Efficient Screening of Diseased Eyes based on Fundus Autofluorescence Images using Support Vector Machine

S. R. Manne* K. K. Vupparaboina G. C. Gudapati R. A. Peddoju C. P. Konkimalla
IIT Hyderabad, India *UPitt. School of Medicine, USA* *IIT Hyderabad, India* *IIT Hyderabad, India* *IIT Hyderabad, India*

S. B. Bashar A. Goud J. Chhablani S. Jana
LVPEI Hyderabad, India *LVPEI Hyderabad, India* *UPitt. School of Medicine, USA* *IIT Hyderabad, India*

2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-0358-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BHI50953.2021.9508560

***Abstract*—A variety of vision ailments are associated with the foveal region of the eye. In current clinical practice, the ophthalmologist manually detects potential presence of such ailments based on fundus autofluorescence (FAF) images, and hence diagnoses the disease, when relevant. However, in view of the general scarcity of ophthalmologists relative to the large number of subjects seeking eyecare, especially in remote regions, it becomes imperative to develop methods to direct expert time and effort to medically significant cases. To serve the interest of both the ophthalmologist and the potential patient, we plan a screening step, where healthy and diseased eyes are algorithmically differentiated with limited input from only optometrists who are relatively more abundant in number. Specifically, an early treatment diabetic retinopathy study (ETDRS) grid is placed by an optometrist on each FAF image, based on which sectoral statistics are automatically collected. Using such statistics as features, healthy and diseased eyes are proposed to be classified by training an algorithm using available medical records. In this connection, we consider support vector machine (SVM) with linear as well as radial basis function (RBF) kernel, and observe satisfactory performance of both variants. Among those, we recommend the latter in view of its slight superiority in terms of classification accuracy (90.55% at a standard training-to-test ratio of 80:20), and practical class-conditional costs.**

***Index Terms*—Fundus autofluorescence (FAF), Early treatment diabetic retinopathy study (ETDRS) grid, Support vector machine (SVM), Monte Carlo cross validation (MCCV).**

I. Introduction

Medical imaging plays a crucial role in managing ophthalmic diseases. Among current modalities [1], fundus autofluorescence (FAF) imaging presents a relatively new technology that exploits the innate fluorescence of the retinal pigment epithelium (RPE), and has proven attractive by not requiring injection of artificial dye [2]. Specifically, the retina is illuminated with blue light which causes the naturally occurring pigment lipofuscin to fluoresce [3]. Consequently, the spatial distribution of intensity in the captured grey-scale FAF image (see Figure 1) indicates that of lipofuscin. Inspecting such distribution, the ophthalmologist is able to visualize the status of the retina (especially, retinal pigment epithelium), and hence diagnose/monitor various ophthalmic diseases. More precisely, for disease diagnosis, the physician generally hypothesizes a disease while inspecting a FAF image, and subsequently confirms it on the basis of quantitative indicators. For follow-up subjects, such indicators are used to monitor disease course. In either scenario, the associated clinical assessment provides the basis for deciding further treatment, if required.

(a) (b)

Fig. 1. FAF images of (a) healthy and (b) diseased eyes.

To facilitate such clinical assessment, significant effort has been directed at developing analysis tools specific to diseases, such as age-related macular degeneration (AMD) and Stargardt disease (STGD), an inherited ailment characterised by macular degeneration [4]. Such tools are based on disease-specific quantitative indicators, whose discovery has also remained an important focus. Indeed, indicators based on sectoral statistics collected from multi-segment grids centered at the fovea have been reported for various diseases including STGD [5], AMD [6] [7], central serous chorioretinopathy (CSCR) [8], and retinitis pigmentosa (RP) [9]. However, all the aforementioned sectoral statistics of FAF images, also possesses general information about diseased conditions, and hence remain relevant even for a generic screening. In this connection, we attempt to explore the usability of said statistics towards building a generic screening tool. Specifically, given a FAF image,

*Corresponding author, Email: ee14resch11006@iith.ac.in

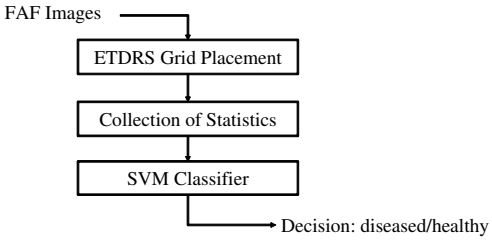

Fig. 2. Schematic diagram of proposed screening method.

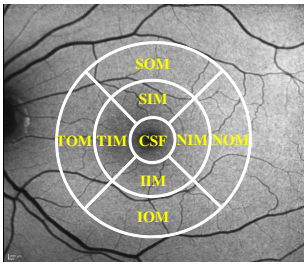

Fig. 3. FAF image with ETDRS grid.

we propose to use an early treatment of diabetic retinopathy study (ETDRS) grid [10], a well accepted tool in ophthalmic studies, centered at the fovea, and compute mean and standard deviation of pixel intensity within each sector [6]. Collecting such sectoral statistics in a feature vector, a classifier is then trained to distinguish FAF images of healthy eyes from the diseased ones in such feature space.

The rest of the paper is organized as follows. FAF image acquisition, ETDRS grid placement and the computation of statistics, SVM classifiers and performance evaluation criteria are described in Section II. In Section III, experimental results on screening performance are presented and Section IV concludes the paper.

II. Materials and Methods

The workflow of the proposed methodology is schematically outlined in Figure 2. First, we begin with the FAF image acquisition.

A. Fundus autoflourescence image acquisition

This is a retrospective study conducted at tertiary eye centers of L V Prasad Eye Institute (LVPEI) in South India. A total of 140 subjects underwent FAF examination where 61 are healthy and 79 are diagnosed to have ophthalmic diseases. Specifically, of the 79 subjects, 21 of them are diagnosed with CSCR, 14 with choroidal neovascular membranes (CNVM), and 44 with STGD. The Heidelberg Retina Angiograph 2 (HRA2; Heidelberg Engineering, Heidelberg, Germany) was used to acquire the FAF images. Standard protocol for image acquisition involved excitation at 488 nm with an optically pumped solid-state laser and emission detected above 500 nm with a barrier filter. Written informed consent was obtained from all the study participants and approval for study protocol was obtained from respective Institutional Review Boards. The procedures performed in this study are conformed to the tenets of Declaration of Helsinki.

B. ETDRS-based statistics as feature set

In the current work, the ETDRS grid was used for computing features from each FAF image with the goal of classifying it as either 'healthy' or 'diseased'. The ETDRS grid, positioned by an optometrist (rather than an ophthalmologist) at the centre of the fovea as shown in Figure 3, has nine segments. The inner most ring, known as central subfield (CSF), is sorrounded by pericentral ring consisting of temporal inner macula (TIM), inferior inner macula (IIM), nasal inner macula (NIM) and superior inner macular (SIM) regions. Further, the peripheral ring in ETDRS grid consists of temporal outer macula (TOM), inferior outer macula (IOM), nasal outer macula (NOM) and superior outer macular (SOM) regions. In each of the nine segments, the mean pixel intensity and the corresponding standard deviation were computed. Such statistics from all nine segments formed a feature vector of length 18 representing each FAF image. These features were used to distinguish between diseased and healthy eyes.

C. Support vector machine (SVM) classifier

The support vector machine (SVM) classifier provides a supervised learning model [11], which upon training, obtains a hyperplanar decision boundary maximally separating two classes. More generally, to handle data that are nonlinearly separable, SVM can incorporate suitable kernel that transforms the original feature space into a higher-dimensional space, where the transformed features become linearly separable. Formally, the decision hyperplane $w^T\phi(x) + b = 0$ (where $\phi(x)$ denotes a point vector in the transformed feature space, ϕ the kernel function, w the weight vector and b the bias) is chosen to maximize overall separation. One equivalently maximizes

$$\sum_{i=1}^{n} \alpha_i \text{ - } 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

over $\alpha_i \geq 0$, $i = 1, 2, ..., n$, subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$. Here, for each data index i, x_i and y_i, respectively, indicate the feature vector and the corresponding class label ($+1$ for diseased; -1 for healthy), and K denotes the kernel function, n the size of the dataset, α_i the Lagrange multiplier. We compare classification performance of two variants of SVM, one with linear kernel (where $K(x_i, x_j) = x_i^T x_j$), and the other with radial basis function (where $K(x_i, x_j) = \exp \frac{|x_i - x_j|^2}{2\sigma^2}$), an ubiquitous nonlinear kernel.

D. Performance evaluation

To evaluate classification performance, the SVM parameters, namely, weight vector w and bias b, are optimized on a subset of data (training subset) and are validated against the complementary (test) subset. More accurately, the dataset is partitioned into training and test subsets such that (i) the ratio of their sizes, called training-to-test (split) ratio, approximately equals a preassigned fraction, and (ii) the proportion of healthy

(as well as diseased) images represented in those subsets also approximately equals the same fraction. In general, the performance of the classifier depends on the subjective choice of the partition. To avoid such subjectivity in performance analysis, one customarily uses Monte Carlo cross validation (MCCV) [12], where, the dataset is randomly partitioned a large number (5000) of times (iterations) keeping the split ratio as well as the aforementioned proportions constant. For each iteration, the SVM parameters are optimized over the training subset, and the mean training as well as test accuracy and corresponding standard deviation are recorded. Noting that the training accuracy indicates the classification performance for seen data, while the test accuracy indicates that for unseen data, classifiers with high average test accuracy assumes practical significance. Further, low standard deviation, indicating low performance variability over random partitions and hence signifying robustness, is desirable. Additionally, we also compute the average confusion matrix over 5000 iterations, which provides class-conditional detection probability for each of healthy and diseased classes. Finally, we train and evaluate classifiers for different split ratios.

III. Experimental Results

As described earlier, placing ETDRS grid on each FAF image, we computed mean and standard deviation of pixel values corresponding to each of the nine sectors (Figure 3), and thus formed a feature vector of length 18. In Figure 4, such sectoral features are illustrated for example FAF images of both healthy and diseased eyes. As mentioned earlier, two variants of SVM classifiers, linear SVM and SVM with RBF kernel (RBF-SVM), were considered. In each case, training and test accuracy values were recorded over a large number (5000) of random partitions for each of the various training-to-test split ratios chosen between 10:90 and 90:10, and the average values (with standard deviation values in parenthesis) are tabulated in Table I.

Clearly, in case of linear SVM, the average training as well as test accuracy level tends to increase with increasing training-to-test ratio but for a few exceptions. In other words, as expected, the said classifier tends to learn better with increased availability of training data. Turning to the RBF-SVM classifier, for each split ratio we chose the scale factor (SF) that maximizes test accuracy. The resulting average test accuracy, which still largely follows the aforementioned increasing trend, slightly improves over that obtained using the linear SVM for each split ratio. This indicates a moderate degree of nonlinearity inherent in the underlying problem. Despite only a slight gain in test accuracy, the gain in average training accuracy is significantly higher, indicating that the RBF kernel is closely modeling certain nonlinear aspects of the training data that do not generalize well. In fact, a similar conclusion can also be drawn by considering standard deviation values as follows. Notice that the standard deviation in training accuracy decreases with increasing split ratio in both cases of linear SVM and RBF-SVM; however, the values are significantly lower for the latter classifier, indicating better

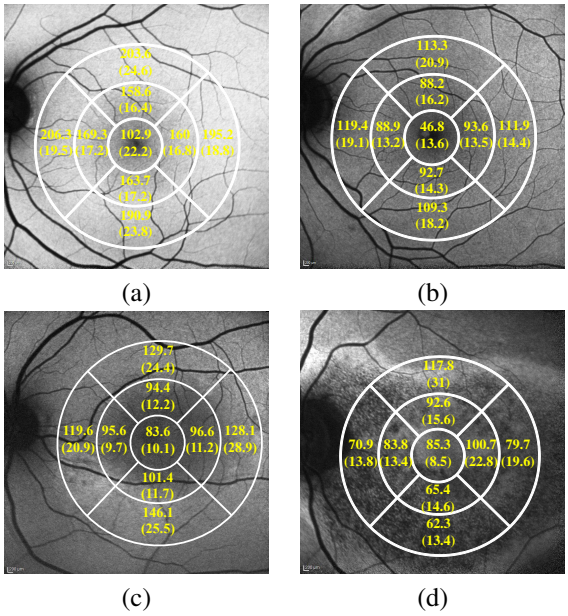

Fig. 4. ETDRS grid placement on and corresponding sectorial statistics (mean with standard deviation in parenthesis) for representative images (a), (b) of healthy eyes and (c), (d) of diseased eyes respectively.

modeling. Yet, turning to test accuracy, the standard deviation is lower in case of RBF-SVM for lower values of split ratio and in case of linear SVM for higher values of split ratio. This phenomenon does not indicate improved generalization. Further, the standard deviation in either case first decreases and then increases with increasing split ratio. Yet a subtle difference can be discerned. While the test accuracy is most reliable (i.e., with the lowest standard deviation) for an evenly split training-to-test ratio in case of linear SVM (also observed elsewhere [13]), highest reliability is observed at the skewed split ratio of 40:60 in case of RBF-SVM.

Having reported performance variation over the entire range of split ratios, we next delve deeper into case of 80:20 split, a general recommendation for various practical applications [14]. In particular, we next study class-conditional performance of these rival classifiers at hand. Practical significance of this study arises from the fact that mistakenly classifying a diseased eye as normal leads to the severe consequence of denial of treatment to a bona fide patient, whereas declaring a normal eye as diseased leads to the comparatively less severe outcome of wasted time and resources. Referring to Table II, RBF-SVM records an average probability of 91.61% in correctly detecting the disease class whereas it is only 89.47% using linear SVM, with respective standard deviations 7.09% and 7.33%. In contrast, conditioned on the healthy class, the healthy class was correctly detected with an average probability of 92.25% by linear SVM and 88.59% by RBF-SVM, with respective standard deviations 7.11% and 8.82%. Thus, given the healthy class, linear SVM slightly outperforms RBF-SVM in terms of both accuracy and reliability. Notice that the advantage of RBF-SVM over linear SVM given the diseased class is comparable to that of linear SVM over

TABLE I
PERFORMANCE OF RIVAL SVM CLASIFIERS FOR DIFFERENT SPLITS (SUPERIOR VALUE BETWEEN LINEAR SVM AND RBF-SVM BOLDFACED)

	Linear SVM		RBF-SVM			Gain (%)	
Split ratio	Train. Acc. a (b)	Test Acc. c (d)	SF	Train. Acc. e (f)	Test Acc. g (h)	Train. Acc. $\frac{e-a}{a}$ ($\frac{b-f}{b}$)	Test Acc. $\frac{g-c}{c}$ ($\frac{d-h}{d}$)
10:90	91.27 (7.01)	75.03 (7.05)	2.85	**99.38 (2.13)**	**78.97 (5.30)**	8.16 (69.61)	4.99 (24.82)
20:80	91.51 (4.82)	82.26 (4.78)	2.80	**98.81 (2.09)**	**84.32 (3.96)**	7.39 (56.65)	2.44 (17.15)
30:70	92.04 (3.66)	85.56 (3.81)	3.20	**97.88 (2.17)**	**86.66 (3.48)**	5.97 (40.71)	1.27 (8.66)
40:60	92.26 (2.91)	87.31 (3.46)	2.70	**98.86 (1.38)**	**88.11 (3.35)**	6.67 (52.27)	0.91 (3.18)
50:50	92.24 (2.44)	88.40 **(3.43)**	3.00	**98.86 (1.38)**	**88.89 (3.43)**	6.69 (43.44)	0.55 (0.00)
60:40	92.27 (2.01)	89.02 (3.68)	2.90	**98.46 (1.20)**	**89.58 (3.47)**	6.28 (40.29)	0.62 (5.71)
70:30	92.31 (1.64)	89.53 **(4.14)**	2.80	**98.62 (0.99)**	**90.10** (4.16)	6.39 (39.63)	0.63 (-0.48)
80:20	92.26 (1.30)	89.88 **(5.12)**	2.75	**98.65 (0.90)**	**90.55** (5.24)	6.47 (30.77)	0.75 (-2.34)
90:10	92.25 (1.01)	89.60 **(7.16)**	2.65	**98.86 (0.71)**	**90.83** (7.21)	6.68 (29.70)	1.35 (-0.70)

TABLE II
TEST CONFUSION MATRICES (%) OF LINEAR SVM AND RBF-SVM FOR SPLIT RATIO 80:20

	Actual / Predicted	Diseased	Healthy
Linear SVM	Diseased	89.47 (7.33)	7.75 (7.11)
	Healthy	10.53 (7.33)	92.25 (7.11)
RBF-SVM	Diseased	91.61 (7.09)	11.41 (8.82)
	Healthy	8.39 (7.09)	88.59 (8.82)

RBF-SVM given the healthy class. However, since making an error incurs higher practical cost given the diseased class than that given the healthy class, one should prefer RBF-SVM over linear SVM as a screening tool. In summary, based on our experiments, we conclude that SVM classifiers of FAF images, attaining close to 90% accuracy, provide attractive screening tools for ophthalmic diseases. Among those, RBF-SVM appears to provide slight advantage over linear SVM in terms of both overall accuracy levels and practical class-conditional costs.

IV. CONCLUSION

In this paper, we developed a screening technique that does not require the participation of ophthalmologists, and semi-automatically identifies the presence of disease based on FAF images with minimal assistance from optometrists. Further, in view of the excellent class separation exhibited by the proposed classifier, our screening service can potentially be integrated with post-diagnosis treatment response monitoring. A detailed discussion on the choice of classifier, possible progress monitoring and a theoretical connection is available in the extended version [15].

ACKNOWLEDGMENT

The work was partly supported by Grant BT/PR16582/BID/7/667/2016, Department of Biotechnology (DBT), Ministry of Science and Technology, the Government of India. S. R. Manne thanks the Ministry of Electronics and Information Technology (MeitY), the Government of India, for fellowship grant under Visvesvaraya PhD Scheme.

REFERENCES

[1] U. Yolcu, O. F. Sahin, and F. C. Gundogan, "Imaging in ophthalmology," in *Ophthalmology-Current Clinical and Research Updates*. IntechOpen, 2014.

[2] D. L. Nickla and J. Wallman, "The multifunctional choroid," *Progress in retinal and eye research*, vol. 29, no. 2, pp. 144–168, 2010.

[3] S. Schmitz-Valckenberg, F. G. Holz, A. C. Bird, and R. F. Spaide, "Fundus autofluorescence imaging: review and perspectives," *Retina*, vol. 28, no. 3, pp. 385–409, 2008.

[4] R. T. Smith, N. Lee, J. Chen, M. Busuioc, and A. F. Laine, "Interactive image analysis in age-related macular degeneration (amd) and stargardt disease (stgd)," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*. IEEE, 2008, pp. 651–654.

[5] T. R. Burke, T. Duncker, R. L. Woods, J. P. Greenberg, J. Zernant, S. H. Tsang, R. T. Smith, R. Allikmets, J. R. Sparrow, and F. C. Delori, "Quantitative fundus autofluorescence in recessive stargardt disease," *Investigative ophthalmology & visual science*, vol. 55, no. 5, pp. 2841–2852, 2014.

[6] M. Marsiglia, S. Boddu, S. Bearelly, L. Xu, B. E. Breaux, K. B. Freund, L. A. Yannuzzi, and R. T. Smith, "Association between geographic atrophy progression and reticular pseudodrusen in eyes with dry age-related macular degeneration," *Investigative ophthalmology & visual science*, vol. 54, no. 12, pp. 7362–7369, 2013.

[7] S. Schmitz-Valckenberg, M. Fleckenstein, A. P. Göbel, K. Sehmi, F. W. Fitzke, F. G. Holz, and A. Tufail, "Evaluation of autofluorescence imaging with the scanning laser ophthalmoscope and the fundus camera in age-related geographic atrophy," *American journal of ophthalmology*, vol. 146, no. 2, pp. 183–192, 2008.

[8] M. Zola, I. Chatziralli, D. Menon, R. Schwartz, P. Hykin, and S. Sivaprasad, "Evolution of fundus autofluorescence patterns over time in patients with chronic central serous chorioretinopathy," *Acta ophthalmologica*, vol. 96, no. 7, pp. e835–e839, 2018.

[9] J. K. Jolly, S. K. Wagner, J. Moules, F. Gekeler, A. R. Webster, S. M. Downes, and R. E. MacLaren, "A novel method for quantitative serial autofluorescence analysis in retinitis pigmentosa using image characteristics," *Translational vision science & technology*, vol. 5, no. 6, pp. 10–10, 2016.

[10] E. T. D. R. S. R. Group *et al.*, "Early photocoagulation for diabetic retinopathy: Etdrs report number 9," *Ophthalmology*, vol. 98, no. 5, pp. 766–785, 1991.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[12] Q.-S. Xu and Y.-Z. Liang, "Monte carlo cross validation," *Chemometrics and Intelligent Laboratory Systems*, vol. 56, no. 1, pp. 1–11, 2001.

[13] S. Tripathy, M. S. Reddy, S. R. K. Vanjari, S. Jana, and S. G. Singh, "A step towards miniaturized milk adulteration detection system: Smartphone-based accurate ph sensing using electrospun halochromic nanofibers," *Food Analytical Methods*, vol. 12, no. 2, pp. 612–624, 2019.

[14] K. Macek, "Pareto principle in datamining: an above-average fencing algorithm," *Acta Polytechnica*, vol. 48, no. 6, 2008.

[15] S. R. Manne, K. K. Vupparaboina, G. C. Gudapati, R. A. Peddoju, C. P. Konkimalla, A. Goud, S. B. Bashar, J. Chhablani, and S. Jana, "Efficient screening of diseased eyes based on fundus autofluorescence images using support vector machine," *arXiv preprint arXiv:2104.08519*, 2021.