we have extended our algorithm originally proposed in [7] to automatically diagnose and grade papilledema from digital retinal images and we have also compared our
proposed system with the state of the art solution proposed in [6] on STARE data set and we obtained better performance in terms of accuracy as shown in Table 7.

| Method           | Accuracy (%) |
|------------------|--------------|
| Proposed system  | 95.62        |
| Proposed in [6]  | 94           |
| Improvement      | 1.62         |

6. Discussion and conclusion

This paper has proposed an automated system for detecting papilledema from fundus images. The input images were first preprocessed (i.e., including optic disc localization, images cropping to focus on optic disc and a small surrounding portion, and vessel segmentation operations). Based on different properties and signs of papilledema, a total of 26 features were extracted and categorized into 4 groups; color features, texture features, vascular features and disc margin obscuration features. The group color features had 3 features and they were calculated from colored fundus images; texture features had 9 features which were extracted from grey-scale fundus images; vascular features had 6 features calculated after vessel segmentation of fundus images and the disc margin obscuration group containing 8 features and calculated from gray-scale fundus images. All the above mentioned features were evaluated using Wilcoxon rank sum test and followed by 10 folds cross validation and classification through SVM classifier. In order to evaluate the system, we calculated accuracy, sensitivity and specificity. To further evaluate the significance of different groups of features, all groups were tested individually as well as with different combinations. The effect of different number of top features on performance measures was also computed. The proposed system was applied on a local data set, acquired from AFIO that had 70 fundus images in which 30 have papilledema symptoms and 40 were healthy. AFIO data set has been annotated by multiple expert ophthalmologists who work in AFIO. Apart from this, we have applied our proposed system on a publically available STARE data set [17] that contained 90 fundus scans in which 40 had papilledema symptoms and 50 were healthy. STARE data set is an online data set that has been annotated by multiple expert ophthalmologists. These annotations act as a ground truth on which our proposed system was validated. We have also applied the proposed system on the combination of both databases. The performance measures of proposed techniques are given in Table 6. We have shown herein and conclude that our proposed methodology gives more accurate results for even higher degrees of papilledema cases. Apart from this, we have compared our proposed system with [6] on the same STARE data set and our proposed algorithm achieved better performance in terms of accuracy as shown in Table 6. In this paper, we have extended our proposed system originally proposed in [7] to incorporate efficient and early diagnosis of low grade papilledema from fundus scans. This has been done by adding a block in our original proposed system to automatically extract vascular features from the candidate retinal image. However, our proposed system is sensitive to the quality of acquired fundus scans and poor quality scan such as shown in Fig. 11 can lead to incorrect classification. Since our local AFIO data set contained such poor-quality scans so that’s why we have lesser performance ratings on our local data set as compared to the STARE data set.
In the future, the grading of papilledema based on Friesen scale [16] can be done for a more detailed analysis of papilledema. Moreover, this work can be extended by incorporating OCT along with fundus scan to give an early and robust classification of papilledema.