# Accuracy of Funduscopy to Identify True Edema versus Pseudoedema of the Optic Disc

*Arturo Carta,*[1] *Stefania Favilla,*[2] *Marco Prato,*[3] *Stefania Bianchi-Marzoli,*[4] *Alfredo A. Sadun,*[5] *and Paolo Mora*[1]

**Purpose.** Differential diagnosis between acute optic disc edema (ODE) and optic disc pseudoedema (PODE) may be a clinical challenge even for well-trained ophthalmologists. Funduscopy remains the first-line investigation. The aim of this study was to assess the accuracy, sensitivity, and specificity of funduscopy in differentiating ODE from PODE.

**Methods.** This was an observational, cross-sectional, two-center study of subjects referred for presumed acute ODE. During funduscopy, each observer completed a form concerning the presence/absence of the 10 conventional signs of ODE. Seventy-four patients with ODE and 48 subjects with PODE were included in the analysis. Accuracy, sensitivity, and specificity from all possible combinations of signs were calculated by support vector machine (SVM) analysis.

**Results.** As a single sign, the swelling of the peripapillary retinal nerve fiber layer had the highest accuracy (0.92; 95% confidence interval [CI], 0.82–0.97). Little variation was observed when more than fours signs were present. The best four-sign combinations were SWELLING, HEMORRHAGES, papilla ELEVATION, and CONGESTION of peripapillary vessels (accuracy, 0.93; 95% CI, 0.83–0.98; sensitivity, 0.95; specificity, 0.89). The presence of retinal or choroidal folds appeared to be a pathognomonic sign of true ODE (100% sensitivity) but had a low rate of presentation (23%).

**Conclusions.** The presence of at least four ophthalmoscopic signs (with the sign "swelling" included) gives the highest accuracy. Furthermore, peripapillary retinal folds seem to be related exclusively to ODE because they were never observed in our PODE group. These data may be useful for clinicians when evaluating patients referred for presumed optic disc edema in the acute phase. (*Invest Ophthalmol Vis Sci.* 2012; 53:1–6) DOI:10.1167/iovs.11-8082

Optic disc edema (ODE) is often considered an alarming finding because of its association with visual and even life-threatening disorders. Bilateral ODE, for instance, is often related to intracranial hypertension and thus is termed papilledema, and it may be secondary to an expanding intracranial mass.[1] Physiological variants of the optic disc, or other benign alterations such as optic disc drusen, myelinated fibers, and even "crowded" optic discs, may mimic elevation of the optic nerve head on the retinal surface. These conditions constitute optic disc pseudoedema (PODE).

Disturbances of axoplasmic transport involving the intraocular portion of the optic nerve can produce mechanical and vascular alterations that, viewed by funduscopy, represent 10 possible ophthalmoscopic signs for ODE in the acute phase.[2,3] Some of these ophthalmoscopic signs, however, are poorly pronounced in patients with initial ODE, and other signs may be observed in patients with PODE. Furthermore, patients with ODE may have presenting symptoms similar to those of some forms of PODE (e.g., optic disc drusen) such as transient visual obscuration, which typically is associated with papilledema.[4–6] For all these reasons, the differential diagnosis between ODE and PODE still represents a clinical dilemma, in particular when ophthalmoscopy is required by physicians who are not trained neuro-ophthalmologists. Improvements in the quality of optical coherence tomography (OCT) and magnetic resonance imaging (MRI) can sometimes support the diagnoses and management of ODE and PODE.[7,8] Such high-technology testing, however, represents laboratory studies that should be performed after establishing the most probable differential diagnosis. Funduscopy remains the most valuable and critical first-level investigation to direct further studies when early ODE is suspected.

To date, no reported study has investigated the accuracy, sensitivity, and specificity of funduscopy for the identification of true ODE versus PODE.

The aim of the present study was to identify the key ophthalmoscopic signs, singly and in combination, to help ensure a high level of diagnostic accuracy for true ODE. To this end, we used a systematic approach using the so-called "machine learning algorithm."[9,10]

## Methods

This was an observational, cross-sectional, two-center study conducted in the neuro-ophthalmology services of the university hospitals involved. The study was approved by the local ethics committees and was conducted in accordance with the principles of the Declaration Helsinki and those of good clinical practice. Written, informed consent was obtained from each patient.

Outpatients referred to the local emergency services for "presumed ODE" in the acute phase were eligible and underwent the study visit at their admission. Modeling analysis, however, considered only each patient for whom a definitive diagnosis had been made after a specific workup. On admission, each eligible subject underwent extensive ophthalmologic examination, including refraction assessment, pupillary reflex testing, slit-lamp examination, intraocular pressure measurement, and fundus evaluation. Funduscopy (under mydriasis with tropicamide 1%) was conducted in both centers by two experienced neuro-ophthalmologists using a 90-diopter lens. None of the specialists at the

From the [1]Ophthalmology Section, Head and Neck Department, University of Parma, Parma, Italy; Departments of [2]Biomedical Sciences and [3]Pure and Applied Mathematics, University of Modena and Reggio Emilia, Modena, Italy; [4]Institute of Ophthalmology, University Vita-Salute, San Raffaele Hospital, Milano, Italy; and [5]Doheny Eye Institute, Keck School of Medicine, University of Southern California, Los Angeles, California.

Submitted for publication June 20, 2011; revised October 21, 2011; accepted November 14, 2011.

Disclosure: **A. Carta**, None; **S. Favilla**, None; **M. Prato**, None; **S. Bianchi-Marzoli**, None; **A.A. Sadun**, None; **P. Mora**, None

Corresponding author: Arturo Carta, Neuro-ophthalmology Service, Ophthalmology Section, University of Parma, Via Gramsci 14, 43126 Parma, Italy; acarta@unipr.it.

**Table 1.** Examination for Definitive Diagnoses and Diagnostic Confirmatory Results

| Group | AF-FA | B-US | Visual Field | Ishihara | VEP | CT | MRI | MRA | LP | ESR-CRP | Infectious | mtDNA | TB |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ODE (*n*) | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A |
| IIH (34) | 2/0 | 0/0 | 34/27 | 34/27 | 0/0 | 20/0 | 34/0 | 34/0 | 34/34 | 34/1 | 34/0 | 0/0 | 4/0 |
| AION (16) | 10/10 | 0/0 | 16/16 | 16/16 | 16/16 | 8/0 | 6/0 | 0/0 | 0/0 | 16/0 | 16/0 | 0/0 | 0/0 |
| AON (12) | 6/6 | 0/0 | 12/12 | 12/12 | 12/12 | 0/0 | 12/12 | 0/0 | 12/4 | 12/2 | 12/2 | 0/0 | 0/0 |
| IH (12) | 0/0 | 0/0 | 10/10 | 10/10 | 0/0 | 12/12 | 12/12 | 10/10 | 0/0 | 12/0 | 12/0 | 0/0 | 0/0 |
| Total ODE (74) | 18/16 | 0/0 | 72/65 | 72/65 | 28/28 | 40/12 | 64/24 | 44/10 | 46/38 | 74/3 | 74/2 | 0/0 | 4/0 |
| PODE (*n*) | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A | D/A |
| ODM (23) | 10/0 | 10/0 | 23/0 | 23/0 | 0/0 | 12/0 | 12/0 | 0/0 | 0/0 | 15/0 | 0/0 | 0/0 | 0/0 |
| Drusen (15) | 15/15 | 15/15 | 6/2 | 15/1 | 0/0 | 10/10 | 0/0 | 0/0 | 0/0 | 10/0 | 0/0 | 0/0 | 0/0 |
| MF (6) | 6/0 | 6/0 | 6/0 | 6/0 | 2/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 4/0 | 0/0 | 0/0 |
| LHON (4) | 0/0 | 0/0 | 4/4 | 4/4 | 4/4 | 4/0 | 4/0 | 4/0 | 0/0 | 4/0 | 4/0 | 4/4 | 0/0 |
| Total PODE (48) | 31/15 | 31/15 | 39/6 | 48/5 | 6/4 | 26/10 | 16/0 | 4/0 | 0/0 | 29/0 | 8/0 | 4/4 | 0/0 |

AF-FA, auto-fluorescence and fluorescein angiography of the fundus; B-US, ocular B-scan ultrasound; Ishihara, Ishihara color plates test; MRA, magnetic resonance angiography; LP, lumbar puncture (opening pressure and cephalorachidian fluid); Infectious, serology for infectious diseases; mtDNA, mitochondrial DNA analysis; TB, temporal biopsy; IIH, idiopathic intracranial hypertension; AION, anterior ischemic optic neuropathy (nonarteritic); AON, anterior optic neuritis; IH, intracranial hypertension; ODM, physiological variants of optic disc morphology; drusen, optic disc drusen; MF, myelinated fibers; LHON, Leber's hereditary optic neuropathy; D/A, number of performed investigation (done)/number of altered results (altered).

centers knew the patients' diagnoses (these became apparent only after workup). They completed a specific forced-choice form for each subject concerning the presence or absence of the 10 reference clinical signs of ODE (five mechanical, five vascular).[2,3,11] After this first ophthalmologic evaluation, the diagnostic workup became specific for each patient, depending on the suspected causative etiology for the presumed ODE. This could include blood tests (routine profile, erythrocyte sedimentation rate [ESR], C-reactive protein [CRP], serology for infectious diseases), autofluorescence and fluorescein angiography of the fundus, ocular B-scan ultrasound, visual field test, Ishihara color plate test, visual evoked potentials, cerebral computed tomography (CT) scan, cerebral MRI and magnetic resonance angiography (MRA), lumbar puncture (considering opening pressure and cerebral spinal fluid composition), mitochondrial DNA analysis, and surgical biopsy. The type and the number of investigations performed, along with the number of confirmatory results, are summarized in Table 1.

Among 131 subjects consecutively referred for presumed ODE in both centers over a period of 26 months, definitive diagnoses were achieved in 122 (60 men, 62 women; mean age, 39 years; range, 9–78 years) and were included in the study analysis. The true ODE group consisted of 74 patients: 34 with idiopathic intracranial hypertension, 16 with nonarteritic anterior ischemic optic neuropathy, 12 with anterior optic neuritis, and 12 with intracranial hypertension secondary to neoplasm or thrombosis. The PODE group consisted of 48 patients: 23 with physiological variants of optic disc morphology (13 "crowded" optic disc, 10 "tilted" optic disc), 15 with optic disc drusen (primarily deep drusen), 6 with myelinated fibers, and 4 with Leber's hereditary optic neuropathy (LHON). Between the groups, differences were evaluated using Student's *t*-test for continuous data and Fisher's exact test for categorical data. $P < 0.05$ was considered statistically significant. For each patient, a single eye was considered for the study: the affected eye in patients with unilateral disease, the eye with the greater number of positive funduscopic signs in patients with bilateral but asymmetric involvement, and the right eye in patients with bilateral, symmetric disorder. The image detailed in Figure 1 was used in both centers as a reference chart to assist the detection of signs during funduscopy. In cases of disagreement between the two observers regarding one particular sign, the case was classified as "absent." Intracenter and inter-center grading agreements were measured by the variation coefficient method and were 2.2% and 2.8% among local examiners and 3.6% between the two centers. The study form envisaged the presence or absence of the following 10 clinical signs:

A. Mechanical signs
   1. BLURRING: blurring of the optic disc edges;
   2. ELEVATION: anterior extension of the optic nerve head on the retinal surface;
   3. FILLING-IN, filling-in of the physiological optic disc cup;
   4. FOLDS: presence of retinal or choroidal folds, or both;
   5. SWELLING: thickening of the peripapillary retinal nerve fiber layer (RNFL).
B. Vascular signs
   1. CONGESTION: enlargement and tortuosity of the arcuate and peripapillary venous vessels;
   2. EXUDATES S/H: "soft" (cotton wool) or "hard" exudates;
   3. HEMORRHAGES: peripapillary hemorrhages;
   4. HYPEREMIA: increased redness of the optic nerve head;
   5. PULSATION: dynamic evidence of venous pulsation at the emergence from the optic disc.

To create clinical archives of data, stereophotographs of the optic nerve head (TRC-50× retinal camera; Topcon Corporation, Tokyo, Japan) were also obtained for each patient on admission. These photographs were considered the gold standard to ascertain the presence/absence of the signs (excluding dynamic signs). The rate of accordance between stereophotographs and observers for single signs ranged from 0.94 to 1; the smallest were for the signs BLURRING and HYPEREMIA and the highest for the signs FOLDS and HEMORRHAGES. Patients also underwent RNFL thickness measurement by optical coherence tomography (OCT; StratusOCT, version 4.0; Carl Zeiss Meditec Inc., Dublin, CA). A skilled optometrist, masked to the clinical status of the subjects, performed the image acquisition using the RNFL thickness 3.4 protocol. OCT data were excluded from analysis when the scans had signal strength <8 or when there was an algorithm failure (common when boundaries for the RNFL layer are not delineated correctly). The average single strength of the accepted scans was 8.2. For each study eye, we considered the mean RNFL thickness (360° measure), as reported elsewhere,[12] without any quadrant comparison because such a qualitative analysis was beyond the purpose of the present study.

## Modeling Analysis

The first step involved converting the clinical data to a vector form (descriptor) that was suitable for analysis with the machine learning algorithm. Thus, the absence/presence responses were translated to a 0/1 category scale. As an internal control for the model's consistency, the GENDER variable, clearly unrelated to ODE diagnosis, was also included in the analysis.

The resultant data set consisted of 122 vectors (one for each subject), each made up of 11 binary variables. ODE and PODE classes were defined according to the result of the medical workup.

The adopted learning methodology was the support vector machine (SVM), which is a supervised learning technique widely used for both classification and regression problems.[13,14] In the specific case of

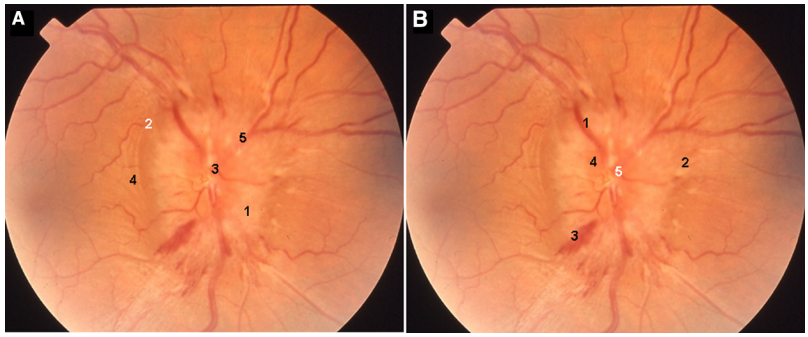


**FIGURE 1.** Funduscopic signs of ODE. (**A**) Mechanical signs: (1) BLURRING: blurring of the optic disc edges; (2, *white*) ELEVATION: anterior extension of the optic nerve head on the retinal surface (the *white* character means that optimal detection is possible only through stereoscopic observation); (3) FILLING/IN: disappearance of the physiological optic disc cup; (4) FOLDS: presence of retinal or choroidal folds, or both; (5) SWELLING: thickening of the peripapillary RNFL, which blurs the profile of the vessels at the optic disc contour. (**B**) Vascular signs: (1) CONGESTION: enlargement and tortuosity of the arcuate and peripapillary venous vessels; (2) EXUDATES S/H: presence of "soft" (cotton wool) or "hard" exudates; (3) HEMORRHAGES: presence of peripapillary hemorrhages; (4) HYPEREMIA: increased redness of the optic nerve head; (5, *white*) PULSATION: dynamic evidence of venous pulsation at the emergence from the optic disc (the *white* character means that detection is possible only through in vivo observation).

our binary classification problem, each vector was marked as belonging to the ODE or the PODE category. The SVM algorithm tries to separate the two classes by drawing a surface whose distance from the nearest data points on each side (the support vectors) is maximized. Because the existence of a surface that perfectly separates the examples of the two classes is not guaranteed, some misclassified training examples are possible.

In our study, the algorithm was asked to learn the classification rule (ODE/PODE eyes) from a subset of examples—half the full data set with known diagnosis (the training set)—and then to predict the class of the remaining (unknown) examples (the test set). In summary, we wanted to predict the presence of the considered pathology in subjects, which had not been considered for the model construction. Additionally, we sought to assess the minimum number of descriptors required to reach a high level of accuracy, specificity, and sensitivity for true ODE detection. In the machine learning framework, this problem is called "feature selection" (a feature is simply one component of the vector that describes each example).[15] We developed a tailored version of a "wrapper" approach[16] by computing all the possible combinations "$Sk$" of $k$ variables (with $k$ = 1–11). For these 11 subsets, $S$ (each composed of all the possible combinations of $k$ variables), the best predictor set was determined by a ranking procedure based on maximum accuracy, sensitivity, and specificity, respectively, calculated from the corresponding "confusion matrix."[17] SVM was implemented using the Classification And REgression Training (CARET) library[18] in the R free software environment.[19] Regarding the choice of the generalized separating surface, we considered a weighted sum of Gaussian functions with constant SD $\sigma$, automatically chosen within the library through the *sigest* function[20] in the kernlab package.[21]

The remaining parameter required by SVM is the so-called regularization parameter (balancing the tradeoff between data fidelity and generalization ability). It was determined through a 50-fold bootstrap procedure[22] in which the highest level of accuracy (the agreement rate averaged over bootstraps iterations) was used to address the best model. In the original data set, no "near-zero-variance" predictors (predictors that have a few unique numerical values that are highly unbalanced), highly correlated predictors (>75%), or linear dependences were present. Two approximately equal, exhaustive, and mutually exclusive subsets were identified by randomly sampling the ODE/PODE subjects. The first subset (61 subjects) was used to train the SVM algorithm, and the resultant classifier was tested on the second one (61 subjects). This procedure was then repeated 20 times, each one considering a different random splitting of the data set into training and test sets. Best combinations for accuracy, sensitivity, and specificity rate were selected by considering the predictions on the test data set only.

Accuracy is conventionally considered the measure of how well a binary classification test correctly identifies or excludes a condition. For our SVM procedure, it represents the overall agreement rate averaged over cross-validation iterations. Sensitivity represents the proportion of true positives correctly identified as such (the percentage of true ODE patients correctly identified as having this condition). Specificity describes the proportion of unaffected subjects correctly identified as such (the percentage of PODE subjects correctly identified as not having ODE).

## RESULTS

Table 1 summarizes the various tests performed to reach the definitive diagnosis for assignment of each subject. The table shows the total number of tests performed and the rate at which tests allowed for the establishment of each type of pathology that led to ODE or PODE. Blood testing was most frequently performed. The absence of abnormal ESRs and CRPs confirmed that all the cases of anterior ischemic optic neuropathy were nonarteritic. Neuroimaging examinations such as CT, MRI, and MRA were performed in 54%, 65%, and 39% of the enrolled patients, respectively. Abnormal imaging ranged, depending on the technology, between 21% and 33% of the examinations. The percentage of patients who received diagnoses of PODE but who were initially referred for presumed ODE was 39% (48/122 subjects). Physiological variants of optic disc morphology (tilted and crowded disc) and optic disc drusen were the most common causes of misdiagnosis (38/48 subjects).

Demographics of the enrolled subjects, visual acuity, and refractive error are reported in Table 2. The two groups were statistically comparable for age, gender, and visual parameters.

The prevalence of each sign, along with the corresponding accuracy, specificity, and sensitivity calculated by univariate binary test, is shown in Table 3. The FOLDS descriptor was never found in the PODE group; thus, it might be considered a pathognomonic sign of true ODE. Its presence in the ODE group, however, was also low (23.0%). Using a conventional statistical approach (univariate), the accuracy ranged from 0.44 to 0.92, the highest value related to the SWELLING descriptor. This latter sign also had good sensitivity, even if it was lower than the four other signs considered. The specificity of SWELLING was instead very low (0.07), as it was for every high-sensitivity descriptor and vice versa. Further implementation of this methodology began by consideration of those signs characterized by the highest rate of presence in the ODE group and by the highest rate of absence in the PODE group. By adding these two values together, four clinical signs were found to be greater than the mean of the sums, which was arbitrarily assigned as a threshold (Table 3, asterisks).

**Table 2.** Demographics and Ocular Parameters in the Two Groups

| Group | M/F | Age (y) | BCVA (logMAR) | SE (D) |
|---|---|---|---|---|
| ODE ($n$ = 74) | 31/43 | 38 ± 19 | 0.28 ± 0.45 | 0.25 ± 1.25 |
| PODE ($n$ = 48) | 25/23 | 41 ± 19 | 0.18 ± 0.28 | −0.07 ± 1.18 |
| $P$ | 0.27 | 0.72 | 0.47 | 0.50 |

Values are mean ± SD. BCVA, best-corrected visual acuity; logMAR, logarithm of minimal angle of resolution; SE, spheric equivalent expressed in diopters (D).

Data from the SVM analysis are reported in Tables 4 and 5. Table 4 shows the mean values of accuracy, sensitivity, and specificity over the 20 different training/test set configurations for the best $k$-combination of descriptors (with $k$ = 1–11). Accuracy, sensitivity, and specificity levels were never lower than 0.92, 0.95, and 0.85, respectively, including the best one-sign ($k$ = 1) combination. All three indices showed little variation when the combination consisted of more than four signs (e.g., 5–11 descriptors). Table 5 provides the clinical signs that constituted the best $k$-combination of descriptors in terms of accuracy. From $k$ = 4 to $k$ = 11, each best combination of features was obtained by adding one descriptor to the previous one (where "previous" means with one less feature). The SWELLING sign (swelling of the peripapillary RNFL) had the highest accuracy as single sign ($k$ = 1), and it invariably formed part of the most accurate combinations with a greater number of descriptors. The best four-descriptor combination was SWELLING, HEMORRHAGES, ELEVATION, and CONGESTION (accuracy, 0.93, sensitivity, 0.95, specificity, 0.89). With more than four descriptors, the accuracy did not improve. The internal control gender had the least influence on combination accuracy.

Based on the requisites previously mentioned, OCT was acceptable (signal strength ≥8 and no algorithm failure) in 48 of 74 patients (65%) included in the ODE group and in 30 of 48 patients (63%) in the PODE group. RNFL thickness was 217 ± 32 $\mu$m (mean ± SD) in the ODE group and 152 ± 12 $\mu$m in the PODE group. This parameter, though higher in the ODE group, did not reach statistical significance ($P$ = 0.12).

## Discussion

Causes of ODE include life-threatening disorders, such as intracranial tumors, obstruction of the ventricular system or shunt failure, subarachnoid hemorrhage, subdural hematoma, encephalitis, and meningitis,[23] along with several sight-threatening disorders, such as ischemic or compressive optic neuropathy, papillitis, and retinal vessel occlusion. Other more benign ocular conditions, however, may mimic ODE. These conditions, including optic nerve head drusen (approximately 2% in adult autopsy eyes),[24] tilted optic disc, myelinated nerve fibers, hyperopic crowded optic disc, and optic nerve hypoplasia, can cause PODE and masquerade as ODE. In all such conditions, funduscopy remains the critical clinical examination best able to characterize the optic nerve head. However, funduscopic differentiation between true ODE and PODE has always represented a dilemma for ophthalmologists, neurologists, and neurosurgeons.[25] Our series corroborates this, as evidenced by our findings that showed that the rate of patients with PODE initially referred for presumed ODE was nearly 40% (48/122). Physiological variants of optic disc morphology and optic disc drusen were the most common causes of misdiagnosis (38/48).

The aim of the present study was to identify the key ophthalmoscopic findings and the minimum combination of signs needed to attain a high level of accuracy in diagnosing any true ODE, regardless of the specific pathology responsible for the ODE itself.

The SWELLING descriptor (clinical evidence of thickening of the peripapillary RNFL) provided the highest level of accuracy as a single sign (0.92) and was present in all the most accurate combinations of signs with greater numbers of descriptors. In particular, our results showed high accuracy in identifying true ODE when three or more conventional ophthalmoscopic signs were observed in addition to SWELLING. The FOLDS descriptor appeared to be a pathognomonic sign of true ODE (100% sensitivity), but the low rate of presentation (23%) limited its contribution to the diagnostic process. The GENDER descriptor notoriously does not discriminate ODE-related pathology, but it was used as an internal control for the reliability of the model; it actually appeared among the least important signs entering the combinations (Table 5).

Several models have been proposed to support diagnosis and prognosis procedures using kernel-based learning methods. The validity of SVM in the identification of glaucoma, for example, has been well confirmed, as has the greater diagnostic capacity of combinations of descriptors compared with the

**Table 3.** Prevalence of Each Sign (Alphabetical Order) in the Two Groups; Accuracy, Sensitivity, and Specificity Calculated Using the Conventional Univariate Binary Test

| Signs | ODE Group Yes-No (%) | PODE Group Yes-No (%) | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Blurring* | 95.9–4.1 | 33.3–66.7 | 0.84 | 0.82 | 0.09 |
| Congestion* | 66.2–33.8 | 2.1–97.9 | 0.79 | 0.98 | 0.35 |
| Elevation | 87.8–12.2 | 60.4–39.6 | 0.69 | 0.69 | 0.32 |
| Exudates | 14.9–85.1 | 2.1–97.9 | 0.44 | 0.55 | 0.65 |
| Filling-in | 86.5–13.5 | 97.9–2.1 | 0.53 | 0.58 | 0.91 |
| Folds | 23.0–77.0 | 0.0–100.0 | 0.53 | 1.00 | 0.54 |
| Hemorrhages | 25.7–74.3 | 2.1–97.9 | 0.48 | 0.92 | 0.57 |
| Hyperemia* | 79.7–20.3 | 41.7–58.3 | 0.71 | 0.75 | 0.35 |
| Pulsation | 29.7–70.3 | 37.5–62.5 | 0.54 | 0.95 | 0.54 |
| Swelling* | 95.9–4.1 | 14.6–85.4 | 0.92 | 0.91 | 0.07 |

* Yes (presence in ODE) + No (absence in PODE) value is greater than the mean of the sum (i.e., >131.4).

**TABLE 4.** Mean Accuracy, Sensitivity, and Specificity for Test Data Set for the Best *k*-Combination of Descriptors Calculated Using the SVM Method

| No. Descriptors (*k*) | Accuracy (95% CI) | Sensitivity | Specificity |
|---|---|---|---|
| 1 | 0.92 (0.82–0.97) | 0.96 | 0.85 |
| 2 | 0.92 (0.82–0.97) | 0.97 | 0.85 |
| 3 | 0.92 (0.83–0.97) | 0.97 | 0.85 |
| 4 | 0.93 (0.83–0.98) | 0.95 | 0.89 |
| 5 | 0.93 (0.83–0.98) | 0.95 | 0.88 |
| 6 | 0.93 (0.83–0.98) | 0.96 | 0.88 |
| 7 | 0.93 (0.84–0.98) | 0.96 | 0.88 |
| 8 | 0.93 (0.84–0.98) | 0.96 | 0.88 |
| 9 | 0.93 (0.83–0.98) | 0.96 | 0.88 |
| 10 | 0.93 (0.83–0.98) | 0.96 | 0.87 |
| 11 | 0.92 (0.83–0.98) | 0.96 | 0.87 |

same descriptors considered singly.[26,27] Recently, machine learning techniques have been used to identify which of 14 algorithms best predicted the genetic risk for development of proliferative vitreoretinopathy in patients who were experiencing primary rhegmatogenous retinal detachment.[28] This study showed that one of the three best predictive models was indeed the radial kernel SVM, and its predictive capability was computed in a manner similar to the one adopted in our series (accuracy, sensitivity, specificity). As far as we know, our study is the first reported attempt to apply the SVM algorithm to the funduscopic diagnosis of ODE. To provide the model with data as close as possible to those from a first-line clinical examination of a subject with suspected ODE, we prepared questionnaires. These questionnaires asked the examiner to indicate the presence or absence of the 10 conventional mechanical and vascular signs.

The diagnostic process can be mathematically modeled within the framework of a pattern classification problem because the challenge is to predict class membership (e.g., health/illness, true/false). Solving the classification problem means reaching a decision to assign one object to a suitable class based on a set of features related to the characteristics of that class. The main advantage of our decision support system (the SVM method) is to permit evaluation of all possible combinations of descriptors in an exhaustive manner, in contrast with conventional univariate binary tests in which each descriptor is considered independently. The SVM feature selection procedure allows a notable increase in the accuracy, sensitivity, and specificity for a "one-sign" test. This can be observed in the case of the SWELLING descriptor, identified as the best $k = 1$ combination. The comparison between values coming from the conventional univariate binary test shown in Table 3 (accuracy, 0.92; sensitivity, 0.91; specificity, 0.07) and the best one-sign SVM results shown in Table 4 (accuracy, 0.92; sensitivity, 0.96; specificity, 0.85), supports the high potential of kernel-based algorithm.

From a trivial prevalence analysis, four signs were present in >85% of true ODE subjects. Using a slightly more complex analysis that considered the high prevalence in the ODE group concomitantly with the high rate of absence in the PODE group, we also found four signs above the arbitrarily chosen threshold. Two of these signs were different from the best SVM combination. Indeed, the two noncorresponding signs were characterized by high prevalence in both the ODE and PODE groups (Table 3). The confounding behavior of such signs (e.g., FILLING-IN) can explain the difficulty of the SVM model in setting up sequential combinations with a number of descriptors <4 (Table 5). With $k \geq 4$, instead, every increasing combination was formed by adding one new descriptor to those previously assessed. Thus, considering accuracy as the leading parameter, the model reached its ideal set by starting from a four-descriptor combination. The best four-descriptor combination was SWELLING, HEMORRHAGES, ELEVATION, and CONGESTION (accuracy, 0.93; sensitivity, 0.95; specificity, 0.89). With more than four descriptors, the accuracy did not improve. There was a decrease in sensitivity, paralleled by an increase in specificity, when passing from the best three-descriptor to the best four-descriptor combination. This was due to the presence in this specific step of only two of four common descriptors instead of "all but one," which happens for the more numerous combinations. However, based on our output data, the decrease in sensitivity from 0.97 to 0.95 signified a change from 29 to 28.5 true-positive subjects.

RNFL measurement by OCT was obtained as a surrogate measure to quantify the SWELLING detected with clinical and photographic observation and to follow the true ODE patients during the treatment phase. Based on the OCT data available, RNFL thickness was higher in the ODE group than in the PODE group. This evidence may reflect the fact that the enrolled subjects had been referred in an acute phase of their disease (e.g., including LHON patients in the PODE group), and it confirms the presence of diffuse thickening of the peripapillary RNFL in the ODE group. The difference between our two groups, however, did not reach statistical significance, as had occurred in previous case-control series.[12] This may be explained by some major limitations concerning the role of OCT in our study. Different etiologies of ODE and PODE were considered as a whole in the two groups, whereas some conditions, especially myelinated nerve fibers and tilted disc, make the RNFL identification troublesome and may significantly reduce the reproducibility of the of RNFL analysis. The peripapillary subretinal fluid described in previous reports contributes to the total retinal thickness and may be very sensitive in detecting initial edema in patients with true ODE.[12,29] Never-

**TABLE 5.** Test Data Set: Best Combination on 20 Partitions

| No. Descriptors (*k*) | Swelling | Hemorrhages | Elevation | Congestion | Blurring | Pulsation | Exudates | Folds | Hyperemia | Gender | Filling-in |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | X | | | | | | | | | | |
| 2 | X | | | | | | | | | | X |
| 3 | X | X | | | | | | | | | X |
| 4 | X | X | X | X | | | | | | | |
| 5 | X | X | X | X | X | | | | | | |
| 6 | X | X | X | X | X | X | | | | | |
| 7 | X | X | X | X | X | X | X | | | | |
| 8 | X | X | X | X | X | X | X | X | | | |
| 9 | X | X | X | X | X | X | X | X | X | | |
| 10 | X | X | X | X | X | X | X | X | X | X | |
| 11 | X | X | X | X | X | X | X | X | X | X | X |

theless, because SWELLING proved to be the most accurate clinical descriptor of true ODE, a readily available and minimally invasive procedure such as OCT is valuable in evaluating presumed ODE. This reinforces the value of studies that formally assess the reproducibility of RNFL thickness measurements by OCT in pathologies that cause ODE and PODE, especially in patients with mild or moderate swelling.[30–32]

In conclusion, this study allowed for a formal investigation of clinical signs of true ODE that have historically proven useful. The presence of retinal folds, even if rare, is a pathognomonic sign of true ODE. Thickening of the peripapillary RNFL (SWELLING) was indispensable for identifying ODE. A combination of four descriptors, in particular SWELLING, CONGESTION, ELEVATION, and HEMORRHAGES, ensured optimal predictive capacity. This approach may assist in the clinical examination of the optic nerve head as part of the diagnosis and treatment of patients with optic disc edema.

## *Acknowledgments*

The authors thank Alessandra Protti for data transcription and manuscript editing.

## *References*

1. Petrohelos MA, Henderson JW. The ocular findings in intracranial tumours. *Am J Ophthalmol.* 1951;34:1387–1394.
2. Frisén L. Swelling of the optic nerve head: a staging scheme. *J Neurol Neurosurg Psychiatry.* 1982;45:13–18.
3. Sadun AA. Optic atrophy and papilledema. In: Albert DM, Jakobiec FA eds. *Principles and Practice of Ophthalmology*. Philadelphia: WB Saunders; 1993:2529–2538.
4. Rizzo JF 3rd, Lessell S. Optic neuritis and ischemic optic neuropathy: overlapping clinical profiles. *Arch Ophthalmol.* 1991; 109:1668–1672.
5. Warner JE, Lessell S, Rizzo JF 3rd, Newman NJ. Does optic disc appearance distinguish ischemic optic neuropathy from optic neuritis? *Arch Ophthalmol.* 1997;115:1408–1410.
6. Sadun AA, Currie JN, Lessell S. Transient visual obscurations with elevated optic discs. *Ann Neurol.* 1984;16:489–494.
7. Heidary G, Rizzo JF 3rd. Use of optical coherence tomography to evaluate papilledema and pseudopapilledema. *Semin Ophthalmol.* 2010;25:198–205.
8. He M, Cestari D, Cunnane MB, Rizzo JF 3rd. The use of diffusion MRI in ischemic optic neuritis and optic neuritis. *Semin Ophthalmol.* 2010;25:225–232.
9. Vapnik VN. In: Vapnik VN, ed. *Statistical Learning Theory*. New York: John Wiley & Sons; 1998.
10. Hastie T, Tibshirani R, Friedman J. In: Hastie T, Tibshirani R, Friedman J, eds. *The Elements of Statistical Learning*. New York: Springer-Verlag; 2001.
11. Sadun AA. Papilledema and raised intracranial pressure. In: Yanoff M, Ducker JS, eds. *Ophthalmology*. 3rd ed. St. Louis, MO: Mosby; 2009:960–963.
12. Savini G, Bellusci C, Carbonelli M, et al. Detection and quantification of retinal nerve fiber layer thickness in optic disc edema using stratus OCT. *Arch Ophthalmol.* 2006;124:1111–1117.
13. Schölkopf B, Burges CJC, Smola AJ. In: Schölkopf B, Burges CJC, Smola AJ, eds. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press; 1999.
14. Cristianini N, Shawe-Taylor J. In: Cristianini N, Shawe-Taylor J, eds. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press; 2001.
15. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003;3:157–182.
16. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell.* 1997;97:273–324.
17. Kohavi R, Provost F. Glossary of terms. *Mach Learn.* 1998;30:271–274.
18. Kuhn M, Wing J, Weston S, Williams A, Keefer C. Caret: Classification and Regression Training. R package version 4.31. http://caret.r-forge.r-project.org. Accessed November 29, 2011.
19. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org. Accessed November 29, 2011.
20. Caputo B, Sim K, Furesjo F, Smola A. Appearance-based object recognition using SVMs: which kernel should I use? Proceedings of NIPS Workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision; December 12–14, 2002; Whistler, Canada.
21. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab, an S4 Package for Kernel Methods in R. *J Stat Softw.* 2004;11:1–20.
22. Efron B, Tibshirani R. The bootstrap method for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci.* 1986;1:54–75.
23. Whiting AS, Johnson LN. Papilledema: clinical clues and differential diagnosis. *Am Fam Physician.* 1992;45:1125–1134.
24. Lam BL, Morais CG Jr, Pasol J. Drusen of the optic disc. *Curr Neurol Neurosci Rep.* 2008;8:404–408.
25. Johnson LN, Hepler RS, Bartholomew MJ. Accuracy of papilledema and pseudopapilledema detection: a multispeciality study. *J Fam Pract.* 1991;33:381–386.
26. Bowd C, Medeiros FA, Zhang Z, et al. Relevance Vector Machine and Support Vector Machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci.* 2005;46:1322–1329.
27. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifier for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci.* 2005;46:4147–4152.
28. Rojas I, Fernandez I, Pastor JC, et al. Development of predictive models of proliferative vitreoretinopathy based on genetic variables: the Retina 4 Project. *Invest Ophthalmol Vis Sci.* 2009; 50:2384–2390.
29. Johnson LN, Diehl ML, Hamm CW, Sommerville DN, Petrosky GF. Differentiating optic disc edema from optic nerve head drusen on optical coherence tomography. *Arch Ophthalmol.* 2009;127: 45–49.
30. Kallenbach K, Simonsen H, Bander B, et al. Retinal nerve fiber layer thickness is associated with lesion length in acute optic neuritis. *Neurology.* 2010;74:252–258.
31. Barboni P, Carbonelli M, Savini G, et al. Natural history of Leber's hereditary optic neuropathy: longitudinal analysis of the retinal nerve fiber layer by optical coherence tomography. *Ophthalmology.* 2010;117:623–627.
32. Scott CJ, Kardon RH, Lee AG, Frisén L, Wall M. Diagnosis and grading of papilledema in patients with raised intracranial pressure using optical coherence tomography vs clinical expert assessment using a clinical staging scale. *Arch Ophthalmol.* 2010;128:705–711.