Title

Full title: Eye-Opening Advances: Automated 3D Segmentation, Key Biomarkers Extraction, and the First Large-Scale MRI Eye Atlas.
Short title: AEye: MRI Eye and Orbit Segmentation.

Authors

Jaime Barranco[1,2,3,4], Adrian Luyken[5], Hamza Kebiri[1,2], Philipp Stachs[6], Pedro M. Gordaliza[1,2], Oscar Esteban[2], Yasser Aleman[2], Raphael Sznitman[7], Oliver Stachs[5,8], Sönke Langner[5,9], Benedetta Franceschiello[3,4†], Meritxell Bach Cuadra[1,2†]

Affiliations

[1]CIBM Center for Biomedical Imaging, Lausanne, Switzerland

[2]Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

[3]School of Engineering, Institute of Systems Engineering, HES-SO Valais-Wallis, Sion, Switzerland

[4]The Sense Innovation and Research Center, Lausanne and Sion, Switzerland

[5]Department of Ophthalmology, Rostock University Medical Center, Rostock, Germany

[6]Karlsruhe Institute of Technology (KIT)

[7]ARTORG Center for Biomedical Engineering, University of Bern, Bern, Switzerland

[8]Department Life, Light & Matter, University of Rostock, Rostock, Germany

[9]Institute for Diagnostic and Interventional Radiology, Pediatric and Neuroradiology, Rostock University Medical Center, Rostock, Germany

[†]These authors provided equal last-authorship contribution

Abstract

This study addresses the need for an automated and accurate 3D segmentation of the healthy human eye and orbit from Magnetic Resonance Images, to allow improved ophthalmic diagnostics and treatments. Past efforts primarily focused on small sample sizes and varied imaging modalities. Here, we leverage a large-scale dataset of T1-weighted MRI of 1245 subjects and the use of the deep learning-based nnU-Net for MR-Eye segmentation tasks. The results showcase robust and accurate segmentations of lens, globe, optic nerve, rectus muscles, and fat. We also present the automated estimation of key ophthalmic biomarkers such as AL and volumetry, while benchmarking correlations between body mass index (BMI) and eye structure volumes. Quality control protocols are introduced through the pipeline to ensure the reliability and clinical relevance of the segmented large-scale data, further enhancing the applicability of our algorithm.

Teaser

This study advances MR-Eye with automated 3D segmentation of human eye and orbit, biomarker extraction, and a large-scale eye atlas in MRI.

Introduction

According to the World Health Organization (WHO), 2.2 billion people have vision impairment or blindness (1) and preventable causes account for 80% of the total global visual impairment burden. The eyes, small, complex, and delicate structures that serve as our primary sensory organ, convey crucial information about our visual health and can be seen as a window to the brain, connecting us to the external world (2). In clinics, an ophthalmic unit is equipped with a variety of devices to perform qualitative and quantitative imaging of the eye. These include, for example, ultrasound (3), funduscopy (4), and optical coherence tomography (OCT) (5, 6), which are capable of extracting anatomical measurements of the eyes. However, these devices are not as effective in imaging the posterior part of the eye, and/or just provide partial information, for example when in presence of volumetric lesions or other pathologies (7, 8, 9, 10). Therefore, the development of diagnostic methods and new imaging analysis techniques remains crucial, and that is where Magnetic Resonance Imaging (MRI) plays a central role. With its non-invasive and penetration characteristics, MRI is particularly a promising technique as it can provide 3D measurements of the complete eye, related to both the tissue and organ structure, and informs about particle deposits within the tissues, such as calcification or tissues deformations. Ophthalmic MRI (7, 8, 9, 10), known as MR-Eye (11, 12, 13, 14, 15), has been proven to be highly effective in oncology, for the evaluation and treatment planning of tumors, as well as for the quantification of orbital inflammation and for refractive surgery planning (10). Furthermore, given that neurodegenerative disorders frequently involve ocular and visual comorbidities (11, 17), and oculomotor dysfunctions can signify underlying brain injuries (18, 19), advancing the current capabilities of MR-Eye technology is paramount. This will help the assessment of the eye-brain pathway.

The integration of MR-Eye is advancing towards a comprehensive understanding and early interception of diseases. Key ophthalmic biomarkers can be manually depicted from MR images, such as the axial length (AL) (20, 21, 22), useful for refractive errors, myopia, hyperopia, glaucoma, retinal detachment… or the volumetry (23, 24, 25), useful for eye growth abnormalities, glaucoma, macular degeneration, and orbital tumors. As of today, the automated extraction of eye biomarkers from MR-Eye remains underdeveloped. To the best of our knowledge, no tools perform automated extraction of AL using 3D MRI. Regarding volumetry analysis, previous works (23, 24), reported only the total orbital volume, around 27.5 cm^3. In (25), they analyzed the orbital muscle fraction in relation to the total orbital volume in patients affected by Graves' orbitopathy (GO). They provided volumetry of the entire globe (globe, lens, sclera, and cornea) and extraocular muscles (four rectus and two obliques) of a specific patient as an example, with 6.95 and 7.92 cm^3, respectively. However, there is currently no ophthalmic technology that can provide an accurate volumetric estimation of the eye and its substructures to the millimeter while their possible correlations with other biomarkers would be highly relevant.

Moreover, the development of eye models, such as eye atlases, would allow for colocalization and navigation in the eye, and it may serve as standardized spatial references for the eye, serving as means for exploring quantitative geometric measurements of eye morphology despite systematic differences within a population. Prior knowledge encoded in the form of anatomical and probability atlases have played a pivotal role in conventional neuroimaging research since many years by providing a standardized framework for spatial normalization and quantitative analysis across diverse populations (26, 27, 28). These atlases serve as reference templates that enable researchers to map and compare structural variations in the brain, facilitating the investigation of neurological disorders and brain function. However, such standardized tools and frameworks are currently lacking in the field of ophthalmic imaging. A recent study (29) has pioneered the development of unbiased MRI eye atlas, available through the HuBMAP project (30), covering various contrasts (T1w pre-contrast, T1w post-contrast, T2w TSE, and T2w FLAIR) based on 100 images. Despite this first effort there is the need for a larger scale atlas and to separate anatomical models tailored to each sex, as sex differences may play a crucial role in various diseases (31, 32, 33, 34), e.g. endocrine orbitopathy.

However, achieving these advancements depends critically on robust and accurate delineation of eye and orbit structures. Pioneer semiautomated methods relied on parametric shape modeling of the eye using prior estimations such as spheres and ellipsoids (35, 36), or, more sophisticated manifolds, i.e. spherical meshes (37). The main drawback of parametric models is that they rely on a deterministic pre-defined geometry ignoring normative stochastic modelling (i.e., image intensity, shape variations, etc.) that could be derived from the anatomical variability within a specific

population. To tackle this, active shape modeling (ASM) algorithms through Machine Learning (ML) optimizations came up with a robust solution to fit each shape to the eye structures via statistically driven deformations. This triggered progress on statistical shape models for semiautomated segmentation of both healthy eye structures and tumors (38, 39). Deep learning methods appeared to address this segmentation task and fully automatize it, using a 2D/3D U-Net (40, 41, 42). Other approaches were combinations of the previous two methods (43, 44), and clustering techniques (45), having similar results. However, these previous works were generally focused on the segmentation of few eye (non-orbit) structures (lens, globe, sclera, cornea, and not so common, optic nerve, only in (39, 40)). Thus, important orbit structures such as rectus muscles (RM) or fat, are yet unexplored, jeopardizing the construction of a comprehensive model of the eye and orbit. Moreover, despite methodological efforts on resampling (e.g. cross-validation, bootstrapping, etc.), they did not count on a big cohort of manually annotated healthy subjects (sample sizes were limited to 24 to 40 subjects) that would properly gather their anatomical variability for the validation of the results and most of them (except for (38, 39)) relied on the availability of multi-contrast MRI setting. Finally, while image quality is a well-known factor that can significantly bias automated results in neuroimaging (46, 47, 48, 49, 50), it is often overlooked in MR-eye analysis. Only a few of the previous studies (39, 41, 44) implemented some form of quality control prior to the segmentation task.

The contribution of this work is three-fold. First, we present (i) a comprehensive, accurate 3D MR-Eye segmentation method of the healthy human adult eye and orbit structures including lens, globe, also known as vitreous humor (VH), optic nerve, RM, and fat, using a T1-weighted (T1w) MRI dataset of 1245 healthy subjects. We evaluated its performance on 74 manually segmented subjects by means of the objective metrics Dice Similarity Coefficient (DSC), Hausdorff distance and volume difference. The state-of-the-art supervised deep learning-based approach, nnU-Net (51), is the method of choice, surpassing our proposed baseline method, namely atlas-based, whose results and method can be found in the Supplementary Materials document. This segmentation method allowed us to (ii) provide automated extraction benchmarks of a large-scale cohort on ocular MR-Eye biomarkers (AL and volumetry) for the first time, and (iii) provide the first large-scale MR-Eye atlases per sex including 594 males and 616 females with their corresponding labels (publicly available at (76)). To ensure the reliability of the segmentation and biomarkers extraction, we introduced a tailored MR-Eye quality control protocol as the existing brain QC approaches failed to assess eye image quality.

Results

Our work presents a deep-learning algorithm (nnU-Net) for automated 3D segmentation of eye and orbital structures, capable of extracting automatically key biomarkers such as AL and volumetric measurements of eye structures (volumetry). Leveraging the extensive scale of our database, we introduce, for the first time to our knowledge, a large-scale probabilistic atlas of the eye.

Automated segmentation

Figure 1 displays a visual representation of the obtained segmentation. To quantitatively assess the performance of our algorithm in anatomically delineating the eye structures as compared to manual expert annotations (referred as ground truth, and more correctly to surrogate truth or reference standard), we used image quality metrics (Dice score – DSC, Hausdorff distance – HD, and volume difference - VD) on a test set of 43 subjects. These 43 subjects (age 38-77, 28 females) have acceptable MR-Eye image quality, i.e. the MR-Eye do not contain major classic artefacts, as rated by MR-Eye experts (subjective ratings, see Material and Methods section). We show that the proposed model produce accurate results in delineating all eye structures (average score across structures: DSC=0.81±0.07, HD=0.35±0.20mm, and VD=0.19±0.14mm3) as compared to the ground truth (scores detailed in Figure 2 and Table 1). As expected, less accuracy was encountered in those structures which are more anatomically variable, such as the fat, and in the superior rectus muscle.

Extraction of biomarkers at large-scale

After automatically delineating the anatomy of eye structures, we developed an automated procedure to compute key ophthalmic biomarkers, including millimeter-scale volumetry of eye structures and AL. This automation allowed us to extract these measurements from a large-scale dataset of 1,157 subjects, following Quality Control (QC; see Quality Control Protocol in the Materials and Methods section).

Our findings show that our large-scale automated measurements of AL from MRI are in line with the reference manual measures (20). A shows boxplots of the automated AL measures per sex, as in (20), on the large-scale cohort of 1157 subjects. The mean AL was close to reported values in the literature (20), which were manually extracted from the T1w

images by trained ophthalmologists, for both males and females. However, in 107/1157 cases, the AL could not be computed due to methodological constraints (e.g., 79 unsegmented lenses), that are further explained in the Materials and Methods section). Specifically, the mean values and standard deviations are:

- NnU-Net: 23.8±1.7 mm (524 M) and 22.9±1.6 mm (526 F)
- Previous studies (20): 23.4±0.8 mm (1059 M) and 22.8±0.9 mm (867 F)

In terms of volumetry extraction, we provide the first large-scale benchmark MR-Eye volumetry of all eye structures. We observed a trend of males having larger eye structures than females (except for the lens), particularly in both intraconal and extraconal fat. Figure 4 illustrates the extracted volumetry per structure, grouped by sex, using violin plots from the large-scale cohort of 1,157 subjects. Table 2 presents median and standard deviation for each structure. Interestingly, we did not find any significant correlation between body mass index (BMI) and the volumes of eye structures. Using correlation analysis and Huber linear regression, we observed Huber scores (R^2) below 0.1, except for intraconal fat in males, which was 0.11. Figure 5 illustrates these findings with scatter plots, Huber regression lines, and Huber R^2 scores, grouped by sex.

Atlas of the eye

Figure 6 presents large-scale, unbiased male and female eye atlases using MRI, constructed from 594 males and 616 females, with their corresponding probability maps of the different labels projected onto the average respective male and female templates, which are publicly released. Additionally, a 3D render of the maximum probability maps is shown. The volumes of these maps indicate similar structure sizes for both sexes, except for the fat, which is larger in males, particularly the extraconal fat.

Discussion

MR-Eye has increasingly gathered interest in the ophthalmic and radiology community (10), due to the incredible tissue contrast that it can achieve in a non-invasive way. Furthermore, and unlike most ophthalmic tools which evaluate the anatomy or the visual performance of the eyes (Ocular Coherence Tomography (OCT) (5, 6), biometry (52), microperimetry (53), eye-tracking, contrast sensitivity), MR-Eye can investigate several pathologies behind the globe, involving nerves paralysis, lesions, tumors and inflammation (7, 10, 54), while exploring the 3D complexity of the eye-shape. In fact, 3T and 1.5T MR-Eye clinical protocols are used regularly in the case of tumor (retinoblastoma (55, 56) or uveal melanoma (57, 58)), or ocular inflammations (10, 20, 57, 59), or pathologies with suspected link to the brain (54), and constitute the current state of the art of clinical practice. Very recent technical advancements propose new ways to deal with the presence of motion artefacts during MR-Eye acquisition (60, 61, 62, 63, 64) or at ultra-high field (7T) (15, 65), increasing the usability and reproducibility of MR-Eye in ophthalmology.

In this rapidly growing field, it is crucial to enable clinicians to extract measurements from MR-Eye and benchmark new metrics, providing them with tools not available before. To address this need, we propose a comprehensive automated pipeline. This pipeline is benchmarked on a large-scale MR-Eye database of 1,157 subjects and introduces a methodology for automated 3D segmentation, see Figure 1, of all eye structures using deep-learning algorithm (nnU-Net). It extracts key ophthalmic biomarkers, such as AL (Figure 3) and volumetry (Figure 4), and allows us to build the first large-scale comprehensive eye atlas for both males and females, complete with their corresponding probability maps (Figure 6).

Our automated 3D segmentation via deep learning (nnU-Net) of all eye structures, once compared with manual segmentation performed by expert ophthalmologists on 43 testing subjects, is optimal with respect to classic image quality metrics, namely DSC, HD, and VD. These results are in line with previous reported values of segmentation performance for lens, globe, and optic nerve (38, 39, 41, 42, 43, 44), but they relied on multi-contrast MRI and healthy and non-healthy eyes, including tumors such as retinoblastoma (41, 42, 43) and uveal melanoma (39, 44). A comparison table of the performances of these previous methods can be found in (42). To our knowledge, this study reports for the first time the anatomical delineation of structures such as fat and rectus muscles'. Moreover, our automated segmentation completes in just one minute per volume. With its high accuracy, it could be seamlessly integrated into MRI console analysis, potentially saving clinicians the 10 to 20 minutes they currently spend on manual segmentation and streamlining the clinical flow. Additionally, we aim to adapt our segmentation to handle variations in contrast and spacing, aligning with the current state-of-the-art MR-Eye protocols, which include T1w imaging, fat-suppressed T1w

and T2w imaging, and contrast injections (7, 8, 10, 54). Incorporating uncertainty quantification for automated predictions can be beneficial to such scopes (66).

To ensure the removal of low-quality images that could compromise the results, we introduced QC protocols at multiple stages of the segmentation pipeline. Inspired by state-of-the-art methods, MRI-QC (49), we discovered a discordance between low-quality image candidates identified by MRI-QC and those identified by our MR-Eye experts. This suggests that QC in MR-Eye requires different metrics and criteria compared to brain imaging, highlighting a crucial new area of investigation. Future development will need to define image quality metrics tailored specifically to eye tissues, incorporate non-tissue metrics, and extend scrutiny to the periorbital region.

To further validate our pipeline, we introduce a novel large-scale automated method to measure AL from segmented MR-Eye volumes. Our automated results closely match the reference manual measures of AL (20, 21, 22) performed by expert ophthalmologists on a database of 1,157 subjects. This reinforces the reliability of our automated approaches for both eye structure segmentation and AL extraction.

We provide the first large-scale benchmark for volumetry of all eye structures at a millimeter scale. Previous work introduced volumetry extraction from MR-Eye volumes (23, 24, 25), but for the entire globe and extraocular muscles in cm^3. By distinguishing between male and female eye anatomy, we observe a general trend where males have larger eye structures than females, with the notable exception of the lens. This sex-wise differentiation in eye structure volumetry could have significant implications for understanding sex-specific ophthalmological conditions and tailoring more personalized medical treatments, particularly as such differentiation is nowadays needed for a better health care (33, 34). Interestingly, despite a previous study (20) found that the exophthalmometric value, defined as the perpendicular distance between the interzygomatic line and the posterior surface of the cornea (20), was significantly associated with AL ($p<0.001$) and that it was also positively correlated with BMI ($p<0.001$), our investigation revealed no significant correlation between body mass index (BMI) and eye structure volumes. This finding suggests that variations in eye structure volumes may be not associated with BMI.

Our study introduces a novel method for automated biomarker extraction, paving the way for benchmarking MR-Eye-derived measurements of the adult human eye. The implications of these findings are vast: potentially enhancing diagnostic precision, informing surgical planning, improving our understanding of eye anatomy across different populations, and saving clinicians' time. Future research should aim to further validate these methods in pathological eyes and explore additional biomarkers. For instance, evaluating changes in rectus muscles is keys in pathology such as strabismus (67, 68), or open to the evaluation of new elements such as cerebrospinal fluid (CSF), whose deposit in the optic nerve plays a crucial role in pathologies such as papilledema and glaucoma (69, 70).

In this study we present pioneering male and female eye MRI atlases, along with their detailed labels. Atlases are crucial in research as reference tools for registration and segmentation in population imaging studies. In clinical practice, they can facilitate the diagnosis and treatment of a wide range of ocular diseases, help to reveal abnormal structural changes, enhance surgical planning, and improve our understanding of sex-specific variations in eye anatomy and physiology (26). These atlases offer a valuable resource for advancing the study of ocular anatomy and can significantly support the accuracy of eye-related research and clinical applications, as has been largely demonstrated for brain studies (26, 27, 28, 71). Furthermore, the atlases enable colocalization and navigation within the eye, serving as a standardized spatial reference. This facilitates the exploration of quantitative geometric measurements of eye morphology and structures, even in the presence of systematic population differences (29).

MR-Eye is indispensable when other ophthalmologic imaging modalities fail (7, 8, 9, 10), and constitutes a rapidly expanding field. Our study sets a new precedent in ophthalmology by demonstrating the feasibility and accuracy of large-scale automated segmentation and biomarker extraction from MR-Eye. Our findings propose a ready-to-use solution to promote the adoption of accurate MR-Eye segmentation, together with its applicability in the clinical and research setting.

Materials and Methods

Experimental Design

We evaluated quantitatively a deep learning-based automated segmentation method on a cohort of manually segmented subjects using similarity metrics (surface overlap, volume error and distance-based error). Then, we performed a large-scale analysis on relevant ophthalmic biomarkers, namely volumetry of eye structures and AL measurements, for which we developed a method for its automated extraction. This allowed further analysis of possible correlation between volumes and BMI grouped by sex (males and females). We introduced eye-quality control checks that are described later in this section. Thus, the major components of our study were large-scale cohort, automated segmentation methods, automated biometry extraction, quality control analysis, and statistical analysis.

Dataset

The cohort was originally acquired within the Study of Health in Pomerania (SHIP) (20) and reused in the context of this study. A total of 3030 healthy subjects underwent whole-body MRI on a 1.5T scanner Magnetom Avanto (Siemens Medical Solutions, Erlangen, Germany) without contrast agent, from which we used 1245 subjects for this study. Subjects were overall aged between 28 and 89 (56±13) years old. T_1-weighted (T1w) images of the head were acquired using a 12-channel head coil, 176 slices per volume, with a slice thickness of 1mm, and a field of view of 256mm, voxel size 1 mm^3, TR=1900 ms, TI=1100 ms, TE=3.37 ms. During the MRI examination, subjects rested their eyes naturally without specific guidelines for viewing or eyelid position. All participants gave informed written consent. The study was approved by the Medical Ethics Committee of the University of Greifswald and followed the Declaration of Helsinki. All data of the study participants were accessed from an anonymized database.

Manual segmentation protocol

Manual annotations on a total of 74 subjects were done, using ITK Snap software (72), by two expert readers independently: one senior ophthalmologist (20 years of experience) and one junior ophthalmologist (1y). The senior one double checked the annotation by the junior and corrected them if needed. These manual annotations included 9 region-of-interest (ROIs) for the right eye: lens, globe, optic nerve, intraconal and extraconal fats, and the four rectus muscles (lateral, medial, inferior, and superior), see Figure 1B.

Subjective quality evaluation

To evaluate the robustness of the nnU-Net, we evaluated possible correlation between the eye-quality of the images and the segmentation method's performance, by means of the DSC. To get the subjective eye-quality of the images, two engineer experts in MR image analysis (20 and 5 years of experience) independently rated them from 0 to 4 (being 0 excluded and 4 excellent quality) making use of adapted MRIQC (49) reports. These reports consist of an html-file per subject in which many thumbnails of the axial view are presented, as well as some sagittal and coronal views, to help the rater evaluate the quality of the image. A rating widget is provided, including several key components to correctly evaluate the quality of the image, such as overall quality, blur, noise, motion, etc. We modified the original reports to meet our needs by changing the field of view for the thumbnails (centered to the right eye) and adding eye-oriented aspects in the rating widget such as open/close, see Figure 7.

Automated segmentation method: nnU-Net

nnU-Net (51) is the state-of-the-art supervised deep learning-based segmentation approach in which data augmentation is extensively used and the hyperparameters are automatically optimized. It has never been evaluated for MR-Eye, but with OCT (73). We split the manual annotated dataset into 31 for training and 43 for testing. Default nnU-Net hyperparameters used: initial learning rate 0.001 with ReduceLROnPlateau scheduler, batch size 2; ADAM optimizer; deep supervision with cross entropy plus dice loss function; data augmentation such as scaling, rotation; patch size [128, 160, 112]; Kaiming-He (0.01) weights initialization; five folds cross validation ; no postprocessing after inference; stop condition 1000 epochs, with an elapsed time of around 140s to 170s per epoch; number of classes 10 (9 ROIs plus background); GPU RTX2080 and RTX3090 (the first available in the cluster), 10 CPUs per fold, RAM 64GB, ran in HPC (High Power Computing) SLURM-based cluster, through Docker accessed by Singularity; PyTorch, Python 3.8. The total training time for the five folds was around 208h 20m. The inference process for one image takes about 1 minute and for the whole non-labeled dataset (1157 subjects), 66185.53s (18h 23m 05s) with GPU RTX 3060Ti.

Evaluation: segmentation similarity metrics

To adequately assess the performance of the segmentation method, we computed similarity and error metrics between the ground truth (manual segmentations) and the method's outputs on the right eye. Based on (74), appropriate metrics to evaluate semantic segmentation of biomedical images are:

- Dice Similarity Coefficient (DSC): it is defined as twice the number of elements common to both sets divided by the sum of the number of elements in each set. The DSC ranges between 0 (indicating no overlap) and 1 (indicating perfect overlap). It is negatively biased by small structures. $DSC = \frac{2|A \cap \hat{A}|}{|A|+|\hat{A}|}$.
- Hausdorff Distance (HD): it measures how far two subsets of a metric space are from each other. It is the greatest of all the distances from a point in one set to the closest point in the other set. It does have units, which are the same as the units of the coordinate space in which the points are defined, mm in our case. The HD can range from 0 to infinity (no overlap between the objects). In Figure 2, this is limited to [0, 3]. $d_H(X,Y) = \max\{\sup d(x \in X, Y), \sup d(X, y \in Y)\}$.
- Volume Difference (VD): it refers to the difference in the amount of three-dimensional space occupied by two objects. The VD can range from -2 (if the second volume is larger) to +2 (if the first volume is larger). In our case, the first volume is the ground truth (manual segmentation) and the second is atlas-based or nnU-Net segmentation volumes. Hence, having a positive VD means that the manual volume is larger than the corresponding method one, and a negative VD means that the method volume is larger than the manual. $VD = \frac{2*(v1-v2)}{v1+v2}$.

Biomarkers extraction

Metadata

We extracted metadata (sex, age, height, weight) from the original DICOM files and computed BMI (kg/m^2) per subject.

Axial length

We developed an algorithm to automatically extract the AL (as defined in (20) and illustrated in Figure 3B). The method inputs both the automated segmented labels and T1w images. First, we computed the centroid of the lens and an orthogonal line to the globe in the 2D axial view, measuring the distance between the extreme intersection points of this line with the lens and the globe. Additionally, we calculated two extra distances: from the globe to the intraconal fat boundary, and from the lens boundary to the cornea. To address the lack of cornea segmentation, we determined the second distance by detecting low-intensity voxels (as the cornea appears black in T1w images) and identifying a significant intensity increase with the help of Sobel filter. This process was applied to multiple slices from the same subject where required segmented structures to compute the measurement were present (lens, globe, optic nerve and intraconal fat). If that was not the case, the AL would be saved as 0 mm. The final AL for a subject was determined by selecting the slice with the highest lens-to-optic nerve ratio (and whose AL was different from 0), indicating the best alignment of these structures to compute the measurement.

Volumetry

The volumetry of the different segmented eye structures in mm^3 was estimated based on the number of voxels per structure, each voxel of 1 mm^3.

Correlation between volumetry and BMI

We fit the volumes and BMIs per structure through a Huber regressor, a linear model robust to outliers. We used scikit-learn library (version 1.1.2). We obtained the slope, the intercept, and the R^2 score.

Atlas of the eye

We performed metric-based registration, consisting of rigid, affine, and then deformable registration, with ANTs toolkit (75) to iteratively create an average mapping of the subjects grouped by sex (594 males and 616 females). We made use of the multivariate template construction tool, using as input images the right-eye-cropped ones obtained from the atlas-based segmentation method (in Supplementary Materials). Therefore, they were much smaller than the initial ones (that included the whole head). The maximum size of these right-eye-cropped images for the three axes were 61 x 70 x 68 and 77 x 95 x 94 voxels for the male and female case, respectively, and the size of the original images was 176 x 256 x 176 voxels. The size of the voxels remained 1mm^3. For the deformable registration, we chose the SyN registration algorithm with the similarity metric of cross-correlation. We chose four resolution levels (8, 4, 2, 1), and iterated over each level for 80, 60, 40, and 10 iterations, respectively. Considering the reduced size of the images, we set the iteration limit (the number of iterations of the template construction) to 15, as we wanted to allow enough iterations for the template to converge and capture the variations present in our dataset. We used a 11th Gen Intel® Core™ i9-11900K × 16 processor with 64GB of RAM. The time spent to construct both atlases were 16h 15m 45s and 32h 16m 45s for the male and female cases, respectively.

To generate the labels on both eye atlases, we first registered them with each subject of its respective group (male or female), and project the labels obtained by the segmentation method, nnU-Net, of each subject to the atlas' space. The whole process lasted for 25m and 39m for males and females, respectively. We then created the maximum probability map of the labels for both atlases based on majority voting. We also generated the probability maps of the labels for both atlases by adjusting the intensity of the color of each voxel per label based on its probability to belong to each one of the classes. More precisely, we assigned an RGB color to every label, converted them to HSV, multiplied the S (saturation) and V (value) components of the color space by the probability per label, reconverted to RGB for visualization, and blended the resulting RGB values for the different labels. This way, low-probability voxels (per label) will appear greyish, showing the uncertainty of those voxels belonging to a single class. The male and female atlases can be downloaded at (76).

Quality Control Protocol

Figure 8 shows a block diagram of this quality control process throughout the pipeline. We passed QC checks at different points of the pipeline (described below) to capture possible excluded-quality subjects, and then manually review those cases, using the previously mentioned reports, to ensure which of them were really excluded. The exclusion criteria for our application are two-fold: first, the quality of the image must be acceptable in terms of noise, blur, motion, and not include heavy artifacts on the area of evaluation (the eyes); and second, all structures intended for segmentation must be visible (i.e. if an image presents no visible lens, it would be removed). We didn't follow further inclusion/exclusion criteria presented in (20), such as including only the images in which the corneal apex and the head of the optic nerve were in the same axial plane, or excluding images where there was a lateral deviation of the subject's viewing direction. Their application (20) was focused on imaging analysis (AL and exophthalmos) whereas ours was mostly focused on image segmentation (followed by imaging analysis).

The QA/QC checks we performed were:

1. Before image segmentation: we ran MRIQC (49), to extract no-reference IQMs, and MRIQC classifier, trained and tested on ABIDE and DS030 datasets, respectively, with updated scikit-learn and NumPy python libraries, to extract candidates as possible excluded-quality images. From 1210 subjects (the first batch of 35 manually annotated subjects was not included in the QA/QC protocols, as they had included quality to be manually segmented in the first place), 29 were selected by the classifier as excluded, and, after manual revision, 10 were really excluded regarding our criteria.
2. After segmentation: we computed the already mentioned similarity metrics but this time between the results of the nnU-Net and the baseline (atlas-based) methods, to then extract the outliers using interquartile approach, as the sets do not follow a normal distribution. The values below and above the lower (Q1-1.5*IQR) and upper (Q3+1.5*IQR) bounds, respectively, were selected as outliers. In total we had 102 outliers, which we manually reviewed, and excluded 20 of them, regarding our criteria.
3. After biomarkers extraction: we extracted the outliers following the same method as before in both AL and volumetry cases. From AL, there were 45 and 150 outliers for atlas-based and nnU-Net methods, respectively, some of them shared between the two. After manual revision, 21 were excluded in total. From volumetry, 25 and 53 subjects popped up as outliers for atlas-based and nnU-Net methods, respectively. Again, some of them were shared between the two. After manual revision, only 2 subjects in total were excluded. In total, in this third step, we removed 23 subjects.

In total, 53/1210 subjects (4.38%) were excluded, having a total of 1157 non-excluded quality subjects remaining.

References

1. Bourne, R. R. A., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Resnikoff, S., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y., Taylor, H. R., Bourne, R., … Zheng, Y. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, *5*(9), e888–e897. https://doi.org/10.1016/S2214-109X(17)30293-0.
2. London, A., Benhar, I., & Schwartz, M. (2013). The retina as a window to the brain—from eye research to CNS disorders. *Nature Reviews Neurology*, *9*(1), 44–53. https://doi.org/10.1038/nrneurol.2012.227.
3. Guthoff, R. F., Labriola, L. T., & Stachs, O. (2013). Diagnostic Ophthalmic Ultrasound. In *Ryan's Retinal Imaging and Diagnostics* (pp. e228–e285). Elsevier. https://doi.org/10.1016/B978-0-323-26254-5.00009-0.
4. Panwar, N., Huang, P., Lee, J., Keane, P. A., Chuan, T. S., Richhariya, A., Teoh, S., Lim, T. H., & Agrawal, R. (2016). Fundus Photography in the 21st Century—A Review of Recent Technological Advances and Their

Implications for Worldwide Healthcare. Telemedicine and E-Health, 22(3), 198–208. https://doi.org/10.1089/tmj.2015.0068.

5. Fujimoto, J. G., Pitris, C., Boppart, S. A., & Brezinski, M. E. (2000). Optical Coherence Tomography: An Emerging Technology for Biomedical Imaging and Optical Biopsy. Neoplasia, 2(1), 9–25. https://doi.org/10.1038/sj.neo.7900071.
6. Meyer, C. H., Saxena, S. & Sadda, S. R. *Spectral Domain Optical Coherence Tomography in Macular Diseases*. (Springer India, 2017). doi:10.1007/978-81-322-3610-8.
7. Townsend, K. A., Wollstein, G., & Schuman, J. S. (2008). Clinical application of MRI in ophthalmology. *NMR in Biomedicine*, *21*(9), 997–1002. https://doi.org/10.1002/nbm.1247.
8. Fanea, L., & Fagan, A. J. (2012). Review: Magnetic resonance imaging techniques in ophthalmology. *Molecular Vision*, *18*, 2538–2560. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3482169/
9. Duong, T. Q. (2014). Magnetic resonance imaging of the retina: From mice to men. *Magnetic Resonance in Medicine*, *71*(4), 1526–1530. https://doi.org/10.1002/mrm.24797.
10. Niendorf, T., Beenakker, J.-W. M., Langner, S., Erb-Eigner, K., Bach Cuadra, M., Beller, E., Millward, J. M., Niendorf, T. M., & Stachs, O. (2021). Ophthalmic Magnetic Resonance Imaging: Where Are We (Heading To)? Current Eye Research, 46(9), 1251–1270. https://doi.org/10.1080/02713683.2021.1874021.
11. Georgouli, T., James, T., Tanner, S., Shelley, D., Nelson, M., Chang, B., Backhouse, O., & McGonagle, D. (2008). High-Resolution Microscopy Coil MR-Eye. *Eye*, *22*(8), 994–996. https://doi.org/10.1038/sj.eye.6702755
12. Tsiapa, I., Tsilimbaris, M. K., Papadaki, E., Bouziotis, P., Pallikaris, I. G., Karantanas, A. H., & Maris, T. G. (2015). High resolution MR eye protocol optimization: Comparison between 3D-CISS, 3D-PSIF and 3D-VIBE sequences. *Physica Medica*, *31*(7), 774–780. https://doi.org/10.1016/j.ejmp.2015.03.009
13. Dobbs, N. W., Budak, M. J., White, R. D., & Zealley, I. A. (2020). MR-Eye: High-Resolution Microscopy Coil MRI for the Assessment of the Orbit and Periorbital Structures, Part 1: Technique and Anatomy. *American Journal of Neuroradiology*, *41*(6), 947–950. https://doi.org/10.3174/ajnr.A6495
14. Fleury E, Trnkova P, Erdal E, Hassan M, Stoel B, Jaarma-Coes M, Luyten G, Herault J, Webb A, Beenakker JW, et al. 3d mri-based treatment planning approach for non-invasive ocular proton therapy. Med Phys. 2020. doi:10.1002/mp.14665.
15. Glarin, R. K., Nguyen, B. N., Cleary, J. O., Kolbe, S. C., Ordidge, R. J., Bui, B. V., McKendrick, A. M., & Moffat, B. A. (2021). MR-EYE: High-Resolution MRI of the Human Eye and Orbit at Ultrahigh Field (7T). *Magnetic Resonance Imaging Clinics of North America*, *29*(1), 103–116. https://doi.org/10.1016/j.mric.2020.09.004.
16. Armstrong, R., & Kergoat, H. (2015). Oculo-visual changes and clinical considerations affecting older patients with dementia. *Ophthalmic and Physiological Optics*, *35*(4), 352–376. https://doi.org/10.1111/opo.12220.
17. Hart, N. J., Koronyo, Y., Black, K. L., & Koronyo-Hamaoui, M. (2016). Ocular indicators of Alzheimer's: exploring disease in the retina. *Acta Neuropathologica*, *132*(6), 767–787. https://doi.org/10.1007/s00401-016-1613-6.
18. Pula, J. H., & Yuen, C. A. (2017). Eyes and stroke: the visual aspects of cerebrovascular disease. *Stroke and Vascular Neurology*, *2*(4). https://doi.org/10.1136/svn-2017-000079.
19. Hunt, A. W., Mah, K., Reed, N., Engel, L., & Keightley, M. (2016). Oculomotor-Based Vision Assessment in Mild Traumatic Brain Injury: A Systematic Review. *The Journal of Head Trauma Rehabilitation*, *31*(4), 252. https://doi.org/10.1097/HTR.0000000000000174.
20. Schmidt, P., Kempin, R., Langner, S., Beule, A., Kindler, S., Koppe, T., Völzke, H., Ittermann, T., Jürgens, C., & Tost, F. (2019). Association of anthropometric markers with globe position: A population-based MRI study. PLoS ONE, 14(2), e0211817. https://doi.org/10.1371/journal.pone.0211817
21. Wiseman, S. J., Tatham, A. J., Meijboom, R., Terrera, G. M., Hamid, C., Doubal, F. N., Wardlaw, J. M., Ritchie, C., Dhillon, B., & MacGillivray, T. (2022). Measuring AL of the eye from magnetic resonance brain imaging. *BMC Ophthalmology*, *22*(1), 54. https://doi.org/10.1186/s12886-022-02289-y
22. Bhardwaj, V., & Rajeshbhai, G. P. (2013). Axial Length, Anterior Chamber Depth-A Study in Different Age Groups and Refractive Errors. *Journal of Clinical and Diagnostic Research : JCDR*, *7*(10), 2211–2212. https://doi.org/10.7860/JCDR/2013/7015.3473
23. Sentucq, C., Schlund, M., Bouet, B., Garms, M., Ferri, J., Jacques, T., & Nicot, R. (2021). Overview of tools for the measurement of the orbital volume and their applications to orbital surgery. Journal of Plastic, Reconstructive & Aesthetic Surgery, 74(3), 581–591. https://doi.org/10.1016/j.bjps.2020.08.101.
24. Senarak, W., Yongvikul, A., Ku, J.-K., Kim, J.-Y., & Huh, J.-K. (2023). Effect of orbital volume in unilateral orbital fracture on indirect traumatic optic neuropathy. International Ophthalmology, 43(4), 1121–1126. https://doi.org/10.1007/s10792-022-02509-w.
25. Steiert, C., Kuechlin, S., Masalha, W., Beck, J., Lagrèze, W. A., & Grauvogel, J. (2022). Increased Orbital Muscle Fraction Diagnosed by Semi-Automatic Volumetry: A Risk Factor for Severe Visual Impairment with Excellent Response to Surgical Decompression in Graves' Orbitopathy. Journal of Personalized Medicine, 12(6), 937. https://doi.org/10.3390/jpm12060937.

26. Dickie, D. A., Shenkin, S. D., Anblagan, D., Lee, J., Blesa Cabez, M., Rodriguez, D., Boardman, J. P., Waldman, A., Job, D. E., & Wardlaw, J. M. (2017). Whole Brain Magnetic Resonance Image Atlases: A Systematic Review of Existing Atlases and Caveats for Use in Population Imaging. *Frontiers in Neuroinformatics*, *11*. https://doi.org/10.3389/fninf.2017.00001
27. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., & Bach Cuadra, M. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, *104*(3), e158–e177. https://doi.org/10.1016/j.cmpb.2011.07.015
28. Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5
29. Lee, H. H., Saunders, A. M., Kim, M. E., Remedios, S. W., Remedios, L. W., Tang, Y., Yang, Q., Yu, X., Bao, S., Cho, C., Mawn, L. A., Rex, T. S., Schey, K. L., Dewey, B. E., Spraggins, J. M., Prince, J. L., Huo, Y., & Landman, B. A. (2024). *Super-resolution multi-contrast unbiased eye atlases with deep probabilistic refinement* (No. arXiv:2401.03060). arXiv. http://arxiv.org/abs/2401.03060.
30. Jain, S., Pei, L., Spraggins, J. M., Angelo, M., Carson, J. P., Gehlenborg, N., Ginty, F., Gonçalves, J. P., Hagood, J. S., Hickey, J. W., Kelleher, N. L., Laurent, L. C., Lin, S., Lin, Y., Liu, H., Naba, A., Nakayasu, E. S., Qian, W.-J., Radtke, A., … Snyder, M. P. (2023). Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). *Nature Cell Biology*, *25*(8), 1089–1100. https://doi.org/10.1038/s41556-023-01194-w.
31. Hierl, K. V., Krause, M., Kruber, D., & Sterker, I. (2022). 3-D cephalometry of the the orbit regarding endocrine orbitopathy, exophthalmos, and sex. *PLOS ONE*, *17*(3), e0265324. https://doi.org/10.1371/journal.pone.0265324.
32. Patra, A., Singla, R. K., Mathur, M., Chaudhary, P., Singal, A., Asghar, A., & Malhotra, V. (2021). Morphological and Morphometric Analysis of the Orbital Aperture and Their Correlation With Age and Gender: A Retrospective Digital Radiographic Study. *Cureus*. https://doi.org/10.7759/cureus.17739.
33. Klinge, I., & Wiesemann, C. (Eds.). (2010). *Sex and gender in biomedicine: theories, methodologies, results*. Göttingen University Press. https://doi.org/10.17875/gup2010-394
34. Zetterberg, M. (2016). Age-related eye disease and gender. *Maturitas*, *83*, 19–26. https://doi.org/10.1016/j.maturitas.2015.10.005
35. Goitein, M., & Miller, T. (1983). Planning proton therapy of the eye. Medical Physics, 10(3), 275–283. https://doi.org/10.1118/1.595258.
36. Dobler, B., & Bendl, R. (2002). Precise modelling of the eye for proton therapy of intra-ocular tumours. *Physics in Medicine and Biology*, *47*(4), 593–613. https://doi.org/10.1088/0031-9155/47/4/304.
37. Singh, K. D., Logan, N. S., & Gilmartin, B. (2006). Three-dimensional modeling of the human eye based on magnetic resonance imaging. *Investigative Ophthalmology & Visual Science*, *47*(6), 2272–2279. https://doi.org/10.1167/iovs.05-0856.
38. Ciller, C., De Zanet, S. I., Rüegsegger, M. B., Pica, A., Sznitman, R., Thiran, J.-P., Maeder, P., Munier, F. L., Kowal, J. H., & Cuadra, M. B. (2015). Automatic Segmentation of the Eye in 3D Magnetic Resonance Imaging: A Novel Statistical Shape Model for Treatment Planning of Retinoblastoma. International Journal of Radiation Oncology*Biology*Physics, 92(4), 794–802. https://doi.org/10.1016/j.ijrobp.2015.02.056.
39. Nguyen, H.-G., Sznitman, R., Maeder, P., Schalenbourg, A., Peroni, M., Hrbacek, J., Weber, D. C., Pica, A., & Bach Cuadra, M. (2018). Personalized Anatomic Eye Model from T1-Weighted Volume Interpolated Gradient Echo Magnetic Resonance Imaging of Patients with Uveal Melanoma. International Journal of Radiation Oncology*Biology*Physics, 102(4), 813–820. https://doi.org/10.1016/j.ijrobp.2018.05.004.
40. Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation* (arXiv:1505.04597). arXiv. https://doi.org/10.48550/arXiv.1505.04597.
41. Nguyen, H.-G., Pica, A., Maeder, P., Schalenbourg, A., Peroni, M., Hrbacek, J., Weber, D. C., Cuadra, M. B., & Sznitman, R. (2018). Ocular Structures Segmentation from Multi-sequences MRI Using 3D U-Net with Fully Connected CRFs. In D. Stoyanov, Z. Taylor, F. Ciompi, Y. Xu, A. Martel, L. Maier-Hein, N. Rajpoot, J. van der Laak, M. Veta, S. McKenna, D. Snead, E. Trucco, M. K. Garvin, X. J. Chen, & H. Bogunovic (Eds.), Computational Pathology and Ophthalmic Medical Image Analysis (pp. 167–175). Springer International Publishing. https://doi.org/10.1007/978-3-030-00949-6_20.
42. Strijbis, V. I. J., De Bloeme, C. M., Jansen, R. W., Kebiri, H., Nguyen, H.-G., De Jong, M. C., Moll, A. C., Bach-Cuadra, M., De Graaf, P., & Steenwijk, M. D. (2021). Multi-view convolutional neural networks for automated ocular structure and tumor segmentation in retinoblastoma. Scientific Reports, 11(1), 14590. https://doi.org/10.1038/s41598-021-93905-2.
43. Ciller, C., De Zanet, S., Kamnitsas, K., Maeder, P., Glocker, B., Munier, F. L., Rueckert, D., Thiran, J.-P., Bach Cuadra, M., & Sznitman, R. (2017). Multi-channel MRI segmentation of eye structures and tumors using patient-specific features. PLOS ONE, 12(3), e0173900. https://doi.org/10.1371/journal.pone.0173900.

44. Nguyen, H.-G., Pica, A., Rosa, F. L., Hrbacek, J., Weber, D. C., Schalenbourg, A., Sznitman, R., & Cuadra, M. B. (2019). A novel segmentation framework for uveal melanoma based on magnetic resonance imaging and class activation maps. https://doi.org/10.7892/BORIS.135253.
45. Hassan, M. K., Fleury, E., Shamonin, D., Fonk, L. G., Marinkovic, M., Jaarsma-Coes, M. G., Luyten, G. P. M., Webb, A., Beenakker, J.-W., & Stoel, B. (2021). An Automatic Framework to Create Patient-specific Eye Models From 3D Magnetic Resonance Images for Treatment Selection in Patients With Uveal Melanoma. *Advances in Radiation Oncology*, *6*(6), 100697. https://doi.org/10.1016/j.adro.2021.100697.
46. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018
47. Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage*, *107*, 107–115. https://doi.org/10.1016/j.neuroimage.2014.12.006
48. Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., & Raznahan, A. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI: Motion Bias in Analyses of Structural MRI. *Human Brain Mapping*, *37*(7), 2385–2397. https://doi.org/10.1002/hbm.23180
49. Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661.
50. Provins, C., MacNicol, E., Seeley, S. H., Hagmann, P., & Esteban, O. (2023). Quality control in functional MRI studies with MRIQC and fMRIPrep. *Frontiers in Neuroimaging*, *1*. https://www.frontiersin.org/articles/10.3389/fnimg.2022.1073734
51. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z.
52. Sheng, H., Bottjer, C. A., & Bullimore, M. A. (2004). Ocular component measurement using the Zeiss IOLMaster. *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, *81*(1), 27–34. https://doi.org/10.1097/00006324-200401000-00007
53. Midena, E. (Ed.). (2014). *Microperimetry and Multimodal Retinal Imaging*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40300-2
54. Al Othman, B., Raabe, J., Kini, A., & Lee, A. G. (2020). Neuroradiology for ophthalmologists. *Eye (London, England)*, *34*(6), 1027–1038. https://doi.org/10.1038/s41433-019-0753-z
55. de Jong, M. C., de Graaf, P., Brisse, H. J., Galluzzi, P., Göricke, S. L., Moll, A. C., Munier, F. L., Popovic, M. B., Moulin, A. P., Binaghi, S., Castelijns, J. A., Maeder, P., & European Retinoblastoma Imaging Collaboration (ERIC). (2015). The potential of 3T high-resolution magnetic resonance imaging for diagnosis, staging, and follow-up of retinoblastoma. *Survey of Ophthalmology*, *60*(4), 346–355. https://doi.org/10.1016/j.survophthal.2015.01.002
56. Graaf, P. de, Göricke, S., Rodjan, F., Galluzzi, P., Maeder, P., Castelijns, J. A., Brisse, H. J., & Collaboration (ERIC), on behalf of the E. R. I. (2012). Guidelines for imaging retinoblastoma: imaging principles and MRI standardization. *Pediatric Radiology*, *42*(1), 2. https://doi.org/10.1007/s00247-011-2201-5
57. Ferreira, T. A., Grech Fonk, L., Jaarsma-Coes, M. G., van Haren, G. G. R., Marinkovic, M., & Beenakker, J.-W. M. (2019). MRI of Uveal Melanoma. *Cancers*, *11*(3), 377. https://doi.org/10.3390/cancers11030377
58. Jaarsma-Coes, M. G., Goncalves Ferreira, T. A., van Haren, G. R., Marinkovic, M., & Beenakker, J.-W. M. (2019). MRI enables accurate diagnosis and follow-up in uveal melanoma patients after vitrectomy. *Melanoma Research*, *29*(6), 655–659. https://doi.org/10.1097/CMR.0000000000000568
59. Mafee, M. F., Karimi, A., Shah, J., Rapoport, M., & Ansari, S. A. (2005). Anatomy and pathology of the eye: role of MR imaging and CT. *Neuroimaging Clinics of North America*, *15*(1), 23–47. https://doi.org/10.1016/j.nic.2005.02.005
60. Demer, J. L., Clark, R. A., Kono, R., Wright, W., Velez, F., & Rosenbaum, A. L. (2002). A 12-year, prospective study of extraocular muscle imaging in complex strabismus. *Journal of AAPOS: The Official Publication of the American Association for Pediatric Ophthalmology and Strabismus*, *6*(6), 337–347. https://doi.org/10.1067/mpa.2002.129040
61. Piccirelli, M., Luechinger, R., Rutz, A. K., Boesiger, P., & Bergamin, O. (2007). Extraocular muscle deformation assessed by motion-encoded MRI during eye movement in healthy subjects. *Journal of Vision*, *7*(14), 5. https://doi.org/10.1167/7.14.5
62. Clark, R. A., & Demer, J. L. (2006). Magnetic Resonance Imaging of the Effects of Horizontal Rectus Extraocular Muscle Surgery on Pulley and Globe Positions and Stability. *Investigative Ophthalmology & Visual Science*, *47*(1), 188–194. https://doi.org/10.1167/iovs.05-0498
63. Sengupta, S., Smith, D. S., Smith, A. K., Welch, E. B., & Smith, S. A. (2017). Dynamic Imaging of the Eye, Optic Nerve, and Extraocular Muscles With Golden Angle Radial MRI. *Investigative Ophthalmology & Visual Science*, *58*(10), 4010. https://doi.org/10.1167/iovs.17-21861

64. Franceschiello, B., Di Sopra, L., Minier, A., Ionta, S., Zeugin, D., Notter, M. P., Bastiaansen, J. A. M., Jorge, J., Yerly, J., Stuber, M., & Murray, M. M. (2020). 3-Dimensional magnetic resonance imaging of the freely moving human eye. *Progress in Neurobiology*, *194*, 101885. https://doi.org/10.1016/j.pneurobio.2020.101885
65. Nguyen, B. N., Cleary, J. O., Glarin, R., Kolbe, S. C., Moffat, B. A., Ordidge, R. J., Bui, B. V., & McKendrick, A. M. (2021). Ultra-High Field Magnetic Resonance Imaging of the Retrobulbar Optic Nerve, Subarachnoid Space, and Optic Nerve Sheath in Emmetropic and Myopic Eyes. *Translational Vision Science & Technology*, *10*(2), 8. https://doi.org/10.1167/tvst.10.2.8
66. Lambert, B., Forbes, F., Doyle, S., Dehaene, H., & Dojat, M. (2024). Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis. *Artificial Intelligence in Medicine*, *150*, 102830. https://doi.org/10.1016/j.artmed.2024.102830
67. Ortube, M. C., Rosenbaum, A. L., Goldberg, R. A., & Demer, J. L. (2004). Orbital imaging demonstrates occult blow out fracture in complex strabismus. *Journal of AAPOS: The Official Publication of the American Association for Pediatric Ophthalmology and Strabismus*, *8*(3), 264–273. https://doi.org/10.1016/j.jaapos.2004.01.011
68. Demer, J. L., Clark, R. A., Kono, R., Wright, W., Velez, F., & Rosenbaum, A. L. (2002). A 12-year, prospective study of extraocular muscle imaging in complex strabismus. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *6*(6), 337–347. https://doi.org/10.1067/mpa.2002.129040
69. Jaganathan, S., Baker, A., Ram, A., Krishnan, V., Elhusseiny, A. M., Philips, P. H., Glasier, C. M., Jayappa, S., Choudhary, A., & Ramakrishnaiah, R. (2024). Collapse or distention of the perioptic space in children - What does it mean to pediatric radiologists? Comprehensive review of perioptic space evaluation. *Clinical Imaging*, *111*, 110150. https://doi.org/10.1016/j.clinimag.2024.110150
70. Sheng, J., Li, Q., Liu, T., & Wang, X. (2022). Cerebrospinal fluid dynamics along the optic nerve. *Frontiers in Neurology*, *13*, 931523. https://doi.org/10.3389/fneur.2022.931523
71. Van Leemput, K. (2009). Encoding Probabilistic Brain Atlases Using Bayesian Inference. *IEEE Transactions on Medical Imaging*, *28*(6), 822–837. https://doi.org/10.1109/TMI.2008.2010434
72. Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015. ITK-SNAP Home. (n.d.). Retrieved February 13, 2024, from http://www.itksnap.org/pmwiki/pmwiki.php?n=Main.HomePage.
73. Valmaggia, P., Friedli, P., Hörmann, B., Kaiser, P., Scholl, H. P. N., Cattin, P. C., Sandkühler, R., & Maloca, P. M. (2022). Feasibility of Automated Segmentation of Pigmented Choroidal Lesions in OCT Data With Deep Learning. *Translational Vision Science & Technology*, *11*(9), 25. https://doi.org/10.1167/tvst.11.9.25.
74. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M. D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M. A., Wiesenfarth, M., Kavur, A. E., Sudre, C. H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Rädsch, A. T., … Jäger, P. F. (2023). Metrics reloaded: Recommendations for image analysis validation (arXiv:2206.01653). arXiv. https://doi.org/10.48550/arXiv.2206.01653.
75. Avants, B., Tustison, N. J., & Song, G. (2009). Advanced Normalization Tools: V1.0. The Insight Journal. https://doi.org/10.54294/uvnhin.
76. Barranco Hernandez, J., Luyken, A., Stachs, P., Esteban, O., Aleman-Gomez, Y., Stachs, O., Sönke Langner, S., Franceschiello B., Bach Cuadra, M., (2024). MR-Eye atlas: a large-scale atlas of the eye based on T1-weighted MR imaging, Zenodo.org, DOI: 10.5281/zenodo.13325371

Acknowledgments

Funding:
Gelbert Foundation and the Swiss National Science Foundation grant 205321-182602. We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by CHUV, UNIL, EPFL, UNIGE and HUG. HK is supported by the Swiss National Science Foundation (grant number 215641).

Author contributions:

Conceptualization: MBC, BF, SL, OS
Dataset: SL, OS, PS, AL
Methodology—Atlas registration: JB, OE, YA, MBC
Methodology—Deep learning, nnU-Net: JB, HK, PMG, MBC
Methodology—QA/QC: JB, OE
Methodology—Biomarkers extraction: JB, BF
Methodology—Clinical relevance: JB, BF, SL, OS, AL

Methodology—Atlas of the eye and maps of the labels: JB, YA
Methodology—Statistics: JB, BF, YA
Supervision: MBC, BF, SL, OS
Writing—review & editing: JB, MBC, BF, SL, OS, AL, OE, HK, PMG

We use generative AI to create code segments based on task descriptions, as well as to debug, edit, and autocomplete code. Additionally, generative AI technologies have been employed to assist in structuring sentences and performing grammatical checks. The conceptualization, ideation, and all prompts provided to the AI originate entirely from the authors' creative and intellectual efforts. We take accountability for the review of all content generated by AI in this work.

Competing interests:
Authors declare that they have no competing interests.

Data and materials availability:
We do provide publicly available the male and female atlases (76). We cannot provide access to the original data due to DTA constraints. We provide materials in the Supplementary Materials document regarding results.

Figures and Tables

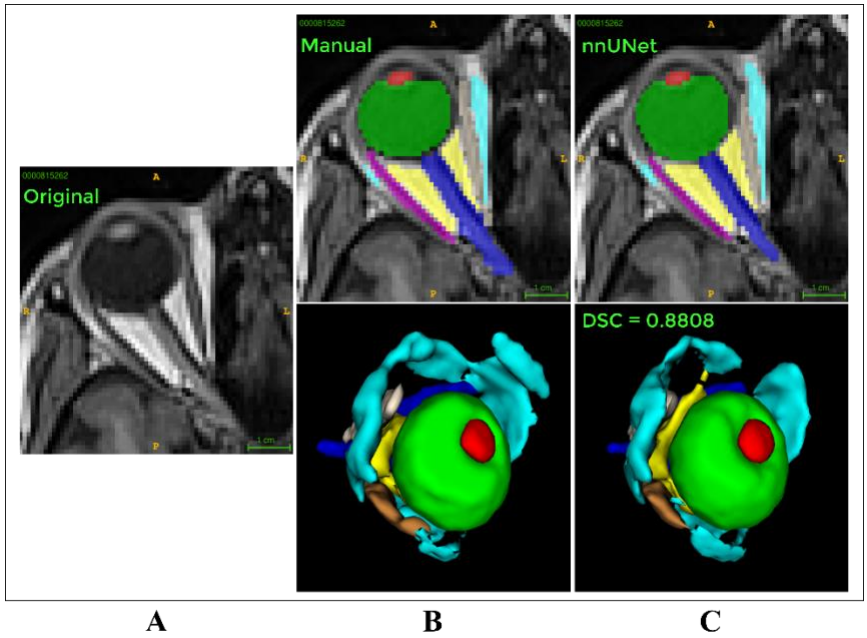

Figure 1. Visual comparison of manual and automated segmentation. ***(A)*** *Original T1w image.* ***(B)*** *Manual segmentation on 9 ROI: lens (red), globe (green), optic nerve (dark blue), intraconal fat (yellow), extraconal fat (cyan), lateral rectus muscle (magenta), medial rectus muscle (ivory), inferior rectus muscle (blue), and superior rectus muscle (brown).* ***(C)*** *nnU-Net segmentation. We provide preliminary overall DSC for nnU-Net compared to the manual segmentation (ground truth).*

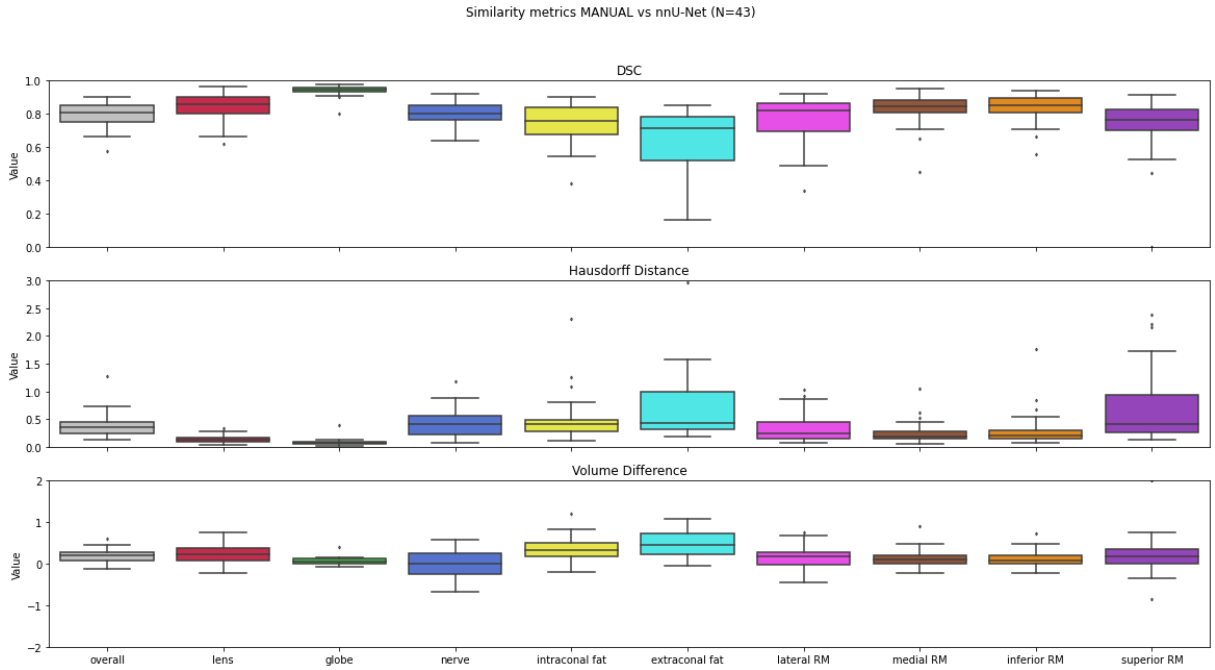

Figure 2. Similarity metrics on 43 subjects. *On the y-axis we have the similarity metric scale (three plots, from top to bottom DSC, HD, VD), and on the x-axis we have the different eye structures.*

Table 1. Similarity metrics' median and standard deviation values per structure on 43 subjects.

Structure	DSC	HD	VD
Average	0.81±0.07	0.35±0.20	0.19±0.14
Lens	0.86±0.08	0.13±0.07	0.20±0.23
Globe	0.94±0.03	0.07±0.05	0.06±0.08
Optic nerve	0.80±0.07	0.41±0.24	-0.006±0.28
Intraconal fat	0.76±0.11	0.40±0.37	0.32±0.27
Extraconal fat	0.73±0.17	0.44±0.82	0.42±0.33
Lateral RM	0.82±0.14	0.25±0.22	0.17±0.26
Medial RM	0.85±0.09	0.19±0.17	0.09±0.2
Inferior RM	0.85±0.08	0.21±0.28	0.07±0.19
Superior RM	0.77±0.10	0.40±0.59	0.17±0.32

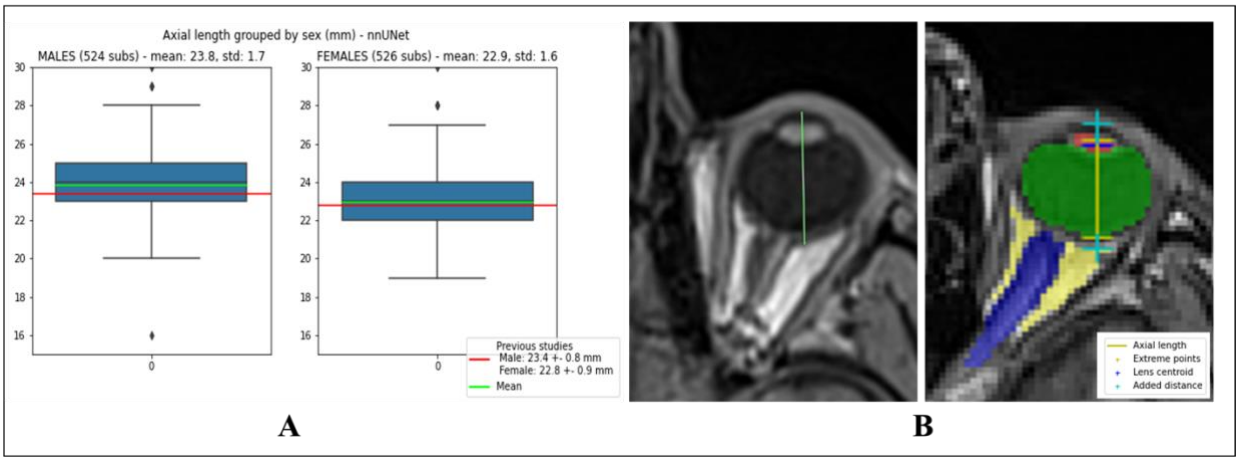

Figure 3. *Axial length (posterior surface of the cornea to the posterior pole of the ocular bulb, at the boundary with orbital fat grouped by sex. (A) Boxplots of the obtained AL per sex. (B) Example of AL extraction in MRI: on the left, manual extraction from (**Error! Reference source not found.**); on the right, automatic extraction using the segmented structures and the T1w image.*

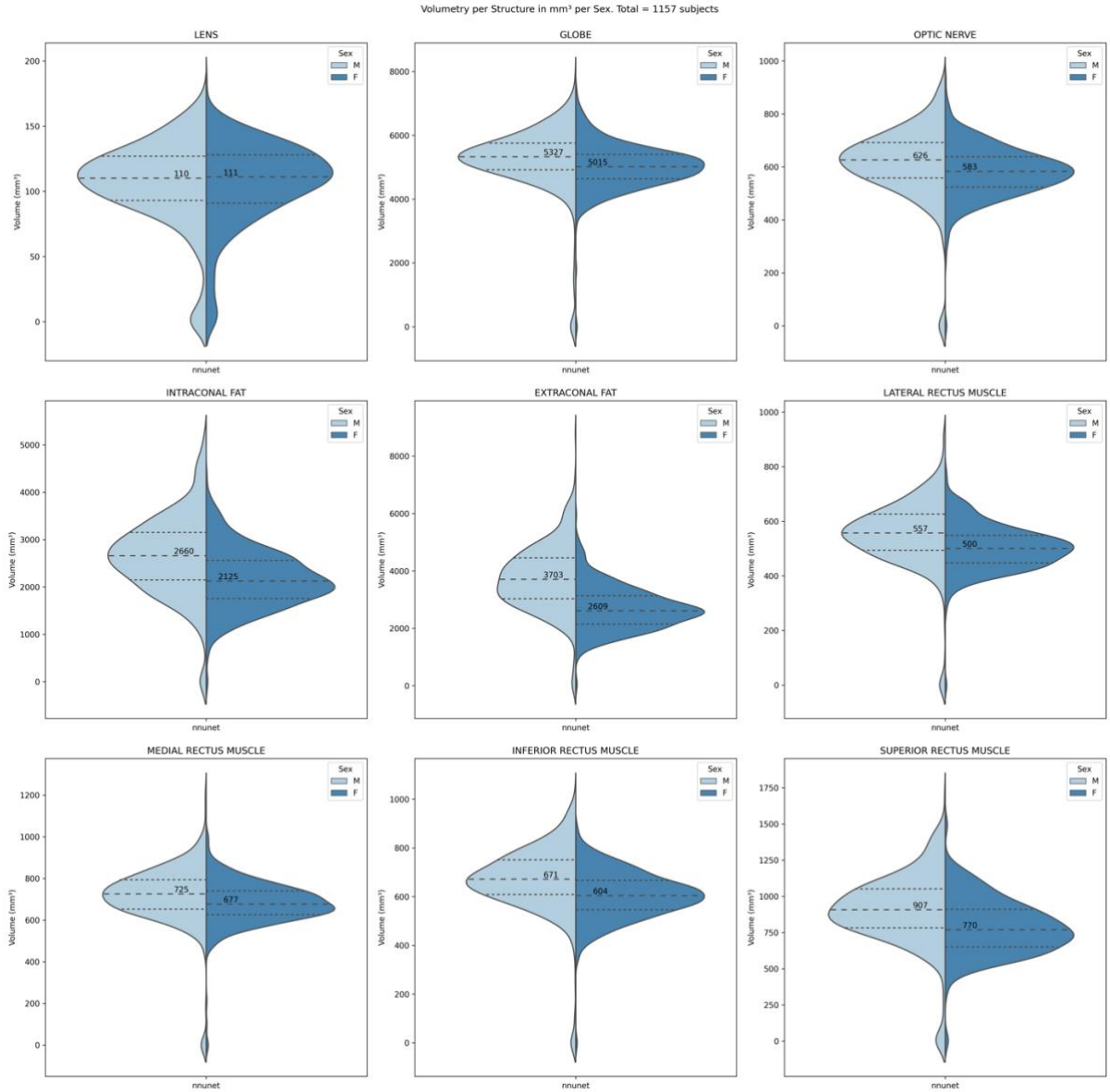

Figure 4. Volumetry per method for each eye structure per sex (568 males and 589 females). *Median values in mm^3 are provided on each plot.*

Table 2. Structures' volumes median and standard deviation grouped by sex.

Structure	Male volume (N=) (mm^3)	Female volume (mm^3)
Lens	110±33	111±32
Globe	5328±1085	5014±703
Optic nerve	626±136	583±95
Intraconal fat	2660±839	2120±610
Extraconal fat	3703±1134	2609±748
Lateral RM	557±124	500±82
Medial RM	726±143	677±104
Inferior RM	671±139	604±97
Superior RM	908±255	770±193

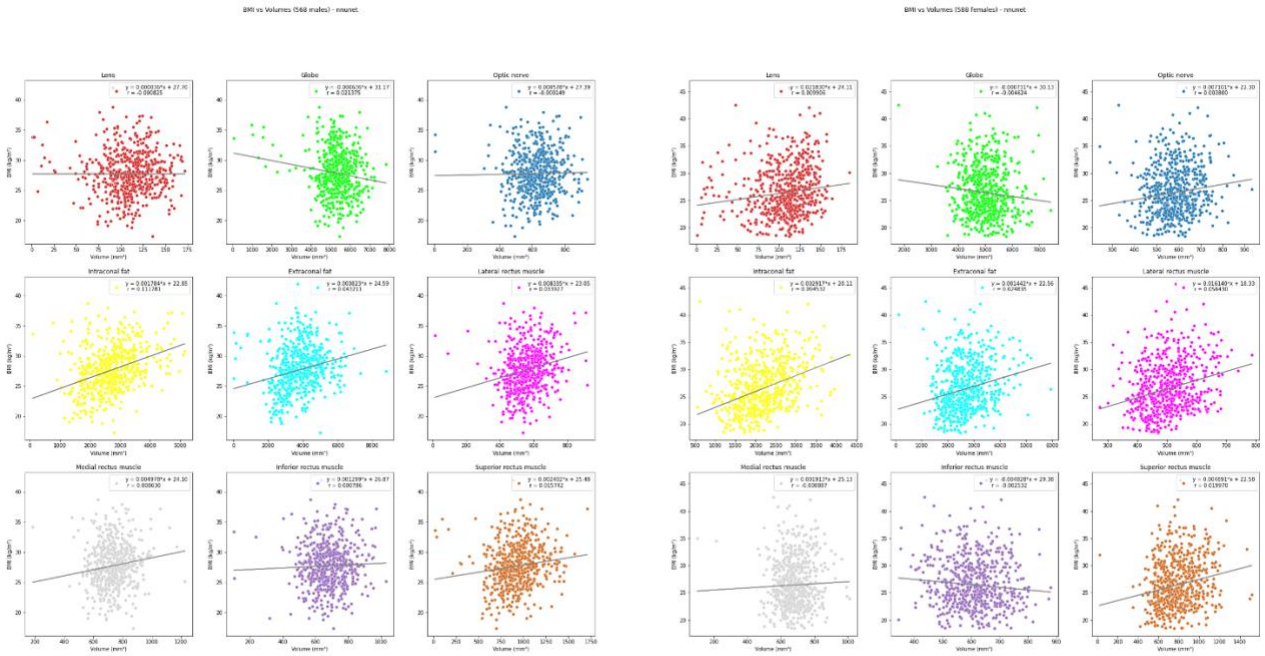

Figure 5. Correlation of volumetry per structure and BMI grouped by sex. *There is no existing correlation between BMI and volumetry per structure based on the Huber R^2 scores for any of both sex cases in any of the eye structures (the scores are lower than 0.3, indicating the lack of correlation).*

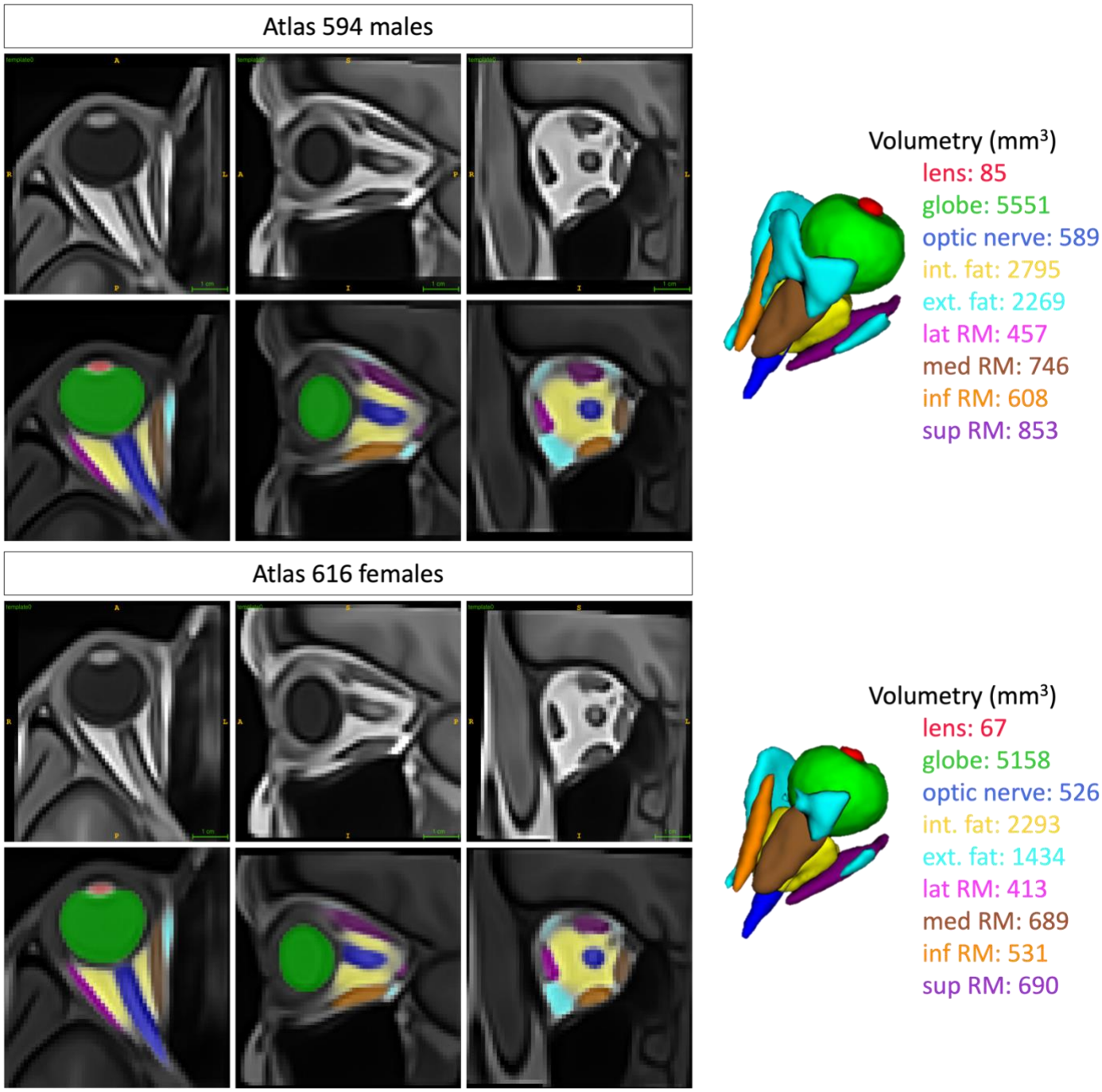

Figure 6. Male and female atlases of the eye. *At the top, the three views of the T1w atlas made of 594 males, below, the probability maps of the labels projected onto the atlas' space, and on the right, the 3D-rendered maximum probability maps of these labels. At the bottom, same for 616 females.*

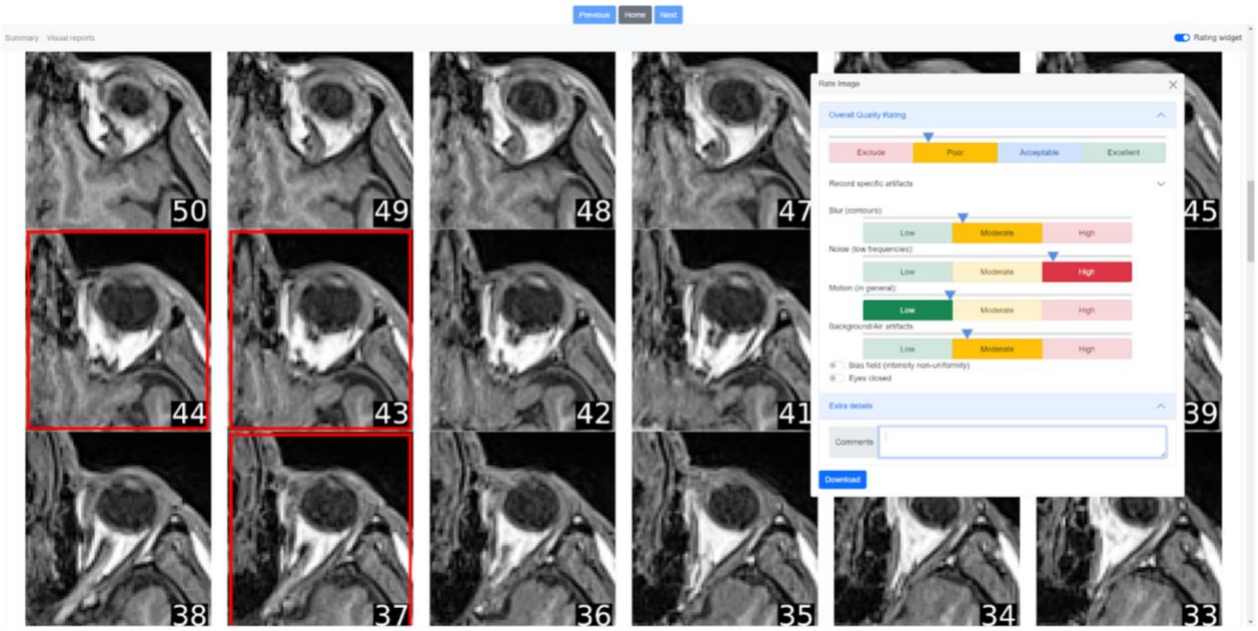

Figure 7. Example of MR-Eye QC report with rating widget. *To assess the quality of the eyes of the MR images, we created an html-based report for each of them: a series of axial slices centered and cropped on the right eye. The rating widget on the right is composed of several sliders regarding overall quality [0-4], blur, noise, motion, and background artifacts. Also, it includes two toggle buttons for bias field and eyes closed/open and a text box for further comments. Additionally, it's possible to select specific slices where heavy artifacts are present (red squares will appear).*

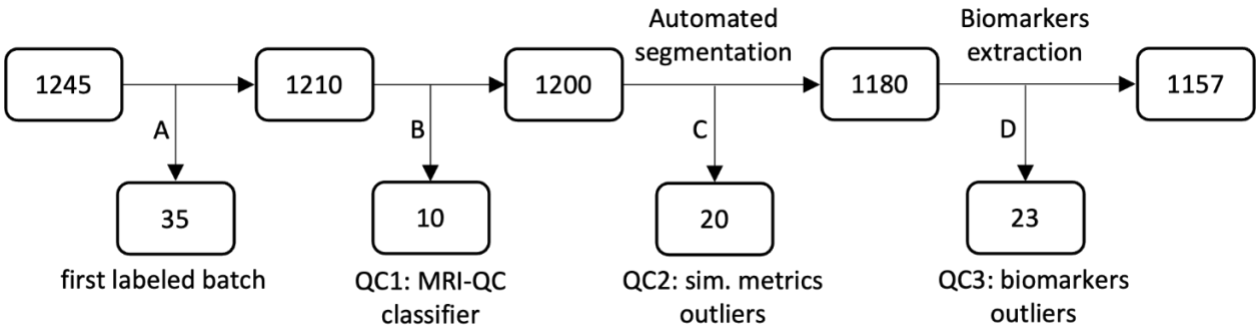

Figure 8. QA/QC integration within a simplified scheme of the A-Eye project's pipeline. ***(A)*** *The first batch of 35 manually annotated subjects are removed from the QC protocol as they all have included quality.* ***(B)*** *Subjects excluded from MRIQC classifier.* ***(C)*** *Subjects excluded from similarity metrics outliers between nnU-Net and the baseline (atlas-based) segmentation results.* ***(D)*** *Subjects excluded from biomarkers outliers (AL and volumetry). In total, 53 subjects were excluded because of its image quality for our application, with 1157 subjects remaining.*