Generalized weighted majority voting with an application to algorithms having spatial output

Henrietta Toman, Laszlo Kovacs, Agnes Jonas,
Lajos Hajdu, and Andras Hajdu

University of Debrecen
Egyetem ter 1, 4032 Debrecen, Hungary
{toman.henrietta,kovacs.laszlo.ipgd}@inf.unideb.hu
jonasagn@gmail.com,hajdul@math.klte.hu
hajdu.andras@inf.unideb.hu

Abstract. In this paper we propose a method using a generalization of the weighted majority voting scheme to locate the optic disc (OD) in retinal images automatically. The location with the maximal sum of the weights of OD center candidates falling into a disc of radius predefined in the clinical protocol is chosen for optic disc. We have worked out a weighted voting scheme, where besides the weights, an additional (e.g. geometrical) condition have to be taken into account in making the final decision. We can achieve better overall performance with this generalized weighted voting system than with the weighted majority voting and each individual algorithm.

Keywords: biomedical imaging, diabetic retinopathy, classifier combination, majority voting, weighted voting

1 Introduction

Diabetic retinopathy (DR) is an eye disease (damage to the retina) that is the most frequent cause of new cases of blindness. In automatic grading of the retinal images and in making diagnosis, determining the exact location of the main anatomical features (e.g. the optic disc and the macula) is among the first steps. Both the optic disc and the macula can be considered as a circular region (disc) on the retinal images. The optic disc is very bright, while the macula is a highly pigmented spot whose center is called fovea which is responsible for the sharpest vision.

In our approach, we organize several different individual OD detector algorithms into a weighted voting system to raise the accuracy of the OD detection ([1],[2]). Each OD algorithm results in a single pixel as output for the OD center. In our application for the OD detection, the majority voting cannot be applied directly since the spatial replacement of each vote also counts in making the final decision. In the generalized weighted voting system, the OD center candidate of each detector has been combined and the minimal bounding circles for all subgroups of the candidates are considered. The radius of the circle must be less than or equal to the radius of the optic disc that is a clinically predetermined constant. In this weighted voting system we choose the circle with the maximal sum of the weights assigned to the candidates falling inside this circle.

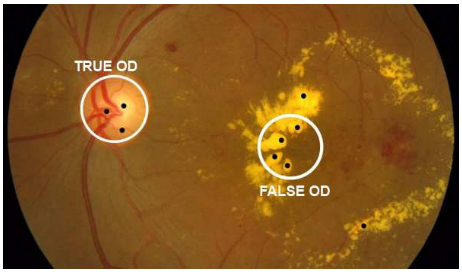


Fig. 1. Results of the different OD detecting algorithms

Weighted majority voting is widely examined in the literature (see e.g. [3], [4]). For characterizing the accuracy of the weighted system to our application a corresponding theoretical model is needed. If we consider the bounding circle with the maximal weight sum, then similarly to a traditional weighted majority setup, we can make a good decision even in the case when the bad candidates have pure majority in number. In our former work [5], we have generalized the classical majority voting to our problem. Now, just as in the traditional case, we check how weighted majority can outperforms classical majority voting. In the non-weighted generalized voting system bad decision can be made only when a subset of bad candidates with larger cardinality than the number of good ones can be bounded by a circle with an appropriate radius such as in the case shown in Fig. 1. In the weighted generalized voting system we make a wrong decision only in that case when a subset of bad candidates having larger sum of weights than the sum of weights assigned to the good ones can be bounded by a circle with an appropriate radius. In the case demonstrated in Fig. 1. good decision is made applying the weighted generalized voting system.

These observations motivated us to work out a corresponding theoretical model, where bad votes can overcome good ones only if a further (e.g. geometrical) condition is fulfilled. This additional condition is the spatial closeness of the candidates in the above application. With this model we generalize the classical non-weighted and weighted majority voting scheme, since in the case of less good votes we may make a good decision. This generalized method can be applied to several problems corresponding to spatial location with additional constraints (e.g. detecting a certain pixel or region).

In the rest of the paper, section 2 presents the classical voting system. In section 3 we recall our results for the generalization of the non-weighted voting system, while section 4 discusses the weighted majority voting and our generalized weighted system. In section 5, our experimental results for the specific

OD detection application are presented. Section 6 gives conclusion and further recommendations.

2 Majority voting

Let $D = (D_1, D_2, \ldots, D_n)$ be a set of classifiers, $D_i : R^k \to \Omega$ $(i = 1, \ldots, n)$ where $\Omega = (\omega_1, \omega_2, \ldots, \omega_c)$ is a set of class labels. If the classifier decisions are combined in the majority voting, then the class label ω_i is assigned to $\mathbf{x}$ that is supported by the majority of the classifiers D_i. In the case of a tie, the decision is usually made randomly.

As a special case, we can consider binary classifiers examined exhaustively in the literature. Let $n (n \in \mathbb{N})$ be odd, $\Omega = (\omega_1, \omega_2)$ (that is, each classifier output is a binary vector) and all classifiers have the same classification accuracy $p (p \in [0, 1])$. An accurate class label is given by the majority vote if at least $[n/2]$ classifiers give correct answers. The overall accuracy of correct classification in majority voting with independent classifier decisions can be computed by the binomial formula:

$$P = \sum_{k=0}^{[n/2]} \binom{n}{k} p^{n-k} (1-p)^k. \quad (1)$$

If the classifiers are independent and $p > 0.5$, then this method is guaranteed to outperform the individual classifiers. Applying the majority voting in pattern recognition, several interesting results can be found in [6] (e.g. about adding one or two new classifiers to the voting system).

3 The generalized majority voting

The classifiers making independent errors are generally considered independent, so under this assumption, the error of the classifiers can be modelled by random variables and their distributions. If we assume initially equal probabilities of errors for all classifiers, the model with Bernoulli distribution is the simplest and for this case the most appropriate one.

In this section, we recall and slightly re-formulate the theoretical and experimental results for the generalized majority voting system [5]. Let $\eta = (\eta_1, \ldots, \eta_n)$ be an n-dimensional random variable (n classifiers). Assume that the coordinates η_i of η are independent random variables with

$$P(\eta_i = 1) = p, \quad P(\eta_i = 0) = 1 - p \quad (i = 1, \ldots, n) \ , \quad (2)$$

where $p \in [0, 1]$ (each classifier has the same accuracy p). Execute the experiment η independently t times (t objects to be classified), and write the outcomes (outputs of the classifiers) in a table of size $n \times t$. The j-th column of the table contains the realization of η in the j-th experiment $(j = 1, \ldots, t)$. Define now the random variables $\chi_1, \ldots, \chi_t$ in the following way: if in the j-th column there are k 1 values (k correct classification for the j-th object) then let

$$P(\chi_j = 1) = p_{nk}, \quad P(\chi_j = 0) = 1 - p_{nk} \quad (j = 1, \ldots, t) \ , \quad (3)$$

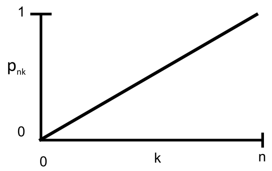
where the p_{nk}-s $(k = 0, 1, \ldots, n)$ are given numbers with monotone increasing property fulfilled in all rows: $0 \le p_{i0} \le \cdots \le p_{ii} \le 1 \quad (i = 1, \ldots, n)$. The p_{nk} describes the probability of the good final decision in case of k correct classifications from n classifiers.

Observe that the χ_j-s are independent. Finally, put $\xi = |\{j \; : \; \chi_j = 1\}|$, that is, ξ is the number of the good final decisions for t objects. We observe that all the individual decisions η_i $(i = 1, \ldots, n)$ are of binomial distribution with parameters (t, p). Then we get that ξ has also binomial distribution, with the appropriate parameters (t, q), where

$$q = \sum_{k=0}^{n} p_{nk} \binom{n}{k} p^k (1-p)^{n-k}. \tag{4}$$

In order to the generalized majority voting outperform the individual decisions, we need only to guarantee that $q \ge p$.

In that case when p_{nk} is linear in k for a given n, that is $p_{nk} = k/n$ $(k = 0, 1, \ldots, n)$, then we get $q = p$.



(a) The curve of p_{nk}

	L = 3	L = 5	L = 7	L = 9
p = 0.6	0.6	0.6	0.6	0.6
p = 0.7	0.7	0.7	0.7	0.7
p = 0.8	0.8	0.8	0.8	0.8
p = 0.9	0.9	0.9	0.9	0.9

(b) System accuracy

Fig. 2. The results of the linear case

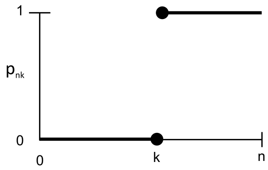
If we suppose that $p_{nk} \ge k/n$ for all $k = 0, 1, \ldots, n$, then $q \ge p$, so in this case the generalized majority voting outperforms the individual decisions.

As a special case of the generalized majority voting, when n is odd, $p \ge 1/2$ and for all $k = 0, 1, \ldots, n$ we have $p_{nk} = 1$, if $k > n/2$, and $p_{nk} = 0$, otherwise, we get the classical majority voting.

We indicate the overall performance P of the voting system in Fig. 2. and Fig. 3. for different L number of classifiers/algorithms at different classifier accuracies p in the linear case when $p_{nk} = k/n$ and in the classical majority voting case, respectively.

We give another example of the matrix p_{nk} that is motivated by our application for OD detection. In this case, the behavior of p_{nk} as a function of k for a fixed n and the system accuracy are illustrated in Fig. 4.

From Fig. 4. we can see, that for a fixed n, p_{nk} increases exponentially in k. This follows from the results of [7] about the diameter d of a point set. The probability that d is not less than a given constant decreases exponentially if the number of points tends to infinity. Note that, this diameter corresponds again to the radius of the OD defined by the clinical protocol.

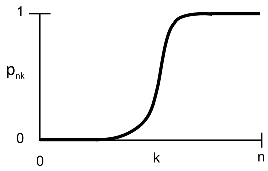


(a) The curve of p_{nk}

	L = 3	L = 5	L = 7	L = 9
p = 0.6	0.6480	0.6826	0.7102	0.7334
p = 0.7	0.7840	0.8369	0.8740	0.9012
p = 0.8	0.8960	0.9421	0.9667	0.9804
p = 0.9	0.9720	0.9914	0.9973	0.9991

(b) System accuracy

Fig. 3. The results of the classical majority voting scheme



(a) The curve of p_{nk}

	L = 3	L = 5	L = 7	L = 9
p = 0.6	0.8208	0.8390	0.8895	0.9247
p = 0.7	0.9163	0.9373	0.9658	0.9823
p = 0.8	0.9728	0.9850	0.9942	0.9980
p = 0.9	0.9963	0.9988	0.9997	0.9999

(b) System accuracy

Fig. 4. The results of our application for OD detection

4 Modifications on the decision rule

In this section, we modify the final decision rule of the ensemble which will result in further improvement of the system accuracy. Our generalization is based on the assignment of weights to the ensemble members (classifiers). First, we recall the necessary procedure for finding the weights in classical majority voting (see e.g. [4]). Then, we derive how the optimal weights can be found for our generalized voting case.

4.1 Weighted voting system

For weighted voting system, first let us consider the classifiers $(D_1, D_2, \ldots, D_n)$ with accuracies $(p_1, p_2, \ldots, p_n)$, respectively. Then, let $d_{i,j}$ be defined in the following way: $d_{i,j} = 1$, if the classifier D_i labels $\mathbf{x}$ in the class ω_j, and $d_{i,j} = 0$, otherwise. In case of weighted voting, the discriminant function for class ω_j is given as:

$$g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,j}, \tag{5}$$

where the weight b_i corresponds to the classifier D_i. Note that the following discriminant functions are equivalent for the given decision rule:

$$g_j(\mathbf{x}) = P(s|\omega_j)P(\omega_j), \quad g_j(\mathbf{x}) = \log(P(s|\omega_j)P(\omega_j)), \tag{6}$$

where $s = [s_1, \ldots, s_n]$ is the vector with the label output of the ensemble. Here $s_i \in \Omega$ is the label suggested for $\mathbf{x}$ by the classifier D_i and $P(\omega_j)$ is the prior probability for class ω_j.

In a weighted majority voting system, the class label ω_k is chosen for $\mathbf{x}$ if

$$g_k(\mathbf{x}) = \max_{j=1,\dots,n} g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,k}. \quad (7)$$

In a weighted majority system a natural question is that how to choose the optimal weights for the classifiers. If we consider independent classifiers , then the system accuracy is maximized by assigning weights (see e.g. [4]):

$$b_i \propto \log \frac{p_i}{1-p_i}, i = 1, \dots, n. \quad (8)$$

Note that, conditional independence is assumed here, that is:

$$P(s|\omega_j) = \prod_{i=1}^{n} P(s_i|\omega_j), \quad (9)$$

where $s = [s_1, \dots, s_n]$ is the same as above.

The weights $b_i \propto log\frac{p_i}{1-p_i}$ do not guarantee the minimum classification error, because the prior probabilities for the classes $P(\omega_j)$ have to be taken into account, too. More precisely, if the individual classifiers are independent, and the a priori likelihood is that each choice is equally likely to be correct, then the decision rule that maximizes the system accuracy is a weighted majority voting rule obtained by assigning weights $b_i \propto \log \frac{p_i}{1-p_i}$.

In contrast to the classical majority voting, we equip each classifier output with different weights b_i, where $0 \leq b_i \leq 1$ $(i = 1, \dots, n)$. It seems natural to give the classifiers with larger accuracies larger importance in making the final decision. Note that the classical majority voting scheme can be considered as a special case of the weighted voting system since in the majority rule the weight of each vote given by a classifier is constrained to be $b_i = 1$ for all $i = 1, \dots, n$.

4.2 Generalized weighted voting system

We can also assign weights to the classifiers within our generalized voting scheme presented in section 3. If we consider the classifiers $(D_1, D_2, \dots, D_n)$ with respective accuracies $(p_1, p_2, \dots, p_n)$ and weights $b_1, \dots, b_n$, then the final decision is made by choosing the maximal sum of weights, where some additional constraints (e.g. a geometrical one for OD detection) have to be fulfilled by the classifier outputs. Let us consider the probability $(1 - p_i)r_i$ with $r_i \in [0, 1]$ for the i-th classifier that means that the i-th classifier makes wrong classification and participates in making a bad decision.

In our application, we choose the maximal sum of those weights of the algorithms whose outputs can be bounded by a circle with an appropriate radius. An algorithm takes part in making a bad decision if its output falling outside the optic disc meets other bad candidates. For the algorithm D_i with accuracy p_i giving a bad candidate (x_i, y_i) for the optic disc, we consider that the distribution of (x_i, y_i) is uniform outside the optic disc for all i $(i = 1, \ldots, n)$. In this case, we have:

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0}, \tag{10}$$

where T_0 and T are the area of the optic disc and the ROI (whole useful image domain), respectively, so r_i is the same predetermined constant for all i $(i = 1, \ldots, n)$. For better understanding, see Fig. 5., where we show how bad candidates can fulfil the geometric constraint by falling inside a disc with OD radius.

Fig. 5. Retinal image and a schematic one with the used notation

The next theorem gives the answer on how to select the weights in our generalized weighted majority voting model.

Theorem 1. *If independent classifiers $(D_1, D_2, \ldots, D_n)$ are given (conditional independence is assumed), then the optimal weight b_i for the classifier D_i with accuracy p_i can be calculated as:*

$$b_i \propto \log \frac{p_i}{(1 - p_i)^2 r_i (1 - r_i)}. \tag{11}$$

Proof. Let $s = [s_1, \ldots, s_n]$ denote the vector with the label output of the ensemble, where $s_i \in \Omega$ is the label suggested for $\mathbf{x}$ by the classifier D_i. A Bayes-optimal set of discriminant functions based on the outputs of the n classifiers is

$$g_j(\mathbf{x}) = \log P((\omega_j)P(s|\omega_j)), \quad (j = 1, \ldots, c). \tag{12}$$

From the conditional independence, for the discriminant functions $g_j(\mathbf{x})$ we get

$$\log P(\omega_j)P(s|\omega_j) = \log\left[P(\omega_j)\prod_{i=1}^{n} P(s_i|\omega_j)\right] = \tag{13}$$

$$\log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} P(s_i|\omega_j) \prod_{i,s_i\neq\omega_j} P(s_i|\omega_j)\right) = \tag{14}$$

$$\log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} p_i \prod_{i,s_i\neq\omega_j} (1-p_i)r_i \prod_{i,s_i\neq\omega_j} (1-p_i)(1-r_i)\right) = \tag{15}$$

$$\log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} \frac{p_i(1-p_i)}{1-p_i} \prod_{i,s_i\neq\omega_j} (1-p_i)r_i \prod_{i,s_i\neq\omega_j} (1-p_i)(1-r_i)\right) = \tag{16}$$

$$\log P(\omega_j) + \log\left(\prod_{i,s_i=\omega_j} \frac{p_i}{1-p_i} \prod_{i,s_i\neq\omega_j} (1-p_i)r_i(1-r_i) \prod_{i=1}^{n}(1-p_i)\right) = \tag{17}$$

$$\log P(\omega_j) + \sum_{i,s_i=\omega_j} \log\frac{p_i}{1-p_i} + \sum_{i,s_i\neq\omega_j} \log((1-p_i)r_i(1-r_i)) + \sum_{i=1}^{n}\log(1-p_i). \tag{18}$$

The last term does not depend on the class label j so we can reduce the discriminant function to

$$g_j(\mathbf{x}) = \log P(\omega_j) + \sum_{i,s_i=\omega_j} \log\frac{p_i}{1-p_i} + \sum_{i,s_i\neq\omega_j} \log((1-p_i)r_i(1-r_i)) = \tag{19}$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j}\log\frac{p_i}{1-p_i} + \sum_{i=1}^{n}(1-d_{i,j})\log((1-p_i)r_i(1-r_i)) = \tag{20}$$

$$\log P(\omega_j) + \sum_{i=1}^{n} d_{i,j}\log\frac{p_i}{(1-p_i)^2 r_i(1-r_i)} + \sum_{i=1}^{n}\log((1-p_i)r_i(1-r_i)). \tag{21}$$

The last term of the summation is also independent from the class label j so it can be omitted. If we consider the equations:

$$g_j(\mathbf{x}) = \log P(\omega_j) + \sum_{i=1}^{n} d_{i,j}\log\frac{p_i}{(1-p_i)^2 r_i(1-r_i)}, \tag{22}$$

and

$$g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,j}, \tag{23}$$

we get that the weights:

$$b_i \propto \log\frac{p_i}{(1-p_i)^2 r_i(1-r_i)} \tag{24}$$

that are supposed to maximize the system accuracy.

Note that, similarly to classical majority voting, the weights given in (24) do not always guarantee the minimum classification error. Only if the individual classifiers are independent and the prior probabilities for the classes $P(\omega_j)$ are equal, the decision rule that maximizes the system accuracy is a weighted majority voting rule, obtained by assigning the above weights.

4.3 Weighted majority voting in OD detection

In our application, the output of each OD detecting algorithm is the OD center given as a single pixel with coordinates (x_0, y_0). In our ensemble-based system we have the set of class labels $\{\omega_{(x,y)} | (x, y) \in ROI\}$. For an OD detector (as a classifier) with its output (x_0, y_0), the class label $\omega_{(x_0,y_0)}$ is assigned to the detector. In other words, the classifier voted to the pixel (x_0, y_0) as OD center. The classification is considered to be correct if the output (x_0, y_0) falls inside the true optic disc on the retinal image. We can define the decision rule as the sum of the weights of the OD detecting algorithms, whose outputs can be bounded by a circle of the OD radius. Such a circle with the maximal sum of weights is accepted as the final decision for the OD.

In this application, the condition for the equal prior probabilities for the classes is fulfilled if we suppose uniform distribution of the candidates both inside and outside the optic disc.

In contrast to the non-weighted systems, less conflicting situations can be obtained when the decision is not exact because of the equal number of outputs falling inside the discs of the predetermined radius. Further improvement of this weighted system on majority voting is that there is no need for accuracy constraints $p > 0.5$ on individual algorithms to achieve larger system accuracy. It can be shown that this weighted voting rule always outperforms the classical majority rule because in case of a conflict (when the same number of votes are densified in different discs of a given radius) majority rule decides randomly between the disc candidates, while the weighted voting system can handle the conflict determining to the sum of the weights corresponding the output votes falling inside the discs.

5 Experimental results

We compare the system accuracies of the classical and the weighted majority voting for different accuracies and different weights. In our tests, we considered three different types of accuracies for the algorithms:

- $A_1 : p_1 = p_2 = \ldots = p_9 = 0.6$,
- $A_2 : p_i = 1 - 0.1i,\ i = 1, \ldots, 9$,
- $A_3 : p_1 = 0.6472, p_2 = 0.9765, p_3 = 0.3205, p_4 = 0.7593, p_5 = 0.3153,$ $p_6 = 0.2276, p_7 = 0.9582, p_8 = 0.7671, p_9 = 0.6432$.

The case A_1 is often examined in the literature with equal weights, A_2 is a theoretical example, while A_3 contains true accuracies of OD detecting algorithms measured on the Messidor test database [9] containing 1200 retinal images.

For comparative studies, we apply the following weights b_i for the i-th algorithm having accuracies p_i $(i = 1, \ldots, 9)$:

- $B_1 : b_i = p_i$,
- $B_2 : b_i = log\frac{p_i}{1-p_i}$,
- $B_3 : b_i = \frac{p_i}{(1-p_i)^2 r_i (1-r_i)}$.

That is, in case B_1 each weight is equal to the accuracy of the individual algorithm (such as taken the i-th algorithm with accuracy p_i, then it participates in the final decision with weight $b_i = p_i$). B_2 is suggested as optimal for the classical weighted majority voting, while B_3 is the proposed assignment for our generalized weighted majority voting. In this way, we give a practical example to confirm the theoretical derivation of the optimal weights given in section 4.2.

We apply OD detecting algorithms as classifiers, so we can test and compare the overall performance of the different voting systems on classifier output generated artificially. In lack of independent OD detecting algorithms providing these accuracies, we are not able to test and compare the voting systems on retinal images. We generate the classifier outputs in the following way: we consider a disc of radius R (ROI) and a disc of radius R_0 inside the ROI (optic disc), where $R = 712$ and $R_0 = 48$ pixels, respectively. We generate 9 output points with coordinates (x_i, y_i) (as outputs the D_i's), where the probability that the point (x_i, y_i) falls inside the optic disc is p_i and the distribution of (x_i, y_i) is uniform outside the optic disc. Now, the probability r_i $(i = 1, \ldots, 9)$ can be determined as:

$$r_1 = \ldots = r_n = \frac{T_0}{T - T_0} = \frac{R_0^2}{R^2 - R_0^2}. \tag{25}$$

In this test we compare the performance of the following voting systems: MV- majority voting, WMV- weighted majority voting, GMV- generalized majority voting, WGMV- weighted generalized majority voting. The system accuracies for the individual accuracy setups A_1, A_2, A_3 with the weight assignments (B_1, B_2, B_3) are given in Fig. 6(a)., Fig. 6(b)., Fig. 6(c)., respectively.

From the tables we can see that if all weights are equal, then it naturally results in the same system accuracy as the non-weighted voting scheme, otherwise, weighted voting outperforms non-weighted voting. Our generalized non-weighted (weighted) voting system has better overall performance than the classical non-weighted (weighted) majority voting scheme.

For the OD detection application, we can test and compare our generalized non-weighted and generalized weighted voting system on a real database of retinal images, as well. The Messidor dataset [9] considered for this aim contains 1200 retinal images. In this test, we assigned the optimal weights derived in section 4.2 to the participating algorithms (classifiers) having individual accuracies $p_1 = 0.6472, p_2 = 0.9765, p_3 = 0.3205, p_4 = 0.7593, p_5 = 0.3153, p_6 =$

A_1	MV	WMV	GMV	WGMV
B_1	0.7323	0.7323	0.9948	0.9996
B_2	0.7380	0.7380	0.9941	0.9991
B_3	0.7326	0.7326	0.9948	0.9989

(a) System accuracies for the set A_1

A_2	MV	WMV	GMV	WGMV
B_1	0.5012	0.8066	0.9889	0.9943
B_2	0.4965	0.9688	0.9901	0.8712
B_3	0.5009	0.7289	0.9877	0.9951

(b) System accuracies for the set A_2

A_3	MV	WMV	GMV	WGMV
B_1	0.8241	0.9526	0.9996	1.0000
B_2	0.8260	0.9926	0.9989	0.9941
B_3	0.8258	0.9481	0.9989	0.9998

(c) System accuracies for the set A_3

Fig. 6. Overall system accuracies for the set of classifier accuracies

$0.2276, p_7 = 0.9582, p_8 = 0.7671, p_9 = 0.6432$ (as given in case $A3$). However, note that we have no information about the dependencies among these algorithms. Despite the unknown dependencies of the algorithms, we found that weighted majority voting with its system accuracy 0.98 outperformed classical majority voting (system accuracy 0.974), and also all the individual accuracies.

6 Conclusion and future plans

We have introduced a new theoretical model that enables the investigation of majority voting systems being more general than the classical majority voting scheme. As for practice, we apply this generalization to set up ensemble of algorithms providing spatial output. This generalized voting system (when some additional geometrical constraints have to be fulfilled) can be applied in that case when weights are assigned to the classifiers, as well. In our specific application, larger overall system accuracy is achieved, than in the case of individual algorithms and weighted voting outperformed the non-weighted one. Same results can be expected for similar image processing problems, where the algorithms vote with a single pixel or region. In our application, adding a new independent algorithm to the system seems to be very effective because of the exponential behavior of the system accuracy. The full characterization of the participating algorithms to achieve the best system performance is still an open issue.

A further issue regarding for the accuracy of the system is the dependence of the algorithms. Though this paper concentrates on the independent case, it can be shown that the accuracy can drop/raise based on the dependencies of the algorithms similarly to the majority voting case [8]. To tune our system, it will be a future research direction to see how the accuracy can be raised by removing/adding algorithms from/to the existing system in consideration to individual accuracies and dependencies.

Acknowledgments. This work was supported in part by the Janos Bolyai grant of the Hungarian Academy of Sciences, and by the TECH08- 2 project DRSCREEN- Developing a computer based image processing system for diabetic retinopathy screening of the National Office for Research and Technology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/ 2008). Research is supported in part by the OTKA grants (K67580, K75566) and by the TÁMOP 4.2.1./B-09/1/KONV-2010-0007 project. The project is implemented through the New Hungary Development Plan, cofinanced by the European Social Fund and the European Regional Development Fund.

References

1. Harangi, B., Qureshi, R.J., Csutak, A., Peto, T., Hajdu, A.: Automatic Detection of the Optic Disc Using Majority Voting in a Collection of Optic Disc Detectors. In: 7th IEEE International Symposium on Biomedical Imaging, 1329–1332 IEEE Press, Rotterdam (2010)
2. Qureshi, R.J., Kovacs, L., Harangi, B., Nagy, B., Peto, T., Hajdu, A.: Combining algorithms for automatic detection of optic disc and macula in fundus images. Computer Vision and Image Understanding 116, Vol. 1, 138–145 (January 2012) (DOI information: 10.1016/j.cviu.2011.09.001)
3. Hansen, L.K., Salamon, P.: Neural Network Ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 993–1001 (1990)
4. Kuncheva, L.I.: Combining Pattern Classifiers, Methods and Algorithms. John Wiley & Sons, Inc., New Jersey (2004)
5. Toman, H., Kovacs, L., Jonas, A., Hajdu, L., Hajdu,A.: A Generalization of Majority Voting Scheme for Medical Image Detectors. Lecture Notes in Artificial Intelligence 6679, Vol. 2, 189–196 (2011)
6. Lam, L., Suen, C.Y.: Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 27, 553–568 (1997)
7. Appel, M.J.,Najim, C. A., Russo, R.P.: Limit Laws for the Diameter of a Random Point Set. Adv. in Appl. Probab. 34, Number 1, 1–10 (2002)
8. Altincay, H.: On Naive Bayesian Fusion of Dependent Classifiers. Pattern Recognition Letters 26, 2463–2473 (2005)
9. http://messidor.crihan.fr