



Clustering Disease Trajectories in Contrastive Feature Space for Biomarker Proposal in Age-Related Macular Degeneration

Robbie Holland¹(✉), Oliver Leingang², Christopher Holmes³, Philipp Anders⁴,
Rebecca Kaye⁶, Sophie Riedl², Johannes C. Paetzold¹, Ivan Ezhev⁷,
Hrvoje Bogunović², Ursula Schmidt-Erfurth², Hendrik P. N. Scholl^{4,5},
Sobha Sivaprasad³, Andrew J. Lotery⁶, Daniel Rueckert^{1,7},
and Martin J. Menten^{1,7}

¹ BioMedIA, Imperial College London, London, UK
robert.holland15@imperial.ac.uk

² Laboratory for Ophthalmic Image Analysis, Medical University of Vienna,
Vienna, Austria

³ Moorfields Eye Hospital NHS Foundation Trust, London, UK

⁴ Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland

⁵ Department of Ophthalmology, Universitat Basel, Basel, Switzerland

⁶ Clinical and Experimental Sciences, University of Southampton, Southampton, UK

⁷ Technical University of Munich, Munich, Germany

Abstract. Age-related macular degeneration (AMD) is the leading cause of blindness in the elderly. Current grading systems based on imaging biomarkers only coarsely group disease stages into broad categories that lack prognostic value for future disease progression. It is widely believed that this is due to their focus on a single point in time, disregarding the dynamic nature of the disease. In this work, we present the first method to automatically propose biomarkers that capture temporal dynamics of disease progression. Our method represents patient time series as trajectories in a latent feature space built with contrastive learning. Then, individual trajectories are partitioned into atomic subsequences that encode transitions between disease states. These are clustered using a newly introduced distance metric. In quantitative experiments we found our method yields temporal biomarkers that are predictive of conversion to late AMD. Furthermore, these clusters were highly interpretable to ophthalmologists who confirmed that many of the clusters represent dynamics that have previously been linked to the progression of AMD, even though they are currently not included in any clinical grading system.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_68.

Keywords: Contrastive learning · Biomarker discovery · Clustering · Disease trajectories · Age-related macular degeneration

1 Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in the elderly, affecting nearly 200 million people worldwide [24]. Patients with early stages of the disease exhibit few symptoms until suddenly converting to the late stage, at which point their central vision rapidly deteriorates [12]. Clinicians currently diagnose AMD, and stratify patients, using biomarkers derived from optical coherence tomography (OCT), which provides high-resolution images of

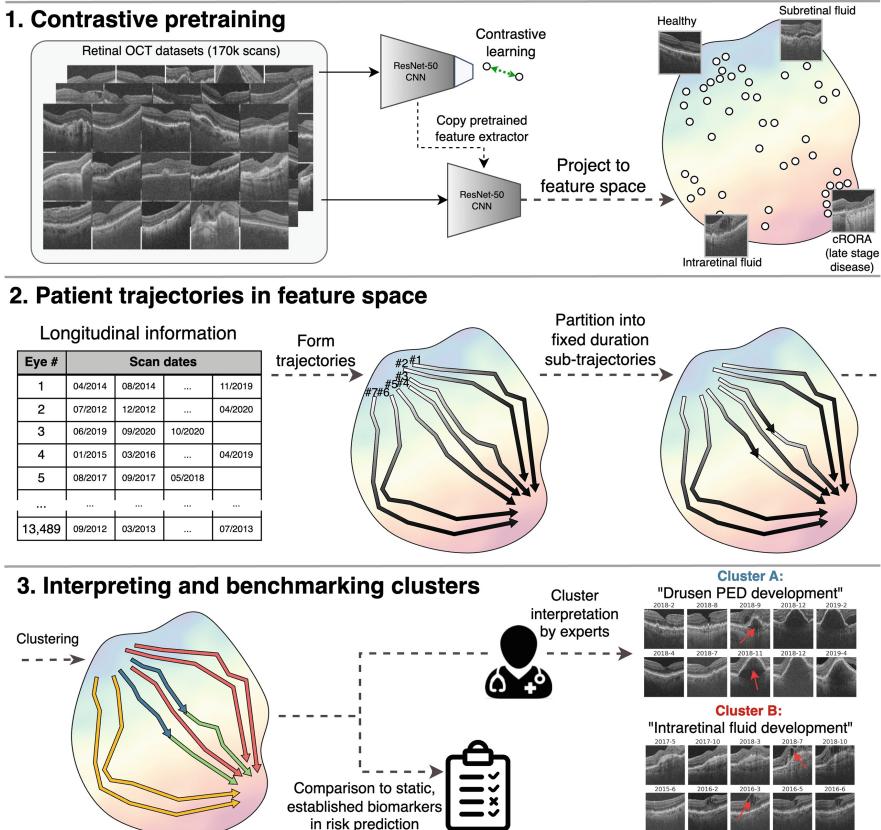


Fig. 1. Our method finds common patterns of disease progression in datasets of longitudinal images. We partition time series into sub-trajectories before introducing a clinically motivated distance function to cluster the sub-trajectories in feature space. The clusters are then assessed by ophthalmologists on their interpretability and ability to capture the progression of AMD.

the retina. However, the widely adopted AMD grading system [7, 13], which coarsely groups patients into broad categories for early and intermediate AMD, only has limited prognostic value for late AMD. Clinicians suspect that this is due to the grading system’s reliance on static biomarkers that are unable to capture temporal dynamics which contain critical information for assessing progression risk.

In their search for new biomarkers, clinicians have annotated known biomarkers in longitudinal datasets that monitor patients over time and mapped them against disease progression [2, 16, 19]. This approach is resource-intensive and requires biomarkers to be known *a priori*. Others have proposed deep-learning-based methods to discover new biomarkers at scale by clustering OCT images or detecting anomalous features [17, 18, 23]. However, these approaches neglect temporal relationships between images and the obtained biomarkers are by definition static and cannot capture the dynamic nature of the disease.

Our Contribution: In this work, we present a method to automatically propose biomarkers that capture temporal dynamics of disease progression in longitudinal datasets (see Fig. 1). At the core of our method is the novel strategy to represent patient time series as trajectories in a latent feature space. Individual progression trajectories are partitioned into atomic sub-sequences that encode transitions between disease states. Then, we identify population-level patterns of AMD progression by clustering these sub-trajectories using a newly introduced distance metric that encodes three distinct temporal criteria. In experiments involving 160,558 retinal scans, four ophthalmologists verified that our method identified several candidates for temporal biomarkers of AMD. Moreover, our clusters demonstrated greater prognostic value for late-stage AMD when compared to the widely adopted AMD grading system.

2 Related Work

Current AMD Grading Systems: Ophthalmologists’ current understanding of progression from early to late AMD largely involves drusen, which are sub-retinal lipid deposits. Drusen volume increases until suddenly stagnating and regressing, which often precedes the onset of late AMD [16]. The established AMD grading system stratifies early and intermediate stages solely by the size of drusen in a single OCT image [1, 6, 7, 10]. Late AMD is classified into either choroidal neovascularisation (CNV), identified by subretinal fluid, or geographic atrophy, signalled by progressive loss of photoreceptors and retinal thinning. The degree of atrophy can be staged using cRORA (complete retinal pigment epithelium and outer retinal atrophy), which measures the width in μm of focal atrophy in OCT [13]. Grading systems derived from these biomarkers offer limited diagnostic value and little to no prognostic capability.

Tracking Evolution of Known Biomarkers: Few research efforts have aimed at quantifying and tracking known AMD biomarkers, mostly drusen, over time

[16,19]. More work has explored the disease progression of Alzheimer’s disease (AD), which offers a greater array of quantitative imaging biomarkers, such as levels of tau protein and hippocampal volume. Young *et al.* [25] fit an event-based model that rediscovers the order in which these biomarkers become anomalous as AD progresses. Vogel *et al.* [21] find four distinct spatiotemporal trajectories for tau pathology in the brain. However, this only works if biomarkers are known *a priori* and requires manual annotation of entire time series.

Automated Discovery of Unknown Biomarkers: Prior work for automated biomarker discovery in AMD explores the latent feature space of encoders trained for image reconstruction [18,23], segmentation [27] or generative adversarial networks [17]. However, these neural networks are prone to overfit to their specific task and lose semantic information regarding the disease. Contrastive methods [3,8,26] encode invariance to a set of image transformations, which are uncorrelated with disease features, resulting in a more expressive feature space. However, all aforementioned methods group single images acquired at one point in time, and in doing so neglect temporal dynamics. The one work that tackles this challenge, and the most related to ours, categorises the time-dependent response of cancer cells to different drugs, measured by the changing distance in contrastive feature space from healthy controls [5].

3 Materials and Methods

3.1 OCT Image Datasets

We use two retinal OCT datasets curated in the scope of the PINNACLE study [20]. We first design and test our method on a *Development dataset*, which was collected from the Southampton Eye Unit. Afterwards, we test our method on a second independent *Unseen dataset*, which was obtained from Moorfields Eye Hospital. All images were acquired using Topcon 3D OCT devices (Topcon Corporation, Tokyo, Japan). After strict quality control, the *Development dataset* consists of 46,496 scans of 6,236 eyes from 3,456 patients. Eyes were scanned 7.7 times over 1.9 years on average at irregular time intervals. The *Unseen dataset* is larger, containing 114,062 scans of 7,253 eyes from 3,819 patients. Eyes were scanned 16.6 times over 3.5 years on average.

A subset of 1,031 longitudes was labelled using the established AMD grading protocols derived from known imaging biomarkers. Early AMD was characterised by small drusen between 63–125 μm in diameter. We also recorded CNV, cRORA ($\geq 250\mu\text{m}$ and $<1000\mu\text{m}$), cRORA ($\geq 1000\mu\text{m}$) [13] and healthy cases with no visible biomarkers. Visual acuity scores, which measured the patient’s functional quality of vision using a LogMAR chart, are available at 83,964 time points.

3.2 Self-supervised Feature Space Using Contrastive Learning

We use the BYOL contrastive loss [8] in Eq. 1 to train a ResNet50 (4x) model f over each batch of twice transformed images x

$$\mathcal{L}(x) = 2 - 2 \frac{\langle f(x), f'(x) \rangle}{\|f(x)\|_2 \cdot \|f'(x)\|_2} \quad (1)$$

where the output of the momentum updated ‘teacher’ network f' is passed through a stop-gradient, so that only the student network f is updated. As several of the contrastive transformations designed for natural images are inapplicable to medical images, such as solarisation, colour shift and greyscale, we use the set tailored for retinal OCT images by Holland *et al.* [9]. Models were trained on the entire dataset for 120,000 steps using the Adam optimiser with a momentum of 0.9, weight decay of $1.5 \cdot 10^{-6}$ and a learning rate of $5 \cdot 10^{-4}$. After training f , we first remove the final linear layer before projecting all labelled images to the feature space of 2048 dimensions.

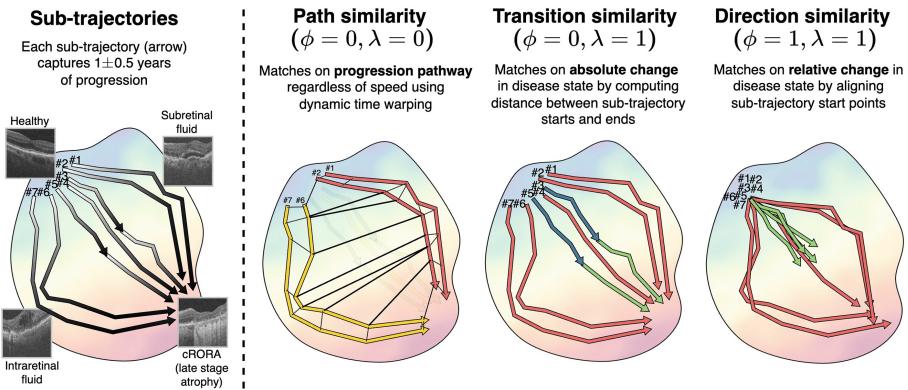


Fig. 2. Illustration of sub-trajectory distance functions which each encode temporal criteria for similarity (see Eq. 4). We illustrate clusters assignments, denoted by colour, resulting from three combinations of ϕ and λ .

3.3 Extracting Sub-trajectories via Partitioning

Naively clustering whole time series of patients ignores two characteristics of longitudinal data. Firstly, individual time series are not directly comparable as patients enter and leave the study at different stages of their overall progression. Secondly, longer time series can record multiple successive transitions in disease stage. Inspired by TRACLUS [11], the state of the art in trajectory clustering, we adapt their *partition-and-group* framework by assuming that trajectories can

be partitioned into a common set of *sub-trajectories* that capture singular transitions between progressive states of the disease.

For each eye, we first form piecewise-linear trajectories by linking points in feature space that were derived from consecutively acquired OCT images. We then extract sub-trajectories by finding all sequences of images spanning 1.0 ± 0.5 years of elapsed time within each trajectory. Next, to avoid oversampling trajectories with a shorter time interval between images, we randomly sample at most one sub-trajectory in every 0.5-year time interval.

3.4 Sub-trajectory Distance Functions and Clustering

In order to find common patterns of disease progression among sub-trajectories we cluster them. To this end we introduce a new distance function between sub-trajectories that incorporates three distinct temporal criteria (see Fig. 2). The first, formulated in Eq. 2, matches two sub-trajectories, U and V , of patients who progress between the same start and end states:

$$D_{transition}(U, V) = \|U_{start} - V_{start}\|_2 + \|U_{end} - V_{end}\|_2 . \quad (2)$$

Since all sub-trajectories cover a similar temporal duration, $D_{transition}$ also differentiates between fast and slow progressors and stable periods of no progression. However, by ignoring intermediary images, this metric does not respect the disease pathway along which patients progress. To incorporate this, we include a second metric that measures path dissimilarity, calculated using dynamic time warping (DTW) [4, 14, 15]. DTW finds the optimal temporal alignment between two time series before computing their distance. This re-alignment allows us to match sub-trajectories that traverse the same disease states in the same order, irrespective of the rate of change between states. We combine $D_{transition}$ with DTW using a $\lambda \in \mathbb{R}, 0 \leq \lambda \leq 1$ coefficient so the overall distance between U and V is

$$D_{path}(U, V) = \lambda D_{transition}(U, V) + (1 - \lambda) DTW(U, V) . \quad (3)$$

The third and final temporal criteria is to match time series that progress in the same relative direction, regardless of absolute disease states. We weight the contribution of this with $\phi \in \mathbb{R}, 0 \leq \phi \leq 1$ in Eq. 4:

$$D_{subtraj}(U, V) = \phi D_{path}(U - U_{start}, V - V_{start}) + (1 - \phi) D_{path}(U, V) . \quad (4)$$

Spectral Clustering: As the non-linearity of $D_{subtraj}$ prohibits the use of k-means for clustering, we instead use spectral clustering [22] to group similar sub-trajectories. Hereby, we construct an affinity matrix \mathcal{A} encoding the negative of the distance $D_{subtraj}$ between all pairs of sub-trajectories. Using \mathcal{A} , we group sub-trajectories into K clusters.

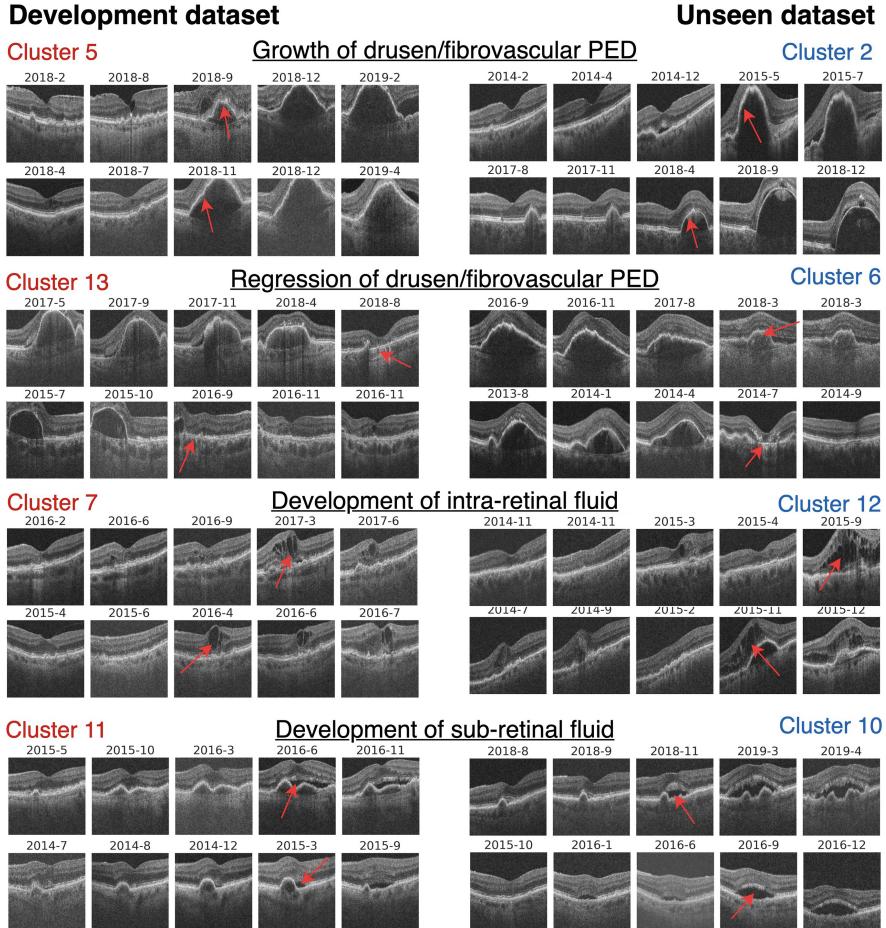


Fig. 3. We show four clusters from the *Development dataset* (left half) and the equivalent clusters in the *Unseen dataset* (right half). Ophthalmologists identified clusters capturing the same progression dynamics in both datasets, providing clinical interpretations (underlined). Clusters show two representative sub-trajectories originating from different patients, each containing five longitudinal images with the time and location of greatest progression marked by arrows.

3.5 Qualitative and Quantitative Evaluation of Clusters

Initially, we tune the hyperparameters, λ , ϕ and K , on the *Development dataset* by heuristically selecting values that result in higher uniformity between sub-trajectories within each cluster. Two teams of two ophthalmologists then review 20 sub-trajectories from distinct patients in each cluster, interpreting and summarising any consistently observed temporal dynamics. Next, using the same hyperparameters we apply the method directly to the *Unseen dataset*. The ophthalmologists then review these clusters and confirm whether they capture the

same temporal biomarkers observed in the *Development dataset*.

In addition to the qualitative evaluation, we also validate the utility of our clusters as biomarkers that stratify risk of disease progression. We test this by predicting the time until conversion to late AMD and its subtypes, CNV and cRORA. Additionally, we predict current visual acuity. For prediction, each sub-trajectory is characterised by a vector of size K that encodes proportional similarity to each cluster. This vector is then used by a Lasso linear regression model. Similarly, we fit an equivalent linear regression model to the static biomarkers from the established grading system detailed in Sect. 3.1. We also include a demographic baseline using age and sex. We also add a temporally agnostic baseline that clusters only single time points. Finally, to demonstrate the performance gap between our interpretable approach and black-box supervised learning algorithms, we include a fully supervised deep learning baseline by fitting an SVR directly to the feature space. Each experiment uses 10-fold cross validation on random 80/20 partitions, while ensuring a patient-wise split. Finally, we repeat the entire method, starting from sub-trajectory extraction, followed by clustering and then regression experiments, using 7 random seeds and report the means and standard deviations.

Table 1. Sub-trajectory clusters were comparable to the established clinical grading systems for AMD in predicting future disease, shown by reduced MAE in years for Late AMD, CNV and cRORA and MAE in LogMAR for visual acuity.

| Development dataset | | | | |
|---------------------------|--------------------|---------------|-----------------|-------------------------|
| | Time to Late AMD ↓ | Time to CNV ↓ | Time to cRORA ↓ | Current visual acuity ↓ |
| Demographic | 0.756±0.01 | 0.822±0.012 | 0.703±0.028 | 0.381±0.007 |
| Current grading system | 0.757±0.01 | 0.819±0.012 | 0.685±0.035 | 0.367±0.008 |
| Single timepoint clusters | 0.747±0.013 | 0.776±0.015 | 0.630±0.05 | 0.230±0.005 |
| Sub-trajectory clusters | 0.739±0.01 | 0.748±0.011 | 0.636±0.031 | 0.375±0.007 |
| Fully supervised | 0.709±0.015 | 0.726±0.012 | 0.609±0.033 | 0.199±0.004 |
| Unseen dataset | | | | |
| Demographic | 1.343±0.027 | 1.241±0.017 | 1.216±0.062 | 0.188±0.007 |
| Current grading system | 1.308±0.018 | 1.244±0.022 | 1.286±0.053 | 0.177±0.008 |
| Single timepoint clusters | 1.325±0.049 | 1.341±0.080 | 1.297±0.096 | 0.136±0.005 |
| Sub-trajectory clusters | 1.322±0.029 | 1.235±0.027 | 1.257±0.056 | 0.188±0.006 |
| Fully supervised | 1.301±0.044 | 1.298±0.08 | 1.255±0.097 | 0.135±0.006 |

4 Experiments and Results

Sub-trajectory Clusters are Candidate Temporal Biomarkers: By first applying our method to the *Development dataset* we found that using $\lambda = 0.75$, $\phi = 0.75$ and $K = 30$ resulted in the most uniform and homogeneous clusters while still limiting the total number of clusters to a reasonable amount. Achieving the same cluster quality with smaller values of ϕ required many more clusters in order to encode all combinations of possible start and end disease states. The expert ophthalmologists remarked that many of the identified clusters capture

dynamics that have already been linked to the progression of AMD, even though they are not currently included in any clinical grading system. Using the same hyperparameters our method generalised to the *Unseen dataset* which yielded clusters with equivalent dynamics and quality (see Fig. 3).

Ophthalmologists identified clusters capturing the same variants of temporal progression in both datasets. They named these as '*rapid growth of drusen pigment epithelial detachments (PED)*', '*regression of drusen PED*', '*development of subretinal fluid*', '*development of intraretinal fluid*', '*development of hypertransmission*' and '*stable state*' (no signs of progression at each disease state).

Sub-trajectory Clusters Predict Conversion to Late AMD: Next, we validated that our clusters are predictive of progression to late AMD. Our sub-trajectory clusters were comparable to, and in some cases outperformed, the current widely adopted grading system in predicting risk of conversion (see Table 1). In all tasks the standard biomarkers are only marginally more indicative of risk than the patient's age and sex. This experiment confirms that our clusters are related to disease progression.

5 Discussion and Conclusion

Motivated to improve inadequate grading systems for AMD that do not incorporate temporal dynamics we developed a method to automatically propose biomarkers that are time-dependent, interpretable, and predictive of conversion to late-stage AMD. We applied our method to two large longitudinal datasets, cataloguing 3,218 total years of disease progression. The found time-dependent clusters were subsequently interpreted by four ophthalmologists. They found them to capture distinct patterns of disease progression that have been previously linked to AMD, but are not currently included in clinical grading systems. Furthermore, we experimentally demonstrated that the found clusters predict conversion to late-stage AMD on par with the established grading system.

In the future, biomarkers identified by our method can be further refined by clinicians. We will also use the full volumetric image to model progression dynamics outside the macular. As late stage patients were overrepresented in our datasets, we also intend to apply our method to datasets with greater numbers of patients progressing from earlier disease stages. Ultimately, we envision that proposals from our method may inform the next generation of grading systems for AMD that incorporate the temporal dimension intrinsic to this dynamic disease.

Acknowledgements. The PINNACLE study is funded by a Wellcome Trust Collaborative Award (ref. 210572/Z/18/Z). This work is also funded by the Munich Center for Machine Learning.

References

1. Bird, A.C., et al.: An international classification and grading system for age-related maculopathy and age-related macular degeneration. *Surv. Ophthalmol.* **39**(5), 367–374 (1995)
2. Chen, K.G., et al.: Longitudinal study of dark adaptation as a functional outcome measure for age-related macular degeneration. *Ophthalmology* **126**(6) (2019)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML, pp. 1597–1607. PMLR (2020)
4. Cuturi, M., Blondel, M.: Soft-dtw: a differentiable loss function for time-series. In: International Conference on Machine Learning, pp. 894–903. PMLR (2017)
5. Dmitrenko, A., et al.: Self-supervised learning for analysis of temporal and morphological drug effects in cancer cell imaging data. In: Medical Imaging with Deep Learning (2021)
6. Ferris, F.L., et al.: A simplified severity scale for age-related macular degeneration. *Arch. Ophthalmol.* **123**(11), 1570–1574 (2005)
7. Ferris III, F.L., Wilkinson, C., Bird, A., Chakravarthy, U., Chew, E., Csaky, K., Sadda, S.R., for Macular Research Classification Committee, B.I., et al.: Clinical classification of age-related macular degeneration. *Ophthalmology* **120**(4) (2013)
8. Grill, J.B., et al.: Bootstrap your own latent-a new approach to self-supervised learning. *NeurIPS* **33**, 21271–21284 (2020)
9. Holland, R., et al.: Metadata-enhanced contrastive learning from retinal optical coherence tomography images. *CoRR* abs/2208.02529 (2022)
10. Klein, R., et al.: Harmonizing the classification of age-related macular degeneration in the three-continent AMD consortium **21**(1), 14–23 (2014)
11. Lee, J.G., et al.: Trajectory clustering: a partition-and-group framework. In: ACM SIGMOD, pp. 593–604 (2007)
12. Mitchell, P., et al.: Age-related macular degeneration. *The Lancet* **392**(10153), 1147–1159 (2018)
13. Sadda, S.R., et al.: Consensus definition for atrophy associated with age-related macular degeneration on oct: classification of atrophy report 3. *Ophthalmology* **125**(4), 537–548 (2018)
14. Sakoe, H.: Dynamic-programming approach to continuous speech recognition. In: 1971 Proc. the International Congress of Acoustics, Budapest (1971)
15. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **26**(1), 43–49 (1978)
16. Schlanitz, F.G., et al.: Drusen volume development over time and its relevance to the course of age-related macular degeneration. *BJO* **101**(2), 198–203 (2017)
17. Schlegl, T., et al.: f-anogan: fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **54**, 30–44 (2019)
18. Seeböck, P., et al.: Unsupervised identification of disease marker candidates in retinal oct imaging data. *IEEE TMI* **38**(4), 1037–1047 (2018)
19. Steinberg, J.S., et al.: Longitudinal analysis of reticular drusen associated with geographic atrophy in age-related macular degeneration. *IOVS* **54**(6), 4054–4060 (2013)
20. Sutton, J., et al.: Developing and validating a multivariable prediction model which predicts progression of intermediate to late age-related macular degeneration-the pinnacle trial protocol. *Eye*, pp. 1–9 (2022)
21. Vogel, J.W., et al.: Four distinct trajectories of tau deposition identified in Alzheimer’s disease. *Nat. Med.* **27**(5), 871–881 (2021)

22. Von Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007)
23. Waldstein, S.M., et al.: Unbiased identification of novel subclinical imaging biomarkers using unsupervised deep learning. *Sci. Rep.* **10**(1), 1–9 (2020)
24. Wong, W.L., et al.: Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob. Health* **2**(2), e106–e116 (2014)
25. Young, A.L., et al.: A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* **137**(9), 2564–2577 (2014)
26. Zhao, A., et al.: Prognostic imaging biomarker discovery in survival analysis for idiopathic pulmonary fibrosis. In: MICCAI Proceedings, pp. 223–233. Springer (2022). https://doi.org/10.1007/978-3-031-16449-1_22
27. Zheng, Q., et al.: Pathological cluster identification by unsupervised analysis in 3,822 UK biobank cardiac MRIS. *Front. Cardiovasc. Med* **7**, 539788 (2020)