Research Article

Evaluation of an Artificial Intelligence-based Detector of Sub- and Intra-Retinal Fluid on a large set of OCT volumes in AMD and DME

Supplementary Material

Methods

Annotation Protocol – details

In the AMD cohort, the annotation protocol required a distinction between intraretinal fluid and intraretinal cysts (IRC), IRC being defined as oval well-defined areas between the internal limited membrane (ILM) and the photoreceptor layer (PR), and IRF as a diffuse darkening between the ILM and PR. Neither of these definitions were tied to numerical value and if a thickening of the retina with cysts was present, then IRF was annotated as present. Therefore, for the rest of this study and considering the grading protocols, we merged both annotations, intraretinal fluid and IRC, into a single one, intraretinal fluid denoted IRF. Indicative of SRF was homogeneous hyporeflective well-defined areas, which are at least 100µm in a horizontal extent, between the retinal pigment epithelium layer (RPE) and PR.

In the DME cohort, IRF was defined as either cystoid spaces or diffuse thickening/darkening, and cystoid macular edema as oval well-defined areas between the ILM and the PR layer within the 3 mm inner ring of the Early Treatment Diabetic Retinopathy Study (ETDRS) grid. SRF was characterized as a well-defined darkening between the RPE and RP, and serous macular detachment as a presence of subretinal fluid under the center of the fovea more than 100µm in height.

Results

Incorrect classifications and biomarkers distribution

The false negative rate for IRF was overall slightly higher in the DME cohort (6.4%) than in the AMD cohort (4.98%), as illustrated in Figure 3. The false positive rate for IRF detection also showed a marked difference between AMD (8.18%) and DME (0.52%) cohorts. The discrepancies of the IRF distribution between both cohorts could explain these results, as nearly 89 % of the DME set was annotated with IRF compared to only 30.7% of the AMD set (Figure 1). This explanation could be also applicable to understand the differences between the false negative rates of SRF in the AMD (2.54%) and DME cohorts (0.39%), given that only 8.5% of the DME set was annotated with SRF versus 34.9% of the AMD set (Figure 1). The false positive rate for SRF detection was 4.75 % in the AMD and 4.30% in the

source: https://doi.org/10.48350/173747 | downloaded: 12.10.2024

DME cohorts. In the context of a lower threshold for SRF in the DME cohort, the explanation could lie in the rarity of the presence of SRF in this group (Figure 1).

Independence of the single detectors and analysis of a combined fluid detector - details

In the AMD cohort, the combined fluid detector allowed the correct prediction of 3220 OCT volume scans (89.0%) while the single detector correctly predicted 2952 OCT volume scans (81.6%). The number of OCT volume scans correctly predicted by the two single detectors and the joint detector was 2909 (80.4%). The fluid detector correctly predicted 311 more OCT volume scans than the single detectors, but missed 43 of them over the 2'952 (1.46%) OCT volume scans correctly predicted when combining the two single detectors. We found that these 43 OCT volume scans corresponded to false negatives of the fluid detector and most of them (41/43) were true positives of the IRF detector and true negatives for the SRF detector. A qualitative analysis showed that the vast majority of these missed OCTs contained only few fluids in form of cysts on less than 3 B-scans and were of low quality. The additional correctly predicted 311 OCT volumes included 232 OCTs wrongly predicted by the IRF detector (182 false positives, 47 false negatives) and 87 OCTs wrongly predicted by the SRF detector (52 false positives and 35 false negatives). A qualitative analysis of 15% of these missed cases demonstrated similar results as shown in Figure 4.

In the DME cohort, the combined fluid detector allowed the correct prediction of 10'703 OCT volume scans (93.1%) while the single detector correctly predicted 10'189 OCT volume scans (88.6%). The number of OCT volume scans correctly predicted by the two single detectors and the joint detector was 10'189 (88.6%). The combined fluid detector correctly predicted all the samples, which were correctly predicted by the single detectors. Furthermore, the fluid detector correctly predicted 514 additional OCTs, where 1 OCT was wrongly predicted by the IRF detector (false negative) and 513 OCTs wrongly predicted by the SRF detector (468 false positives and 45 false negatives). A qualitative analysis of 15% of these cases confirmed our previous findings, summarized in Figure 4.