Enhanced ensemble-based object detection in medical image analysis

Doktori (Ph.D.) értekezés

KOVÁCS LÁSZLÓ
TÉMAVEZETŐ: DR. HAJDU LAJOS

DEBRECENI EGYETEM
TERMÉSZETTUDOMÁNYI DOKTORI TANÁCS
INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA
DEBRECEN, 2018.

Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Informatikai Tudományok Doktori Iskola Diszkrét matematika, képfeldolgozás és komputergeometria programja keretében készítettem a Debreceni Egyetem természettudományi doktori (Ph.D.) fokozatának elnyerése céljából.
Debrecen, 2018.

.....
a jelölt aláírása

Tanúsítom, hogy Kovács László doktorjelölt 2010 - 2013 között a fent megnevezett Doktori Iskola Diszkrét matematika, képfeldolgozás és komputergeometria programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javaslom.
Debrecen, 2018.

.....
a témavezető aláírása

Enhanced ensemble-based object detection in medical image analysis

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
az informatikai tudományágban
Írta: Kovács László okleveles programtervező matematikus
Készült a Debreceni Egyetem Informatikai Tudományok Doktori Iskolája, Diszkrét matematika, képfeldolgozás és komputergeometria programja keretében

Témavezető: Dr. Hajdu Lajos

A doktori szigorlati bizottság:

elnök:	Dr. Kruppa András	.....
tagok:	Dr. Palágyi Kálmán	.....
	Dr. Fazekas Gábor	.....

A doktori szigorlat időpontja: 2013. október 25.

Az értekezés bírálói:

.....	.....
.....	.....

A bírálóbizottság:

elnök:	.....	.....
tagok:	.....	.....
	.....	.....
	.....	.....
	.....	.....

Az értekezés védésének időpontja:

Acknowledgement:

I would like to express my gratitude to my thesis supervisor Professor Lajos Hajdu and my coworker Professor András Hajdu for their encouragement and guidance during my PhD studies and doctoral candidate period. I am honored to have worked with the co-authors of my papers too.

I would like to thank to my family, especially my wife for their support, patience and understanding throughout the years.

Finally, some elements of this thesis were supported and financed by different programs and projects: GINOP-2.1.7-15-2016-01641, FUNDUSDOMUS: Automatic recognition and characterization of diseases by cellular analysis of cellular fundus images; H2020 HU-MATHS-IN, the Hungarian Service Network for Mathematics in Industry and Innovations; GINOP-2.2.1-15-2017-00055: Research on osteosynthesis of implants and development of trabecular structure using Additive Manufacturing; GINOP 2.1.1-15-2015-00376: Research and development of digital identification and automatic classification of visual TUMOR MARKERS; Hybrid Small Size HPC research project; VKSZ 14-1-2015-0072 SCOPIA: Development of software supported clinical devices based on endoscope technology; PRACE SoHPC - 3D CT segmentation and photorealistic visualization at IT4Innovations national supercomputing center, VSB - Technical University of Ostrava, Ostrava, Czech; OTKA, NK101680, Mathematical modeling of clinical observations for improved melanoma detection; TÁMOP-4.2.2.C-11/1/KONV-2012-0001: Future Internet Research, Services, Technology; 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; involving Nvidia, National Excellence Program of the Ministry of Human Capacities (Hungary) and the European Social Fund.

Contents

List of Figures

List of Tables

1 Introduction

Medical imaging has improved considerably and contributed to the creation of remarkable achievements in disease treatment over the past decades. Nowadays it is completely general to use medical imaging in everyday medical practice to support non-invasive or semi-invasive medical interventions. Thanks to the technical progress of computers, digitization of imaging brought in a new era in the field of medical imaging. The stored digitized data and improved technology allow creating more accurate and precise treatments. At the same time, the larger world population causes an increased demand for medical treatment, which stimulates the era of automatization in medical imaging. In recent years the number of the technology-driven developments are increasing for better treatments becoming stoppable in the age of small and mobile computers. Despite of the clear progress, medical image processing is a non-trivial task, and is one of the most widely researched fields. Due to the increasing size and diversity of data obtained by better and better sensors, the most important challenges in this field are the complexity of the image processing problem and the computation time of its algorithmic solutions. To reduce complexity and raise the accuracy of the solutions, splitting the problem into subtasks is a known answer. The detection of the location of an object on an image is one of the most common tasks in the field of medical image processing. In this task a classification procedure of the image pixels is solved by the help of classifiers. The reason for composing ensembles from individual classifiers is to make joint classification from the aggregated data instead of using only one classifier. These ensembles are popular to use since they can raise the accuracy of the final answer and make it possible to apply existing research results. For decision making, majority voting is a usually applied model. Although the ensembles can reduce the complexity in the level of sub-tasks, the complete ensemble-based software solution mostly has great computational demand.

Under my thesis work, I focused on the investigation of ensemble-based image processing involving optimization which increases the accuracy and decreases the computing time. In my dissertation, novel results are reported about ensemble-based enhanced object detection covering extension, optimization of the ensembles. I gave solutions to respond high-performance computing demand also.

The structure of the dissertation is the following. After a short introduction in chapter 2, a novel ensemble-based combined object detection algorithm is introduced. It is studied in chapter 3 how to improve further the performance of ensemble-based object detection with adding an additional

classifier based on the its theoretical model. Chapter 4 contains comprehensive examinations about optimization of ensemble-based methods. Namely, in section 4.1, the generalized diversity-based ensemble optimization results are described. To follow the ensemble optimization consideration, a new approach is presented in section 4.2 to compose an ensemble with maximal performance under certain constraints. Finally, in section 4.3 a hardware-based solution with optimal task scheduling is given in a high-performance computing environment for ensemble-based solutions. Finally, a summary is given in chapter 5 in English and in chapter 6 in Hungarian.

1.1 Object detectors for practical applications

The retina is a delicate light-sensitive lining located at the back of our eye that "takes pictures" and sends the images to the brain. Diabetic retinopathy (DR) is the damage to the eye's retina that occurs with long-term diabetes, which can eventually lead to blindness. The blood vessels that provide nourishment to the retina may weaken and leak, forming small, dot-like hemorrhages in case of a person with diabetes. These leaking vessels often lead to swelling or edema in the retina and decreased vision. The fluorescein angiography and retinal photography are the most commonly used medical imaging of eye that provides ophthalmologists the possibilities to monitor the progression of the disease and to make decisions for the appropriate treatment. Screening programs for DR are being introduced, and automation of image grading would have some benefits. However, an important prerequisite for automation is the accurate location of the main anatomical features in the image, notably the optic disc (OD) and the macula. The OD is a circular shaped anatomical structure with a bright appearance. It is the location where the optic nerve enters the eye. If the position and the radius of the optic disk are detected correctly, then they can be used as references for approximating other anatomical parts, e.g., the macula and the fovea. The macula is located roughly in the center of the retina, temporal to the optic nerve. It is a small and highly sensitive part of the retina responsible for detailed central vision. The macula allows us to appreciate details and perform tasks that require central vision such as reading. The fovea is the very center of the macula, the site of our sharpest vision, see Figure 1.

Information about the locations of these features is necessary because the severity and characterization of abnormalities in the eye partially depend on their distances to the fovea and OD. The precise identification of the macula is important to enable successful grading of macular pathology such as diabetic maculopathy (clinically significant macular edema, macular ischemia). In the followings we introduce the collected set of OD and macula detectors to benefit from their predictions for finding a more accurate OD and macula center.

1.1.1 Description of algorithms used for detecting optic disc

Although numerous OD algorithms have been developed, a major concern in constructing ensembles is done how to select appropriate algorithms as ensemble components. The challenge is that there is no single algorithm that can outperform any other algorithms in all respect, i.e., to reduce the complexity and computational time while achieving better performance. De-

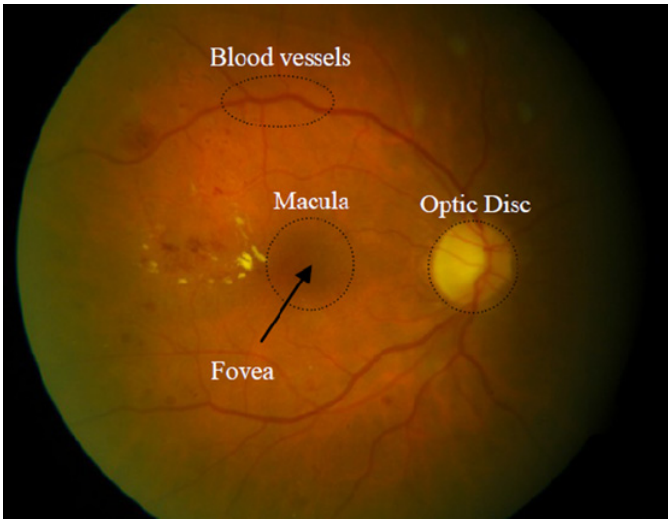


Figure 1: A retinal fundus image and main anatomical features.

spite of these fact, We propose adopting the following OD detectors found to achieve better predictive accuracy when combined. Nevertheless, much effort has been devoted to this area, and it needs further attention as well. We discuss now the key OD detector algorithms that have been improved and implemented.

1.1.1.1 Optic disc detection based on pyramidal decomposition (OD_{pd})

This algorithm relies on three assumptions. First, the image is centered on the macula or OD, second, the OD represents a bright region (not necessarily the brightest) and finally, the form of the OD is approximately circular. Based on the hypothesis that the OD is roughly a circular patch of bright pixels surrounded by darker pixels Lalonde et al. [44] proposed to locate the candidate OD regions on the green plane of the original image by mean of pyramidal decomposition (Haar-based discrete wavelet transform). In the low-resolution image pixels which have the highest intensity values compared to the mean pixel intensity over the search area were selected as possible candidates. Next, smoothing is done within each of these regions, and the brightest pixel is selected as a possible OD center point, and its confidence value is computed as the ratio of average pixel intensity inside a

circular region centered at the brightest pixel and the average intensity in its neighborhood.

1.1.1.2 Optic disc detection based on edge detection (OD_{ed})

In this method, introduced by Lalonde et al. [44], we search the area identified by the pyramidal decomposition as described above for a circular shape. To reduce the number of regions of interest, contiguous regions were aggregated into a single zone. A binary edge map is obtained by performing Canny edge detection [51] firstly in the region of interest, and then a thresholded image I_T is obtained with a special threshold value computed from noisy edge map. The search for the OD contour is performed using an algorithm based on Hausdorff distance [51]. The Hausdorff distance provides a degree of mismatch between two sets of points, defined as

$$H(A,B) = \max\left(h(A,B), h(B,A)\right) with\ h(A,B) = \max_{a\in A}\min_{b\in B} \parallel a-b \parallel . \tag{1}$$

Several circular templates of variably sized diameters were used to compute the Hausdorff distance between the templates and thresholded image I_T containing edges. Hence, a percentage of matches are computed, and if the certain proportion of the pixels template is found to overlap edge pixels in I_T then the location is retained as the center point of a potential OD candidate.

1.1.1.3 Optic disc detection based on entropy filter (OD_{ef})

Sopharak et al. [10] presented the idea of detecting the OD by entropy filtering. The original RGB image is transformed into HSI color space, median filtering is applied to remove possible noise and for contrast enhancement contrast limited adaptive histogram equalization (CLAHE) [51] is done after. After preprocessing, OD detection is performed by probability filtering, using the following equation:

$$H(I_x) = \sum_{i=0}^{255} P_{I_x}(i) \log P_{I_x}(i) \tag{2}$$

where P_{I_x} is the probability mass function of the pixel intensities I_x in a local neighborhood of x. Binarization is done with Otsu's algorithm [49]

to separate the complex regions from the smooth ones, and the largest connected region with an approximately circular shape is marked as a candidate for the OD.

1.1.1.4 Optic disc detection based on feature vectors and uniform sample grids (OD_{fv})

Niemeijer et al. [46] defines a set of features based on vessel map and image intensity, like the number of vessels, the average width of vessels, standard deviation, orientation, maximum width, density, average image intensity etc., measured under and around a circular template to determine the location of the OD. After pre-processing, each image is scaled so that the width of its field of view (FOV) is 630 pixels. The binary vessel map obtained [45] is thinned until only the centerlines of the vessels remain and all the centerline pixels that have two or more neighbors are removed. Next, the orientation of the vessels is measured by applying principal component analysis on each centerline pixel with its neighboring pixel on both sides. A two-step sampling process is launched to get the training database. First, a circular template of radius $r = 40$ pixels having manually selected OD center within the radius are used, and all features are extracted for each sample location (a uniform grid spaced 8 pixels apart) of the template including distance d to the true center. In the second step, 500 randomly selected location (i.e., not on a grid) in the training image were sampled similarly. During the OD location process, a sample grid (grid points spaced 10 pixels apart) is overlaid on top of the complete FOV and feature vectors are extracted. A rough location of the OD is found containing pixels having the lowest value of d. The process is repeated with a 5x5 pixel grid centered on the rough OD location to get the more accurate OD center.

1.1.1.5 Optic disc detection based on fuzzy convergence of blood vessels (OD_{fu})

Hoover et al. [9] developed a procedure to localize OD via a voting map of fuzzy segments. The voting map accumulates the fuzzy segments gathered from the detected segments of blood vessel systems. The input parameter of the method is binary image foregrounded blood vessels. From this input, the line-shape segments are modeled with fuzzy segments. The segments are gathered from the vessel network by the help of the procedure of thinning and fracturing the vessel network into smaller pieces. The voting map has the same resolution as the original image - on which the OD should be located - and will contain probabilities in the case of all pixels based on the strength

of its convergence. The final candidate of the algorithm is the strongest convergence detected via the computation of blurred and threshold voting map.

1.1.1.6 Optic disc detection based on Hough transformation (OD_{ht})

Ravishankar et al. [53] tried to track the OD by combining the convergence of the only thicker blood vessel initiating from it and high disk intensity properties in a cost function. On the initially resized image standard resolution (768 x 576), a grayscale closing operation is performed on the green channel image. This step is followed by thresholding and median filtering to obtain the binary image of the blood vessels. The segments of the thicker blood vessels skeleton are modeled as lines found by the Hough transform. The dataset of lines generated is reduced by removing those lines with slopes $\theta < 30°$. This reduced dataset of lines is intersected pairwise to generate an intersection map. The map is dilated to make the region of convergence more apparent. A weighted image is produced by combining this dilated intersection map and a preprocessed green channel image. A cost function is defined to obtain the optimal location of the OD that is a point which maximizes the cost function.

1.1.1.7 Optic disc detection based on temporal arcade (OD_{ta})

Fleming et al. [3] proposed to identify the macula region based on the information of the temporal arcade and OD center. First, the arcade was found by using semielliptical templates having a range of sizes, orientations, and eccentricities and having right- and left-handed forms. Next, the OD has to be detected by using a Hough transform with circular templates having diameters from 0.7 to 1.25 OD diameter (DD). Finally, the fovea was detected by finding the maximum correlation coefficient between the image and a foveal model. The search was restricted to a circular region with diameter 1.6 DD centered on a point that is 2.4 DD from the OD and on a line between the detected OD and the center of the semi-ellipse fitted to the temporal arcades.

1.1.1.8 Optic disc detection based on circle Hough transformation (OD_{ch})

Similarly to the published algorithm in the article Zhu at al. [63] we developed an own algorithm based on the outer border of the OD is determined by the help of the circle Hough transformation. This Hough transformation

maps the 2D image space into the 3D image space where the components of the dimensions are the coordinates of a center and a radius of a fitted circle. The main goal is to determine the best-fitted circle via 3D Gaussian filtering. As sub-procedures, the methods also contain grayscale conversions, and Canny edge detectors to localize the changing intensity on the image are also computed in advance. As the final step after the normalization of the candidate center is computed. The middle point algorithm using this candidate center determines the OD plate.

1.1.2 Description of algorithms used for detecting macula

In this section, we present an overview of those macula detector algorithms that were selected for the ensemble-based system.

1.1.2.1 Macula detection based on intensity (MC_i)

In [58] a region of interest (ROI) is defined to process macula detection. A Gaussian low pass filter is applied to smooth the intensity of the image. The statistical mean and standard deviation of the ROI area is used to compute a threshold for segmentation to get binary objects. The object that is located nearest to the center of the ROI is labeled as the macula of the retina. Its center of mass is considered to be the center of the macula. However, we did some modification to this approach, because it is not mentioned in the article how this ROI is defined; therefore instead of restricting the process to a specific ROI we applied it to the whole image. The smoothing of the image is done using a large kernel (70 x 70 pixels with Sigma = 10) so that vascular network and small patches do not interfere in detection. Then, the thresholding based segmentation process is launched which generates a set of binary images corresponding to different threshold values in the iteration. The object satisfying the area and distance from the center constraints are identified in each binary image. Finally, the object found nearest to the center with the minimum surface area is marked as the macula.

1.1.2.2 Macula detection based on spatial relationship with the optic disc (MC_{sr})

In [54] a region of interest for macula is defined using its spatial relationship with the optic disk. That is, the portion of a sector subtended at the center of the optic disk by an angle of 30° above and below the line between this center and the center of the retinal image disk. The macula is identified within this ROI by iteratively applying a threshold, and then applying morphological

opening (erosion followed by dilation) on the resulting blob. The value of the threshold is selected such that the area of the smoothed macula region is not more than 80% of that of the detected optic disk. The fovea is simply determined as the centroid of this blob.

1.1.2.3 Macula detection based on temporal arcade (MC_{ta})

Fleming et al. [3] proposed to identify the macular region based on the information of the temporal arcade and OD center. First, the arcade has to be found by using semielliptical templates having a range of sizes, orientations, and eccentricities and having right- and left-handed forms. Next, the OD was detected by using a Hough transform with circular templates having diameters from 0.7 to 1.25 OD diameter (DD). Finally, the fovea was detected by finding the maximum correlation coefficient between the image and a foveal model. The search was restricted to a circular region with diameter 1.6 DD centered on a point that is 2.4 DD from the OD and on a line between the detected OD and the center of the semi-ellipse fitted to the temporal arcades.

1.1.2.4 Macula detection based on Watershed and morphological operators (MC_{wm})

Zana et al. [25] presented a region merging algorithm based on watershed cell decomposition and morphological treatments for macula recognition. After noise removal, morphological closing followed by opening is performed to remove the small dark holes and white spots. A watershed-based decomposition of the gradient image into cells is done, and the cell with darkest gray level inside the macula is selected as the first step of a merging algorithm. A complex criterion based on the gray values and on edges of the filtered image is calculated to merge the cells of the macula while rejecting perifoveal inter-capillary zones to produce the contour of the macula.

1.1.2.5 A novel approach (MC_n)

Besides the algorithms discussed so far, we have also tested our novel contribution for macula detection which has been not published yet. We used state-of-the art image processing tools in a new fashion but similarly in OD_ef thanks to the similarity between OD and macula. First, we extract the green channel form the color fundus image (RGB). Then, we generate the background image by applying an AxA median filter and subtract this background image from the green plane, resulting in a shade corrected

image. Next, we binarize the image by considering all-non zero pixels as foreground pixels, and others as background. Finally, we apply an image labeling procedure and select the largest component as the macula.

1.2 Databases

To take advantage of the theoretical foundations, we used different datasets. The most used databases contain images with different field of view (FOV) resolutions. For instance, the MESSIDOR dataset [19] contains 1200 digital images, with 45° FOV and 1440×960, 2240×1488, and 2304×1536 resolutions in pixels. For the evaluation purposes, all the images and related descriptive data were rescaled to 1500×1152 in the normalization procedure. The ground truth information was manually created by a specialist from the relevant field. The used databases are listed in Table 1.

Table 1: Databases used for experimental results.

DB Names	Number of data
Diaretdb0	130
Diaretdb1	89
Drive	40
Hei-med	169
Kaggle	53576
Messidor	1200
Moorfields	3000
Programming lab test	6750

After this introductory chapter, we discuss our combined object detection motivated by medical applications. At the beginning of the chapter we review shortly the basic results of that theoretical foundation from the literature which are generalized by us. In this manner we give a solution which can be analyzed and evaluated experimentally. Moreover the theoretical studies are also possible to conclude the limits of the system in real applications.

2 Combined object detection of optic disc and macula based on anatomical features

2.1 Majority voting in object detection

The data fusion is becoming increasingly frequent not exclusively in the field of image processing. In the field of image processing in most cases is realized the majority voting as a decision mechanism based on the data generated by different algorithms as an answer to the same question. Mathematically this can be described as follows.

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. In the majority voting, the class label ω_i is assigned to $\mathbf{x}$ regarding the majority votes of the classifiers from the set D.

As a special case, we can consider binary classifiers examined exhaustively in the literature. Let n be odd, $\Omega = (\omega_1, \omega_2)$ (each classifier output is a binary vector) and all classifiers have the same classification accuracy p. 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:

$$P = \sum_{k=0}^{[n/2]} \binom{n}{k} p^{n-k} (1-p)^k. \tag{3}$$

Applying the majority voting in pattern recognition, several interesting results can be found in [39] (e.g., adding one or two new classifiers to the voting system). It is also prooved if the classifiers are independent and $1 > p > 0.5$, then this method is guaranteed to outperform the individual classifiers.

In the following section we will present our algorithmic solution for a real image processing object detection problem. As it can be seen in later chapters, the basis of the algorithm has a model based mathematical background, and there are general theoretical results also for extending, optimizing via measuring diversity among classifiers or taking into account execution time.

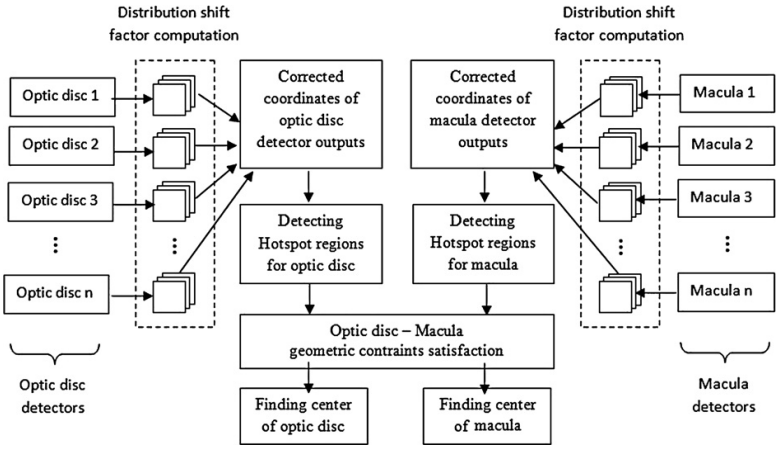


Figure 2: Flowchart showing the steps of the proposed technique.

2.2 Combining algorithms for automatic detection of optic disc and macula in fundus images

As discussed (also in [21]), an ensemble classifier is more accurate than any of its individual members if the individual classifiers are doing better than random guessing. Because, if the algorithms are complementary, then when one or a few algorithms make an error, the probability is high that the remaining algorithms can correct it. The main steps involved in the proposed system are depicted in the flowchart see Figure 2, which presents a compact summary of our approach. Extensive tests have shown that combining the predictions of multiple detectors is more accurate than any of the individual detectors in the ensemble.

2.2.1 Computing distribution shift factors

The macula/OD centers detected by a particular algorithm in all test images are mapped onto a single image to check the distortion of its distribution help by training it in advance. We observed that the outputs generated by the algorithms are quite dispersed and deviated from the manually selected macula center. In Figure 3a, the Gaussian kernel density estimation of the MC_I [58] algorithm outputs are shown for the macula and of the ODfv [46] algorithm for the OD, respectively. Therefore, we propose to compute the distortion in the data and applying a shift operation before an actual

combination of outputs for finding macula center is performed to make the individual algorithms unbiased. To compute the shift factor, we calculate the average difference between the candidate of the algorithm and the manually selected macula center for those n images, where the algorithm successfully found the macula/OD region. That is, there are shifting vectors for the x and y coordinates:

$$d(x) = \frac{\sum_{i=1}^{n} (x_i - x_{mi})}{n}, \quad d(y) = \frac{\sum_{i=1}^{n} (y_i - y_{mi})}{n} \tag{4}$$

where (x_i, y_i) stands for the fovea/OD center candidate of the algorithm on the i-th image, while (x_{mi}, y_{mi}) is for the manually selected fovea/OD center. The new output distribution is generated by applying the distortion shift factor on each output pixel coordinate. The improvement on the result can be seen in Figure 3.

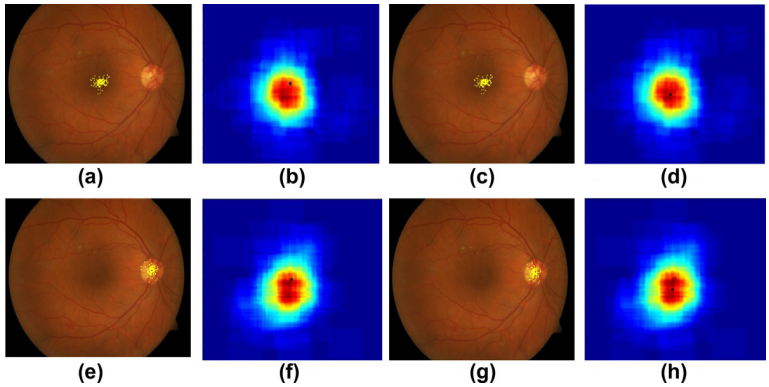


Figure 3: a-d) The distortion in macula detector outputs [58] before and after applying a shift factor, e-h) the distortion in OD algorithm [46] outputs before and after applying a shift factor.

2.2.2 Detecting hotspot region for OD/macula

2.2.2.1 OD and macula detection separately

We devised a circular template voting scheme to determine the hotspot region [13, 14], i.e., an area in the image where the majority of the outputs lies. A circular template of radius R is fitted on each pixel in the image

and outputs of candidates of the algorithms that fall within the radius of the predefined circular template are counted. The circular template covering the maximum number of OD detector outputs in its radius is considered to be a hotspot. There can be more hotspots covering the maximum number of detector outputs; hence they together define a hotspot region. That pixels have the highest probability value to be the final candidate which are covered most by the hotspots inside the hotspot region. The radius R of the template was set to 102 pixels, keeping in view the fact that clinically this is the average OD radius at the investigated resolution (field of view FOV is equal to 1432). A circular template voting scheme has an intuitive appeal, and our results show it works well in practice. Overall, we have used the outputs of the five algorithms for OD detection and five algorithms for macula detection, as discussed previously, in our combined system. Following the principle of majority voting, the center of the circular templates covering the maximum number of OD detectors and macula detector outputs in its radius is considered to be the OD and macula hotspots, respectively. If there is a tie then such conflict is handled using additional post-processing, e.g., for OD detection, the Gaussian filter is applied on the green channel of the image with a large sigma ($\sigma = 300$). The smoothened image is then subtracted from the original image to get the resultant image, a rather darker image in which OD appears as a brighter patch as compared to the background image. The average intensity around the detectors' outputs is computed using the same circular template, and the template with the highest average intensity is selected as the best OD hotspot region.

In case of macula detection, we observed and tested that the vessel based macula detection algorithm MC_{ta} [3] influenced the combination result, therefore, whenever there is a conflicting situation or ties among hotspot regions, the hotspot region containing the output of algorithm MC_ta is selected for further processing. If the output of MC_{ta} is lying isolated and not as part of any hotspot region detected, then the situation is handled by computing distance between hotspot centers and the center of the image followed by selecting the closest region.

Sometimes, majority voting may fail to detect the correct OD or macula center, especially in case of the diseased retina (see Figure 4, where the template with three outputs is assumed to be the best hotspot for OD).

Therefore, we proposed a more sophisticated alternative method to avoid these shortcomings, explained in the following section.

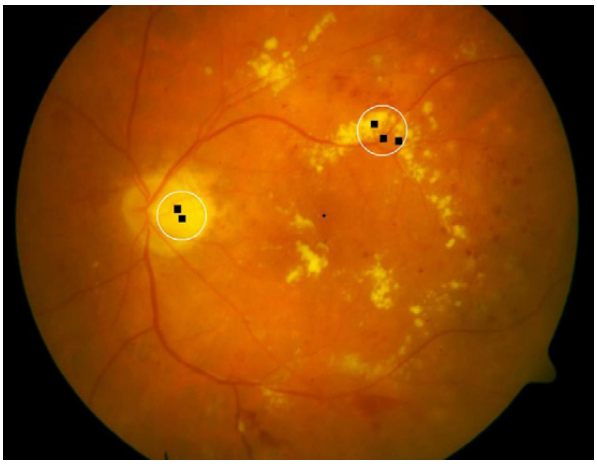


Figure 4: False positive reported in the majority voting during OD detection.

2.2.2.2 Optic disc and macula detection based on mutual information

Our idea was to raise the accuracy by exploiting the anatomical constraints between OD and macula determined by doctors, particularly the information about distance and angle criteria can be used to choose the best candidate hotspot region pair. The minimum radius bounding circle around a set of points is a simple measure of the area they occupy, as well as a useful tool in graphical applications. Welzl [22] proposed a simple randomized algorithm for the minimum covering circle problem that runs in provabilty linear running time. Therefore to reduce the image processing needs instead of scanning the whole image, a minimum radius bounding is used to locate the hotspot region as being the candidate of OD/macula. For this purpose an unordered collection of different OD detectors and macula detectors are produced separately using the following equation:

$$C(n, r) = \frac{n!}{r!(n-r)!} \tag{5}$$

where n, r are nonnegative integers and $r \leq n$. Here, n is the number of detectors, and r is the number of items taken from n elements. The minimal enclosing circle algorithm considers every circle defined by these set of detectors generated using combination without repetition. The algorithm returns the center and radius of the circle defined by detectors. There can be many

such circular regions, however, to validate a region as being hotspot region (candidate for OD/macula), the OD radius is used as threshold criteria to accept or discard a circle,. i.e., if the radius of the minimal enclosing circle is less than or equal to the OD radius, then such a circle is considered as a candidate for OD/macula. For all hotspot regions for macula and OD detected in this way, we calculate the score using their mutual information and the candidate hotspot pair with maximum score is selected as the best hotspot region for macula and OD. The score of a pair is based on the cardinalities, i.e., a number of outputs present in a hotspot and penalties in term of distance and angle errors:

$$Score(MC_i, OD_j) = |MC_i| \times |OD_j| - \frac{D_{err}}{D_{avg}} - \frac{A_{err}}{A_{avg}} \tag{6}$$

Here, MC_i is the i-th macular candidate hotspot and OD_j is the j-th OD candidate hotspot region, $|.|$ stands for the cardinality of a set. D_{err} and A_{err} is the distance and angle error respectively, obtained as a difference between computed values and average values. Using manually marked macula and OD centers, we found that the average distance (D_{avg}) is 114 pixels and the average angle (A_{avg}) is $\pm 6°$. Figure 5a shows the macula and OD candidates representing triangles and squares respectively. So the dashed circles represent hotspots of the OD, while the circle with continuous line are for macula candidates. In Figure 5b, all possible OD/macula pairs are shown (dashed lines), the solid line between OD_1 and M_1 has the maximum score, i.e., they satisfy best the cardinality, distance and angle criteria.

2.2.3 Center of the hotspot based on a weighted combination

The center of the final hotspot regions could be found by averaging algorithms' outputs, however, for more accurate estimation, weights can be associated with detector outputs to determine the final location. The principal difficulty is how to choose the 'right' weights. By following classical statistical recommendations, we assumed the observed data set to be a linear combination of a certain basis. Therefore, weights w_i can be computed for the x-coordinates and y-coordinates of the detector outputs.

If we denote the output centers of the individual algorithms by random variables $S_1(x_1, y_1), S_2(x_2, y_2), \ldots, S_N(x_N, y_N)$ with distinct variances $Var(S_i) = \sigma_i^2 > 0$, then the problem of output combination is to reduce these N outputs to one final center $S_w(x_w, y_w)$. An appropriate combination of the outputs can be the weighted linear combination

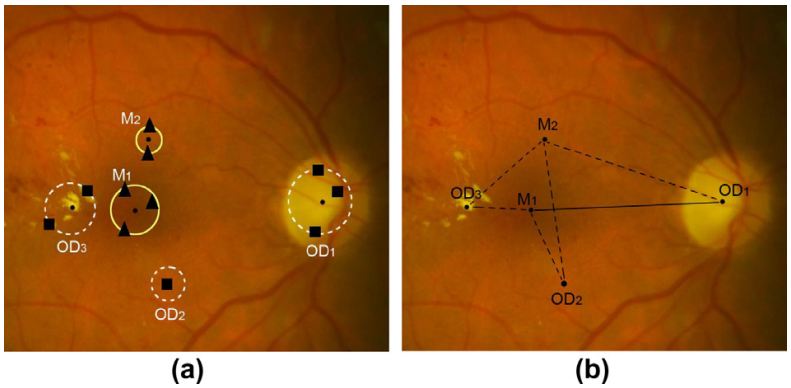


Figure 5: A retinal image with spotted hotspot regions for macula and OD (a), and M1 and OD1 satisfying geometric constraints with highest score (b).

$$S_w = \sum_{i=1}^{N} w_i S_i \tag{7}$$

where $w_1, w_2, \ldots, w_N$ are non-negative weights, constrained to sum to 1. This constraint guarantees the combined estimate to remain unbiased. The variance of the estimator S_w is determined by choice of weights and the variances of the individual outputs. When the random variables are independent, then the variance of the estimator S_w is determined by choice of weights and the variance of individual outputs. In this case, we have to minimize the expression below

$$Var\left(S_w\right) = \sum_{i=1}^{N} w_i^2 Var\left(S_i\right). \tag{8}$$

In this way the choice of weights that minimizes the variance of the combined estimate is

$$w_i = \frac{\sigma_i^2}{\sum\limits_{j=1}^{n} \sigma_j^{-2}} \tag{9}$$

as a result of [62]. The choice of these weights leads to small variance. Hence, the expected Euclidean distance from the true center of the combined

estimate can be minimized. The higher the weight for output, the more that detector is trusted to provide the correct answer. In the other case, when the random variables are dependent, then we have to consider the pairwise covariances, as well [41]. We minimize

$$Var\left(S_w\right) = \sum_{i=1}^{N} w_i^2 Var\left(S_i\right) + \sum_{1 \leq i < j \leq N}^{N} w_i w_j Cov\left(S_i, S_j\right) \tag{10}$$

to get the optimal combinations for the weights. We can also write equation (10) as $Var\left(S_w\right) = w^T D w$, where w is a vector containing the weights and D is a matrix containing the pairwise covariances. This minimum problem can be solved via derivation i.e., $\frac{\partial Var(S_w)}{\partial w_i} = 0$ should be solved for the N variables $w_1, w_2, \ldots, w_N$ to determine the proper weights.

In our setup, we have compared all the macula and OD detectors, respectively, with χ^2 – test to check their independence. Then, according to the above description we removed the term $Cov\left(S_i, S_j\right)$ from equation (10), when the i-th and j-th algorithms were proven to be independent.

2.2.4 Results and discussion

We have evaluated our proposed combination of algorithms to localize OD and macula on the publically available three benchmarks listed in Table 2.

Table 2: The training and test databases used.

Benchmark	Resolution	No. of images
Diaretdb0 [9]	1500×1152	130
Diaretdb1 [10]	1500×1152	89
Drive [11]	565×584, 730×490	40

The proposed algorithm should be trained on the training database (without the influence of the test database). Therefore, the set of images from the three benchmarks listed in Table 2 has been split into a training database and a test database. The training database contains 40% of the images from each benchmark, and the test database contains the remaining 60% of the images. To reduce variability, multiple rounds of cross-validation are performed using different partitions, performing the analysis on one subset (the training set), and validating the analysis on the other subset (the validation set or testing set).

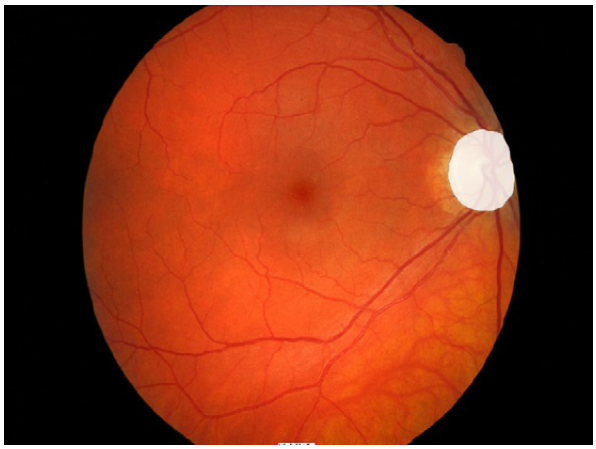


Figure 6: A retinal image with manually selected OD patch.

2.2.4.1 Optic disc detection results

The methods have been evaluated on the basis of two criterions, i.e., to fall inside the manually selected OD patches (see Figure 6), and to be close to the manually selected OD center. Table 3 shows the correct detection of the OD location based on OD patch; the percentage detection rate of the proposed approach is higher than any of the individual algorithms. In Table 4 we can see the secondary error term that measures the average Euclidean distance of the candidates from the manually selected OD centers. We can observe that the combined system led to more accurate center localization than any of the individual algorithms.

Table 3: Candidates falling inside manually selected OD patch in percentage.

Test database	OD detector algorithms					
	OD_{pd}	OD_{ed}	OD_{fv}	OD_{ef}	OD_{ht}	Simple OD majority voting
Diaretdb0	89.52	77.56	78.20	95.29	80.12	96.79
Diaretdb1	88.99	75.46	77.04	93.70	76.41	94.02
Drive	80.55	97.22	67.35	98.61	86.10	100
Total avegare	87.95	79.88	76.12	95.26	79.78	96.34

Table 4: Average Euclidean distant error of the OD candidates from the correct in pixels.

Test database	OD detector algorithms					
	OD_{pd}	OD_{ed}	OD_{fv}	OD_{ef}	OD_{ht}	Simple OD majority voting
Diaretdb0	15.89	20.63	18.55	11.89	11.41	12.23
Diaretdb1	17.38	20.84	20.41	13.05	11.95	11.95
Drive	23.31	13.21	20.69	16.98	14.67	15.95
Total avegare	17.54	19.55	19.52	13.07	12.10	12.71

Table 5: Candidates falling inside MC region in percentage.

Test database	OD detector algorithms					
	MC_i	MC_{sr}	MC_{ta}	MC_n	MC_{wm}	Simple MC majority voting
Diaretdb0	67.94	70.29	82.47	86.10	60.67	93.58
Diaretdb1	60.373	77.66	78.29	91.50	71.69	99.68
Drive	72.06	81.07	58.55	81.07	81.07	91.88
Total avegare	65.77	74.24	78.03	87.39	67.10	95.53

2.2.4.2 Macula detection results

To localize macula and detect fovea, we tested the five algorithms listed previously on the same three databases of Table 2.

The methods have been evaluated on the basis of two criterions, i.e., macula error and fovea error. The macula error is the number of times the algorithm's output falls within the $0.5DD$ radius of the manually selected macula center (Table 5), and fovea error, i.e., the average Euclidean distance of these candidates and the manually selected centers (see Table 6). Results are given in pixels at the resolution of $ROIdiameter = 1432$ (average ROI diameter of Diaretdb0 and Diaretdb01). Weights are calculated by considering the algorithm's dependencies. Just like the OD, the combined system provided more accurate results than any of the individual algorithms both for the primary error and the secondary error terms.

2.2.4.3 Joint OD-macula detection results

Research in ensemble methods has largely revolved around designing ensembles consisting of competent yet complementary models. The goal of the proposed approach is to create a more stable and achieve performance im-

Table 6: Average Euclidean distant error of the fovea candidates from the correct in pixels.

Test database	OD detector algorithms					
	MC_i	MC_{sr}	MC_{ta}	MC_n	MC_{wm}	Simple MC majority voting
Diaretdb0	29.98	31.63	52.13	35.57	29.28	29.64
Diaretdb1	27.05	29.93	47.88	28.86	27.75	24.26
Drive	16.52	24.19	61.68	36.08	22.19	26.80
Total avegare	27.27	30.11	51.81	33.25	27.86	27.38

Table 7: OD and MC detection based on mutual information.

Test database	Joint OD-MC detection	
	MC	OD
Diaretdb0	96.79	97.64
Diaretdb1	98.74	97.79
Drive	91.73	100
Total avegare	96.87	98.06

provement with a more precise algorithm via a precisely defined decision mechanism. The results demonstrate that the use of mutual geometric constraints between OD and macula have improved upon traditional approaches like simple averaging or voting. Weighted averaging has proven to be very effective and has achieved better prediction accuracy than any of the individual algorithms could have on their own (Figure 7). The use of geometric constraints between OD and macula is a unique contribution which has been rewarded as it outperforms all individual decisions and separate combination as shown in Table 7. Therefore, practitioners in diabetic retinopathy should stay tuned for further developments in the vibrant area of multiple algorithm-based decision making.

2.2.4.4 Conclusion

We developed a framework for combining state-of-the-art OD/fovea detector in a flexible way. The aim was to give an intuitive understanding of the potential advantage of ensemble methods in term of improving performance as compared to individual algorithms. We showed that our combination outperformed all the individual algorithms. We have used a majority voting and weighted linear combination based scheme that counts the number of outputs of the algorithms falling in a circle of a specified radius is marked

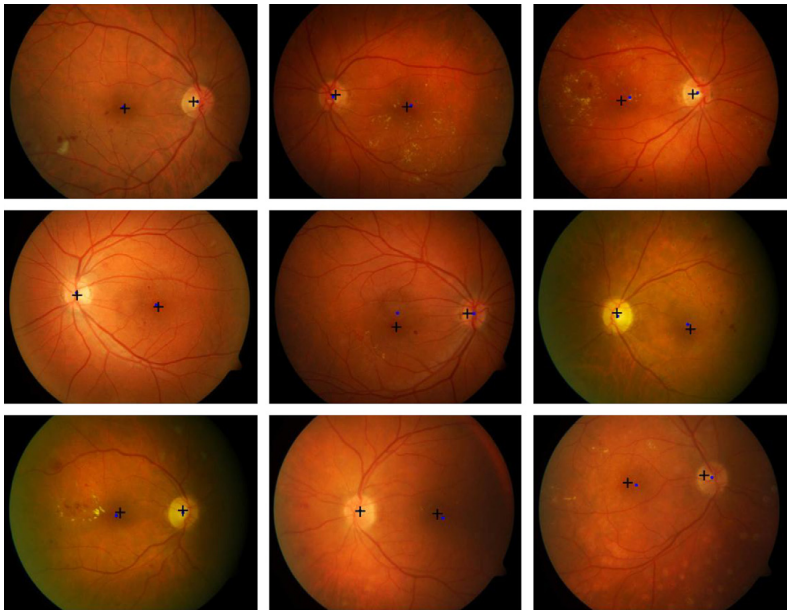


Figure 7: Joint macula-OD detection results, the plus signs indicates system outputs against dots (manually choosen by an ophthalmologist).

as a hotspot. Our method achieved the highest performance, and closest to the manually selected OD center chosen by a retinal specialist. This proposed combination is just one possibility; there is still room for further improvement through future research by combining other methods. We tried to check the influence of each algorithm present in our combination system by excluding it and running the test again with the remaining of algorithms. In both cases, i.e., OD and macula we found that all the algorithms were potentially useful and excluding any of them did not improve the results, rather it decreases the detection rate.

In the following chapters based on our generalized model for object detection we will investigate how the behavior of the algorithms influences the performance of a fusion system. In the next chapter, we will concentrate on what one can expect when extending the object detection fusion system with additional detectors.

3 Extending an ensemble-based system with an additional classifier

As it was demonstrated in the previous chapter, a well-built fusion system adjusted to the investigated problem can lead to a robust and high-performance level object detection. Although the object detectors as classifiers with adequate performance level are essential, they need to have other exploitable features in group work environment also. That is, beside natural abilities like performance and running time; the behavior conducted dependency in a complex system is important, too. It is not hard to see that a brute-force way extension could lead a fusion system which performs below the expectations regarding computational complexity, running time and system performance. In this context we address the problem to find an adequate dependency among classifiers with low running time. Instead of the highest performance of the algorithms, the goal will be to optimize our complete system to the direction of *pattern of success* and *pattern of failure* [41] which are such realizations of the votes in a series of experiments that lead to the possible highest and lowest accuracy of the ensemble.

To achieve our aim we need a deeper mathematical model-based investigation. The final performance in our application is based on the how successful is the component detection. So the followings performance investigations concern that how could be enhanced the single detection effectively.

From a practical point of view, we tried to organize different individual detector algorithms for example [11] into a weighted voting system to raise the accuracy of the object detection. In the example application, the result of the algorithm is a single pixel which is the center of the searched object. It inducts a majority voting where the spatial constraint of each vote also counts in making the final decision. In the generalized weighted voting system, the object center candidates of each detectors have been combined, and the minimum bounding circles for all subgroups of the candidates are considered. While the searched objects have rounded shape, a center and radius of a circle is an adequate feature to describe final positions. The radius of the circle must be less than or equal to the radius of the object (macula, OD) 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 the circle.

In the following section we will briefly summarize the basic concepts and theorems of the majority voting and our generalized majority which was applied above in previous chapter. More details about the basic model and

its experimental results can be found in [30], [31], [6] and [32]. In general, we show the improvability of the ensemble methods both theoretically and experimentally.

3.1 Introduction and basic concepts of applied majority voting

We recall first the previously published theoretical results for the generalized majority voting system [6]. The figures, tables, data etc. in this section are from paper [6]. We note that some of them also appears in [32]. However, here these objects serve as a starting point only and/or are used for comparison. It will be clear that the results discussed here are completely different from those in [32]. For a further detailed discussion, see also [8].

As introduced above 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, and let us assume that classifier D_i has accuracy p_i. 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 most often made randomly.

The classifiers making independent errors are generally considered independent, so under this assumption, the error of the classifiers can be modeled 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.

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 of Bernoulli distribution with

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \ldots, n) \ , \tag{11}$$

where $p \in [0, 1]$ and $\eta_i = 1$ means correct classification by D_i. In particular the accuracy of D_i is the expected value of η_i: $E\eta_i = p_i (i = 1, \ldots, n)$. Define now the random variables ξ in the following way:

$$P(\xi = 1) = p_{n,k}, \quad \textit{and} \quad P(\xi = 0) = 1 - p_{n,k} \tag{12}$$

where $k = |\{i : \eta_i = 1\}|$. Let the $p_{n,k}$-s are given numbers with monotone increasing property:

$$0 \leq p_{n,0} \leq \cdots \leq p_{n,n} \leq 1 \quad . \tag{13}$$

Here the $p_{n,k}$ describes the probability of the good final decision in case of k correct classifications from n classifiers and so ξ represents the modified majority voting. In this context ξ is a random variable of Bernoulli distribution, where the parameter is q, so in the special case of equal accuracies for the members ($p = p_1 = \ldots = p_n$) are assumed, so the ensemble accuracy is:

$$q = \sum_{k=0}^{n} p_{n,k} \binom{n}{k} p^k (1-p)^{n-k}. \tag{14}$$

In order to have that the generalized majority voting outperforms the individual decisions, we need only to guarantee that $q \geq p$ and consequently $E\xi \geq p$.

If the assumption about accuracies for the members ($p = p_1 = \ldots = p_n$) are equal is left, the accuracy term takes the following form regarding the majority of correct votes ($\mathrm{k} = \lceil \frac{n}{2} \rceil$):

$$q_{classical} = \sum_{k=\lceil \frac{n}{2} \rceil}^{n} \Bigg(\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1-p_j) \Bigg), \tag{15}$$

where $0 \leq p_i \leq 1$ $(i = 1, \ldots, n)$ stand for the individual accuracy of the i-th independent member [41].

Moreover, if the geometric restriction transformation is taken into account we get:

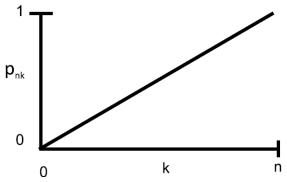
$$q = \sum_{k=0}^{n} p_{n,k} \Bigg(\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1-p_j) \Bigg). \tag{16}$$

In that case when $p_{n,k}$ is linear in k for a given n, that is $p_{n,k} = k/n$ $(k = 0, 1, \ldots, n)$, then we get $q = p$ (see Figure 8).

If we suppose that $p_{n,k} \geq k/n$ for all $k = 0, 1, \ldots, n$, then $q \geq p$, 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, $p \geq 1/2$ and for all $k = 0, 1, \ldots, n$ we have $p_{n,k} = 1$, if $k > n/2$, and $p_{n,k} = 0$, otherwise, we get the classical majority voting (see: Figure 9).

We would like to mention, as it also can be seen on Figure 9 the decision is not clear if the number of voters is even.

From a practical point of view, the $p_{n,k}$ could be encoded into the mechanism the factors influencing decisions. We give now an example of the

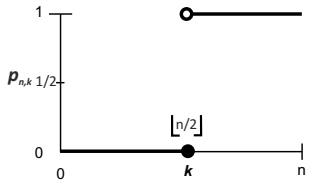


(a) The curve of $p_{n,k}$ when $p = q$

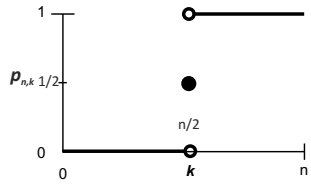
	n = 3	n = 5	n = 7	n = 9
p = 0.6	0.6	0.6	0.6	0.6
p = 0.7	0.7	0.7	0.7	0.7
p = 0.8	0.8	0.8	0.8	0.8
p = 0.9	0.9	0.9	0.9	0.9

(b) System accuracy when $p = q$

Figure 8: The linear case.



(a) n is odd



(b) n is even

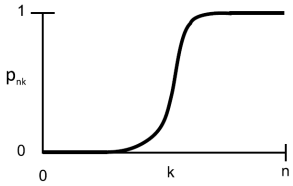
	n = 3	n = 5	n = 7	n = 9
p = 0.6	0.6480	0.6826	0.7102	0.7334
p = 0.7	0.7840	0.8369	0.8740	0.9012
p = 0.8	0.8960	0.9421	0.9667	0.9804
p = 0.9	0.9720	0.9914	0.9973	0.9991

(c) System accuracy

Figure 9: The classical majority voting scheme.

matrix $p_{n,k}$ of independent algorithms in connection with a medical imaging application. In this case the behavior of $p_{n,k}$ as a function of k for a given n and the system accuracy are illustrated in Figure 10.

Figure 10 shows that if n is given then the $p_{n,k}$ increases exponentially in k. Based on the results of [12] about the diameter d of a point set, indeed, 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 object defined in advance in our case by the clinical protocol.



(a) The curve of $p_{n,k}$

	n = 3	n = 5	n = 7	n = 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

Figure 10: The system accuracy and $p_{n,k}$ curve of experiment with geometric constraint.

As a modification of the decision rule, the weighted majority also could apply in majority voting model, where the choice of the appropriate weighting model is necessary. The class label ω_k is chosen for $\mathbf{x}$ in a weighted majority voting system, if

$$g_k(\mathbf{x}) = \max_{j=1,\ldots,n} g_j(\mathbf{x}) = \sum_{i=1}^{n} b_i d_{i,k} \tag{17}$$

where the weight b_i, decision $d_{i,k} \in [0,1]$ corresponds to the classifier D_i.

The optimal weights for independent classifiers in a classical majority voting system [41], which maximize system accuracy:

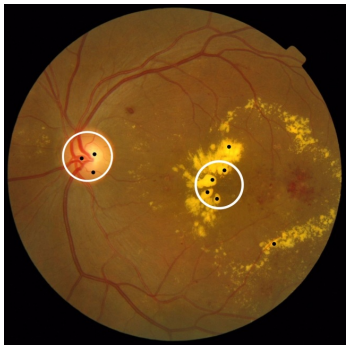
$$b_i = \log \frac{p_i}{1-p_i}, i = 1, \ldots, n \tag{18}$$

cannot guarantee this kind of optimality in our application thanks to the $p_n k$ and non-linear behaviour of the q function. Instead, we can use for independent classifiers the following calculations

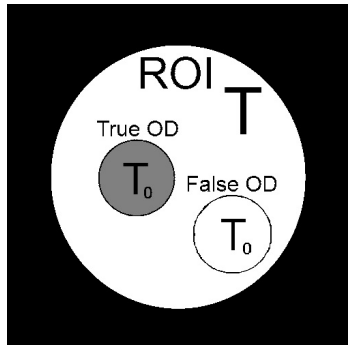
$$b_i = \log \frac{p_i}{(1-p_i)^2 r_i (1-r_i)} \tag{19}$$

where b_i denotes the optimal weight for the classifier D_i with accuracy p_i [31] so that $r_i \in [0,1]$ for classical case and in our application $r_i = \frac{T_0}{T-T_0}$. In the last equation the T_0 and T are the area of the OD and the ROI. Note the relationship among this optimal weight and the sample $p_{n,k}$ given in [10]. From Figure 11. one can see the relationship between the model and object detection.

If we check our theoretical comparison tables (see below Figure 12, regarding majority voting (MV), weighted majority voting (WMV), generalized majority voting (GMV), weighted generalized majority voting (WGMV))



(a) Example of OD detection.



(b) Schematic visualization of a generalized model.

Figure 11: Relationship between the model and object detection.

Table 8: Applied weights in our comparative studies, where B_3 is optimal for OD detection in case of independent classifers.

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

in the case of single object detection (the applied weights for OD detection are in Table 8), we can realize, that our application (2.2) based on this theoretical model works with a slightly weaker performance.

The object detectors in real application are hardly independent. Therefore our primary objective is to study and extend our basic model: how changes the performance of our object detection model if a new classifier is added to the system. As dependency plays a key role, we will investigate this at the level of model.

3.2 Extending the ensemble adding a new classifier

As we could see the generalized decision model can characterize well a single object detection problem if the algorithms are independent. Nonetheless, the applied classifiers are slightly independent, so it is not obvious that how an additional algorithm will affect the system accuracy. Below we study the possibility to improve system accuracy with adding a new classifier D_{n+1} with accuracy p_{n+1}. The change of the system accuracy q and the interval

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) Overall system accuracies for the set of classifiers equal accuracies, that is $p_1 = p_2 = \ldots = p_9 = 0.6$

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) Overall system accuracies for the set of classifiers with accuracies $p_i = 1 - 0.1i$, $i = 1, \ldots, 9$

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) Overall system accuracies for the set of real classifier with the following measured accuracies in Messidor [19] database: $p_1 = 0.6472, p_2 = 0.9765, p_3 = 0.3205, p_4 = 0.7593, p_5 = 0.3153, p_6 = 0.2276, p_7 = 0.9582, p_8 = 0.7671, p_9 = 0.6432$.

Figure 12: Overall system accuracies for the set of classifiers, in case of different accuracies.

$[q_{min}, q_{max}]$ for the minimal and maximal system accuracy are also studied. The ambiguity about dependency/independency we will consider different cases, given in a numerated list below for reference in the next chapter:

A. we fix the individual accuracies and outputs of the algorithms of the current ensemble for an experiment in terms of a contingency table, and:

   1. add a new independent algorithm and check how the ensemble

accuracy (q) changes,

2. add a new dependent algorithm and check how the minimal (q_{min}) and maximal (q_{max}) ensemble accuracy change,

B. we fix the individual accuracies, but ignore the outputs of the algorithms of the current ensemble.

Beside the performance of the algorithms $p_1, \ldots, p_n$, the values of $p_{n+1,k}$ have significant influence on the system accuracy, if a new algorithm is added to an existing system. Therefore based on the definition of $p_{n,k}$ and that fact a wrong decision will be added to the system, our estimation for the related $p_{n,k}$ values are:

$$p_{n,k} \geq p_{n+1,k}, \tag{20}$$

$$p_{n,k} \leq p_{n+1,k+1}. \tag{21}$$

According to the interpretations the added vote is expected to be false so that the probability of a right decision cannot increase in case of (20). However the probability of a right decision cannot decrease in (21).

After adding a new algorithm to the existing system, the new system accuracy depends not only on the accuracies $p_1, \ldots, p_{n+1}$, but also on the values $p_{n+1,k}$.

As a conclusion we obtain from the inequalities (20) and (21) the properties for $p_{n+1,k}$ to estimate $p_{n+1,k}$ from $p_{n,k}$:

$$p_{n,k-1} \leq p_{n+1,k} \leq p_{n,k}. \tag{22}$$

We briefly discuss the existing results. For classical majority voting the so-called *pattern of success* and *pattern of failure* [41] are defined to measure the extremal, or otherwise the highest and lowest accuracies of an ensemble. However, these results have a severe restriction considering discretization of the model. More precisely, in the experiments besides the accuracies of the individual classifiers, the precise numbers of the successful decisions are also fixed. That is, a classifier with performance $p = 0.7$ gives seven correct votes in 10 repeated experiments on average. It is easy to see that this implies the following, a lower performance classifier ($p < 0.7$) could be the best-added component if its correct votes are arranged carefully based on the tie or one-vote-missing cases.

So, if a new member is added to an existing ensemble, the accuracy of the extended ensemble is affected by two main properties of the new voter:

its accuracy and its correlation with the members of the existing system. We will investigate the joint distribution of the output of the detectors as a dependency measurement. In the equations the $*$ denotes 0 or 1 and η_i $(i = 1, \ldots, n)$ are random variables, so that $c_{a_1,\ldots,a_n}$ can be considered as the elements of the contingency table of $\eta_1, \ldots, \eta_n$:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n), \tag{23}$$

so that $a_i \in 0, 1, *$. If $(\eta_1, \ldots, \eta_n)$ were totally independent, we would get:

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1), \ldots, P(\eta_n = a_n), \tag{24}$$

Now let η_{n+1} be also a random variable with $E\eta_{n+1} = p_{n+1}$. To determine the best/worst choice for the new member to achieve the best (q_{max})/worst (q_{min}) performance for the extended ensemble the following linear optimization problem has to be solved in the general case B. Maximize/Minimize the function:

$$q(c_{a_1,\ldots,a_{n+1}}) = \sum_{k=0}^{n+1} \left(p_{n+1,k} \sum_{\substack{a_1+\ldots+a_n+ \\ +a_{n+1}=k}} c_{a_1,\ldots,a_{n+1}} \right) \tag{25}$$

under the following conditions:

$$\begin{aligned} &\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n+1), \\ &\sum_{a_1,\ldots,a_{n+1}} c_{a_1,\ldots,a_{n+1}} = 1, \\ &c_{a_1,\ldots,a_{n+1}} \geq 0, \; a_i \in \{0, 1\} \quad (i = 1, \ldots, n+1), \end{aligned} \tag{26}$$

where $E\eta_i = p_i$ $(i = 1, \ldots, n+1)$, so the accuracy of the i-th classifier is p_i. In case A.2., besides the objective function in (25) and the conditions in (26) being the same, we have extra conditions thanks to the dependency:

$$c_{a_1,\ldots,a_n} = c_{a_1,\ldots,a_n,0} + c_{a_1,\ldots,a_n,1}. \tag{27}$$

$$\begin{aligned} &\sum_{a_1+\ldots+a_{n+1}=k} c_{a_1,\ldots,a_{n+1}} = \\ &\sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n,0} + \sum_{a_1+\ldots+a_n=k-1} c_{a_1,\ldots,a_n,1}. \end{aligned} \tag{28}$$

From the definition of $c_{a_1,\ldots,a_n}$ given in (23) it follows that the term containing $c_{a_1,\ldots,a_{n+1}}$ in (25) can be split.

Without having any further information about $p_{n+1,k}$, we can give an interval for q_{min} and q_{max}. Let $q^{\ominus}_{min}/q^{\ominus}_{max}$ and $q^{\oplus}_{min}/q^{\oplus}_{max}$ be the minimal/maximal value of the objective function (25) if we have the equations $p_{n,k-1} = p_{n+1,k}$ and $p_{n+1,k} = p_{n,k}$, respectively.

From (22), we get:

$$q^{\ominus}_{min} \leq q_{min} \leq q^{\oplus}_{min}, \quad \text{and} \quad q^{\ominus}_{max} \leq q_{max} \leq q^{\oplus}_{max}. \tag{29}$$

In the special case, when η_{n+1} is totally independent from $(\eta_1, \ldots, \eta_n)$, the entries of the extended contingency table can be determined by $c_{a_1,\ldots,a_n}$ and p_{n+1}:

$$\begin{aligned} c_{a_1,\ldots,a_n,1} &= p_{n+1} c_{a_1,\ldots,a_n}, \\ c_{a_1,\ldots,a_n,0} &= (1-p_{n+1}) c_{a_1,\ldots,a_n}. \end{aligned} \tag{30}$$

Considering the equations (25), (28) and (30) we get that the linear optimization problem which can be solved by maximizing/minimizing the function:

$$\begin{aligned} &q(c_{a_1,\ldots,a_{n+1}}) = \\ &\sum_{k=0}^{n+1} p_{n+1,k} \Bigg(\sum_{a_1+\ldots+a_n=k} (1-p_{n+1}) c_{a_1,\ldots,a_n} + \\ &\qquad + \sum_{a_1+\ldots+a_n=k-1} p_{n+1} c_{a_1,\ldots,a_n} \Bigg) \end{aligned} \tag{31}$$

under the following conditions:

$$\begin{aligned} &\sum_{a_i=1} c_{*,\ldots,*,a_i,*,\ldots,*} = p_i \quad (i = 1, \ldots, n), \\ &\sum_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n} = 1, \\ &c_{a_1,\ldots,a_n} \geq 0, a_i \in \{0,1\} \quad (i = 1, \ldots, n), \end{aligned} \tag{32}$$

If we consider that the entries of the contingency table of $\eta_1, \ldots, \eta_n$ remain the same after adding an independent variable η_{n+1} to the ensemble

(case A.1.), the solution of the problem in (31) under the conditions (32) depends only on p_{n+1} and $p_{n+1,k}$.

In the same way as in (29), from (22) we get:

$$q^{\ominus} \leq q \leq q^{\oplus}, \tag{33}$$

where $q^{\ominus}$ and $q^{\oplus}$ denote the minimal/maximal value of the objective function (31) for a fixed p_{n+1} if we consider the equations $p_{n,k-1} = p_{n+1,k}$ and $p_{n+1,k} = p_{n,k}$, respectively.

For the improvability of the system, we have the following lemma.

Lemma 3.1. *For the accuracy of the extended ensemble in case of the independence we have:*

$$q(c_{a_1,\ldots,a_{n+1}}) \geq q(c_{a_1,\ldots,a_n}),$$

provided that,

$$p_{n+1} \geq \frac{\sum_{k=0}^{n} \left(\sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n}(p_{n,k} - p_{n+1,k}) \right)}{\sum_{k=0}^{n} \left(\sum_{a_1+\ldots+a_n=k} c_{a_1,\ldots,a_n}(p_{n+1,k+1} - p_{n+1,k}) \right)}$$

holds for the accuracy of the added member.

Proof. First, note that the value of this fraction is non-negative, since $p_{n,k} \geq p_{n+1,k}$ and $p_{n+1,k} \leq p_{n+1,k+1}$. Moreover, from (22) and (30) the statement follows. □

In the next section, we will show some experimental results for the improvability of the accuracy of our OD detector ensemble with adding a new algorithm based on the established theoretical results.

3.3 Experimental results

After we have laid the theoretical background to extend the ensemble by adding a new algorithm, we have a tool to finalize the connected investigation experimentally. Namely, we have formulated the ways of the calculation of ensemble accuracy for the cases, when the new member is dependent or independent from the ensemble, respectively. Besides the simple ensemble accuracy, we have also explained how the minimal and maximal accuracies of the ensemble would change.

To understand more deeply how to change the accuracy in our model and application during extension, as comparison, we recall some experimental measurements about the fundamental behavior of our basic model (which were introduced in section 3.1) and related the applications from [30], [31], [6] and [32]. (Similarly as in section 3.1), the figures and tables are from [6]. Some of there are also in [8], but in a different context.)

In our real application the joint detection is deterministic thanks to the geometric constraint among OD and macula detection.We will concentrate only on the improvement of the model based single-object detection. Moreover, the detection based on majority voting in both cases are the same since we apply a majority voting with the same geometric constraint. As a demonstration, we will use the OD detectors in the experiments. The reason is simple, the number of OD detectors is higher.

The applied model for object detection is a generalized majority voting model. The generalization is due to the use of the probabilities $p_{n,k}$. In fact, the probabilities $p_{n,k}$ are restricted by geometric constraints based on anatomic rules in our real application. Namely, in our application the output of the detectors as votes required to fall within a circle of radious not larger then (d_{OD},d_{MA}) defined by anatomical features. As a matter of fact, the calculation of $p_{n,k}$ for our proposed method is guided by the shape characteristic, that is the k pieces of correct votes have to fall inside the correct object region. However, the remaining $n-k$ votes are spread within the complete ROI. While the ROI is significantly larger, a false region could be formed. This means that the correct votes need not necessarily to be in majority to make the correct decision since the condensation of the incorrect ones is not guaranteed inside only one false region. On Figure 13 there is a sample situation in the case of the OD.

To take advantage of the theoretical foundations for the experiments we used the MESSIDOR databases [23]. To characterize and compare our $p_{n,k}$ model, both the independent and dependent cases, we made empirical calculations where the values of $p_{n,k}$ characterize our ensemble accuracies as a function of k. As the simulation of the independent case, we make simulation by dropping a large number of random points. Figure 14 shows the exponential behaviour of $p_{n,k}$ for both cases.

In this experiment we also confirm an existing result known in the literature [12, 47]: if the number of points tends to infinity, the probability that the diameter of a point set is not less than a given constant decreases exponentially, where in our case the diameter was the diameter of the object (for instance d_{OD}). The obtained data for the ensemble accuracy from empirical measurement are shows in Table 9, where the independent classifiers

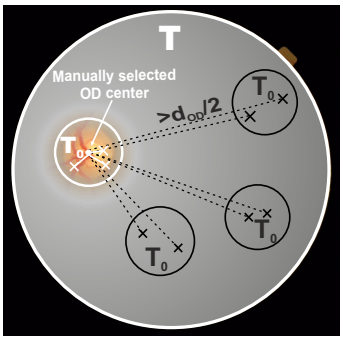
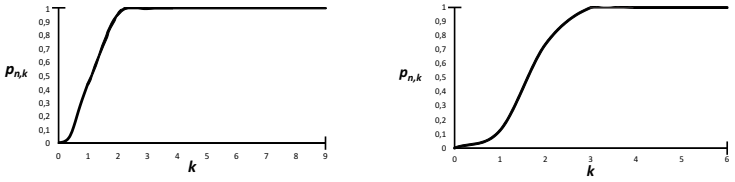


Figure 13: The geometric constraint applied to the candidates of the algorithms: they should fall inside a disc of a fixed diameter d_{OD} to vote together.



(a) The graph of $p_{n,k}$ for $n = 9$ and $p = 0.9$ with our geometric constraint to fall within a disc of diameter d_{OD}.

(b) The graph of $p_{n,k}$ for $n = 6$ in our OD detector ensemble.

Figure 14: The exponential behaviour of $p_{n,k}$ in independent (left) and dependent (right) situations.

(n) have the same accuracy (p). Based on the table it is quite clear, that the geometric constraint has a significant effect on the ensemble accuracy. The ratio of the areas (T_0 and T) is decisive. Thanks to the independency, the rapid increase is visible because of the uniformly distributed false votes. Moreover, there is a significant performance decrease if the number of votes is too small paired with low performance.

To start our experimental discussion of extending the ensemble-based system, we check the behavior of our current OD detector ensemble during its compilation. Namely, we measure the change of the ensemble accuracy,

	n=3	n=5	n=7	n=9
$p = 0.6$	0.6435	0.9076	0.9654	0.9893
$p = 0.7$	0.7889	0.9631	0.9938	0.9985
$p = 0.8$	0.9029	0.9906	0.9986	0.9997
$p = 0.9$	0.9697	0.9994	1.0000	1.0000

Table 9: Measured ensemble accuracy for independent classifiers under the geometric constraint.

when the sixth member is added to the ensemble of five members. For this aim, we calculate the accuracy of each ensemble of five individual algorithms with the corresponding figures enclosed in Table 10. Thus, Table 10 contains the accuracies of the six possible ensembles of five members, wherein the i-th column the i-th member is excluded having individual accuracy p_i, for $i = 1, \ldots, 6$.

Index of excluded member	1	2	3	4	5	6
Ensemble accuracy (5 members)	0.980	0.957	0.979	0.976	0.961	0.976
Ensemble accuracy (6 members)	0.981					

Table 10: Change of the ensemble accuracy, when a sixth member is added to the ensemble of five algorithms.

As we have given in (34), the accuracy of the ensemble of the six given members has been measured to be the highest (0.981) from all the possible ensembles that we could generate from the eight implemented algorithms. Thus, each ensemble of five members has lower accuracy, as well. From Table 10 we can see that the largest increase in accuracy (from 0.957 to 0.981) is reached not by adding the most accurate ($p_5 = 0.977$) member, but a slightly less accurate ($p_2 = 0.958$) one. Similarly, the smallest improvement (from 0.980 to 0.981) is found not by adding the least accurate ($p_3 = 0.315$) member, but by adding an individually more accurate ($p_1 = 0.767$) one. To understand these results we should realize that there are specific dependencies among the members. Thus, in general, it is not sufficient to compose an ensemble-based on the individual accuracies only.

Next, we adopt the results from section 3.2 to investigate how our current OD detector ensemble consisting of six algorithms is going to behave if a new detector algorithm is added. We start with the case A.1 from section 3.2 when the dependencies of the current ensemble members are considered as

known in terms of a contingency table belonging to our experimental test on the dataset MESSIDOR ([23]) and the new algorithm is considered to be independent of the ensemble. For this case, through the solution of (31), we gain the numeric results enclosed in Table 11. Note that, in this case we can check the interval $[q^{\ominus}, q^{\oplus}]$ introduced in (33) where the ensemble accuracy will fall based on the lower and upper estimations that can be derived for $p_{n+1,k}$ as given in (22).

Accuracy of the new algorithm	$q^{\ominus}$	$q^{\oplus}$
$p_7 = 0.6$	0.957	0.989
$p_7 = 0.9$	0.975	0.995

Table 11: The interval for the OD detector ensemble accuracy if a new independent algorithm is added to a dependent system.

From Table 11, we can see that in our application a new (independent) algorithm with accuracy approximately 0.9 is highly expected to improve the current system accuracy based on our experimental test:

$$q = 0.981. \tag{34}$$

The case A.1 in section 3.2 also includes the special scenario, when the existing ensemble contains independent members, and we add an independent algorithm, as well. For this scenario, we can investigate the minimal and maximal accuracies of the new system by solving the problem in (31) under the new condition (35):

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1) \ldots P(\eta_n = a_n). \tag{35}$$

In Table 12, we enclosed the respective accuracy figures regarding the lower and upper estimations of the values $p_{n+1,k}$.

Accuracy of the new algorithm	$q^{\ominus}$	$q^{\oplus}$
$p_7 = 0.6$	0.975	0.997
$p_7 = 0.9$	0.984	0.999

Table 12: The interval for the OD detector ensemble accuracy if a new independent algorithm is added to an independent system.

By comparing Table 11 and Table 12, we can see that if we assume complete independency among the algorithms, we can expect higher ensemble

accuracy. Since the original ensemble would lead to very high accuracy with independent algorithms that is $q_{ind} = 0.998.$, only in case of a very accurate new algorithm can we expect any improvement. The calculation of the system accuracy we used (31) for n only under the conditions (32) and with the assumption (24) on the independence of the detectors.

Next, we analyse the case A.2 from section 3.2, when the dependencies of the algorithms are still considered,but the new algorithm should not be independent. In this setup, we can determine the accuracy interval introduced in (29) for the minimal (q_{min}) and maximal (q_{max}) ensemble accuracies, respectively, based on the estimation for the values $p_{n+1,k}$ as given in (22). The corresponding figures presented in Table 13 can be determined by the solution of (25) under the conditions (26), (27).

Accuracy of the new algorithm	$q^{\ominus}_{min}$	$q^{\oplus}_{min}$	$q^{\ominus}_{max}$	$q^{\oplus}_{max}$
$p_7 = 0.1$	0.920	0.981	0.981	0.995
$p_7 = 0.7$	0.920	0.981	0.981	0.995
$p_7 = 0.9$	0.942	0.981	0.981	0.995

Table 13: The interval for the minimal and maximal OD detector ensemble accuracy if a new dependent algorithm is added to a dependent system.

Table 13 shows that an individually very weak, but diverse algorithm could lead to a remarkable improvement of the ensemble, however, this possibility is rather unrealistic. Moreover, since the current ensemble is not optimal regarding dependencies, even with a very diverse and accurate algorithm, we cannot reach accuracy 100%. It is also visible from Table 13 that the original system accuracy (34) cannot be outperformed with the lower estimation for $p_{n+1,k}$, and cannot be degraded with its upper estimation, either.

Another point which is worth considering is that since the retinal databases are quite heterogeneous, we cannot go for sure regarding the dependencies of the algorithms of the ensemble found for a specific (in our case for the MESSIDOR [23]) database. Thus, if we keep the individual accuracies of the ensemble members, but drop the dependency relations, it would be useful to know to what extent a new algorithm may ruin or improve the ensemble accuracy. Consequently, we investigate case B from section 3.2, when a new algorithm with accuracy p_7 is added to our current ensemble with no constraints are given for the dependencies. In other words, we check the intervals for the minimal and maximal accuracies of the extended system regarding the lower and upper estimation of the values $p_{n+1,k}$, respectively.

The corresponding figures enclosed in Table 14 can be determined by the solution of (25) under the conditions (26).

Accuracy of the new algorithm	$q_{min}^{\ominus}$	$q_{min}^{\oplus}$	$q_{max}^{\ominus}$	$q_{max}^{\oplus}$
$p_7 = 0.7$	0.764	0.899	1	1
$p_7 = 0.9$	0.908	0.934	1	1

Table 14: The interval for the minimal and maximal OD detector ensemble accuracy if a new dependent algorithm is added to a system with no dependency constraints.

Table 14 indicates the natural fact that if the dependencies are unknown, the minimal and maximal accuracy can profoundly differ, and, e.g., the ensemble performance can be worse than that of some of its members. However, it is also worth considering for our specific OD detector ensemble that a new algorithm of accuracy $p_7 = 0.9$, by all means, will raise the minimal system accuracy ($q_{min} = 0.899, q_{max} = 1$) calculated by a linear programming technique for n described in equation (31). A comparison with Table 13 shows that if we do not assume any dependencies for the original ensemble, we can reach higher maximal and lower minimal system accuracies.

For the strict majority voting approach, an ensemble with an even number of members is meaningless, since as it is also known from classical theory [39] ensemble accuracy always drops for even number of L members with respect to the $L - 1$ case.

So we have analyzed the change in accuracy when the ensemble containing five members is extended to seven members. First of all, we have determined the most accurate ensemble with seven members from all the implemented eight algorithms. This ensemble includes the same six algorithms as listed before plus the one described in [43] having individual accuracy $p_7 = 0.320$. Then, we have selected the most/least accurate ensembles with five members, respectively, and checked which members were added to compile the ensemble with seven members. The corresponding quantitative results are given in Table 15.

Indices of excluded members	2,5 (lowest acc.)	3,7 (highest acc.)
Ensemble accuracy (5 members)	0.626	0.944
Ensemble accuracy (7 members)	0.853	

Table 15: Change of the ensemble accuracy for strict majority, when the sixth and seventh member is added to the ensemble of five algorithms.

The results of Table 15 are quite obvious, since two individually highly accurate (p_2, p_5) and also two rather inaccurate (p_3, p_7) algorithms are present. Thus, their joint removal leads to a strong drop/increment regarding the ensemble accuracy, respectively.

3.4 Conclusion

In section 2.2 we introduced our joint object detection fusion based system. As decision rule, we applied a generalized spatial constrained majority voting for each detected object extended with a combined final decision driven by anatomical features. In this section, the first aim was to investigate how the system performance is affected if a new classifier is added to the system. While the joint model is deterministic thanks to the anatomical features and obviously depend in particular on the performance of the separate detection, we studied the separate behavior. We described our generalized model working behind the separate object detection. The basis of the generalization is the introduced $p_{n,k}$ model for handle spatial majority voting regarding independency and dependency. Although the $p_{n,k}$ is entirely general, we produced it's values as experimental results both int the dependent and in the independent cases to characterize the performance of our application. We would do this in order to investigate and characterize what is the possibility to make some further improvement including it's level by help of more classifiers. We had results which characterize how the system performance is affected if a new classifier is added to the system. Based on the results it is obvious that our classifiers cannot work in a fusion system as independent voters. Our system works effectively since, as the results show, the performance is very close to the optimum. Nevertheless, if it contains completely independent classifiers it could work more effectively. However, it is unclear whether it is possible to create the complete independence while all the classifiers nominate candidates based on the same data source (input images) only.

Whereas we identified the limits of the performance focusing mainly on the extension capability, the level of dependency among classifiers remains unknown. Therefore in the first part of the next section, we will introduce generalizations of well-known statistical diversity measures for our generalized majority voting as a general tool for ensemble optimization. Namely, we will evaluate the diversity level among the classifiers, and based on the results all of those will be selected for majority voting, which together makes the maximum performance. The second part of the next section we will consider system optimization also if there is a constraint, for instance, computational

time or resources. The primary goal is the same; we try to compose the best performance fusion system using generalized majority voting, regarding computational limitations.

4 Ensemble-based system optimization

4.1 Composing an ensemble-based system guided by diversity measures

We have already presented how the classical majority voting model can be extended to the spatial domain, e.g., to solve object detection problems. As we have also pointed out, the detector algorithms cannot be considered as independent classifiers in applications, so a good ensemble cannot be composed by simply selecting the individually most accurate members. In classical theory, diversity measures are recommended to help to explore the dependencies among the classifiers. In this section, we generalize the classical diversity measures in two steps. First, we show how classical pairwise measures can be extended to explore larger sets of classifiers via letting the contingency table entries of an ensemble-based system be freely weighted. Second, we introduce these extended measures also to the spatial domain within a majority voting framework. We show that these measures fit better to spatial applications with a specific example of OD detection. Moreover, we show how a more efficient descriptor can be found in terms of a weighted combination of diversity measures which correlates better with the accuracy of the ensemble.

4.1.1 Generalized diversity measures

In this section we introduce some basic concepts and notations and explain our approach to generalize commonly used diversity measures. We also exhibit how existing measures can be generalized with the new model.

4.1.1.1 Basic concepts and notations

Let $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n\}$ be a set (ensemble) of classifiers (voters) $\mathcal{D}_i : \Lambda \subseteq \mathbb{R}^m \to \mathbb{R}^M_{\geq 0}$ $(i = 1, \ldots, n)$, and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ is a set of finite class labels.

The classifier $\mathcal{D}_i$ assigns the support values

$$\mathcal{D}_i(\boldsymbol{\chi}) = (d_{i,1}(\boldsymbol{\chi}), \ldots, d_{i,M}(\boldsymbol{\chi}))$$

to a so-called feature vector $\boldsymbol{\chi} \in \Lambda$ describing the opinion of the classifier on what degree $\boldsymbol{\chi}$ should be labeled by $\omega_1, \ldots, \omega_M$, respectively.

The simple majority voting based classical ensemble classifier can be derived by restricting the support of the individual classifiers with $d_{i,j}(\boldsymbol{\chi}) =$

1, if the classifier $\mathcal{D}_i$ labels $\boldsymbol{\chi}$ in the class ω_j and $d_{i,j}(\boldsymbol{\chi}) = 0$, otherwise. The final labeling of the ensemble is based on determining the class ω_k received the largest support in terms of the number of votes:

$$D_{maj}(\boldsymbol{\chi}) = \omega_k \iff \sum_{i=1}^{n} d_{i,k} = \max_{j=1}^{M} \left(\sum_{i=1}^{n} d_{i,j} \right). \tag{36}$$

In our model, we consider classifiers $\mathcal{D}_i$ with accuracies p_i as random variables η_i of Bernoulli distribution, i.e.,

$$P(\eta_i = 1) = p_i, \quad P(\eta_i = 0) = 1 - p_i \quad (i = 1, \ldots, n). \tag{37}$$

Here $\eta_i = 1$ means correct classification by $\mathcal{D}_i$. In particular, the accuracy of $\mathcal{D}_i$ is just the expected value of η_i, that is, $\mathbb{E}\eta_i = p_i$ $(i = 1, \ldots, n)$.

In our practical applications, we consider different algorithms as voters. These algorithms cannot be assumed to be independent in all cases, because it can happen that the operations of the algorithms are based on very similar principles. In case of dependent algorithms, we have to decide how to measure the dependencies of the algorithms. For this aim, we can investigate the joint distribution of the outputs of the algorithms. So let

$$c_{a_1,\ldots,a_n} = P(\eta_1 = a_1, \ldots, \eta_n = a_n), \tag{38}$$

where $a_i \in \{0, 1, *\}$. The star denotes any of the possible correctness values, that is, $* = 0$ or 1, but it is not specified. The probabilities $c_{a_1,\ldots,a_n}$ can be considered as the entries of the contingency table $\mathcal{C}$ of $\eta_1, \ldots, \eta_n$.

From practical point of view, the performance of the ensemble and the individual classifiers are always evaluated on a finite dataset $\Delta = \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_N\}$ having cardinality $N \in \mathbb{N}$. Such a dataset can be a sequence of questions or images to be classified. Using this realization, the entries of the contingency table $\mathcal{C}$ can be written as:

$$c_{a_1,\ldots,a_n} = N^{a_1 a_2 \ldots a_n}/N, \tag{39}$$

where $N^{a_1 a_2 \ldots a_n}$ denotes the number of cases from the whole dataset, when $\eta_1 = a_1, \ldots, \eta_n = a_n$ hold. When we are interested in only a subset of the classifiers $\mathcal{D}_S = \{\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_K}\}$ of cardinality $|\mathcal{D}_S| = K$ with $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$, we will shortly write $N^{a_{i_1} a_{i_2} \ldots a_{i_K}}$ instead of $N^{*\cdots *a_{i_1} *\cdots *a_{i_2} *\cdots *a_{i_K} *\cdots *}$. For example, N^{00} will denote the number of cases, when two fixed classifiers both made incorrect decisions. Moreover, we write $N^{\{0\}_{K_1}\{1\}_{K_2}}$ to collect all the possible $N^{a_1 a_2 \ldots a_K}$ values with $K_1 + K_2 = K$ and $|\{a_i \in \{a_1, a_2, \ldots, a_K\} \ : \ a_i = 0\}| = K_1$, $|\{a_i \in \{a_1, a_2, \ldots, a_K\} \ : \ a_i =$

$1\}| = K_2$. Especially, $N^{\{0\}_K}$ stands for the number of cases, when all the K classifiers made incorrect decisions, while $N^{\{0\}_{K_1}\{*\}_{K_2}}$ collects all those values, where at least K_1 among the K classifiers made incorrect decisions.

Among existing diversity measures we find such ones that explore ensemble dependency only based on the individual performances of the members. As for the realization of the classifier $\mathcal{D}_i$ for the j-th element of the dataset we write $\eta_i^{(j)}$. That is, $\eta_i^{(j)} = 1$ if $\mathcal{D}_i$ classified $\boldsymbol{\chi}_j$ correctly, while $\eta_i^{(j)} = 0$ in the case of wrong classification. Notice that, with this notation for the individual classifier accuracies we also have $p_i = \sum_{j=1}^{N} \eta_i^{(j)}/N$ for $i = 1, \ldots, n$.

4.1.1.2 Weighing the contingency table

As an essential consideration for our approach for the generalization of diversity measures, notice that, widely used diversity measures are simply derived from special entries of the contingency table $\mathcal{C}$. Thus, we introduce generalized diversity measures with letting arbitrary weights to be assigned to the entries of $\mathcal{C}$. More formally, a generalized diversity measure DM has the form

$$DM = \mathcal{T} : (DM_1, DM_2, \ldots, DM_T) \in \mathbb{R}^T \to \mathbb{R} \tag{40}$$

with

$$DM_t = \sum_{a_i \in \{0,1\}, i=1,\ldots,n} w^{(t)}_{a_1,\ldots,a_n} c_{a_1,\ldots,a_n}, \; t = 1, \ldots, T, \tag{41}$$

where $w^{(t)}_{a_1,\ldots,a_n} \in \mathbb{R}$ is the weight assigned to the contingency entry $c_{a_1,\ldots,a_n}$ in the t-th component DM_t. As for the transformation $\mathcal{T}$ in (40), its usual role is to combine the differently weighted contingency entries via common arithmetic operators, like addition, subtraction, multiplication, division, square root, etc.

Commonly used diversity measures can be obtained by applying specific weights in the general model as will be shown in the forthcoming section.

4.1.1.3 classical measures in our model

In this paper, we consider the generalization of the currently widely used diversity measures summarized in Table 16. The table also includes the formulas for their calculations.

Now, we shortly exhibit each of these measures and present how they can be derived with an appropriate weigthing applied to some specific elements of the contingency table. We note that all of the generalized number introduced below are reducible with the appropriate parameters to the originals.

Table 16: classical diversity measures and their calculations.

Diversity measure	**Calculation**
Pairwise measures	
Double fault (F)	$F_{ij} = N^{00}/N$
Disagreement (D)	$D_{ij} = (N^{01} + N^{10})/N$
Product-moment correlation (PM)	$PM_{ij} = N^{00}/\sqrt{N^{*0}N^{0*}}$
Correlation coefficient (C)	$C_{ij} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{N^{1*}N^{0*} - N^{*1}N^{*0}}}$
Q statistics (Q)	$Q_{ij} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}$
Measures based on individual performances	
Minimum individual error (MIE)	$MIE = \min_{i=1}^{n}(1 - p_i)$
Mean error (ME)	$ME = \sum_{i=1}^{n}(1 - p_i)/n$
Majority rule error (MRE)	$MRE = \left\lvert \left\{ j \ : \ j \in \{1, \ldots, N\}, \sum_{i=1}^{n} \eta_i^{(j)} \geq \frac{N}{2} \right\} \right\rvert /N$
System improvement (SI)	$SI = MRE - ME$

4.1.1.3.1 Double fault (F)

Originally, the double fault measure N^{00}/N tries to estimate the probability of the coincident error between two classifiers $\mathcal{D}_i, \mathcal{D}_j$ $(i, j \in \{1, \ldots, n\}, i \neq j)$. The desired different behavior of the classifiers is observed related to their bad decisions only by this measure. In our model, it can be written as

$$F_{ij} = N^{00}/N = \sum_{a_i=0, a_j=0} c_{*,\ldots,*,a_i,*,\ldots,*,a_j,*,\ldots,*} \tag{42}$$

with applying the following weighing in (41):

$$w_{a_1,\ldots,a_n} = \begin{cases} 1, & \text{if } a_i = 0 \text{ and } a_j = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{43}$$

Moreover, notice that this simple diversity measure has only one component considering (40). The double fault measure has been introduced also within

the whole set of base classifiers as

$$\overline{F} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} F_{ij}. \tag{44}$$

4.1.1.3.2 Disagreement (D)

This measure tries to explore the degree of diversity based on pairwise disagreement with summarizing the cases, where the voters $\mathcal{D}_i, \mathcal{D}_j$ $(i, j \in \{1, \ldots, n\}, i \neq j)$ disagree. It is also normalized with the volume of the dataset (see Table 16).

In our model, it can be written as

$$D_{ij} = \mathcal{T}(DM_1, DM_2) = (DM_1 + DM_2)/N \tag{45}$$

with applying the following respective weighings for DM_1 and DM_2 in (41):

$$w^{(1)}_{a_1,\ldots,a_n} = \begin{cases} 1, & \text{if } (a_i = 0 \text{ and } a_j = 1), \\ 0, & \text{otherwise,} \end{cases} \tag{46}$$

$$w^{(2)}_{a_1,\ldots,a_n} = \begin{cases} 1, & \text{if } (a_i = 1 \text{ and } a_j = 0), \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

Notice that, the disagreement measure is composed from two components with $\mathcal{T}$ in terms of an addition and normalization with $1/N$. The disagreement measure is defined also for the whole set of base classifiers as

$$\overline{D} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} D_{ij}. \tag{48}$$

For shorter formulation of the remaining diversity measures for higher dimensions, from now on we will use the alternative definitions based on

$$N^{*\cdots*a_{i_1}*\cdots*a_{i_2}*\cdots*a_{i_k}*\cdots*}$$

instead of the forms (40) and (41). However, notice the underlying equivalent weigthing schemes.

4.1.1.3.3 Product-moment correlation (PM)

The product-moment correlation measure has been created to support the selection of the most diverse algorithms in neural network-based scenarios. The original pairwise measure for binary input focuses on the simultaneous failures divided by the multiplication of the individual errors:

$$PM_{ij} = \frac{N^{00}}{\sqrt{N^{*0}N^{0*}}} \tag{49}$$

Similarly to the previous measures, product-moment correlation is defined also for the whole set of classifiers as

$$\overline{PM} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} PM_{ij}. \tag{50}$$

4.1.1.3.4 Correlation coefficient (C)

The correlation coefficient, which is one of the most widely applied statistical measure for binary scenarios has the following form:

$$C_{ij} = \frac{N^{11}N^{00} - N^{01}N^{10}}{\sqrt{N^{1*}N^{0*} - N^{*1}N^{*0}}}. \tag{51}$$

Its variant for the whole set of classifiers is written as

$$\overline{C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} C_{ij}. \tag{52}$$

4.1.1.3.5 Q-statistics (Q)

As the last investigated measure, the pairwise Q-statistics measure has the range $[-1, 1]$, where a value 0 shows independence, 1 suggests similar predictions, while -1 different ones:

$$Q_{ij} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}}. \tag{53}$$

To include the whole set of classifiers we write

$$\overline{Q} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} Q_{ij}. \tag{54}$$

4.1.1.3.6 Measures based on individual performances

In Table 16 we have enclosed such simple measures that are based only on the individual performances of the voters. The minimum individual error (MIE) selects the largest individual accuracy, the mean error (ME) calculates the average of the individual accuracies, the majority rule error (MRE) determines the error of majority-rule based decision, while the system improvement (SI) checks MRE and ME simultaneously. The proper formulation of these measures are contained in Table 16.

4.1.1.4 Generalization of classical measures

Using the flexible weighting model introduced in section 4.1.1.2, now we generalize the classical measures introduced in section 4.1.1.3. The basis of generalization is letting to observe the classifiers not only pairwise, but also larger sets and subsets of them having cardinality 3 or more. This approach allows a deeper discovery of underlying dependencies. Moreover, our weighting model provides the possibility to weigh the entries of the contingency table in a non-binary (0/1) way introducing free weight parameters that can be adjusted afterwards for optimal performance for specific applications. Next, we show how can be generalized the five measures described in section 4.1.1.3 in this way.

4.1.1.4.1 K-fault (F_K)

The K-fault measure ($2 \leq K \leq n$) is a natural generalization of the double fault one with discovering the mutual behavior of more than two classifiers. Accordingly, for any subset of classifiers $\mathcal{D}_S$ with $|\mathcal{D}_S| = K$ we introduce this measure in the following way:

$$F_K = \frac{N^{\{0\}_K}}{N} \tag{55}$$

Moreover, to aggregate for all the possible subsets of classifiers for cardinality K we introduce the weighted average of the K-fault measures as

$$\overline{F_K} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1,i_2,\ldots,i_K=1 \\ i_j \neq i_k \ (1 \leq j < k \leq K)}}^{n} W_K F_K, \tag{56}$$

where $W_K \in \mathbb{R}$ for $K \in \{2, \ldots, n\}$. As for (56), notice that a non binary weighting became possible to be applied to the selected contingency entries. Possibilities on the optimal adjustment of these weights will be discussed in later chapters.

4.1.1.4.2 Full K-fault (F_{K*})

To improve the descriptive property of K-fault measure, we introduce the full K-fault one, which besides a subset $\mathcal{D}_S$ of K members ($2 \leq K \leq n$) with $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$ analyzes smaller collections of it. Namely, the full K-fault measure is introduced as

$$F_{K*} = \frac{\sum\limits_{i=K'}^{K} \left(W_i \sum\limits_{S' \subseteq S,\ |S'|=i} N^{\{0\}_i} \right)}{n\,(K - K' + 1)}, \tag{57}$$

with some $K' \in \{2, \ldots, K\}$. The main idea behind using this measure is that if in an application we have insufficient number of samples in the K-dimensional contingency cells, smaller dimensional ones can be considered instead; the smallest such subset size is given with K'.

4.1.1.4.3 K-disagreement (D_K)

The pairwise disagreement measure can be generalized to K components ($2 \leq K \leq n$) in the following way:

$$D_K = \frac{N - N^{\{1\}_K} - N^{\{0\}_K}}{N}. \tag{58}$$

Similarly to the previous measures, all the possible subsets with the same cardinalities can be aggregated via a weighted sum:

$$\overline{D_K} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1, i_2, \ldots, i_K = 1 \\ i_j \neq i_k\ (1 \leq j < k \leq K)}}^{n} W_K D_K, \tag{59}$$

4.1.1.4.4 Full K-disagreement (D_{K*})

Based on a similar argument discussed for the full K-fault measure F_{K*}, we introduce the full K-disagreement one D_{K*} as

$$D_{K*} = \frac{\sum\limits_{i=K'}^{K} \left(W_i \sum\limits_{S' \subseteq S,\ |S'|=i} N - N^{\{1\}_i} - N^{\{0\}_i} \right)}{N} \tag{60}$$

with some $K' \in \{2, \ldots, K\}$.

4.1.1.4.5 Product-moment correlation measure for K - (PM_K)

We generalize the pairwised product-moment correlation measure to K classifiers ($2 \leq K \leq n$) in a quite straightforward way as

$$PM_K = \frac{N^{\{0\}_K}}{\sqrt[K]{\prod_{\mathcal{N} \in N^{\{0\}1\{*\}_{K-1}}} \mathcal{N}}}. \tag{61}$$

As usual, to aggregate for all the possible subsets of classifiers we define the weighted average of the product-moment correlation measures as

$$\overline{PM_K} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1, i_2, \ldots, i_K = 1 \\ i_j \neq i_k \ (1 \leq j < k \leq K)}}^{n} W_K PM_K, \tag{62}$$

where $W_K \in \mathbb{R}$ for $K \in \{2, \ldots, n\}$.

4.1.1.4.6 Full product-moment correlation measure for K - (PM_{K*})

Similarly to the previous measures, we introduce the full product-moment correlation measure for K as

$$PM_{K*} = \sum_{i=K'}^{K} W_i \sum_{S' \subseteq S, \ |S'|=i} \frac{N^{\{0\}_i}}{\sqrt[i]{\prod_{\mathcal{N} \in N^{\{0\}1\{*\}_{i-1}}} \mathcal{N}}} \tag{63}$$

with some $K' \in \{2, \ldots, K\}$.

4.1.1.4.7 Correlation coefficient for K - (CK)

As for our generalization to cross correlation, the main structure of the original formula is preserved, however, the members are expressed in a completely different way. Namely, let $\mathcal{D}_S = \{\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_K}\}$ be a subset of classifiers of cardinality $|\mathcal{D}_S| = K$ with $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$, and $K' \in \{2, \ldots, K\}$. Then, the correlation coefficient for K is introduced as

$$CK = CK'/CK'', \tag{64}$$

where

$$CK' = \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=1,\ S'\subseteq S \\ \mathcal{D}_{i_k}\in S',\ |S'|=j+1, \\ \mathcal{N}\in N^{\{1\}_{j+1}}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=0,\ S'\subseteq S \\ \mathcal{D}_{i_k}\in S',\ |S'|=j+1 \\ \mathcal{N}\in N^{\{0\}_{j+1}}}} \mathcal{N} \right) - \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=1,\ S'\subseteq S \\ \mathcal{D}_{i_k}\in S',\ |S'|=j+1, \\ \mathcal{N}\in N^{\{1\}_{K-j}\{0\}_j}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=0,\ S'\subseteq S \\ \mathcal{D}_{i_k}\in S',\ |S'|=j+1 \\ \mathcal{N}\in N^{\{1\}_j\{0\}_{K-j}}}} \mathcal{N} \right), \tag{65}$$

and

$$CK'' = K \sqrt{\left(\sum_{k=1}^{K} \sum_{\substack{a_{i_k}=1 \\ \mathcal{N}\in N^{\{1\}_1\{*\}_{K-1}}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{\substack{a_{i_k}=0 \\ \mathcal{N}\in N^{\{0\}_1\{*\}_{K-1}}}} \mathcal{N} \right) - \left(\sum_{k=1}^{K} \sum_{j=K'+1}^{K-1} \frac{1}{j} \sum_{\substack{S'\subseteq S, \mathcal{D}_{i_k}\in S' \\ |S'|=j+1 \\ \mathcal{N}\in N^{\{1\}_j}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'+1}^{K-1} \frac{1}{j} \sum_{\substack{S'\subseteq S, \mathcal{D}_{i_k}\in S' \\ |S'|=j+1 \\ \mathcal{N}\in N^{\{0\}_j}}} \mathcal{N} \right)}. \tag{66}$$

4.1.1.4.8 Q statistics for for K - (QK)

The same technique presented for the correlation coefficient can be applied to the generalization of the Q statistics, as well. Namely, with also comparing the original formulas of the Q statistics and the correlation coefficient in Table 16, we introduce the Q statistics for K as

$$QK = CK'/QK'', \tag{67}$$

where

$$QK'' = \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=1,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1, \\ \mathcal{N} \in N^{\{1\}_{j+1}}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=0,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1 \\ \mathcal{N} \in N^{\{0\}_{j+1}}}} \mathcal{N} \right) +$$

$$\left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=1,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1, \\ \mathcal{N} \in N^{\{1\}_{K-j}\{0\}_j}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=0,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1 \\ \mathcal{N} \in N^{\{1\}_j\{0\}_{K-j}}}} \mathcal{N} \right). \tag{68}$$

The next seven diversity measures have already been introduced in the literature. These measures also accumulate the subsets of a given set of classifiers, so our generalization approach considered for the former measures are not applicable for these ones; they are enclosed for the sake of completeness, since we have found them efficient to discover better dependencies within an ensemble.

4.1.1.4.9 Measure of difficulty (DI)

This measure is based on the discrete error distribution $Z = [p_z(0), \ldots, p_z(K)]$ for a subset of classifiers $\mathcal{D}_S = \{\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_K}\}$ of cardinality $|\mathcal{D}_S| = K$ with $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$, where

$$p_z(j) = \sum_{|\{a_{i_t} : a_{i_t}=1, t=1,\ldots,K\}|=j} c_{a_{i_1},\ldots,a_{i_K}}$$

for $j = 0, \ldots, K$. Namely, the measure of difficulty is defined as

$$DI = \frac{1}{K} \sum_{j=0}^{K} (p_z(j) - \overline{p_z})^2, \tag{69}$$

where $\overline{p_z} = \sum_{j=0}^{K} p_z(j)/K$. For more details see [17].

4.1.1.4.10 The entropy measure (EN)

This measure considers a subset of voters $\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_K}$ and tries to calculate the level of dependency based on entropy via

$$EN = \sum_{i=1}^{N} \frac{min\{m(\chi_i), K - m(\chi_i)\}}{N\lfloor K/2 \rfloor}, \tag{70}$$

where $m(\chi_i)$ is the number of correct classifications for the sample χ. Thus, its range is between 0 and 1, where 1 means full diversity. For more details see [17].

4.1.1.4.11 Kohavi-Wolpert variance (KW)

This measure is based on a similar idea to the entropy one. Namely, it calculates the average variance of the outputs for all the classifier from a binomial distribution as

$$KW = \frac{1}{NK^2} \sum_{i=1}^{N} m(\chi_i)(K - m(\chi_i)). \tag{71}$$

Notice that $KW = K \cdot DI$ holds for independent classifiers. For more details see [17].

4.1.1.4.12 Interrater agreement measure (IA)

Another modification of the same idea is used for the interrater agreement measure via

$$IA = 1 - \frac{\sum_{i=1}^{N} m(\chi_i)(M - m(\chi_i))}{NK(K-1)\overline{e}(1-\overline{e})}, \tag{72}$$

where $\overline{e} = \sum_{i=1}^{K} \sum_{j=1}^{N} (1 - \eta_i^{(j)})/KN$ is the average error rate of the classifiers. For more details see [17].

4.1.1.4.13 The fault majority measure (FM)

This measure is based on partial error distribution (PDED) to utilize the degree of contributing to the different levels of ensemble error coincidences for each voters. Only those PDED components are aggregated that could

contribute to majority voting error for a given subset of classifiers. That is, for error coincidences with at least $\lceil K/2 \rceil$ errors and coming from $\lceil K'/2 \rceil$ locally best classifiers with $1 \leq K' \leq K$. More precisely, the measure is defined as

$$FM = \sum_{j=\lceil K'/2 \rceil}^{K'} \sum_{i^*=1}^{\lceil K'/2 \rceil} z_{i^*,j}, \tag{73}$$

where i^* stands for the classifiers sorted according to their values of $z_{i,j}$ for the fixed coincidence level j, and

$$z_{i,j} = \begin{cases} \sum_{k=1}^{N} [1 - \eta_i^{(k)} \mid m(x_k) = 0]/N & \text{for } j = 0, \\ M \sum_{k=1}^{N} [1 - \eta_i^{(k)} \mid m(x_k) = 0]/Nj & \text{for } j \neq 0. \end{cases} \tag{74}$$

For more information see [18].

4.1.1.4.14 Generalized diversity (GD)

The basic idea of the GD diversity is based on measuring the simultaneous failures $p(K')$ of K' voters which can be shown to be

$$q(K') = \sum_{i=1}^{K} \frac{i}{K} \frac{i-1}{K} \cdot \ldots \cdot \frac{i-K'+1}{K-K'+1} q_i, \tag{75}$$

where q_i denotes the probability that exactly i voters fail among the total K ones.

It can be easily verified that maximum diversity is observed for $q(2) = 0$, while minimum diversity for $q(2) = q(1)$. That is, the simple generalized diversity is introduced as

$$GD = 1 - \frac{q(2)}{q(1)} \tag{76}$$

with falling in the interval $[0, 1]$. For more details see [17].

4.1.1.4.15 Coincident failure diversity (CFD)

This measure is similar to the GD one described above, however, it considers higher order failure coincidences, as well. Namely, coincident failure diversity is defined as

$$CFD = \begin{cases} \frac{1}{1-q_0} \sum_{i=1}^{K} \frac{K-i}{K-1} q_i, & q_0 < 1, \\ 0, & q_0 = 1. \end{cases} \tag{77}$$

Again, we have maximum diversity, when $CFD = 1$. For more details see [17].

4.1.2 C-constrained diversity measures

The diversity measures described in the previous section give useful information on how to select the members to achieve the highest ensemble accuracy. More specifically, efficient diversity measures should correlate with the system accuracy. In the literature, the case when the classifier decision making method is not the majority rule is rarely examined. Our aim is to modify these measures to be applicable in spatial voting scenarios, as well. The main difference is that in our case we must observe the incorrect votes further when checking the dependency of two (or more) member algorithms.

The investigation of the spatial domain was motivated by a retinal image processing problem – the detection of the OD, which appears as a bright circular patch within the region of interest (ROI) in a retinal image, see Figure 15. We have observed that organizing more individual OD detector algorithms into an ensemble may raise detection accuracy. In this voting system, each individual OD algorithm votes in terms of a single pixel as its candidate for the OD center. The application of existing majority voting models are not completely adequate here, since they consider only the correctness of the votes, which concerns falling into the true OD region in this scenario. However, in our case, the spatial behavior of the votes is also important, since they vote together for a specific location of the OD, only if they fall within a region matching the OD geometry. Consequently, we should consider discs of diameter of the OD $d_{OD} \in \mathbb{R}_{>0}$ covering the candidates of the individual detector algorithms as shown in Figure 15.

The diameter d_{OD} can be derived by averaging the manual annotations made by clinical experts on a dataset and can be adjusted to the resolution of the image. As a final decision, the disc having diameter d_{OD} with maximal number of candidates is chosen for the OD location. In this combined system, we can make a good decision even if the false candidates have majority such as in the case illustrated in Figure 15. A wrong decision is made only when the number of correct ones is smaller than the number of the false candidates covered by a disc having diameter d_{OD}

Thus, when two algorithms miss, the chance of the final wrong decision increases only if these incorrect votes meet the geometric constraint with e.g., falling in a disc of diameter d_{OD} in the OD detection task. By this consideration, we label the member outputs with 1 if the vote is correct, 0^- if the vote is wrong and there is another wrong vote close to it meeting the

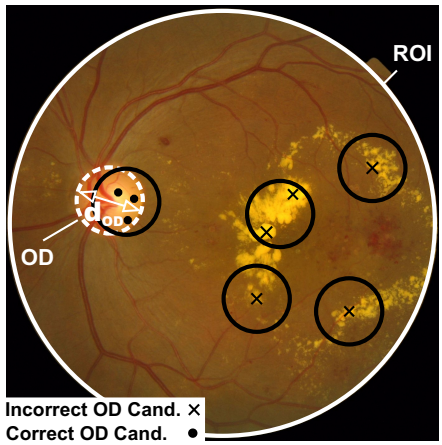


Figure 15: The OD of diameter d_{OD} in a retinal image and the OD center candidates (3 correct, 5 false) of individual detector algorithms. Candidates inside the black circles can vote together for possible OD locations.

geometric constraint, and 0^+ if the vote is wrong, but no other wrong vote can be found close to it. If we do no mind how worse an incorrect vote is, we simply write 0, that is, a 0 vote is either 0^+ or 0^- Thus, we generalize the term N^{ab} to let $a, b \in \{0, 0^+, 0^-, 1\}$, where e.g. $N^{0^-0^-}$ is the number of cases, when both algorithms miss and their votes meet also the geometric constraint. With this generalization, we have $t = N^{0^+0^+} + N^{0^-0^-} + N^{0^-1} + N^{10^-} + N^{0^+1} + N^{10^+} + N^{11}$ for the total number of cases for any two classifiers $\mathcal{D}_i, \mathcal{D}_{i'}$. Now, the diversity measures listed in the previous section can be generalized as follows.

4.1.2.1 C-constrained double fault

The pairwise C-constrained double fault measure estimates the diversity of two classifiers $\mathcal{D}_i, \mathcal{D}_j$ $(i, j \in \{1, \ldots, n\}, i \neq j)$ according to their outputs which satisfy the C-constraint. That is, we define it as

$$F_{ij}^C = N^{0^-0^-}/N. \tag{78}$$

The C-constrained double fault measure is introduced also within the whole set of base classifiers as

$$\overline{F^C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} F_{ij}^C. \tag{79}$$

The C-constrained double fault can be simply extended to a subset of classifiers $\mathcal{D}_S = \{\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_K}\}$ of cardinality $|\mathcal{D}_S| = K$ with $S = \{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$ to define the C-constrained K-fault via

$$F_K^C = N^{\{0^-\}_K}/N, \tag{80}$$

while for all the possible subsets of classifiers of cardinality K a weighted average can be calculated as

$$\overline{F_K^C} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1,i_2,\ldots,i_K=1 \\ i_j \neq i_k \ (1 \leq j < k \leq K)}}^{n} W_K F_K^C. \tag{81}$$

Moreover, analogously to the K-fault measure, the full K-fault measure is generalized to its C-constrained one as

$$F_{K*}^C = \frac{\sum_{i=K'}^{K} W_i \sum_{S' \subseteq S,\ |S'|=i} N^{\{0^-\}_i}}{n\,(K - K' + 1)} \tag{82}$$

using the common notations.

4.1.2.2 C-constrained disagreement

All the variants of the disagreement measures are generalized to their C-constrained variants similarly to the K-fault ones. That is, the pairwise C-constrained disagreement measure has the form

$$D_{ij}^C = \frac{N^{0^-1} + N^{10^-}}{N}, \tag{83}$$

while all the pairs can be aggregated via

$$\overline{D^C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} D_{ij}^C. \tag{84}$$

Next, the C-constrained K-disagreement has the form

$$D_K^C = \frac{N - N^{\{1\}_K} - N^{\{0^-\}_K}}{N}, \tag{85}$$

while all the possible subsets with the same cardinalities can be aggregated as

$$\overline{D_K^C} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1,i_2,\ldots,i_K=1 \\ i_j \neq i_k \ (1 \leq j < k \leq K)}}^{n} W_K D_K^C. \tag{86}$$

Furthermore, the C-constrained full disagreement is introduced as

$$D_{K*}^C = \frac{\sum_{i=K'}^{K} W_i \sum_{S' \subseteq S, \ |S'|=i} N - N^{\{1\}_i} - N^{\{0^-\}_i}}{N}. \tag{87}$$

4.1.2.3 C-constrained product-moment correlation

The original product-moment correlation measure aims to explore diversity via focusing on the coincident errors, so the C-constraint holds. Thus, the C-constrained variants can be easily generalized as

$$PM_{ij}^C = \frac{N^{0^-0^-}}{\sqrt{N^{*0}N^{0*}}} \tag{88}$$

for the pairwise measure. For the whole set of classifiers

$$\overline{PM^C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} PM_{ij}^C \tag{89}$$

can be considered. The C-constrained product-moment correlation measure for K has the form

$$PM_K^C = \frac{N^{\{0'\}_K}}{\sqrt[K]{\prod_{\mathcal{N} \in N^{\{0\}_1\{*\}_{K-1}}} \mathcal{N}}}, \tag{90}$$

and aggregation on subsets can be considered with

$$\overline{PM_K^C} = \frac{1}{\binom{n}{K}} \sum_{\substack{i_1,i_2,\ldots,i_K=1 \\ i_1 \neq i_2 \neq \cdots \neq i_K}}^{n} W_K PM_K^C. \tag{91}$$

Moreover, the full product-moment correlation measure for K is introduced as

$$PM_{K*}^{C} = \sum_{i=K'}^{K} W_i \sum_{S' \subseteq S,\ |S'|=i} \frac{N^{\{0'\}_i}}{\sqrt[i]{\prod_{\mathcal{N} \in N^{\{0\}_1 \{*\}_{i-1}}} \mathcal{N}}}. \tag{92}$$

4.1.2.4 C-constrained correlation

Compared with the extension of the previous measures to the spatial domain, the generalization of the correlation and Q statistics measure needs further considerations. Namely, starting with the pairwise correlation, while the N^{00} terms can be mapped to $N^{0^-0^-}$ as before, the 0^+ cases should be incorporated in the original N^{01} and N^{10} disagreement terms to reflect that if the answer is incorrect, it cannot significantly influence the final decision in the constrained case for larger ensembles. That is, the pairwise C-constrained correlation is introduced as

$$C_{ij}^{C} = \frac{N^{11}N^{0^-0^-} - \left(N^{01} + \frac{N^{0^+0^+}}{2}\right)\left(N^{10} + \frac{N^{0^+0^+}}{2}\right)}{\sqrt{N^{1*}N^{0*} - N^{*1}N^{*0}}} \tag{93}$$

with the possible aggregation for all the cases via

$$\overline{C^C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} C_{ij}^{C}. \tag{94}$$

Naturally, the same considerations are to be applied to the extensions to K classifiers. That is, the C-constrained correlation coefficient for K is introduced as

$$CK^{C} = CK^{C'}/CK'', \tag{95}$$

where

$$CK^{C'} = \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=1,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1, \\ \mathcal{N} \in N^{\{1\}_{j+1}}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{\substack{a_{i_k}=0^-,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1 \\ \mathcal{N} \in N^{\{0^-\}_{j+1}}}} \mathcal{N} \right) -$$
$$\left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{l=2}^{j} W_l \sum_{\substack{a_{i_k}=1,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1, \\ \mathcal{N} \in N^{\{1\}_{K-j-1}\{0\}_{j-l}\{0^+\}_l}}} \mathcal{N} \right) \left(\sum_{k=1}^{K} \sum_{j=K'}^{K-1} W_j \sum_{l=2}^{K-j-1} W_l \sum_{\substack{a_{i_k}=0,\ S' \subseteq S \\ \mathcal{D}_{i_k} \in S',\ |S'|=j+1, \\ \mathcal{N} \in N^{\{0\}_{K-j-1}\{1\}_j\{0^+\}_l}}} \mathcal{N} \right). \tag{96}$$

Notice that similarly to the pairwise case, the denominator of CK^C coincides with that of CK.

4.1.2.5 C-constrained Q-statistics

Considerations similar to the correlation measures are applied to derive the C-constrained versions of the Q-statistics related measures. Accordingly, the pairwise C-constrained Q-statistics is introduced as

$$Q_{ij}^C = \frac{N^{11}N^{0^-0^-} - \left(N^{01} + \frac{N^{0^+0^+}}{2}\right)\left(N^{10} + \frac{N^{0^+0^+}}{2}\right)}{N^{11}N^{0^-0^-} + \left(N^{01} + \frac{N^{0^+0^+}}{2}\right)\left(N^{10} + \frac{N^{0^+0^+}}{2}\right)} \tag{97}$$

with possible aggregation via

$$\overline{Q^C} = \frac{2}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} Q_{ij}^C. \tag{98}$$

Moreover, the Q-statistics for K is defined as

$$QK^C = CK^{C'}/QK'', \tag{99}$$

with noticing that the nominator is the same as that of the C-constrained correlation, while the denominator coincides with the non-constrained Q-statistics for K.

4.1.2.6 More C-constrained measures

As for the C-constrained variants of the measure of difficulty (DI^C), entropy measure (EN^C), Kohavi-Wolpert variance (KW^C), interrater agreement measure (IA^C), fault majority measure (FM^C), generalized diversity (GD^C), coincident failure diversity (CFD^C) listed at the end of section 2, all in these seven measures the 0^- votes should be considered instead of the original 0 ones in all the aggregations.

4.1.2.7 C-constrained measures based on individual performances

The measures presented in section 4.1.1.3.6 are based on the individual performances of the voters, so it does not make sense to interpret them as pairwise or K-wise ones. However, their generalization to C-constrained variants is meaningful just following the same way given in section 4.1.2.6

with considering only the 0^- votes instead of the 0 ones. That is, we can introduce the C-constrained minimum individual error (MIE^C), the mean error (ME^C), the majority rule error (MRE^C), and the system improvement (SI^C).

4.1.3 Empirical results

As a basic consideration of our approach, most of the generalized diversity measures incorporate weight terms to provide larger flexibility. These weights can also be considered as free parameters adjustable according to the specific application.The generalized measures were composed to return the original ones with a specific weight adjustment (as we have shown for e.g., double fault), however, for specific applications practical considerations can be useful. The determination of the appropriate weights has to be based on the definition of the diversity measure and the decision model collectively. Namely, we have followed the strategy that larger weights were assigned to that cases which have positive influence on the increment of the system accuracy and vice versa there are proportional penalties for cases with a contrary effect. In our practical application, the votes should fall within a region matching the OD geometry, as it was explained in section 4.1.2. For example, in our application the weights W_i of C-constraint full K-fault measure (4.1.2.1) are decreasing negative numbers. As far as weighted average is concerned, the performance of the classifiers from the subsets (p_j or $1 - p_j$, where $j = 1 \dots K$) can be considered to be a basis in finding the weights. If there is a representative training database, the weights can be learnt by the help of this database. The strategy to follow is that larger weights are assigned to such situations whose original versions show larger correlation with the system accuracy.

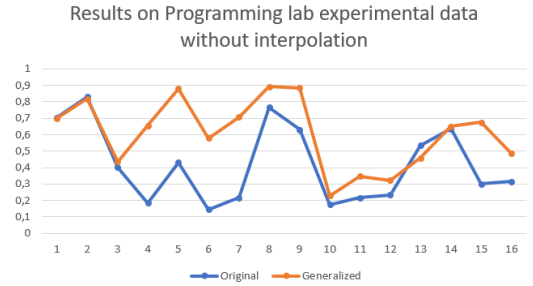
To check the efficiency of the generalized diversity measures, we have checked their correlations with the ensemble performance and how they can be used to compose accurate ensembles. Firstly, we checked the correlation with system accuracy of the generalized diversity measures in an experiment on the data from the test papers of programming lectures, where the voters were students, who have to pass a multiple-choice driving test. The number of incorporated students was 135, and the number of the questions was 50. Secondly, we checked the correlation with system accuracy of the c-constraint diversity measures in an experiment for the OD detection task described in section 4.1.2. Eight OD detector algorithms were involved in this analysis, namely [44], [10], [9], [53], [46], [63], [3]. As the experimental data of the retinal application, the publicly available database KAGGLE [20]

has been used which contains 53576 images for the study of the generalized C-constraint diversity measures.

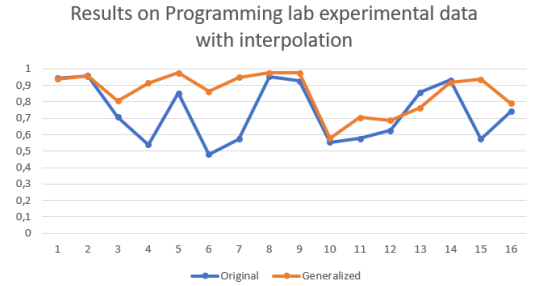
4.1.3.1 Correlation with ensemble performance

In case of the multiple-choice driving test experiment, the system accuracy is calculated with the help of its answer key. For the practical problem of OD detection, the system accuracy is measured on the database [20] described above. For the final decision on the location of the OD, constrained majority voting-based rule has been applied where only those outputs of the algorithms vote together, which fall in the same disc of diameter d_{OD}. More details are given in section 4.1.2.

We have calculated the ensemble performances for all the possible ensembles and also all the diversity measures. In our experiment with the classical generalized diversity measures the number of possible ensembles is given by $\sum_{k=1}^{n=60} \binom{n}{k}$. However, there is no real point for the ensembles having only one or two members to be included into the investigation for practical reasons, since substituting $n = 2$ into the generalized diversity measures we get back the original formulas. While this size effects unmanageable pieces of different of ensembles, we reduced the number of voters by grouping the voters into the disjunct sets with ten cardinalities. Finally, we performed the experiment on these disjunct sets separately without and with interpolation (discussed later), and the final results are calculated by averaging them. In Figure 16(a) we present the correlation between the diversity measures and the ensemble performance. To make the visualization feasible, we have averaged these values for each possible ensemble cardinality $K = 3, \ldots, 60$.



(a) Results on Programming lab experimental data without interpolation



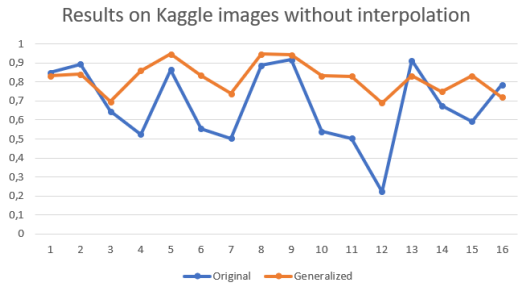
(b) Results on Programming lab experimental data with interpolation

Figure 16: Result of evaluating the generalized diversity measures based on multiple-choice driving test experiment.

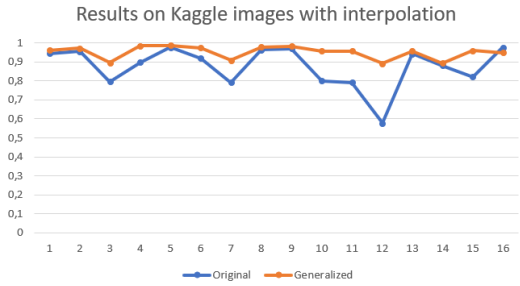
The Figure 16(a) and 16(b) show the correlation between system performance and generalized diversity measures without and with applying interpolation, respectively. The values of the ordinate are the correlation values, and the individual generalized diversity measures are listed on the abscissa. Based on the plot values, the generalized diversity measures estimate by 18.7% and by 12% more accurately the system performance, respectively.

To test the C-constraint diversity measures exhaustively, all the possible ensembles and also all the diversity measures were considered. In our experiment with the generalized C-constraint diversity measures the number of possible ensembles is given by $\sum_{k=1}^{n=60} \binom{n}{k}$. However, because of practical

reasons, only the cases $n > 2$ were investigated. In the experiment we have calculated the ensemble performances for 219 variously composed ensembles. In Figure 17(a) we present the correlation between the diversity measures and the ensemble performance. To make the visualization feasible, we have averaged these values for each possible ensemble cardinality $K = 3, \ldots, 8$.



(a) Results on Kaggle images without interpolation



(b) Results on Kaggle images with interpolation

Figure 17: Result of evaluating the generalized C-constrained diversity measures based on KAGGLE data experiment.

The Figure 17(a) and 17(b) show the correlation between system performance and generalized C-constraint diversity measures without and with applying interpolation on Kaggle images, respectively. The values of the ordinate are the correlation values, and the individual generalized C-constraint diversity measures are listed on the abscissa. Based on the plot values, the generalized C-constraint diversity measures estimate by 14.1% and by 7.5% more accurately the system performance, respectively.

From the figures 16 and 17, it can be clearly seen that the (C-constrained) generalizations of the measures correlate much better with the ensemble accuracy, so they have a stronger descriptive dependency behavior than the existing measures.

Another practical issue arises for incomplete contingency tables, when some realizations are missing. Technically, it makes the corresponding contingency entries 0, which may ruin some diversity measures (e.g. where a product is considered). To resolve this problem we may apply an interpolation technique – specifically we have tested the 5-point moving average one – to the contingency table to eliminate the 0 entries. This solution also makes the correlation of the diversity measures with the system performance smoother even if the rate of correlation may slightly drop. These observations have been made in our OD detection scenario with the correlation between the diversity measures and the ensemble accuracy depicted in Figure 18.

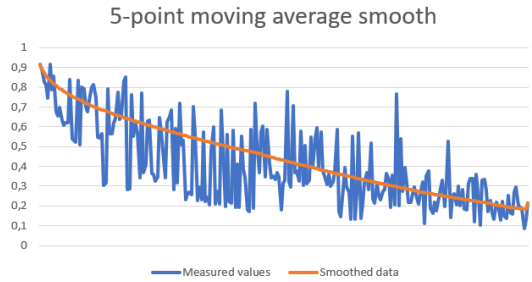


Figure 18: The result of the interpolation applying 5-point moving average kernel on the measured data in case one of the diversity measures. The plot shows how smooth the correlation values are after the interpolation.

Moreover in the figures 17(b) and 16(b) there are the plots about interpolated experiment results for comparison reasons to the non-interpolated investigations. It is clearly shown, that the performance superiority of the generalized diversity for both cases is maintained. The positive effect of the interpolation step is also obvious. Moreover, an important feature of diversity measures is observable too. Namely, the generalized diversity measure is less vulnerable to the incompleteness of the contingency tables.

4.1.3.2 Application to composing ensembles

Besides the diversity measures discussed so far, we can introduce an aggregated (super) measure that incorporates all the dependency observations supplied from the simple measures. To make this super measure adoptable to different applications and to compensate the possible less descriptive behavior of some measures – caused e.g. by the lack of specific samples causing 0 or close to 0 entries in the contingency table – we propose a weighted linear combination of the diversity measures.

More formally, suppose that n classifiers and h diversity measures $\mathcal{DM}_1$, $\ldots, \mathcal{DM}_h$ are given and the aim is to compose a system from the classifiers having the highest accuracy using the diversity measures. For all the possible $\binom{n}{k}$ $(k = 1, \ldots, n)$ ensembles, we calculate a weighted combination of the diversity measures as

$$GD_j = \sum_{i=1}^{h} \alpha_{i,j} \mathcal{DM}_{i,j}, \quad j = 1, \ldots, \binom{n}{k}, \quad k = 1, \ldots, n, \tag{100}$$

where $\alpha_{i,j} \in \mathbb{R}_{\geqslant 0}$ are some weights, $\mathcal{DM}_{i,j}$ is the value of the i-th diversity measure on the j-th ensemble. With assuming stronger correlation between the weighted combination and system accuracy, the ensemble with the maximal GD_j is selected

$$GD = \max_j(GD_j) = \sum_{i=1}^{h} \alpha_i \mathcal{DM}_i. \tag{101}$$

The appropriate selection of the weights α_i are well-known from the literature for independent feature selectors. Namely, the optimal weights can be determined from the individual accuracies of the feature selectors [41]. In this special case, the correlation values show the performance of the diversity measures as feature selectors. If we consider independent members $\mathcal{DM}_1, \ldots, \mathcal{DM}_h$ with accuracies $dp_1, \ldots, dp_h$ then GD can be maximized by assigning the weights

$$\alpha_i = \ln \frac{acorr_i}{1 - acorr_i}, \quad i = 1, \ldots, h. \tag{102}$$

In our application, the accuracy $acorr_i$ is defined as the average correlation of the i-th diversity measure with the system accuracy regarding all possible assembled ensembles having the same number of members. For our OD detection task, in [7] we have found that an ensemble composed using the GD figure outperforms the one collecting simply the individually most accurate

members, or by considering only a single generalized diversity measure for this aim. That is, to compose the ensemble we can rely on the combined diversity measure GD, as it correlates quite well with the system accuracy demonstrated also in Table 17.

System accuracy	**Ensemble of detectors**								***GD***
	OD_{pd}	OD_{ed}	OD_{ef}	OD_{fv}	OD_{ht}	OD_{fu}	OD_{ch}	OD_{ta}	
0.9774		X	X	X		X			85.68
0.9765	X	X	X	X	X	X	X		86.35
0.9783	X	X		X	X	X	X	X	86.36
0.9774	X	X	X	X		X		X	89.88
0.9800	X	X	X	X	X	X			89.88

Table 17: Selecting ensembles by using a weighted linear combination of generalized diversity measures.

4.1.4 Conclusion

In classical theory, diversity measures are recommended to help to explore the dependencies among the classifiers. As we have also pointed out, the detector algorithms cannot be considered as independent classifiers in applications. Moreover, in the previous section it was proved by heretical based investigations that there is dependency in our example object detection application. Our aim was to describe the dependency to improve the selection efficiency of classifiers towards composing more performance optimized ensemble compared with that it is composed by simply selecting the individually most accurate members. We showed how classical pairwise measures could be extended to explore larger sets of classifiers via letting the contingency table entries of an ensemble-based system be freely weighted. These generalized diversity measures fit to use with classical majority voting and other multidimensional investigations about diversity. As experimental proof, we showed how much they work better in case of the classical majority. As a more specific use of diversity measures, we introduced extended measures for spatial constraint majority voting framework. We showed experimentally that these measures fit better to spatial applications with a specific example of OD detection. Moreover, we showed also how a more efficient descriptor could be found in terms of a weighted combination of diversity measures which correlates better with the accuracy of the ensemble. In our experimental test, we use the publicly available KAGGLE database [20].

In the following chapter, our primary goal remains similar. We try to compose the best performance fusion system. For this purpose, we extend

our model by adding a natural constraint regarding a time limit within which the ensemble should make the decision. For this aim, we consider both the execution time and accuracy of each member. Then, we solve the problem on how to find the most accurate ensemble, where the sum of the execution times of its members remains belows the limit. As a decision rule, we analyze a majority voting-based one generalized to be applicable in single object detection scenarios. This optimization task leads to a non-separable Knapsack problem, which is addressed using stochastic considerations.

4.2 Composing an ensemble-based system guided by a constraint

There are many possibilities to improve further the performance of ensemble-based systems. One possibility, e.g., is to adjust the free parameters of the members to provide maximal ensemble accuracy [35]. High ensemble accuracy is a generally fundamental requirement. However, other performance issues might be reasonable. One specific such point is the execution time since ensembles are more resource demanding in this respect, as they have to execute all their members before making their decision. Over a specific total execution time, we cannot achieve a considerable increase in system accuracy by adding new members to the ensemble. In this section, we address this issue with solving the problem to compose the most accurate ensemble from certain possible members under a constraint on execution time. The primary challenge of this optimization problem is that the selection of ensemble accuracy q in (103) as an energy function leads to a non-separable task, which cannot be solved easily using, e.g., classical dynamic programming approaches:

$$q_{classical} = \sum_{k=0}^{n} \Bigg(\sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\dots,n\} \setminus I} (1 - p_j) \Bigg), \tag{103}$$

After formulating the problem as a knapsack one, we give a theoretical approach for optimization and also validate it in a single object detection problem.

The rest of the section is organized as follows. A proper formulation of the problem in terms of a knapsack problem is given in section 4.2.1. In section 4.2.2, we consider several specific cases for accuracies and execution times of the potential members of the ensemble to have a better insight of the problem and discuss the cases when the optimization can be performed easily.

Our approach to solving the general problem is presented in section 4.2.3. In section 4.2.4, we enclose some experimental results on the specific detection problem to locate the OD in retinal images. Finally, some conclusions are drawn in section 4.2.5.

4.2.1 Formulation of the knapsack problem

First, we present the classical knapsack problem and then formulate it for our ensemble composition task with a short discussion on the difficulties of selecting q from (104) as the target (energy) function:

$$q = \sum_{k=0}^{n} p_{n,k} \Bigg(\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{1,\ldots,n\} \setminus I} (1 - p_j) \Bigg). \tag{104}$$

4.2.1.1 The classical knapsack problem

Let n items be given, with values $v_1, \ldots, v_n$ and weights $w_1, \ldots, w_n$, respectively. Let x_i be the number of the i-th item to be packed (zero or one). The maximum weight being carried in the knapsack is W. It is common to assume that all values and weights are non-negative. The aim is to maximize the function $\sum_{k=1}^{n} x_k v_k$ under the following conditions:

$$\sum_{k=1}^{n} x_k w_k \leq W, \quad x_k \in \{0, 1\} \quad (k = 1, \ldots, n). \tag{105}$$

According to a vast number of applications, the corresponding properties of the objective function and/or constraints led to many variations of the original knapsack problem. The major examples from the related literature are about real or integer variables, linear or non-linear, separable or non-separable, convex or non-convex objective functions, and additional specially structured constraints such as bounds on the variables or generalized upper bound constraints. Although some varieties of nonlinear knapsack problems are investigated in the literature [36], [59], most of the research efforts have been focused on knapsack problems with a linear objective and a linear constraint [57], and knapsack problems with a separable convex non-linear objective function and a simple linear equality constraint [61].

4.2.1.2 Knapsack problem for majority voting target function

In the investigated ensembles, each member is an object detector with its accuracy p_i as value and with its running time t_i as weight in a constrained majority voting system. The problem is to find the most accurate ensemble from these algorithms with bounded total execution time. That is, now we have to maximize the ensemble accuracy q given in (104).

More specifically, the aim is to get the optimal solution q_T for the system accuracy, where:

$$q_T = \max_{\{i_1,\ldots,i_s\}} \left\{ \sum_{k=0}^{s} p_{s,k} \Bigg(\sum_{\substack{I \subseteq \{i_1,\ldots,i_s\} \\ |I|=k}} \prod_{i \in I} p_i \prod_{j \in \{i_1,\ldots,i_s\} \setminus I} (1-p_j) \Bigg) \right\} \tag{106}$$

under the following conditions:

$$\sum_{j=1}^{s} t_{i_j} \leq T, \quad \{i_1,\ldots,i_s\} \subseteq \{1,\ldots,n\} \quad (s=1,\ldots,n). \tag{107}$$

The optimization task formalized above is far from being trivial to solve. The primary challenge can be easily recognized with noticing that the target function q of the majority voting rule is non-linear and non-separable. As we have mentioned before, such functions are rarely investigated in the literature with a strict restriction on their functional structure (e.g., the exponential one is analyzed in [59]). That is, for a proper analysis we need a model for the optimization of the specific target function (104) within the knapsack framework.

4.2.2 Relation of member accuracies and execution times

There can be several natural relationships between the running time t_i and accuracy p_i of the members of the ensemble. Regarding this issue, we formulate some specific cases besides the most general scenario, when both the execution times and accuracy values may be different within the system. In all of our examples, we will consider a fixed number $n = 7$ for the number of algorithms from which the ensemble can be composed. This selection leads to better comparability with our practical application (see section 4.2.4) also considering seven algorithms. As for the time limit T from now on we will

suppose that $T \leq T_{total} = \sum_{i=1}^{n} t_i$, since the total computational time of the algorithms is equal to T_{total}.

- *CASE1: Both the execution times and accuracies are constants.* The intuitive interpretation of this case is that all the members have the same accuracy and execution time. As a concrete example, we show some corresponding experimental results for ensembles having the following parameters:

  - EXAMPLE1: $p{=}p_1{=}\ldots{=}p_7{=}0.6$; $t{=}t_1{=}\ldots{=}t_7{=}1$,
  - EXAMPLE2: $p{=}p_1{=}\ldots{=}p_7{=}0.7$; $t{=}t_1{=}\ldots{=}t_7{=}2$,
  - EXAMPLE3: $p{=}p_1{=}\ldots{=}p_7{=}0.8$; $t{=}t_1{=}\ldots{=}t_7{=}3$,
  - EXAMPLE4: $p{=}p_1{=}\ldots{=}p_7{=}0.9$; $t{=}t_1{=}\ldots{=}t_7{=}4$.

  The setup of these cases emulates the situations when more accurate members require more time. The knapsack problem can be trivially solved for *CASE1* for a given time limit T with noticing that because of the equal member accuracies the best strategy is to immediately add a new member to the ensemble till the time limit is reached leading to $\left\lfloor \frac{T}{t} \right\rfloor$ for the number of possible ensemble members. Now, the accuracy of the system can be easily calculated using (108):

$$q_{equal} = \sum_{k=0}^{n} p_{n,k} \binom{n}{k} p^k (1-p)^{n-k}. \tag{108}$$

  The maximum ensemble accuracy that can be reached for different time limits is shown in Fig. 19, for the examples above. The corresponding curves show that after a while the ensembles of slower, but more accurate members overtake the ones consisting of faster, but less accurate algorithms.

- *CASE2: The execution times are constant, the accuracies are different.* This case interprets such a scenario, where the members are not equally accurate, but their execution times are the same. As a concrete example we show some corresponding results for one ensemble having the following parameters:

  - EXAMPLE5: $p_1{=}0.6$, $p_2{=}0.65$, $p_3{=}0.7$, $p_4{=}0.75$, $p_5{=}0.8$, $p_6{=}0.85$, $p_7{=}0.9$; $t{=}t_1{=}\ldots{=}t_7{=}4$.

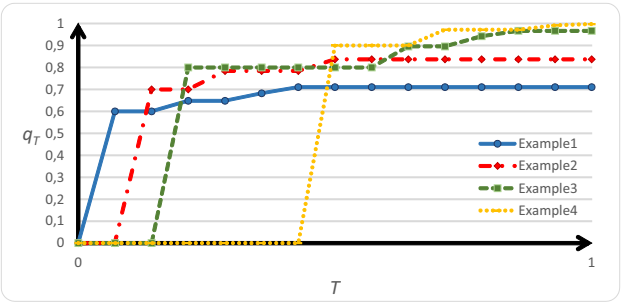


Figure 19: The maximum ensemble accuracies in *CASE1*. The graph plotted the system accuracy q_T over the normalized time T.

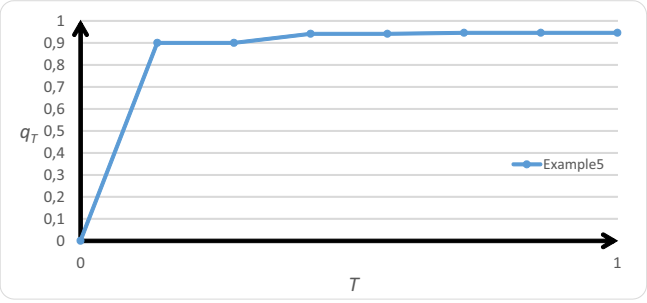


Figure 20: The maximum ensemble accuracy in *CASE2*.

The knapsack problem can be trivially solved for *CASE2* by observing that because of the equal execution times the best strategy is to add the most accurate one first, then the second most accurate one, etc. till reaching the time limit. For the number of possible ensemble members we have $\left\lfloor \frac{T}{t} \right\rfloor$ with the accuracy of the system calculated by (104). The maximum ensemble accuracy q_T versus the normalized time limit T is shown in Fig. 20.

- *CASE3: The accuracies are constant, the execution times are different.* Just as a reversed version of *CASE2* now the system members are equally accurate, but take different times to run. As concrete examples we show some corresponding results for four ensembles having the following parameters:

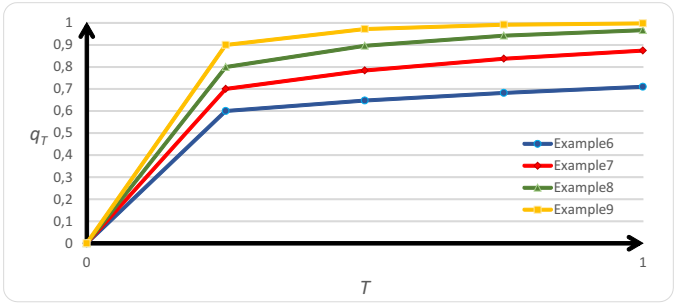


Figure 21: The maximum ensemble accuracies in *CASE3*.

– EXAMPLE6: $p{=}p_1{=}\ldots{=}p_7{=}0.6$; $t_1{=}1,\ldots,t_7{=}7$,
– EXAMPLE7: $p{=}p_1{=}\ldots{=}p_7{=}0.7$; $t_1{=}1,\ldots,t_7{=}7$,
– EXAMPLE8: $p{=}p_1{=}\ldots{=}p_7{=}0.8$; $t_1{=}1,\ldots,t_7{=}7$,
– EXAMPLE9: $p{=}p_1{=}\ldots{=}p_7{=}0.9$; $t_1{=}1,\ldots,t_7{=}7$.

Because of the same accuracies, the solution is quite simple since now we have to put the fastest algorithm first into the knapsack, then the second fastest one, etc. till reaching the time limit. For the number of possible ensemble members we have the maximal m for which $\sum_{i=1}^{m} t_i \leq T$ with the accuracy coming from (108). The maximum ensemble accuracy q_T versus T is plotted in Fig. 21. It can be observed how the more accurate members naturally lead to better performance in this case.

- *CASE4: Both the accuracies and execution times are different.* As the most general and natural scenario we turn to discuss on the case, when the members have both different execution times and accuracies. As concrete examples we show some corresponding results for two ensembles having the following parameters:

  – EXAMPLE10: $p_1{=}0.6$, $p_2{=}0.65$, $p_3{=}0.7$, $p_4{=}0.75$, $p_5{=}0.8$, $p_6{=}0.85$, $p_7{=}0.9$; $t_1{=}1,\ldots,t_7{=}7$,
  – EXAMPLE11: $p_1{=}0.6$, $p_2{=}0.65$, $p_3{=}0.7$, $p_4{=}0.75$, $p_5{=}0.8$, $p_6{=}0.85$, $p_7{=}0.9$; $t_1{=}7,\ldots,t_7{=}1$.

  The determination of the most accurate ensemble for a given time limit T is much more challenging in this case. In the next section,

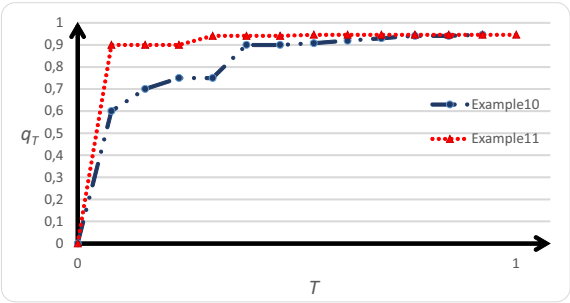


Figure 22: The maximum ensemble accuracies in *CASE4*.

we propose a stochastic algorithm to find an efficient solution for this specific knapsack problem. The maximum ensemble accuracy q_T regarding the time limit T is shown in Fig. 22. EXAMPLE10 stands for a natural case when a more accurate algorithm needs more execution time. However, EXAMPLE11 illustrates well that it is highly worth to put some effort to try to make more accurate algorithms as a massive improvement of performance can be gained.

4.2.3 Stochastic approach for knapsack problems with majority voting-based energy function

A greedy approximation algorithm was proposed by G. Dantzig to solve the unbounded knapsack problem [57]. First, the items are sorted in decreasing order of value per unit of weight, $\frac{v_i}{w_i}$. Then, they are inserted into the knapsack in this order with as many copies as possible until there is no longer space in the knapsack. Greedy algorithms are usually the most straightforward and most efficient algorithms available when they result in optimal solutions. Provided that there is an unlimited supply of each kind of item, then the greedy algorithm is guaranteed to achieve at least the half of the maximum possible value. However, when the supply of each kind of item is limited, the algorithm may be very far from being optimal. Moreover, in contrast to the classical knapsack problem, our optimization task cannot be solved using dynamic programming because of the non-linear, non-separable objective function. These results show that several significant modifications are needed to find the optimal solution.

In our proposed approach, the selection of the items to the ensemble is based on the efficiency of the members determined in the following way: for

the i-th item with accuracy p_i and execution time t_i, the system accuracy $q(p_i, t_i)$ of the ensemble containing maximal number of i-th items:

$$q(p_i, t_i) = \sum_{k=0}^{\left\lfloor \frac{T}{t_i} \right\rfloor} p_{\left\lfloor \frac{T}{t_i} \right\rfloor, k} \binom{\left\lfloor \frac{T}{t_i} \right\rfloor}{k} {p_i}^k (1 - p_i)^{\left\lfloor \frac{T}{t_i} \right\rfloor - k} \tag{109}$$

characterizes the efficiency of the i-th kind of item, instead of the values $\frac{p_i}{t_i}$ considered in the classic greedy method.

A greedy algorithm for an optimization problem always chooses the item that seems to be the best at that moment. In our selection method, a discrete random variable depending on the efficiency values of the remaining items is applied in each step to determine the probability of choosing an item from the remaining set to add to the ensemble. Namely, in the k-th selection step, if the items $i_1, \ldots, i_{k-1}$ are already in the ensemble, then the efficiency values $q^{(k-1)}(p_i, t_i)$ of the remaining items are updated to the maximum time of $T - \sum_{j=1}^{k-1} t_{i_j}$. The i-th item is selected as the next member of the ensemble with the following probability:

$$(p_{ens})_i^{(k)} = \frac{q^{(k-1)}(p_i, t_i)}{\sum\limits_i q^{(k-1)}(p_i, t_i)}, \tag{110}$$

where $i \in \{1, \ldots, n\} \backslash \{i_1, \ldots, i_{k-1}\}$. This discrete random variable reflects that the more efficient the item is, the more probable it is selected to the ensemble in the next step.

If $t_i > T - \sum_{j=1}^{k-1} t_{i_j}$ for all $i \in \{1, \ldots, n\} \backslash \{i_1, \ldots, i_{k-1}\}$, then our stochastic process ends because there is not enough remaining time for any algorithms to be executed.

The variation of the efficiency values in (109) and the selection probabilities of the items in (110) derived from the terms $p_{n,k}$ for the classical majority voting rule is shown step by step in Table 18 for EXAMPLE10, when $T = 0.85 \sum_{j=1}^{n} t_i$, if the most efficient item is supposed to be selected at each step.

For the further characterization of the results that can be achieved by the proposed method, it is a beneficial task to investigate the distribution of the system accuracy q considered as a random variable. To determine its distribution function $F_q(x)$, the distribution of the accuracy values of the algorithms needs to be described. For an approach solving an optimization problem, several fundamental questions are raised: when it is able to find the optimal solution, how far the final result of the method can be from the optimal solution or rather how often this result is above a particular value.

To answer these questions, it is expedient to examine the complementary cumulative distribution function $\bar{F}_q(x) = P(q \geq x) = 1 - F_q(x)$.

Here the system accuracy q is defined as a linear combination of the random variable ξ derived from the accuracy values. The distribution function of ξ^k $(k = 1, \ldots, n)$ is determined by the distribution of ξ because the power functions are bijective for the non-negative random variable ξ. It is a very challenging task to describe the distribution function $F_q(x)$ in the next step because the random variable ξ^k has different kind of distributions for $k = 1, \ldots, n$. By applying the moment generating function technique (that is $M_\eta(x) = E(e^{x\eta})$), the distribution of a linear combination of independent random variables can be determined uniquely. Because of the special and complex form of the system accuracy q, it can frequently occur that we cannot find an exact and close form for its distribution function $F_q(x)$. In this case, we can either give a numerical approximation for the distribution function by applying advanced numerical methods or Markov's Inequality is adaptable to estimate the values of the complementary cumulative distribution function $\bar{F}_q(x)$:

$$P(q \geq x) \leq \frac{E(q)}{x}, \tag{111}$$

because q is a non-negative random variable. Without knowing the exact value of the expected value $E(q)$, it can be estimated by a sufficient number of samples of the system accuracy q.

However, $E(q)$ can be calculated if we have some knowledge on the distribution of the accuracy values of the algorithms. Accordingly, we have checked the literature and some online challenges, where the accuracy of the participating algorithms are made publicly available [1, 48]. Naturally, some distortions may appear as consequences of the competitive behavior. For example, several authors try to beat state-of-the-art results by over-training, or competitors submit relatively weak solutions, as well. However, if we eliminate these outliers, the remaining accuracy values often have a uniform or normal distributions, which behaviors seem to be reasonable also intuitively. Without the sake of completeness, we can mention the sources [48] or [1] as specific examples. Since this experiment suggested us that the uniform or normally distributions of the accuracies can have practical importance, we investigated these cases more closely. In these special (normally or uniformly distributed) instances, the expected value $E(q)$ of the system accuracy is determined as the distribution of ξ^k is known. For example, if ξ has a standard uniform distribution, then ξ^k has a beta distribution with parameters $\frac{1}{k}$ and 1, that is $E(\xi^k) = \frac{1}{1+k}$. Furthermore, the distribution of a product of nor-

mally (or uniformly) distributed variables is known, that can be given, e.g. by a modified Bessel function or in terms of Meijer G-functions.

The results of the proposed method in Table 18 and Table 19 can be compared, when the distribution of the accuracy values is uniform in EXAMPLE10 and when a normal distribution is given in the next object detection problem in 4.2.4.

Table 18: The efficiency values and the selection probabilities of the items respectively, and the remaining normalized time for EXAMPLE10.

	(p_1,t_1)		(p_2,t_2)		(p_3,t_3)		(p_4,t_4)		(p_5,t_5)		(p_6,t_6)		(p_7,t_7)		T
$q^{(0)}$	0.836	0.133	0.851	0.136	0.874	0.139	0.897	0.144	0.896	0.143	0.94	0.150	**0.972**	**0.155**	0.85
$q^{(1)}$	0.787	0.157	0.8	0.160	0.837	0.167	0.844	0.168	**0.896**	**0.179**	0.850	0.170	-	-	0.6
$q^{(2)}$	0.753	0.193	0.765	0.196	0.784	0.201	0.75	0.192	-	-	**0.85**	**0.218**	-	-	0,421
$q^{(3)}$	0.683	0.245	0.65	0.233	0.7	0.252	**0.75**	**0.270**	-	-	-	-	-	-	0.207
$q^{(4)}$	**0.6**	**1**	0	0	0	0	-	-	-	-	-	-	-	-	0.064
$q^{(5)}$	-	-	0	0	0	0	-	-	-	-	-	-	-	-	0.028

Table 19: The efficiency values and the selection probabilities of the items respectively, and the remaining normalized time for OD detection.

	(p_1,t_1)		(p_2,t_2)		(p_3,t_3)		(p_4,t_4)		(p_5,t_5)		(p_6,t_6)		(p_7,t_7)		T
$q^{(0)}$	0.9994	0.198	**0.9999**	**0.199**	0.169	0.033	0.994	0.197	0.189	0.036	0.976	0.194	0.709	0.141	0.85
$q^{(1)}$	**0.9998**	**0.248**	-	-	0.169	0.042	0.992	0.246	0.189	0.047	0.978	0.242	0.709	0.176	0.771
$q^{(2)}$	-	-	-	-	0.169	0.057	**0.989**	**0.334**	0.189	0.064	0.978	0.329	0.643	0.216	0.745
$q^{(3)}$	-	-	-	-	0.169	0.085	-	-	0.189	0.096	**0.978**	**0.494**	0.643	0.325	0.705
$q^{(4)}$	-	-	-	-	0.304	0.24	-	-	0.319	0.252	-	-	**0.643**	**0.508**	0.372
$q^{(5)}$	-	-	-	-	0	0	-	-	0	0	-	-	-	-	0.115

4.2.4 Experimental results – optic disc detection

As a single object localization problem, we consider the detection of the OD in retinal images to test our approach under real circumstances. The proper detection of the OD is useful from several aspects including vessel segmentation and bright lesion localization tasks, but also the recognition of the presence of glaucoma. We have already shown that the performance of individual OD detector algorithms can be outperformed with organizing them into an ensemble [52]. In this realization, the output of each OD detector is a single pixel indicating its candidate for the center of the OD. The member's decision is considered to be correct if its candidate falls inside the disc-shaped OD region. Majority voting rule can be applied to have the decision of the ensemble by taking the shape constraint of OD also into consideration. Namely, we aggregate those outputs of the members that can be bounded by a circle of a given diameter d_{OD}. The disc containing the maximal number of outputs is accepted as the final decision for the OD location, see also Figure 23. In [5] we have shown how to calculate the $p_{n,k}$ parameters in (104) to determine the accuracy of the ensemble.

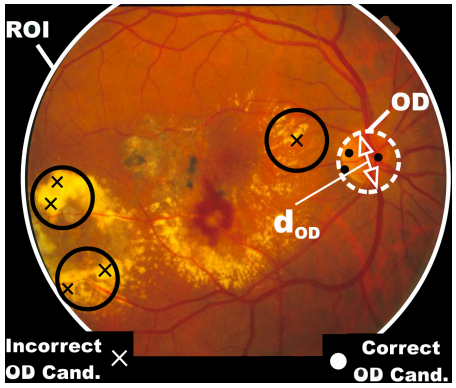


Figure 23: Majority voting rule for OD detection in retinal images.

In this specific scenario, we have seven member OD detectors with the following respective individual accuracies and execution times (in seconds): $p_1 = 0.765, p_2 = 0.958, p_3 = 0.304, p_4 = 0.754, p_5 = 0.319, p_6 = 0.976, p_7 = 0.643$, and $t_1 = 7, t_2 = 21, t_3 = 38, t_4 = 11, t_5 = 34, t_6 = 90, t_7 = 69$. These figures have been calculated offline with no additional time using the publicly available manually annotated dataset MESSIDOR [23]. We have the accuracy of a detector as the portion of the correct decisions in the

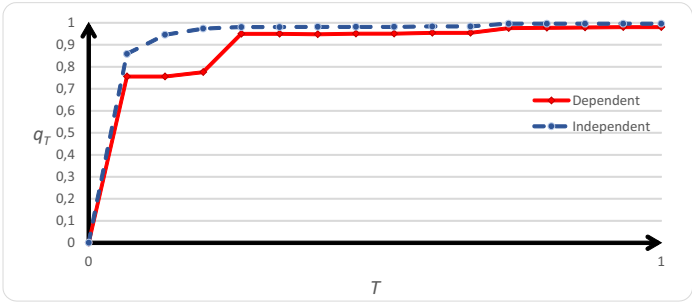


Figure 24: The maximum ensemble accuracies for the real object detection scenario.

whole dataset.

To see how our approach operates on this problem, we have determined the most accurate ensembles regarding all possible bounds on the total execution time $T = 1, \ldots, \sum_{i=1}^{7} t_i$.

The efficiency values, see (109), and the selection probabilities, see (110), are derived from the generalized terms $p_{n,k}$ is shown step by step in Table 19 for the OD detection problem, when $T = 0.85 \sum_{j=1}^{n} t_i$, if the most efficient item is supposed to be selected at each step. The results are shown for seven detectors, but this stochastic method is not dependent on changes in the number of the members.

As we have mentioned earlier, both the target function (104) and our ensemble composition approach suppose the members to be independent. However, in practical applications, the detectors show certain levels of dependencies, which can influence the system accuracy, as well. On the other hand, our stochastic approach can guarantee optimality only a given level of confidence. Thus, the efficiency of our approach cannot be wholly validated by checking its outcome in comparison with a brute-force ensemble composition strategy. However, since the dependency issue is hardly distinguishable from the error of the stochastic approach, this comparison is still a natural one. Accordingly, in Figure 24, besides the maximal ensemble performance found by brute-force, we also plot the maximum ensemble accuracies q_T versus the time limit T found by our approach (supposing independent members).

As a general behavior, it can be nicely observed that in real applications the algorithms with having the same aims are usually dependent in such a way that ruins ensemble performance. Moreover, the ensembles found by our

stochastic approach for the specific time limits indeed supplied the largest system accuracies. However, in general, we cannot guarantee this feature without a complementary exhausting dependency check of the members.

4.2.5 Conclusion

Ensemble-based systems are beneficial to obtain a higher accuracy than that could be gained by any of its members. However, more computation resources are needed for an ensemble, since the execution times of its members cumulate. Thus, it is a meaningful condition to consider a time limit for the total execution time of an ensemble together with trying to keep up accuracy. This limitation leads to an optimization problem of identifying those members which compose the most accurate ensemble together with having their cumulated execution times below the required limit. In this chapter, we solve this combinatorial optimization task by a stochastic approach through formulating knapsack problems with an energy function based on the majority voting rule. We plot the ensemble accuracy against the time limit for several cases, which may help a user to select an appropriate time limit, where the accuracy reasonably increases. Moreover, it also becomes possible to see whether the acceleration of specific components may raise ensemble accuracy. As a real experimental study, we consider a single object detection scenario for the localization of the OD in retinal images. In this way, we also compare our models for independent systems with a real one having dependencies among its members. Though the independent models help to measure up the expected behavior, the extension of the study seems to be reasonable for dependent systems, as well.

In previous sections we have made serious efforts to improve the performance of medical image analysis methods with using ensemble-based systems. In the following section, we present a novel hardware-based solution for the efficient adoption of our complex, fusion-based approaches for real-time applications. Even though most of the image processing problems and the increasing amount of data have high-performance computing (HPC) demand, there is still a lack of corresponding dedicated HPC solutions for several medical tasks. To widen this bottleneck, we have developed a Hybrid Small Size high-performance computing Resource (abbreviated by HuSSaR) which efficiently alloys CPU and GPU technologies, mobile and has an own cooling system to support easy mobility and wide applicability. Besides a proper technical description, we include several practical examples from the clinical data processing domain in this work. For more details see also: `https://arato.inf.unideb.hu/kovacs.laszlo/research_hybridmicrohpc.html`

4.3 Ensemble-based systems in HPC environment

4.3.1 Introduction

Information technology has developed significantly, especially in the last decades. New hardware-based achievements establish never seen capabilities for more precise detection and personalized treatment in healthcare. General-purpose computing on graphics processing units (GPGPU)[27], the Intel Many Integrated Core Architecture (MIC)[55], the field-programmable gate array (FPGA)[33] and more recently the Neural Processing Unit (NPU) architectures are the examples for the hardware related progress. Several papers have been published about performance comparisons in different applications to determine the superiority of the Processing Units [37]. Though comparative studies [26] suggest it, the performance superiority of this platform is not evident in several cases. Namely, though GPUs are capable of running faster than other current PUs for instance deep learning algorithms, it should be noticed that GPUs are specialized vector processors on which convolutional operations are very efficient. PUs can take advantage of using fusion-based approaches with higher computing efficiency instead of considering competitive individual algorithms. Although the area of medical image analysis is full of high-performance computing problems, including machine learning, deep learning, and ensemble-based systems, access to efficient, corresponding HPC systems for the daily routine is limited. Thus, solutions like the proposed Hybrid Small Size high-performance computing Resource (HuSSaR) can be a precious tool.

4.3.2 The HuSSaR machine

Our primary aim was to build a hybrid small size mobile high-performance computer (see Figure 25) which could be placed in an operation room or place on a truck for mobile screening to support different detection or treatment-related computational issues [38]. While the number of processor units (PU) is limited on the mainboards, the co-processor cards are the only options to be built a high PU density machine in a regular size case. When there is no particular algorithm predefined, it is hard to tell in advance which hardware technology of PU should be followed. It is proved that the aggregation of different solutions can solve a complex problem more efficiently compared with any individual approach. HPC hardware components generate a high amount of heat. Thus, they are designed to work in an air-conditioned server room. The maximum working temperature of these PUs is around 90°C. The HPC components are overheated during heavy loaded computations at

Figure 25: The HuSSaR machine

normal room temperature (20-25°C). To achieve our objective, we have constructed a high PU density machine equipped with combined liquid/airflow cooling, which is sufficiently efficient for HPC demands. Moreover, it prevents overheating, performance drop and noise intolerability caused by fans.

4.3.2.1 Built-up

Although the field of hybrid computing is a new domain, there are several attempts to exploit PU architectures (CPU, GPU, MIC, FPGA) simultaneously. To guarantee full usability of the CPUs, HuSSaR has an additional co-processor for each CPU. The number of the co-processors could be determined based on the required bandwidth and connection type between the CPU and the co-processors, as well. We have considered a dual processor motherboard with two different types (GPGPU and MIC) of co-processor cards for each CPU to support hybrid computing on PCI-E sockets. The prototype was designed carefully to establish direct wired connections between the CPUs and their co-processors. This architecture supports the optimal assignment of the specific subtasks to the necessary co-processor. Moreover, all the PUs can work as a separate computing node if needed, but can also form a CPU-GPGPU-MIC set to enhance performance [4, 6, 26].

4.3.2.2 Cooling

The maximum operating temperatures of PUs is about 90°C, which limit is quickly reached if, e.g. a strong GPU card is inserted into a typical PC. In such situations, the driver usually decreases performance. In a HPC environment, this problem is avoided with maintaining constant temperature via air-conditioning, so that the significant part of the heat is conducted by a liquid-based cooling system. Liquid(water)-based cooling is accessible also for regular PCs primarily for optical tunning purposes. A customized, water-cooled PC only optical tunning reasons, combining with air-based ventilation, can reach HPC-level cooling.

During our stress tests under a full load of computations, the temperatures of the CPUs were around 40°C, GPUs 50°C, and Xeon Phi CPUs 60°C in a closed office with room temperature 24°C.

4.3.3 Applications with optimal task scheduling

The idea of hybrid applications is twofold. First, the types of HPC environments based on different architectures increase according to the rapidly developing software and hardware technologies. Thus, attempting optimal usage of the available tools is a natural motivation. Second, complex systems that aggregate different approaches could outperform single solution based ones [6]. However, such fusion-based approaches have large computational needs with including more algorithms or complex workflows. As for modeling this approach, we can consider a set of components to be executed with their corresponding computational needs. If we have limited computational resources and an execution time constraint for execution, the optimal selection of a subset of the components can be solved as a non-linear knapsack problem. For more details on the theoretical results on this topic, see [4]. As an outcome of this step, we have the components that need computational resources to be assigned to them.

To control the access to the computing resources, HuSSaR considers state-of-the-art schedulers to support distributed processing, e.g., for a fusion-based application consisting of many algorithms. Technically, the system has master and computing nodes, where the master receives and schedules the computing task via waiting queues. The free computing node (PU resources) can pick and execute the jobs from the queue. Since there is no shared memory access, this technique is used mainly for running the jobs with different inputs. If there is a need to increase the performance of a single HuSSaR machine, two or four of them can be piled up, where all the member HuSSaRs have from one to four CPUs equipped with the co-processors. The commu-

nication among the HuSSaR is realized via UTP network using scheduling or computing library software solutions. Overall, a HuSSaR configuration of maximum performance includes four motherboards, 16 CPUs, and 32 co-processor cards providing operation speed 350 TFlops.

Machine learning technologies nowadays have a leading role in the field of image processing. However, the application of deep learning-based methods is only one example. As a computational task, deep learning is just similar to other traditional, e.g., detector algorithms or an ensemble of them for a dedicated image processing task. Thus, similarly to distributed deep learning, they can be efficiently executed on the HuSSaR system.

We demonstrate the operation of the hybrid processing unit environments and HuSSaR also on the specific tasks corresponding to the detection of skin lesions, CT segmentation and visualization in medical images [29] including a not own solution of pairwise alignment of very long DNA sequences from the field of bioinformatics.

4.3.3.1 Skin Lesion Detection

4.3.3.1.1 Introduction

Malignant melanoma is a cancerous lesion from the pigment-bearing basal layer of the skin. Melanoma is one of the deadliest forms of skin cancer in the developed countries, though it can be treated by surgical excision if detected early. One of the visual features of melanoma is the presence of streaks. Streaks are linear extensions of a mole/pigmented lesion that are arranged in linear structures in the radial growth directions of the mole at its borders. This clinical definition can be translated to an abstract model, whose parameters can be determined by image processing techniques: 1) Streaks are three or more linear structures co-radially oriented at the border of the mole having thickness 1/3 of the minor axis of the lesion. 2) Streaks are darker than other mole regions in their neighborhood. 3) Streaks are shorter than the 1/3 of the minor axis of the lesion, and they should be longer than 1% of the major axis. 4) Streaks do not branch and have small curvature.

4.3.3.1.2 Streak detector algorithm

As an example to detect the streaks on melanoma images, we can use conventional image processing tools as follows. First, the lesion is segmented, rotated to align its major axis to be horizontal as it represents the growth direction of the mole. To detect dermatoscopic structures, a Laplace of Gaussian operator is applied to the image. A Gabor filter with orientation

and tuned ridge frequency removes noise while preserves correct ridges and valleys. Since streaks appear in more radial directions, Gabor filters are considered in more directions (0°, 45°, 90°, 135°). A proper threshold function is used to get the main structures of the mole with 0 for valleys and 1 for ridges. As a final step, skeletonization is applied to get the streaks with a fast parallel algorithm [16] including a removal step for the too short/long skeleton segments, as well. A sample output can be observed in Figure 26.

Figure 26: Detected streaks in a mole.

4.3.3.1.3 Implementation and compilation

To implement the above method to optimize performance, different parts of the corresponding codes run on different PUs to be computed at the same time. Namely, an image is processed using a CPU including Intel graphics cards (GFX) if exists, a GPU by Nvidia CUDA libraries, and an Intel MIC PU architecture, typically on Xeon Phi processors by Intel Parallel Libraries [34]. As Gaussian blur and minimum/maximum pixel intensity extraction run sufficiently fast on CPU, it is worthless uploading and downloading data to/from co-processor cards. However, a Canny edge detector is fast on GPU. Thus it is worth uploading/downloading. Figure 28 explains which part of the code has implemented on which architecture. The figure also shows how distributed parallelization is applied to each computing node; our prototype has two computing subnodes (two CPU/GPU/MIC sets).

Because of the heterogeneous hardware setup, there are no direct compiler options available. However, it is possible to combine the native compilers of the individual architectures, that is, Intel ICC and Nvidia Cuda.

During compilation, there is a check to discover existing Intel and CUDA compatible architectures. The parameters and compiler commands are assembled in a makefile. The compilation procedure (see Figure 27) for coprocessor architectures is executed before compiling the main functions of the code, which links the precompiled PU kernels directly.

In the next subsection, we will give another example of using hybrid HPC environment from the field of CT segmentation and visualization. The following result is achieved during the Summer of HPC 2015 program as granted one of twenty research assistant student from Europe student applicant. This work was a part of the project titled "Medical image segmentation and visualization" (MEDIVIS) at IT4I Czech National Supercomputing Center at VSB – Technical University of Ostrava, Czech Republic.

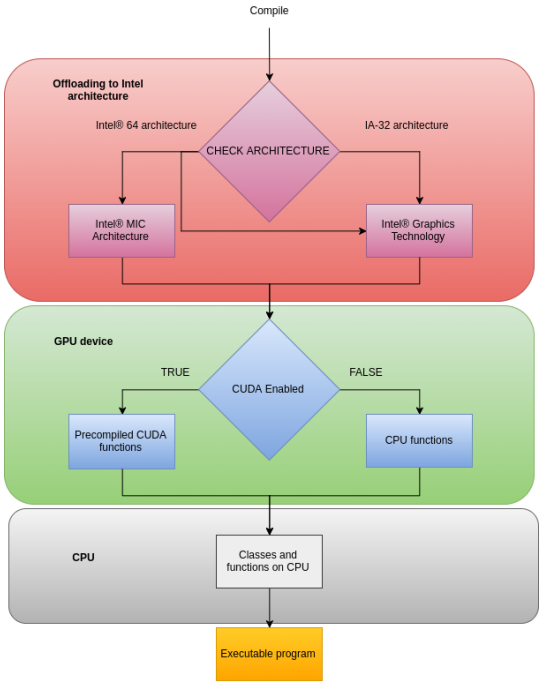


Figure 27: Compiler architecture.

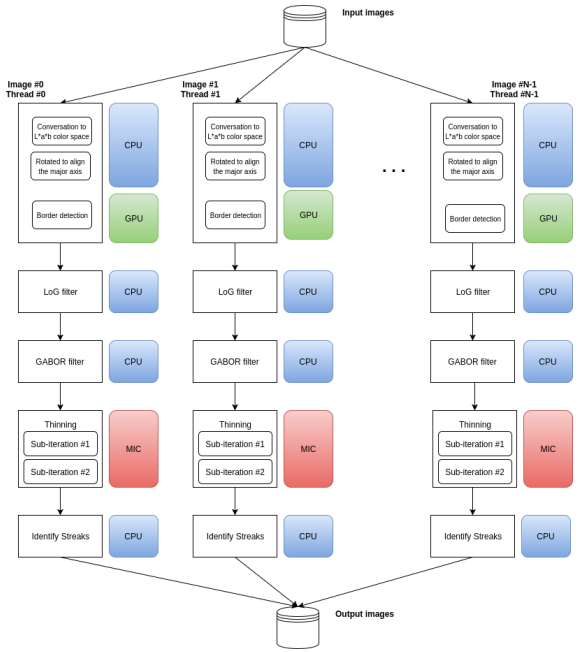


Figure 28: Executional architecture.

4.3.3.2 Liver 3D visualization for CT images

In this subsection, we demonstrate how to segment the human liver using the hybrid HPC environment from the CT images and create a photo-realistic visualization to support more precise and faster medical treatments. We note that our 3D viewer supports the hand-gestures control by the Leap sensor and Virtual Reality by the Oculus glasses in HPC environment.

4.3.3.2.1 Introduction

Computed Tomography (CT) is one of the fields which generates a significant amount of data in medicine. The CT scanning or also so-called X-ray computed tomography (X-ray CT) could be considered as a non-invasive procedure since it allows us to see inside the human body without clinical surgery. The only limitation of the technology is the x-ray radiation dose during the process and its effects on the human body. The risk that comes from the radiation is much lower nowadays thanks to the new types of CT machines. Moreover, a punctual and more easily understandable visualization of the data can prevent from repeating the scanning intervention.

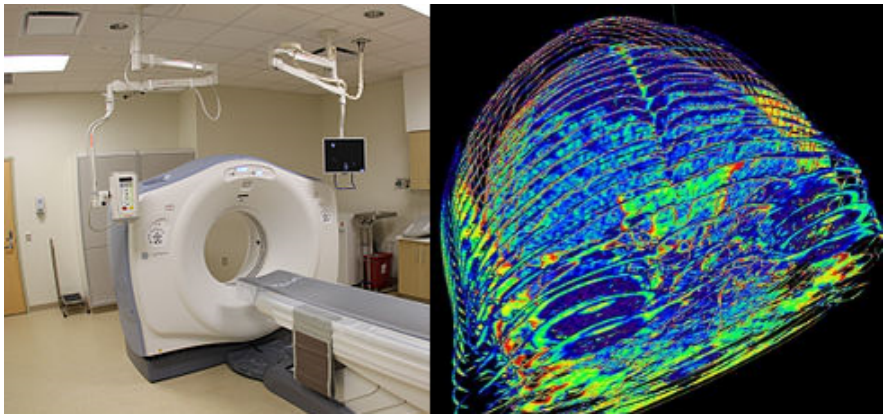


Figure 29: CT machine and the results of the scanning

During the scanning process, the CT machine recreates the inner image of the body in sequential axial slices, see Figure 29. The CT images are stored in a special Digital Imaging and Communications in Medicine (DICOM) format. It is a well-known and widely adopted standard by hospitals and medical device industries for handling, storing, and transmitting all the

necessary information in medical imaging. In the viewpoint of the medical treatments, the precise and realistic image processing and visualization are important since they can obviate the repeated scanning and directly help doctors with the evaluation and decision-making process. A typical application of it is the liver carcinoma, which is frequently a malignant epithelial tumor. Based on reports of World Health Organisation (WHO) the human cancer is a major health problem worldwide. Over a million deaths per year, which is about 10% of all deaths among the adults, can be contributed to hepatocellular carcinoma all over the world. On the Figure 30 it can be seen how the hepatocellular carcinoma looks like.

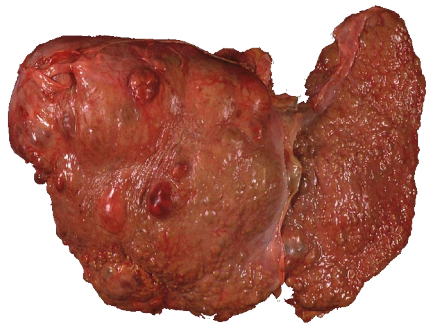


Figure 30: Liver carcioma

Regarding this type of cancer, one of the most important information coming from the segmentation is the exact evaluation of the organ volume. Based on this information exact liver resection is made. Without the information, the surgery can be lethal. For the punctual liver segmentation and visualization, we exploit the HPC technology extended with coprocessors to provide the solution. This process consists of four main steps: preprocessing, segmentation, post-processing, and 3D visualization [15],[50].

4.3.3.2.2 Pre-processing

In the pre-processing step, the first task is to retrieve the data from DICOM standardized medical image data format. In our research, we try to use detailed CT images with mutual axial slices of 0.6 mm to support accurate 3D reconstruction. The next important sub-step is noise reduction by applying the denoising filters in all the acquired images. During the scanning process, random noise appears in the images. The noise is tightly connected with the

physical principle of the scanning. We have to mention that this is just a half true because during scanning process many different kinds of noise could distort the final output. Nonetheless, based on the state-of-the-art studies, the application of our model can suppress the majority of the noise very effectively, Figure 31. The images with a suppressed level of noise support the segmentation algorithms to achieve better results. In the current stage of our research, we focus on the state-of-the-art methods of Gauss smoothing and BM3D.

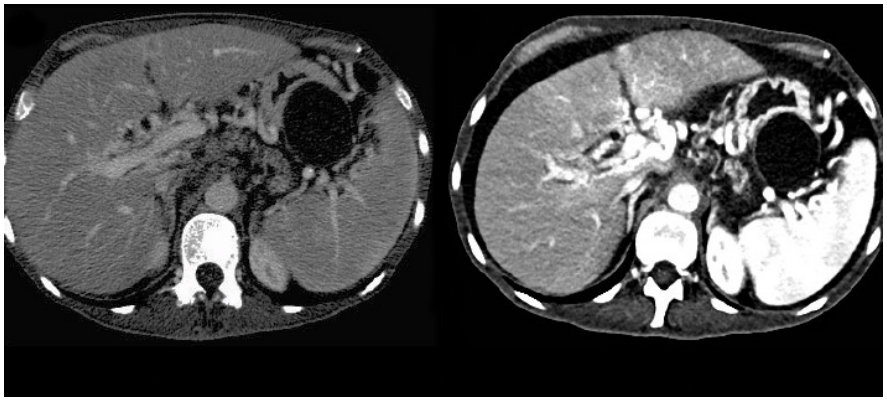


Figure 31: Results of the denoising filter

Before the segmentation step, there is a mapping, which is, in fact, a pixel intensity transformation from CT to Hounsfield Units (HU). During the calculation, we apply the following linear equation, in which the inputs come from only the DICOM format. This mapping is useful since the different organs (in the human body) have different HU values. For instance, the bones have typically +700 HU or higher, the liver has +40 to +60 HU, and the air has -1000 HU. As the last substep, we applied to image data conversion to store all the image slices in one image vector.

4.3.3.2.3 Segmentation

In the case of the segmentation process, we focus on the state-of-the-art k-means method. The generalized algorithm can partition n pixels into k clusters, where k is an integer value and $k < n$. In our case, the k-means algorithm classifies the pixels regarding the similarity features. The pictures

below show the segmentation results in the case of one image slice with and without prefiltering.

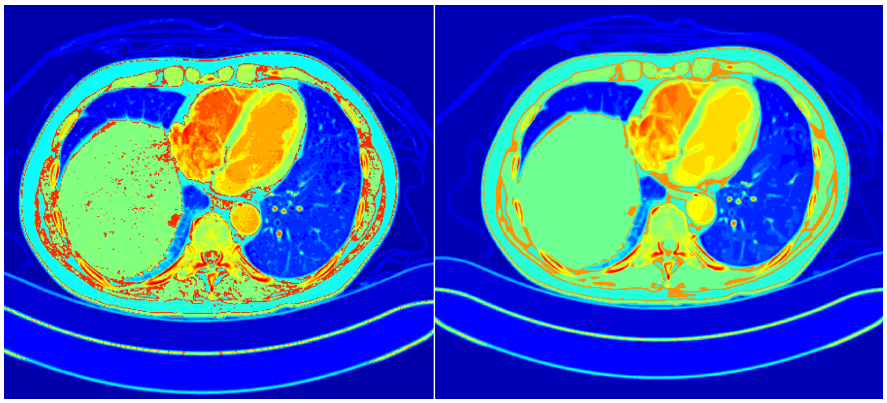


Figure 32: Results of the segmentation without/with denoising

4.3.3.2.4 Post-processing

As a post-processing step, we proceed to the reconstruction of the surface based on the gained individual segments. These boundary segments are calculated with the flooding algorithm. As for the 3D reconstruction algorithm we used the Marching Cubes method. Figure 33 shows the result of the reconstruction.

4.3.3.2.5 3D Visualisation

The final step is the creation of the 3D object and its visualization in a suitable environment based on the segmentation results. With these results we can provide the expected measurements such as volume measurement. We concentrate on the liver. It simplifies the final solution, but on the other hand, it is remains possible to easily extend the solution to other organs in the future. Considering the high performance and portability demands we have choosen the OpenSceneGraph (OSG) as a visualization environment. The OSG is an open source LGPL licensed high-performance 3D visualization toolkit. It is widely used as scientific visualization, virtual reality or modeling environment. OSG is written in C++ and supports the full range of operating systems like Windows, OSX, Linux. The cross-

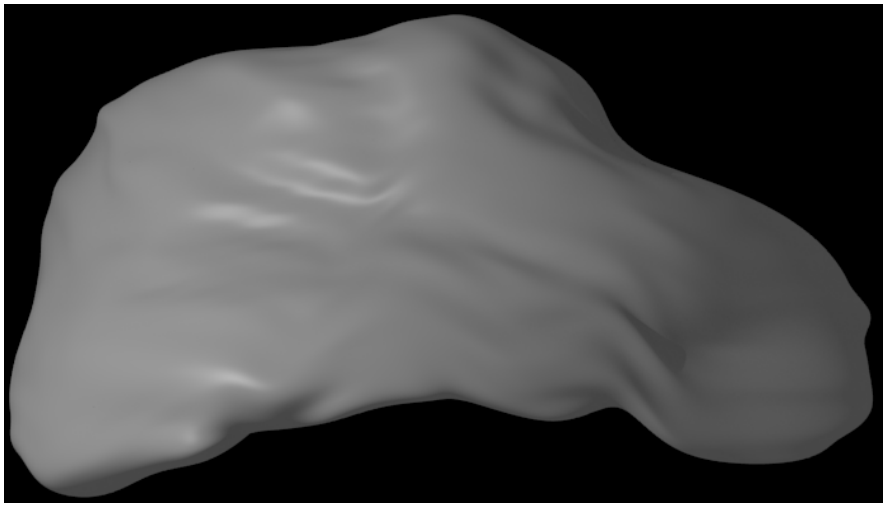


Figure 33: Results of the 3D reconstruction

platform environment is convenient since in this way it is possible to show the segmentation results not only on a computer but also on a smart-phone or tablet. It can be useful not only for medical treatments but also for the education of medical students.

In fact the visualization part, we also focus on the newest real 3D visualization technology and unconventional LEAP Motion controller for interactive control of the visualized scene. The Leap Motion Controller tracks the hands at up to 200 frames per seconds using two infrared cameras. The field of view of the sensor is 150 degree. We integrate the leap sensor into our viewer to replace the usage of the regular mouse. We add the following features into our program: object grabbing with one hand; object rotating with one and two hands; zooming with two hands; pointer function with one and two hands. While the project focuses on real 3D visualization, we also work with OCULUS Virtual Reality Glasses. These glasses are a virtual reality head-mounted display developed by Oculus VR®. The glasses track the head movement and put the viewer in the middle of the scene, while the head mounted display renders the picture to view everything in real 3D from the viewpoint of the person. We successfully integrate the Oculus Rift into our viewer. The program can recognize the connected devices and renders the scene on the glasses monitor in a right form. With the help of the

glasses, moving the head to follow the object movement more easily. We pay considerable attention to create the feeling of the virtual reality as real as possible. We successfully integrate the LEAP sensor and the OCULUS Rift to work together. It means both tools can be used at the same time.

While the project also focuses on the photorealistic post-processing of the acquired data, we chose the Blender project as an open source software to create adequate materials, textures, and scenes. It is a modeling and rendering program that can create photo realistic pictures and animations. The 3D model of the liver is imported into the Blender, and by the help of our texture materials and by using both Blender's renderers a real photo-realistic liver model is created, see Figure 34.

Figure 34: Results of 3D visualization

4.3.3.2.6 Post-processing

This model is exported and used in our virtual reality visualizer. Due to the massive data from the organs, the HPC computers have enough computational power only to execute the computation in time. During computations the Intel Xeon Phi were used to compute the image segmentation and reconstruction of the organsThe texturing and photorealistic visualization were run on NVIDIA Tesla coprocessors.

4.3.3.2.7 Results

Our solution can use the automatic image segmentation of medical images to visualize the liver in a precise and realistic 3D way. The application uses the latest 3D visualization technology involving the photo-realistic post-processing of the acquired data, and it is enabled for the users to control the visualized scene by an unconventional LEAP controller interactively even in the virtual reality environment. To achieve the results, we exploit all types of coprocessors so that the task was optimized for the adequate GPU and CPU architectures.

Finally, we briefly discuss the applied scheduling for HPC environment in our fusion-based solutions. A pairwise alignment of very long DNA sequences from the field of bioinformatics as another application of hybrid coprocessor HPC environment is also discussed as proof of relevance.

4.3.3.3 Task scheduling for an ensemble-based system including HPC and analysis of DNA sequences

The HPC environments including hardware and software have increased attention year after year. The exploitation of hybrid environments come to mind of other researchers at the same time, which increases the relevance of existing HuSSaR type hardware environments. During the construction of HuSSaR, another study has been released [28] about a hybrid solution of optimal pairwise alignment of very long DNA sequences, and protein sequence database search (SWhybrid). This work has been shared on Github, and the hybrid implementation was successfully tested on HuSSaR without any problems with good running performance.

The task scheduling is an essential part of the solution for ensemble-based application and exploitation of HPC hardware. There is more software scheduler developed over the past decade. The best-known maybe are Grid Engine and Slurm. The responsibility of a scheduler in HPC is to control the access to the computation resources. It is a core issue what percentage of the complete system are accessible for a process. In most cases the HPCs are complex resources so distributed parallel processing technology is applied. In this context the master node of the scheduler controls sharing the resources between the tasks, while the task are running on the computing nodes. Namely, the master node receives and stores in queues the jobs with demands to run, and after proper scheduling, the jobs run on computing nodes. During our experimental results we use Grid Engine and Slurm scheduler involving our own setup and maintain grid (approx. 800 CPU core with 800GB memory), our own build HuSSaR HPC, and NIFI

managed HPCs. An ensemble-based solution fits perfectly the distributed computing while the task separation could be self-evident. Namely, one of the convenient options is to run components/classifiers parallelly but separately on computing nodes. It is a good idea when the input of each task is the same. When there is a sequential demand for the execution because of the next input generated from another output, the image based parallelism are optimal. Of course, the mixture of these is also feasible. During our experiments, the image based, that is the second version was applied mostly. The reason is simple: first of all, the data migration was minimal among master computing nodes (one picture was sent one time for all tasks), moreover in most cases ensemble-based solutions, like ours, are complex enough to be necessary to create an own inside control algorithm to execute all the component guaranteeing the final answer of the ensemble. Our created ensembles (embedded into our DRSCREEN system [2]) including our grid and own hardware-based solutions competed in an international online challenge also: last time on ISBI 2018 Diabetic Retinopathy – Segmentation and Grading Challenge passing out with good marks.

4.3.4 Conclusion

The ensemble-based system has high performance computing demand. In the last section, we have presented a Hybrid Small Size high-performance computing Resource (HuSSaR) dedicated mainly for digital image analysis. Weexplained how we used our grid and HuSSaR HPC, and how we optimized our computations to fit HPC environment. The optimal exploitation of the system was demonstrated on a skin lesion classification, CT segmentation and 3D visualization including virtual reality tasks and the third party application so-called SWhybrid.For all of the applications, a task scheduler (grid engine/slurm) manages task assignment to hardware. The main goal is to maximize the utilization of the hardware and reduce computing time. So the hardware acceleration is driven by optimal task scheduling based on hardware requirement of the concrete tasks. Our distributed parallelized complex fusion based retinopathy screening system included the described model above also competed on ISBI 2018 Diabetic Retinopathy – Segmentation and Grading Challenge and passed out with good marks.

Summary

In this section we summarize the contents of the dissertation in thesis points.

Thesis 1. I developed a combined object detection method based on anatomical features, using a generalized majority voting model. I constructed the weight function for its application to detect optic disc and macula. By the help of experiments using public databases, I showed that the method outperforms both the individual algorithms, and their combination using the classical majority voting.
Corresponding publications: [1–9]

Thesis 2. I developed a method to characterize and measure the improvability of the accuracy of a generalized ensemble-based object detection system. I adopted the theoretical results for a medical application.

Thesis 2.1. Based upon the theoretical results, by the help of experiments using a public image database, I showed that a specific set of optic disc classifiers is dependent. I described the accuracy of every subset of the elements of the ensemble.

Thesis2.2. Based on the theoretical results, by experiments on publicly available image databases, I obtained the minimal and maximal values of the accuracy of a generalized ensemble-based optic disc detection method in the following cases:

- a new independent algorithm is added to a dependent system;
- a new independent algorithm is added to an independent system;
- a new dependent algorithm is added to a dependent system.

Corresponding publications: [1, 3–7]

Thesis 3. I generalized the standard two-dimensional diversity measures of the classical majority voting to the higher dimensional case. I showed experimentally that the generalized diversity measures outperform the original ones, i.e., they provide better estimates for the ensemble accuracy.
Corresponding publications: [10]

Thesis 4. I generalized the higher dimensional diversity measures of the classical majority voting to be applicable in the case of spatial constraint voting. By the help of experiments, I justified that the generalized diversity measures outperform the original ones, i.e. they give better estimates for the ensemble accuracy.
Corresponding publications: [10]

Thesis 5. I described a method to compose an ensemble with maximal performance, based on the generalized diversity measures. Using the theoretical results, I constructed the best performing combination of a given set of classifiers in a concrete application concerning the detection of the optic disc.
Corresponding publications: [10]

Thesis 6. I introduced a method for the optimal composition of ensembles under the constraint that the total computational time is limited. For this, I generalized the knapsack problem using a majority voting energy function and solved the combinatorial optimization task by a stochastic approach. Applying the theoretical results, I developed an algorithm to compose an ensemble with maximal performance regarding execution time constraint in the optic disc medical application.
Corresponding publications: [11]

Thesis 7. I provided a hybrid hardware-based solution for optimal task scheduling of complex algorithms and composed ensembles in HPC environment. I introduced a method to find the optimal scheduling with respect to the computational requirements of the individual algorithms, in various applications.
Corresponding publications: [12–18]

Összefoglás (summary in Hungarian)

Ebben a fejezetben, tézispontokba szedve összegezzük a disszertációban bemutatott eredményeket:

1. Tézis. Kifejlesztettem egy többségi szavazás elvén működő, anatómiai képletek tulajdonságait is figyelembe vevő több objektum együttes detektálását megvalósító eljárást. Kidolgoztam az eljárás végső jelöltpár (sárgafolt - vakfolt) súlyfüggvényét. Publikus adatbázisokon történő kísérletek segítségével megmutattam, hogy a kidolgozott módszer nemcsak az egyéni algoritmusoknál, hanem a klasszikus többségi szavazást használó összetett rendszernél is hatékonyabb.
Kapcsolódó publikációk: [1–9]

2. Tézis. Módszert adtam egy objektumdetektáló általánosított összetett rendszer kibővítésekor történő teljesítményváltozás mérésére és a teljesítmény javítására. Az elméleti eredményeket sikerrel alkalmaztam egy valós orvosi képfeldolgozási problémában is.

2.1. Tézis. Az elméleti eredmények alapján, egy publikus adatbázison végrehajtott kísérletek segítségével igazoltam egy konkrét osztályozórendszer függőségét. Sikerült a rendszer valamennyi részrendszerének teljesítményét is meghatároznom.

2.2. Tézis. Az elméleti eredmények alapján, publikus adatbázison végrehajtott kísérletek segítségével meghatároztam az általánosított vakfolt detektáló összetett rendszer teljesítményének minimális és maximális értékét a következő esetekben:

- független algoritmussal bővítünk egy függő rendszert;
- független algoritmussal bővítünk egy független rendszert;
- függő algoritmussal bővítünk egy függő rendszert.

Kapcsolódó publikációk: [1, 3–7]

3. Tézis. Kiterjesztettem a klasszikus többségi szavazó modellek esetén használatos kétváltozós diverzitás mérőszámokat magasabb dimenziós esetre. Kísérletek segítségével igazoltam, hogy a kiterjesztett mérőszámok teljesítményben felülmúlják az eredeti mérőszámokat, azaz pontosabb becslést adnak az összetett rendszer pontosságára.
Kapcsolódó publikációk: [10]

4. Tézis. Kiterjesztettem a magasabb dimenziós diverzitás mérőszámokat térbeli többségi szavazó modellekre. Kísérletek segítségével igazoltam, hogy a kiterjesztett mérőszámok térbeli szavazás esetén teljesítményben felülmúlják

az eredeti mérőszámokat, azaz pontosabb becslést adnak az összetett rendszer pontosságára.
Kapcsolódó publikációk: [10]

5. Tézis. Kidolgoztam egy, az általánosított mérőszámokon alapuló módszert maximális teljesítményű összetett rendszer összeállítására. Az elméleti eredményeket valós orvosi alkalmazásba átültetve, egy konkrét algoritmusrendszerből kiválasztottam a vakfolt meghatározására szolgáló maximális teljesítményű algoritmuskombinációt.
Kapcsolódó publikációk: [10]

6. Tézis. Módszert adtam a maximális teljesítményű részrendszer meghatározására a számítási erőforrások korlátozása mellett. Ehhez a többségi szavazó energiafüggvény segítségével általánosítottam a hátizsák problémát, és arra egy sztochasztikus megközelítésű kombinatorikus optimalizációs megoldást adtam. Az elméleti eredmények felhasználásával algoritmust fejlesztettem egy korlátozott futási idejű, maximális teljesítményű, vakfolt megkeresésére szolgáló összetett rendszer konstrukciójára.
Kapcsolódó publikációk: [11]

7. Tézis. Hardver alapú megoldást adtam a komplex algoritmusok és összetett rendszerek optimális ütemezésére hibrid szuperszámítógépes környezetben. Több alkalmazás esetén megmutattam, hogyan valósítható meg az egyes algoritmusok számítási igénye szerinti optimális ütemezés.
Kapcsolódó publikációk: [12–18]

Publications of László Kovács

[1] A. Hajdu, L. Hajdu, A. Jonás, L. Kovács, and H. Tomán: Generalizing the majority voting scheme to spatially constrained voting, IEEE Transactions on Image Processing, Volume: PP, Issue: 99, DOI: 10.1109/TIP.2013.2271116, IF: 3.042, 2013, Q1

[2] R. J. Qureshi, L. Kovács, B. Harangi, B. Nagy, T. Pető, A. Hajdu: Combining algorithms for automatic detection of optic disc and macula in fundus images, Computer Vision and Image Understanding (CVIU), 116, 138–145, IF=2.404, 2012, Q1

[3] H. Tomán, L. Kovács, A. Jonás, L. Hajdu, A. Hajdu: Generalized weighted majority voting with an application to algorithms having spatial output, 7th International Conference on Hybrid Artificial Intelligence Systems (HAIS), Salamanca, Spain, 2012, Volume Part II, Lecture Notes in Computer Science 7209: pp. 56-67, 2012, Q2

[4] H. Tomán, L. Kovács, A. Jonás, L. Hajdu, A. Hajdu: A generalization of majority voting scheme for medical image detectors, 6th International Conference on Hybrid Artificial Intelligence Systems (HAIS), Wroclaw, Poland, in Lecture Notes in Artificial Intelligence 6679/2: pp.189-196., 2011, Q2

[5] L. Kovács, B. Harangi, B. Nagy, R.J. Qureshi, A. Hajdu: Gráf alapú vakfolt és sárgafolt detektálás retina felvételeken, 8th Conference of the Hungarian Association for Image Processing and Pattern Recognition, (KÉPAF) Szegedi Egyetem, Szeged, 329-341, 2011.

[6] Rashid Jalal Qureshi, Laszlo Kovacs, Brigitta Nagy, Balazs Harangi, Andras Hajdu: Automatic detection of the fovea and optic disk in digital retinal images by combining algorithms, 8th International Conference on Applied Informatics (ICAI), Eger, Hungary, January 27–30, Vol. 1. pp. 175–184., 2010

[7] L. Kovács, R. J. Qureshi, B. Nagy, B. Harangi, A. Hajdu: Graph based detection of optic disc and fovea in retinal images, 4th IEEE International Workshop on Soft Computing Applications (SOFA), Arad, Romania, 143-148., 2010

[8] L. Kovács, B. Nagy: Látógödör automatikus detektálása digitális retina képeken algoritmusok kombinálásával, XXX. Jubileumi OTDK Informatika Tudományi Szekció, Budapest, Magyarország, 87., 2011

[9] R. J. Qureshi, T. Pető, L. Kovács, B. Nagy, B. Harangi, A. Hajdu: Detection of the optic disc and the macula through combining algorithms, University of Debrecen, Faculty of Informatics, Preprints No. 380 (Technical Reports No. 7/2010.), 2010

[10] A. Hajdu, L. Hajdu, L. Kovács, H. Tomán: Diversity measures for majority voting in the spatial domain, 8th International Conference on Hybrid Artificial Intelligence Systems (HAIS), Salamanca, Spain, Lecture Notes in Computer Science 8073: pp. 314-323, 2013, Q2

[11] A. Hajdu, H. Tomán, L. Kovács, L. Hajdu: Composing ensembles of object detectors under execution time constraint, 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, pp. 222-227., 2016

[12] L. Kovács, A. Hajdu, B. Harangi, A. Sós: HuSSaR – Hybrid small Size HPC resource – HuSSaR [mintaoltalom], Hungarian Intellectual Property Office, role: inventor, under review, 2018

[13] L. Kovács, R. Kovács, A. Hajdu: High Performance Computing in Medical Image Analysis – HuSSaR (arXiv: 1806.06171), 2018

[14] B. Antal, MKGS Tavares, L. Kovács, B. Harangi, T. Pető, A. Csutak, A. Hajdu: Data Analysis Applied to Diabetic Retinopathy Screening: Performance Evaluation, European Journal of Ophthalmology 27:(3)124-125. 27th Meeting of the European Association for the Study of Diabetes Eye Complications Study Group (EASDec), IF: 1.897, 2017, Q2

[15] L. Kovács, S. Petr, L. Riha: HPC GPU/CPU társkártyákkal támogatott fotorealisztikus 3D CT vizualizáció, VIII. Magyar Számítógépes Grafika és Geometria Konferencia / szerk. Szirmay-Kalos László, Renner Gábor, Neumann János Számítógép-tudományi Társaság, Budapest, 175-180, 2016.

[16] L. Kovács: 3D Visualisation from CT, Summer of High Performance Copmuting program (SoHPC) Medical image segmentation and visualization Project, BSC – Barcelona, IT4I Natinal Supercomputing Center – Ostrava, PRACE, pp. 43-45, (https://summerofhpc.prace-ri.eu/final-reports-2015/), 2015

[17] L. Kovács: Hybrid Small Size hpcResource, 7. GPU Day 2017 - The Future Of Many-Core Computing In Science, Budapest, Hungary, 2017

[18] L. Kovács: Photorealistic 3D CT visualization supported by HPC GPU and CPU coprocessors, GPU Day 2016 – 6. The Future Of Many-Core Computing In Science, Budapest, Hungary, 2016

[19] Additional optics equipped internal flashlight illuminator to mobile devices, role: inventor, under review (2016)

[20] Endoscope RGB LED light source with variable wavelength. University of Debrecen Technology Transfer Center, role: inventor, under review (2015)

[21] L. Kovács, Á. Pintér: Stirling számok vizsgálata HPC környezetben, Tavaszi Szél - Spring Wind III. Konferenciakötet / szerk. Keresztes Gábor, Líceum Kiadó ; Budapest : Doktoranduszok Országos Szövetsége, Eger, 181-187, 2015.

[22] L. Kovács, Á. Pintér: Stirling számok vizsgálata HPC környezetben, Tavaszi Szél - Spring Wind 2015 : Absztraktkötet / szerk. Keresztes Gábor, Publio Kiadó, Budapest, 302-303, 2015

[23] J. Tóth, L. Kovács, B. Harangi, Cs. Kiss, A. Mohacsi, Z. Orosz, A. Hajdu: An Online Benchmark System for Image Processing Algorithms, 5th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2014), Vietri sul Mare, Italy, 2014, 5th IEEE International Conference on Cognitive Infocommunications, pp. 377-382., 2014

[24] J. Tóth, L. Kovács, B. Harangi, Cs. Kiss, A. Mohácsi, Z. Orosz, A. Hajdu: An Online Benchmark System for Image Processing Algorithms, 5th IEEE International Conference on Cognitive Infocommunications / ed. Baranyi Péter, CogInfoCom, Vietri sul Mare, Italy, 383, 2014

[25] L. Kovács, L. Hajdu: High Performance Computing With Mathematical Applications, WSPS 2013.: 1st Winter School of PhD Students in Informatics and Mathematics, Tradeorg Nyomdaipari és Kereskedelmi Kft, Balatonalmádi, 13, 2013

Bibliography

[1] "Netflix Prize," http://www.netflixprize.com/, 2009.

[2] 3TECH08-2 grant of the Hungarian National Office for Research and Technology (NKTH), DRSCREEN - Developing a computer-based image processing system for diabetic retinopathy screening.

[3] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P.F. Sharp: Automated Assessment of Diabetic Retinal Image Quality Based on Clarity and Field Definition, Investigative Ophthalmology and Visual Science. 47(2006), 1120-1125.

[4] A. Hajdu, H. Toman, L. Kovacs, L. Hajdu: Composing ensembles of object detectors under execution time constraint, 23rd International Conference on Pattern Recognition (ICPR), pp 217-222, Cancun, Mexico, 2016

[5] A. Hajdu, L. Hajdu, A. Jonas, L. Kovacs, and H. Toman: Generalizing the majority voting scheme to spatially constrained voting, IEEE Transactions on Image Processing, vol. 22, no. 11, pp. 4182–4194, Nov 2013.

[6] A. Hajdu, L. Hajdu, A. Jonas, L. Kovacs, and H. Toman: Generalizing the majority voting scheme to spatially constrained voting, IEEE Transactions on Image Processing, Volume: PP 4182-94, Issue: 99, DOI: 10.1109/TIP.2013.2271116, 2013.

[7] A. Hajdu, L. Hajdu, L. Kovacs, and H. Toman: Diversity measures for majority voting in the spatial domain, in Hybrid Artificial Intelligent Systems, ser. Lecture Notes in Computer Science, J. Pan, M. Polycarpou, M. Woźniak, A. de Carvalho, H. Quintián, and E. Corchado, Eds. Springer Berlin Heidelberg, 2013, vol. 8073, pp. 314–323.

[8] A. Hajdu: Discrete geometric and fusion based techniques for object detection and decision support, Dissertation Dissertation for the Doctoral Degree of the Hungarian Academy of Sciences, 2015

[9] A. Hoover and M. Goldbaum: Locating the Optic Nerve in aRetinal Image using the Fuzzy Convergence of the Blood Vessels, IEEE Trans. Med. Imag., Vol. 22, pp 951-958, 2003.

[10] A. Sopharak, K. Thet New, Y. Aye Moe, M. N. Dailey, B. Uyyanonvara: Automatic Exudate Detection with a Naive Bayes Classifier, International Conference on Embedded Systems and Intelligent Technology, Bangkok, Thailand, 2008, pp. 139–142.

[11] A.E. Mahfouz, A.S. Fahmy: Ultrafast Localization of the Optic Disc Using Dimensionality Reduction of the Search Space. Med. Image Comput. Assist. Interv. 12, 985–992 (2009)

[12] 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)

[13] B. Harangi, R. J. Qureshi, A. Csutak, T. Peto, A. Hajdu: Automatic Detection Of The Optic Disc Using Majority Voting In A Collection Of Optic Disc Detectors, IEEE International Symposium on Biomedical Imaging (ISBI 2010), Rotterdam, The Netherlands, 2010, pp. 1329-1332

[14] B. Harangi: Single and Multiple Object Detection Problems in Medical Image Analyses, PhD Thesis, PhD School of Informatics, Faculty of Informatics, University of Debrecen, 2015

[15] Block-matching and 3D filtering (BM3D) algorithm and its extensions. (http://www.cs.tut.fi/ foi/GCF-BM3D/), 2015.

[16] C. Y. Suen, Wen-Hsing Hsu: A modified fast parallel algorithm for thinning digital patterns, Pattern Recognition Letters,Volume 7, Issue 2, February 1988, Pages 99-106

[17] D. Ruta, B. Gabrys: Analysis of the correlation between majority voting error and the diversity measures in multiple classifier systems. In Proceedings of the 4nd International Symposium on Soft Computing, number ISBN: 3-906454-27-4, Paper No. 1824-025, Paisley, UK, 2001.

[18] D. Ruta, B. Gabrys: Classifier Selection for Majority Voting. Information Fusion 6, 63–81 (2005)

[19] Dataset MESSIDOR provided by the Messidor program partners [Online]. Available: http://messidor.crihan.fr.

[20] Diabetic Retinopathy Detection (2016) (October 18. Retrieved from https://www.kaggle.com/c/diabetic-retinopathy-detection)

[21] Dietterich T. G.: Ensemble Methods in Machine Learning, Multiple Classifier Systems, LNCS. 1857(2008), 1-15.

[22] E. Welzl: Smallest Enclosing Disks (Balls and Ellipsoids), New Results and New Trends in Computer Science, LNCS. 555(1991), 359–37.

[23] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein: Feedback on a publicly distributed database: the messidor database, Image Analysis & Stereology, vol. 33, no. 3, pp. 231–234, 2014.

[24] E. Kong and T. Dietterich: Error-correcting output coding corrects bias and variance, in International Conference on Machine Learning, 1995, pp. 313–321.

[25] F. Zana, I. Meunier , J. C. Klein: A region merging algorithm using mathematical morphology: application to macula detection, Proceedings of the fourth international symposium on Mathematical morphology and its applications to image and signal processing, Amsterdam, The Netherlands, 1998, pp. 423 - 430.

[26] G. Teodoro, T.Kurc, G. Andrade, J. Kong, R. Ferreira, J. Saltz: Application Performance Analysis and Efficient Execution on Systems with multi-core CPUs, GPUs and MICs: A Case Study with Microscopy Image Analysis, J High Perform Comput Appl. 2017 32–51.

[27] H. Gamaarachchi, M. Fawsan, F. Fasna, D. Elkaduwe: User-friendly interface for GPGPU programming IEEE 6th National Conference on Technology and Management (NCTM) 2017, Malabe, Sri Lanka

[28] H. Lan, W. Liu, Y. Liu, B. Schmidt, SWhybrid: A Hybrid-Parallel Framework for Large-Scale Protein Sequence Database Search, IEEE International Parallel and Distributed Processing Symposium, 2017

[29] H. Soyer, G. Argenziano, et al. Dermoscopy of pigmented skin lesions, An atlas based on the consensus net meeting on dermoscopy Milan Edra, 2001.

[30] H. Toman, L. Kovacs, A. Jonas, L. Hajdu, A. Hajdu: A Generalization of Majority Voting Scheme for Medical Image Detectors, Lecture Notes in Artificial Intelligence 6679, Vol. 2, 189–196 (2011)

[31] H. Toman, L. Kovacs, A. Jonas, L. Hajdu, A. Hajdu: Generalized Weighted Majority Voting with an Application to Algorithms Having Spatial Output, Lecture Notes in Artificial Intelligence, Vol. 2, 55–67 (2012)

[32] H. Toman: Geometric Investigation in Loop Theory and in Image Processing, PhD Thesis, PhD School of Informatics, Faculty of Informatics, University of Debrecen, 2014

[33] J. Canilho, M. Véstias, H. Neto: Multi-core for K-means clustering on FPGA, IEEE 26th International Conference on Field Programmable Logic and Applications (FPL), 2016, Lausanne, Switzerland

[34] J. Dokulil, S. Benkner: Automatic Tuning of a Parallel Pattern Library for Heterogeneous Systems with Intel Xeon Phi, 2014 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), Milan, Italy

[35] J. Toth, L. Szakacs, and A. Hajdu: Finding the optimal parameter setting for an ensemble-based lesion detector, in IEEE International Conference on Image Processing (ICIP), Oct 2014, pp. 32 532–3536.

[36] K. M. Bretthauer and B. Shetty: The nonlinear knapsack problem – algorithms and applications, European Journal of Operational Research, vol. 138, no. 3, pp. 459 – 472, 2002.

[37] L. Fiack, L. Rodriguez, B. Miramond: Hardware design of a neural processing unit for bio-inspired computing, 2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS), 2015, Grenoble, France

[38] L. Kovacs: 3D Visualisation from CT, Summer of High Performance Copmuting program (SoHPC) Medical image segmentation and visualization Project, BSC – Barcelona, IT4I Natinal Supercomputing Center – Ostrava, PRACE (2015) (https://summerofhpc.prace-ri.eu/final-reports-2015/)

[39] L. Lam and C.Y. Suen: Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance, IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 27, no. 5, pp. 553-568, Sep. 1997, doi:10.1109/3468.618255.

[40] L. Hansen and P. Salamon: Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993–1001, Oct 1990.

[41] L. Kuncheva: Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience, 2004.

[42] L. Xu, A. Krzyzak, and C. Suen: Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems, Man and Cybernetics, vol. 22, no. 3, pp. 418–435, May 1992.

[43] M. Lalonde, M. Beaulieu and L. Gagnon: Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching, IEEE Transactions on Medical Imaging, vol.20, no.11, pp.1193-1200, Nov. 2001, doi: 10.1109/42.963823.

[44] M. Lalonde, M. Beaulieu, and L. Gagnon: Fast and Robust Optic Disk Detection using Pyramidal Decomposition and Hausdorff-Based Template Matching, IEEE Trans. Medical Imaging. 20(2001), 1193-1200.

[45] M. Niemeijer, J.J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff: Comparative study of retinal vessel segmentation methods on a new publicly available database, in: SPIE Medical Imaging. 5370(2004), 648-656.

[46] M. Niemeijer, M. D. Abramoff, B. van Ginneken: Fast detection of the optic disc and fovea in color fundus photographs, Medical Image Analysis. 13(2009), 859–870.

[47] M.J. Appel, C.A. Najim and R.P. Russo: Limit Laws for the Diameter of a Random Point Set, Advances in Applied Probability, vol. 34, no. 1, pp. 1-10, Mar. 2002, doi:10.1239/aap/1019160946.

[48] Maria-Elena Nilsback and Andrew Zisserman: Automated flower classification over a large number of classes, 08. Sixth Indian Conference on Computer Vision, Graphics & Image Processing Conference proceedings, pp. 722–729, 2008.

[49] N. Otsu: A Threshold Selection Method from Gray-Level Histograms, IEEE Trans. Syst. Man and Cybern. 1979, pp. 62-66.

[50] P. Strakos, M. Jaros, T. Karasek, L. Riha, M. Jarosova, T. Kozubek, P. Vavra and T. Jonszta: Parallelization of the Image Segmentation Algorithm for Intel Xeon Phi with Application in Medical Imaging, Proceedings Of The Fourth International Conference On Parallel, Distributed, Grid And Cloud Computing For Engineering, doi:10.4203/ccp.107.7, 2015.

[51] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 4th Edition, Pearson, 2018

[52] R. Qureshi, L. Kovacs, B. Harangi, B. Nagy, T. Peto, and A. Hajdu: Combining algorithms for automatic detection of optic disc and macula in fundus images, Computer Vision and Image Understanding, vol. 116, no. 1, pp. 138–145, 2012.

[53] S. Ravishankar, A. Jain, A. Mittal: Automated Feature Extraction for Early Detection of Diabetic Retinopathy in Fundus Images, IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 210-217.

[54] S. Sekhar, W. Al-Nuaimy, A. K. Nandi: Automated localization of optic disc and fovea in retinal fundus images, 16th European Signal Processing Conference, Lausanne, Switzerland, 2008.

[55] S. Xu, Z. Wu, Y. Hong, Q Xue, S Liao, B Liu: Optimization of High Performance Computing Cluster based on Intel MIC, 2nd IEEE International Conference on Computer and Communications (ICCC), 2016, Chengdu, China

[56] S. Cho and J. Kim: Combining multiple neural networks by fuzzy integral for robust classification, IEEE Transactions on Systems, Man and Cybernetics, vol. 25, no. 2, pp. 380–384, Feb 1995.

[57] S. Martello and P. Toth: Knapsack Problems: Algorithms and Computer Implementations, New York, NY, USA: John Wiley & Sons, Inc., 1990.

[58] T. Petsatodis, A. Diamantis, G.P. Syrcos: A Complete Algorithm for Automatic Human Recognition based on Retina Vascular Network Characteristics, 1st International Scientific Conference e RA, Tripolis, Greece, 2004, pp. 41- 46.

[59] T. C. Sharkey, H. E. Romeijn, and J. Geunes: A class of nonlinear nonseparable continuous knapsack and multiple-choice knapsack problems, Mathematical Programming, vol. 126, no. 1, pp. 69–96, 2009.

[60] T. Ho, J. Hull, and S. Srihari: Decision combination in multiple classifier systems, IEEE Transactions on Pattern Analysis and MachineIntelligence, vol. 16, no. 1, pp. 66–75, Jan 1994.

[61] T. Ibaraki and N. Katoh: Resource Allocation Problems: Algorithmic Approaches, Cambridge, MA, USA: MIT Press, 1988.

[62] W. G. Cochran: Problems arising in the analysis of a series of similar experiments, J. of the Royal Statistical Society. 4(1937), 102–118.

[63] X Zhu, RM Rangayyan, AL Ells: Detection of the optic nerve head in fundus images of the retina using the hough transform for circles. J. Digit. Imag. 23 (3), 332–341 (2010)

[64] Y. Huang and C. Suen: A method of combining multiple experts for the recognition of unconstrained handwritten numerals, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 1, pp. 90–94, Jan 1995.

UNIVERSITY AND NATIONAL LIBRAY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/268/2018.PL
Subject: PhD Publikációs Lista

Candidate: László Kovács
Neptun ID: EV3E3L
Doctoral School: Doctoral School of Informatics
MTMT ID: 10036360

List of publications related to the dissertation

Hungarian book chapters (1)

1. **Kovács, L.**, Pintér, Á.: Stirling számok vizsgálata HPC környezetben.
   In: Tavaszi Szél - Spring Wind III. Konferenciakötet. Szerk.: Keresztes Gábor, Líceum Kiadó ; Budapest : Doktoranduszok Országos Szövetsége, Eger, 181-187, 2015. ISBN: 9786155509964

Foreign language international book chapters (6)

2. Hajdu, A., Tomán, H., **Kovács, L.**, Hajdu, L.: Composing ensembles by stochastic approach under execution time constraint.
   In: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR 2016), Cancun, Mexico, 2016, IEEE, [Piscataway], 222-227, 2016.

3. Tóth, J., **Kovács, L.**, Harangi, B., Kiss, C., Mohácsi, A. I., Orosz, Z., Hajdu, A.: An Online Benchmark System for Image Processing Algorithms.
   In: 5th IEEE International Conference on Cognitive Infocommunications. Ed.: Baranyi Péter, CogInfoCom, Vietri sul Mare, Italy, 377-385, 2014. ISBN: 9781479972807

4. Hajdu, A., Hajdu, L., **Kovács, L.**, Tomán, H.: Diversity measures for majority voting in the spatial domain.
   In: Hybrid Artificial Intelligent Systems : 8th International Conference, HAIS 2013, Salamanca, Spain, September 11-13, 2013. Proceedings, Springer-Verlag, Berlin Heidelberg, 314-323, 2013, (Lecture Notes in Artificial Intelligence, ISSN 0302-9743 ; 8073) ISBN: 9783642408458

5. Tomán, H., **Kovács, L.**, Jónás, Á., Hajdu, L., Hajdu, A.: Generalized Weighted Majority Voting with an Application to Algorithms Having Spatial Output.
   In: Hybrid Artificial Intelligent Systems. Ed.: Emilio Corchado, Václav Snášel, Ajith Abraham, Michał Woźniak, Manuel Graña, Sung-Bae Cho, Springer-Verlag Berlin Heidelberg, Heidelberg, 56-67, 2012, (Lecture Notes in Computer Science ; 7209) ISBN: 9783642289309

6. Tomán, H., **Kovács, L.**, Jónás, Á., Hajdu, L., Hajdu, A.: A Generalization of Majority Voting Scheme for Medical Image Detectors.
In: Hybrid Artificial Intelligent Systems : 6th international conference, HAIS 2011, Wroclaw, Poland, May 23-25, 2011; proceedings. Eds.: Emilio Corchado, Marek Kurzynski, Michal Wozniak, Springer, Heidelberg [etc.], 189-196, 2011, (Lecture Notes in Computer Science ; 6679.) ISBN: 9783642212215

7. **Kovács, L.**, Qureshi, R. J., Nagy, B., Harangi, B., Hajdu, A.: Graph Based Detection of Optic Disc and Fovea in Retinal Images.
In: Proceedings of IEEE 4th International Workshop on Soft Computing Applications, Arad-Romania, IEEE 4th International Workshop on Soft Computing Applications, Arad-Romania, 143-148, 2010. ISBN: 9781424479856

Foreign language scientific articles in international journals (3)

8. **Kovács, L.**, Kovács, R., Hajdu, A.: High Performance Computing in Medical Image Analysis HuSSaR.
arXiv 1, 1-4, 2018.
DOI: http://dx.doi.org/1806.06171

9. Hajdu, A., Hajdu, L., Jónás, Á., **Kovács, L.**, Tomán, H.: Generalizing the majority voting scheme to spatially constrained voting.
IEEE Trans. Image Process. 22 (11), 4182-4194, 2013. ISSN: 1057-7149.
DOI: http://dx.doi.org/10.1109/TIP.2013.2271116
IF: 3.111

10. Qureshi, R. J., **Kovács, L.**, Harangi, B., Nagy, B., Pető, T., Hajdu, A.: Combining algorithms for automatic detection of optic disc and macula in fundus images.
Comput. vis. image underst. 116 (1), 138-145, 2012. ISSN: 1077-3142.
DOI: http://dx.doi.org/10.1016/j.cviu.2011.09.001
IF: 1.232

Hungarian conference proceedings (1)

11. **Kovács, L.**, Petr, S., Riha, L.: HPC GPU/CPU társkártyákkal támogatott fotorealisztikus 3D CT vizualizáció.
In: VIII. Magyar Számítógépes Grafika és Geometria Konferencia. Szerk.: Szirmay-Kalos László, Renner Gábor, Neumann János Számítógép-tudományi Társaság, Budapest, 175-180, 2016. ISBN: 9786155036118

Foreign language conference proceedings (2)

12. **Kovács, L.**: 3D Visualisation from CT.
In: PRACE Summer of HPC Final Report. Ed.: Leon Kos, PRACE SoHPC, [s.l.], 43-45, 2015.

13. Qureshi, R. J., **Kovács, L.**, Nagy, B., Harangi, B., Hajdu, A.: Automatic detection of the fovea and optic disk in digital retinal images by combining algorithms.
In: Proceedings of the 8th International Conference on Applied Informatics. Ed.: Attila Egri-Nagy, Emőd Kovács, Gergely Kovásznai, Gábor Kusper, Tibor Tómács, Eszterházy K. College, Eger, 175-184, 2011. ISBN: 9789639894723

Hungarian abstracts (3)

14. **Kovács, L.**, Pintér, Á.: Stirling számok vizsgálata HPC környezetben.
In: Tavaszi Szél - Spring Wind 2015 : Absztraktkötet. Szerk.: Keresztes Gábor, Publio Kiadó, Budapest, 302-303, 2015.

15. **Kovács, L.**, Harangi, B., Nagy, B., Qureshi, R. J., Hajdu, A.: Gráf alapú vakfolt és sárgafolt detektálás retina felvételeken.
In: 8th Conference of the Hungarian Association for Image Processing and Pattern Recognition January 25-28, 2011 Szeged, Hungary : Proceedings of KÉPAF 2011. Szerk.: Zoltán Kató, Palágyi Kálmán, Erdőhelyi Balázs, Szegedi Egyetem, Szeged, 329-341, 2011.

16. **Kovács, L.**, Nagy, B.: Látógödör automatikus detektálása digitális retina képeken algoritmusok kombinálásával.
In: XXX. jubileumi OTDK Informatika Tudományi Szekció : [2011. április 18-20.] : [konferencia kiadvány], PPKE Információs Technológiai Kar, Budapest, 87, 2011.

Foreign language abstracts (5)

17. Antal, B., Tavares, M., **Kovács, L.**, Harangi, B., Pető, T., Csutak, A., Hajdu, A.: Data Analysis Applied to Diabetic Retinopathy Screening: Performance Evaluation.
Eur. J. Ophthalmol. 27 (3), E124-E125, 2017. ISSN: 1120-6721.
DOI: http://dx.doi.org/10.5301/ejo.5000979

18. **Kovács, L.**: Hybrid Small Size hpcResource.
In: GPU Day 2017 - The Future of Many-Core Computing in Science : absztrakt kötet, [s.n.], Budapest, 1, 2017.

19. **Kovács, L.**: Photorealistic 3D CT visualization supported by HPC GPU and CPU coprocessors.
In: GPU Day 2016 - The Future of Many-Core Computing in Science : absztrakt kötet, [s.n.], Budapest, 8, 2016.

20. Tóth, J., **Kovács, L.**, Harangi, B., Kiss, C., Mohácsi, A. I., Orosz, Z., Hajdu, A.: An Online System for Algorithm Benchmarking.
In: 5th IEEE International Conference on Cognitive Infocommunications. Ed.: Baranyi Péter, CogInfoCom, Vietri sul Mare, Italy, 383, 2014. ISBN: 9781479972807

21. **Kovács, L.**, Hajdu, L.: High Performance Computing With Mathematical Applications.
In: WSPS 2013.: 1st Winter School of PhD Students in Informatics and Mathematics, Tradeorg Nyomdaipari és Kereskedelmi Kft, Balatonalmádi, 13, 2013. ISBN: 9789638800121

UNIVERSITY AND NATIONAL LIBRAY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Patents (1)

22. **Kovács, L.**, Hajdu, A., Harangi, B., Sós, A.: Hybrid Small Size hpcResource - HuSSaR [mintaotalom]. 2018
Country: Hungary
Application Date: 2017.06.08-
Application number: U1700106/5
Under consideration

List of other publications

Foreign language Hungarian books (1)

23. Hajdu, A., Qureshi, R. J., Pető, T., Nagy, B., **Kovács, L.**, Harangi, B.: Detection of the optic disc and the macula through combining algorithms. Faculty of Informatics, Debrecen, 17 p., 2010. (Technical Report ; 7. Preprints ; 380.)

Total IF of journals (all publications): 4,343
Total IF of journals (publications related to the dissertation): 4,343

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

12 July, 2018