Comparative and Behavioural Analysis of a Diffuse Paradigm for the Evaluation of Diabetic Macular Edema in OCT images

Plácido L. Vidal*†, Joaquim de Moura*†, Macarena Díaz*†, Jorge Novo*† and Marcos Ortega*†
*Centro de investigación CITIC, Universidade da Coruña,
Campus de Elviña s/n 15071 A Coruña, Spain
†Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña,
Xubias de Arriba, 84, 15006 A Coruña, Spain
Email: {placido.francisco.lizancos.vidal, joaquim.demoura, macarena.diaz1, jnovo, mortega}@udc.es

***Abstract*—Nowadays, Diabetic Macular Edema (DME) is one of the leading causes of blindness in developed countries, and its characterized by the presence of pathological fluid accumulations inside the retinal layers. Currently, the main way to detect these fluid accumulations (as well as their severity) is through the use of Optical Coherence Tomography (OCT) imaging. In particular, this ophthalmological image modality allows a precise non-invasive analysis of the morphology of the retina and its structures.**

Due to the complexity of attempting to successfully segment these fluid accumulations, an alternative paradigm for their detection has been recently proposed. This paradigm, based on a diffuse representation of the pathological regions, creates an intuitive representation of the pathological regions based on a confidence map.

Currently, there are only two approaches for this paradigm: one based on a predefined library of texture and intensity features with established machine learning algorithms and other based on deep learning methods. Both approaches have proven to offer satisfactory results, but each one of them performs better in different scenarios.

In this work, we perform a complete analysis and comparison on the behaviour and performance of both strategies in a clinical screening scenario to evaluate the suitability of both approaches for the clinical practice as well as their performance as computer vision strategies.

I. Introduction

Diabetic Macular Edema (DME) is characterized by the pathological accumulation of fluid in the internal layers of the retina. These accumulations progressively form cystoid bodies that destroy the morphology of the retina. If left unchecked, they may derive in a problematic vision loss and, eventually, irreversible blindness [1]. This pathology, direct consequence of the diabetes mellitus, represents one of the main causes of blindness in developed countries [2].

To effectively diagnose this pathology, Optical Coherence Tomography (OCT) scans are commonly used [3]. This medical image modality allows to obtain a non-invasive representation of the retinal tissues [4]. Thanks to this capture technique, a diagnosis of the severity of the damage, treatment and possible clinical follow-up can be established. However, factors like the capture conditions and the subjectivity of the expert in the image analysis greatly affect the diagnosis and this posterior follow-up. For this reason, over the recent years, Computer-Aided Diagnosis (CAD) methodologies emerged to palliate these complications and help the clinicians to satisfactorily study the disease condition and progression.

Currently, in the context of intraretinal fluid accumulation identification methodologies, there are two complementary paradigms. The first paradigm aims to obtain a precise segmentation of the fluid accumulations. For this precise segmentation, works have been developed that use both classical computer vision techniques (like in the works of Wilkins *et al.* [5] with a thresholding and posterior rule-based filtering, Xu *et al.* [6] with a voxel classification or Girish *et al.* with a watershed transform and more advanced deep learning approaches (most of them based on the U-Net architecture [7] or networks derived from it, like in the works of Roy *et al.* [8], Chen *et al.* [9] and Lu *et al.* [10], this last one with a multiclass approach).

However, since this paradigm relies on a precise detection of the limits of the fluid accumulations, it suffers in cases where these accumulations do not have defined edges (situations that are frequent in this issue). For this reason, a paradigm based on a regional analysis and diffuse representation of the detections was proposed by Vidal *et al.* [11], using both classical [12] as well as deep learning approaches [13] to that aim. While both approaches have proven its applicability and robustness in a clinical environment over a precise segmentation approach,

This research was funded by Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project, Ayudas para la formación de profesorado universitario (FPU), grant ref. FPU18/02271; Ministerio de Ciencia e Innovación, Government of Spain through the research project with reference PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Grupos de Referencia Competitiva, grant ref. ED431C 2020/24; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

given the diffuse nature of the representations, it is difficult to establish differences beyond mere visual inspection. Both approaches present satisfactory results, but excel in different fields.

Nowadays, it is critical to explain the behaviour of automatic diagnostic support systems for an actual implementation in clinical practice. Thus, in this work, we perform a detailed analysis on the behaviour of both approaches to this paradigm by simulating a complete medical screening scenario, being able to test the real capabilities of both strategies.

II. Intraretinal fluid identification and confidence map generation

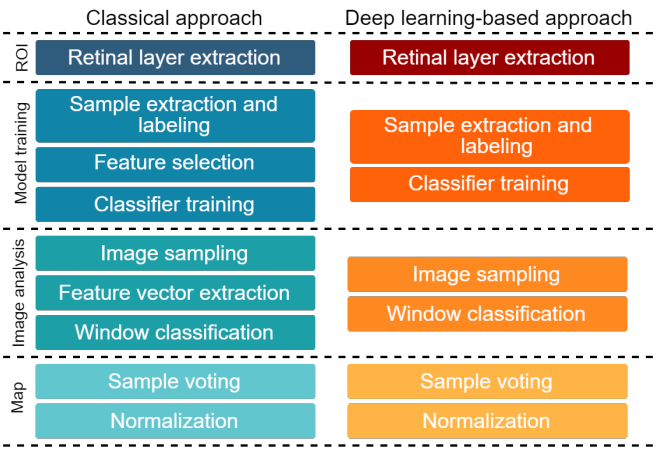

Fig. 1. Methodological steps for each of the approaches.

In Figure 1 we can graphically see the involved steps of both analyzed approaches. To create the confidence maps, both divide the retinal region into heavily overlapping samples of a fixed size, classify them, and calculate the probabilities of each pixel of belonging to a fluid region using the classification results and a voting strategy (finding, this way, an estimate of the confidence in the system for a pixel of belonging to a pathological fluid region).

A. Classical learning-based approach

This approach [11] requires a complete and heterogeneous set of 312 features, including intensity and texture characteristics. Thus, mainly characterized by extracting, before training, a complete set of characteristics from all the available datasets. From this matrix of a considerable size, a feature selection analysis is carried out by means of different complementary strategies, thus extracting those which provide more information to the subsequent discriminating system. Finally, the system is trained with this subset of characteristics in a progressively incremental way. This progressive and incremental analysis allows us to study in more detail the contribution of each of the characteristics, and to find the point where the improvement differential is insignificant (thus simplifying the model to the maximum and allowing a greater capacity of generalization).

B. Deep learning-based approach

As shown in Figure 1, this approach [13] follows the same map generation strategy, but it does not need any feature definition, selection or extraction. It directly trains a densely connected convolutional neural network [14] to classify the retinal region samples into pathological or not, greatly simplifying the training process and promoting its sensitivity and generalisation capability.

III. Performance and comparative analysis

To compare both approaches, given that both are intended for the clinical environment, we designed a medical screening experiment similar to those that the system could be subjected to during its use in clinical practice routine. In particular, the objective of this experiment is to classify an image as clinically relevant or not, evaluating two factors in the maps: the *confidence threshold* from which we consider a detection as relevant and the *area* occupied by this regional confidence. Since the images present different resolutions and micron-px relationships, we opted for using a relative measure of the occupied area of the intraretinal fluid accumulations. This measurement is the proportion of area occupied by the fluid accumulations identified by a methodology versus the total identified retinal surface in the OCT scan (in px). This way, we study the proportion of region of interest considered as pathological, given a confidence threshold from which a pixel is considered part of a fluid region. In this case, the confidence is the value returned by a methodology for its positive class.

Therefore, in a range of $[0, 1]$ of possible confidence values present in the maps and a possible pathological area/region of interest ratio range of $[0, 1]$, a given image x is considered pathological as:

$$F(x) = \frac{|M(x) > Ct|}{|ROI(x)|} > Rt \tag{1}$$

where *M(x)* represents the confidence assigned to the pixels in the ROI by the algorithms, *ROI(x)* a binary mask determining which image pixels belong to the region of interest and *Ct* & *Rt* the thresholds in a given iteration for the confidence and pathological area ratio, respectively.

To evaluate the results of this clinical screening experiment we used as metric the area under the ROC curve (AUC) that, for a binary score classification, can be defined as:

$$AUC = \frac{\frac{TP}{TP+FN} + \frac{FP}{FP+TN}}{2} \tag{2}$$

where TP indicate the True Positives, FN the False Negatives, FP the False Positives and TN the True Negatives.

For these experiments, both deep learning-based and classical approaches were trained using the same set of samples from two of the most representative capture devices of reference: a Cirrus Meditec HD-OCT from Carl Zeiss and a Spectralis OCT confocal laser ophthalmoscope from Heidelberg Engineering. From the OCT images of these devices, a total of 3,247 samples where extracted and manually labeled by an expert, including

both healthy, fluid and other pathological regions. In both cases, the samples were taken with a size of 61×61.

The screening experiment, on the other hand, was performed using a total of 323 OCT images. These images were separated into both pathologically relevant or not, considering clinically relevant images with a significant accumulation of intraretinal fluid (these being both severe fluid accumulations and dense groups of small fluid accumulations) and not clinically relevant OCT images with sparse small fluid accumulations or less [11]. In that sense, 92 Spectralis and 51 Cirrus OCT images were deemed as clinically relevant and 131 Spectralis and 49 Cirrus OCT images where deemed as not clinically relevant. All the maps, for both approaches, were generated using an image sampling configuration with an overlap of 55px.

In Figures 2 & 3, we can see the results of this experiment. As global conclusion, we can see how both strategies present a configuration where both classes can be easily separated, considering that the images include the most complex cases with small fluid accumulations. In the case of the classical approach (Figure 2), this optimal separation of 0.9240 AUC is achieved if we consider a positive case whenever any region (proportion threshold higher than 0) surpasses a 50% of confidence in the cystoid class. On the other hand, the deep learning-based approach needs to include at least a 10% of relative area marked as pathological with a confidence higher than 20% to reach its optimal separation of 0.9436 AUC.

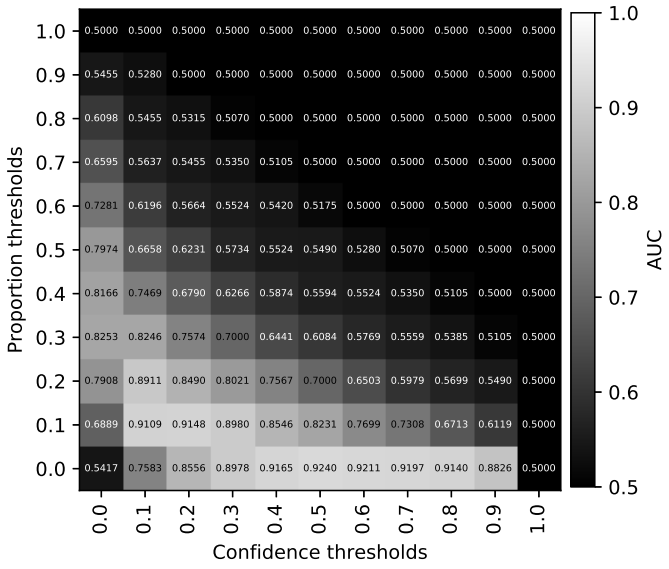

Fig. 2. Results of the screening experiment for the classical learning-based approach.

The fact that the classical system obtains its best results from any detection with a confidence greater than 40-50% is a clear sign that it is not excessively sensitive to these minuscule cysts, and either detects them with little confidence (hence the fact its best results are around the 50% threshold of confidence) or directly does not detect them.

Considering this, the fact that the centre of mass of the

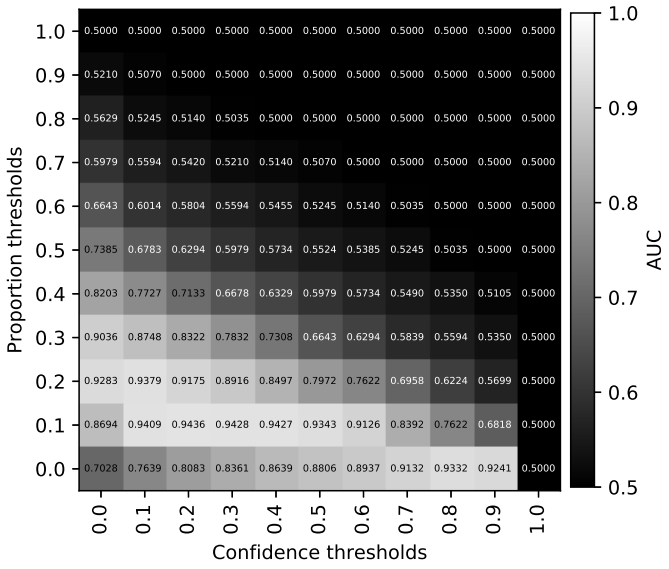

Fig. 3. Results of the screening experiment for the deep learning-based approach.

deep learning matrix is diverted to higher proportion and lower confidence values is indicative that it is considerably more sensitive and robust to smaller fluid accumulations than the classical approach. In Figure 4, two representative examples of this case are shown, where images with minuscule fluid accumulations (deemed as non-clinically relevant) are completely ignored by the classical learning approach, while the deep learning one gives a mid-to-high confidence value to even the smallest sign of intraretinal fluid.

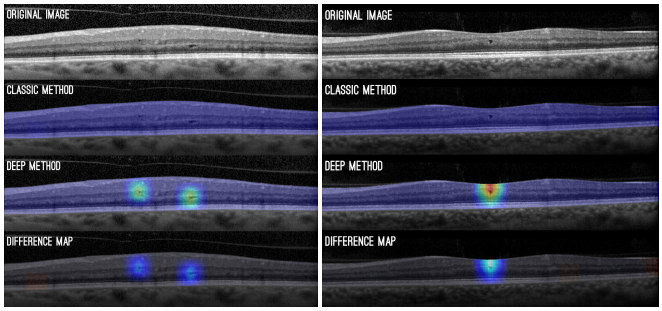

Fig. 4. Examples of OCT images with minuscule and minor fluid accumulations. The difference map shows in blue and red the areas where the deep and classical approaches returned higher confidence values, respectively.

Related to this, if we study the behaviour of both approaches in the point at which any detection is accepted whatever the confidence and proportion of the detection, we observe that, compared to the AUC obtained in the deep approach of 0.7028, the classical approach offers results close to randomness (an AUC of 0.5417). This implies that the approach based on classical learning is presenting a greater number of false detections in lower levels of confidence while, at the same time, the deep learning-based method is more likely to only generate detections where its connected component is over a significant

accumulation of fluid (thus, the FPs of the deep learning approach are negligible, as they are only slight overextensions of true detections).

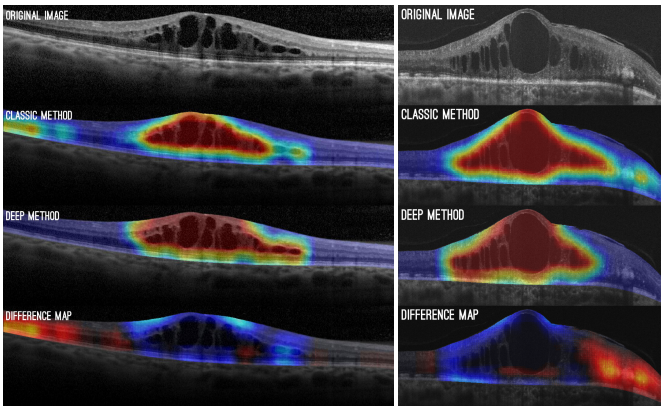

Fig. 5. Examples of OCT images with diverse fluid accumulations and complications. The difference map shows in blue and red the areas where the deep and classical approaches returned higher confidence values, respectively.

Figure 5 shows two OCT images representing the aforementioned cases. The leftmost image shows a darkened pattern product of the capturing process, while the rightmost image includes another pathological structure in the retinal tissues. In these examples, the classical approach shows FPs in these darkened regions and other abnormal pathological non-fluid structures, while the deep learning approach shows no detection whatsoever correctly detecting only the fluid accumulations.

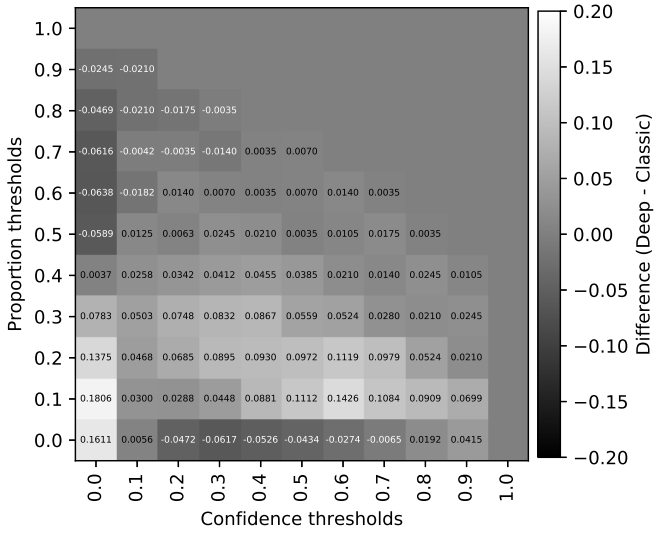

Fig. 6. Map of differences between the deep learning and classical approach screening performances. A negative difference (white values) indicates the configurations where the classic approach surpasses the deep learning one.

Finally, in Figure 6, we can observe how, as we increase the proportion of considered area and the minimum necessary confidence, it is the deep learning approximation the one that obtains a better performance. However, at extreme confidence-minimum proportion and higher proportion-minimum confidence threshold bands, it is the classical approach the one that surpasses the deep learning methodology (represented in the figure as white).

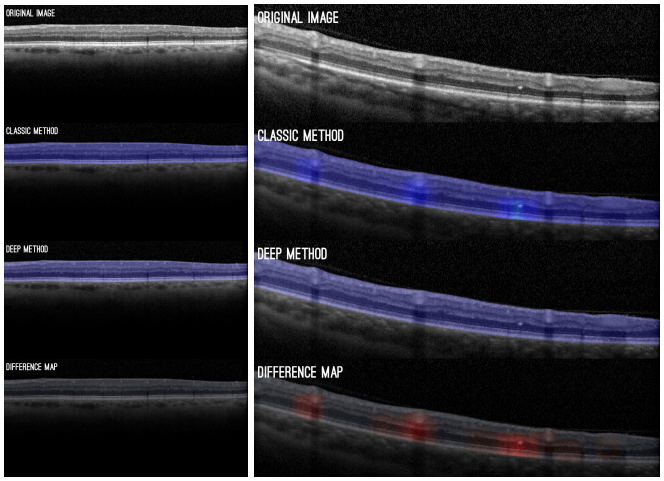

Fig. 7. Examples of OCT images with no fluid accumulations at all. These images contain different structures (like vessels) that present similar features with the studied fluid accumulations. The difference map shows in blue and red the areas where the deep and classical approaches returned higher confidence values, respectively.

This implies that the deep learning approach presents a smoother confidence transition than the classical learning approach, the later resulting in more adjusted confidence levels to the fluid accumulations. This does not necessarily imply worse results. As mentioned before, these broader detections of the deep approach tend to be connected to a component that overlaps a fluid region, while the classical approach is more prone to FPs. As shown in Figure 5, the approach based on deep learning returns measurements significantly more widespread than the approach based on classical learning, but only around the fluid regions. Also, in this same figure, it can also be observed how the maps generated with the support of deep learning techniques are able to indicate with high confidence some fluid regions that the classical approach does not.

In order to facilitate a thorough assessment, Figure 7, also presents a comparison between two healthy cases, from lesser to higher complexity regarding normal retinal patterns that could be confused with pathological fluid bodies. In the image on the left, we can see a simple case. In this image, there is no accumulation of fluid whatsoever, and no detection was made by any of the methodologies. That is, no complications and no patterns are present that could be confused with a fluid accumulation. In the same figure on the right, we can observe a high presence of shadows with high contrast and with a higher width. It is precisely in these patterns where a system based on a window sampling strategy can fall short. A partial capture of one of these regions can be easily mistaken for a pathological fluid accumulation, thus resulting in FPs.

By observing how both strategies reacted to these different cases, we can further understand what the Figure 6 was showing us. If we look at that table, we see how the classical system only begins to outperform the deep method from a certain

consideration of size (in this case, 0.5 times the size of the region of interest), but closing in to the deep learning based system as we increase the size of the region considered as pathological. The same way, it is shown how the confidence threshold necessary for the classical method to surpass the deep method is very low (20% of confidence).

Having this in mind along with previous section considerations, we can understand the behavior of the system (that is, apparently, paradoxical) observed in Figures 5 and 7. The classical system presents a much higher sensitivity than the system based on deep learning, and returns detections even with the slightest "suspicion" that the pattern could belong to an accumulation of pathological fluid. However, the classical method is completely blind to patterns smaller than a certain sample size. This is because the texture characteristics were designed and optimized for a dataset with larger fluid accumulations, so small patterns (such as those present in the microcysts in Figure 4) go completely unnoticed. On the other hand, larger patterns (such as those seen in Figures 5 and 7) are detected from the point of view of the texture characteristics extracted and selected by the system. Thus, while smaller texture patterns go unnoticed (since the system is blind to them and perceives them as normal layer patterns), the system returns detections to the minimum signal of similarity with a pathological body if its size is large enough to overcome that "blindness threshold".

It is interesting to note the inherent separation of the size of the cysts made by the system, since in the clinical literature these cysts are treated as a separate case from the DME disease. Microcystic Macular Edema (as are called these smaller fluid accumulations) is not only the precursor of larger accumulations of pathological fluid byproduct of diabetes and age-related macular degeneration [15] but, depending on their location in the layers of the retina, it can also be an indicator of other pathologies such as multiple sclerosis [16], [17]. Therefore, and taking into account the inherent separation that the system has made *per se* (thus, suggesting that there are features in the image that clearly separate both types of fluid accumulations), it is an indicator that an independent study of characteristics that helps to identify MME fluid accumulations is both desirable and necessary.

This "blindness threshold" issue is represented in the examples presented in Figure 8, in which we see how, in spite of having an accumulation of smaller cysts, the system returns only a slight detection in the classic methodology. However, if we examine closer the detection patterns of the classic strategy in this figure, we can see that it is not really detecting fluid accumulations: the contrast pattern generated by the mixture of sponge-like fluid (which has led to microcysts) has resulted in the appearance of a bubble of relatively high contrast. Thus, although we have a fluid accumulation with detection, it is clear that the classical system is detecting the joint pattern of multiple microcysts rather than the individual detection of each one of them.

On the other hand, since the network used presents a considerable size, it is capable of retaining discrimination potential, learning texture characteristics at different scales and maintaining its effectiveness even with smaller fluid patterns. Thus, despite the smaller fluid accumulations shown in Figures 4 and 8, all are marked with high confidence where the fluid bodies are present.

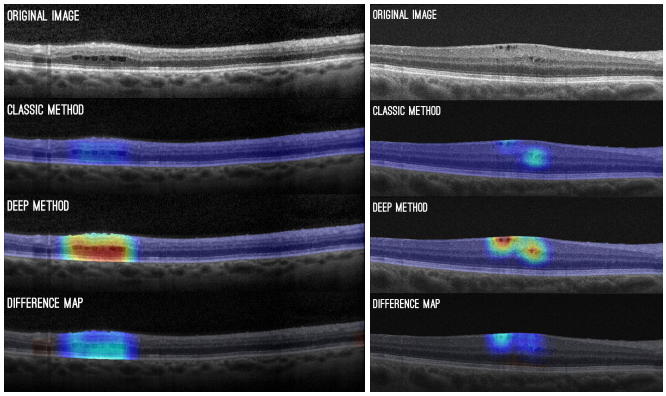

Fig. 8. Examples of small fluid accumulations where the classic methodology barely makes a detection. The difference map shows in blue and red the areas where the deep and classical approaches returned higher confidence values, respectively.

IV. Conclusions

In this work, we performed a study and exhaustive analysis between three approaches designed for the detection of intraretinal fluid to measure the suitability and capabilities in this complex issue. Due to the importance of the pathology (being one of the main causes of blindness in developed countries) and the relevance of an early diagnosis for the complete and satisfactory recovery of patients, numerous segmentation methodologies have emerged for its identification. However, due to the complexity of detection and instance extraction, a complementary paradigm to the classic segmentation has been alternatively proposed: a detection based on an analysis of overlapping regions and its supporting visualization strategy.

Currently, there are two main strategies employed for the application of this algorithm: one based on a complex analysis of texture characteristics, intensity, relevance ranking; and other based on the same sampling algorithm; but replacing the analysis of characteristics and classification by a deep convolutional network, thus bypassing the necessity of a predefined feature library and configurations.

To analyze these methodologies, since both systems are based on a non-segmentation paradigm, we have designed an adapted screening test. This test analyzes all possible fluid accumulations and system responses. Thanks to this analysis methodology we can study in a comprehensive way the behavior of the systems and evaluate their contribution to the clinical domain: their explanability.

Overall, the experiments indicated that the deep learning-based approach obtained higher global results (0.9436 AUC of the deep approach compared to the 0.9240 AUC of the classical approach). However, the behavioural analysis shows

that there are cases in which the use of one approach or another is desirable, depending on the goal of the search.

From what we have been able to observe, the core based on a deep learning strategy presents greater sensitivity to any type of fluid in the retina, returning a high level of confidence even in the smallest fluid accumulations at the cost of lower specificity, slightly extending the detections (but only around the correct true positive regions with fluid).

On the other hand, the system with the core based on an exhaustive analysis of texture and intensity features and machine learning techniques already predominant in the domain has achieved a higher level of sensitivity, but only to fluid accumulations that exceed a certain threshold of size. Also, the limitation of having to use a predefined library of characteristics is the main drawback of this approach, causing it to be partially blind to patterns of a size smaller than what its point of view allows. However, this behavior (as we have already seen in the analysis of the results) is contemplated as a separate case to DME by the clinical literature (the Microcystic Macular Edema or MME), with its own characteristics, implications and casuistry that should be analyzed independently. Additionally, its predefined library allows this methodology to reach a greater specificity in the cases of fluid that exceed that already mentioned "threshold of blindness", adjusting better to the limits of pathological accumulations of fluid than its deep learning based counterpart even versus the regularized deep learning approach.

Finally, in reference to healthy cases, all systems are able to satisfactorily distinguish regions with healthy tissue. However, of the three studied strategies, the ones based on deep learning are the ones that best distinguish these regions. The classical learning-based approach suffers in healthy regions with high contrast patterns of considerable size, returning low-confidence detections in these perfectly healthy zones. Nonetheless, the original deep learning approach only returns a detection in healthy tissue if its connected component is also part of a section with true pathological fluid. In other words, the original deep learning strategy tends to expand true detections a bit towards healthy tissue, but never returns a positive in an isolated healthy region. Thus, never returning a true false positive, but rather an overextension of a true positive. This implies that, although all approaches return similar results from AUC, thanks to our analysis, we can see that their behavior is noticeably different.

The approach based on deep learning is the most interesting for the clinical domain because of its independence from a library of characteristics and its improved sensitivity in terms of smaller detections. However, a solution involving both approaches is recommended: as proven by this analysis, no approach is strictly better than the other. They complement each other.

REFERENCES

[1] S. J. Bakri, J. D. Wolfe, C. D. Regillo, H. W. Flynn, and C. C. Wykoff, "Evidence-based guidelines for management of diabetic macular edema," *Journal of VitreoRetinal Diseases*, vol. 3, no. 3, pp. 145–152, 2019. [Online]. Available: https://doi.org/10.1177/2474126419834711

[2] P. Romero-Aroca, "Managing diabetic macular edema: The leading cause of diabetes blindness," *World journal of diabetes*, vol. 2, no. 6, pp. 98–104, Jun 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/21860693

[3] M. R. Hee, C. A. Puliafito, C. Wong, J. S. Duker, E. Reichel, B. Rutledge, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto, "Quantitative assessment of macular edema with optical coherence tomography," *Archives of ophthalmology*, vol. 113, no. 8, pp. 1019–1029, 1995.

[4] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito *et al.*, "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[5] G. R. Wilkins, O. M. Houghton, and A. L. Oldenburg, "Automated segmentation of intraretinal cystoid fluid in optical coherence tomography," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1109–1114, 2012.

[6] X. Xu, K. Lee, L. Zhang, M. Sonka, and M. D. Abràmoff, "Stratified sampling voxel classification for segmentation of intraretinal and subretinal fluid in longitudinal clinical oct data," *IEEE transactions on medical imaging*, vol. 34, no. 7, pp. 1616–1623, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[8] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network," *CoRR*, vol. abs/1704.02161, 2017.

[9] Z. Chen, D. Li, H. Shen, H. Mo, Z. Zeng, and H. Wei, "Automated segmentation of fluid regions in optical coherence tomography b-scan images of age-related macular degeneration," *Optics & Laser Technology*, vol. 122, p. 105830, 2020.

[10] D. Lu, M. Heisler, S. Lee, G. W. Ding, E. Navajas, M. V. Sarunic, and M. F. Beg, "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network," *Medical Image Analysis*, vol. 54, pp. 100 – 110, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841519300167

[11] P. L. Vidal, J. de Moura, J. Novo, M. G. Penedo, and M. Ortega, "Intraretinal fluid identification via enhanced maps using optical coherence tomography images," *Biomed. Opt. Express*, vol. 9, no. 10, pp. 4730–4754, Oct 2018. [Online]. Available: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-9-10-4730

[12] J. Moura, P. L. Vidal, J. Novo, J. Rouco, and M. Ortega, "Feature definition, analysis and selection for cystoid region characterization in optical coherence tomography," in *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference KES-2017, Marseille, France, 6-8 September 2017.*, 2017, pp. 1369–1377.

[13] P. L. Vidal, J. de Moura, J. Novo, and M. Ortega, "Cystoid fluid color map generation in optical coherence tomography images using a densely connected convolutional neural network," in *2019 International Joint Conference on Neural Networks (IJCNN)*, July 2019, pp. 1–8.

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[15] S. Haldar, R. Mukherjee, and J. Elston, "Microcystic macular edema in a case of tobacco–alcohol optic neuropathy," *Canadian Journal of Ophthalmology*, vol. 52, no. 1, pp. e19–e22, Feb 2017.

[16] A. Lang, A. Carass, E. K. Swingle, O. Al-Louzi, P. Bhargava, S. Saidha, H. S. Ying, P. A. Calabresi, and J. L. Prince, "Automatic segmentation of microcystic macular edema in oct," *Biomed. Opt. Express*, vol. 6, no. 1, pp. 155–169, Jan 2015. [Online]. Available: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-6-1-155

[17] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Deep learning based topology guaranteed surface and mme segmentation of multiple sclerosis subjects from retinal oct," *Biomed. Opt. Express*, vol. 10, no. 10, pp. 5042–5058, Oct 2019. [Online]. Available: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-10-10-5042