Machine Learning Can Predict Anti–VEGF Treatment Demand in a Treat-and-Extend Regimen for Patients with Neovascular AMD, DME, and RVO Associated Macular Edema

Mathias Gallardo, PhD, Marion R. Munk, MD, PhD, Thomas Kurmann, PhD, Sandro De Zanet, PhD, Agata Mosinska, PhD, Isil Kuthturk Karagoz, MD, PhD, Martin S. Zinkernagel, MD, PhD, Sebastian Wolf, MD, PhD, Raphael Sznitman, PhD

Purpose: To assess the potential of machine learning to predict low and high treatment demand in real life in patients with neovascular age-related macular degeneration (nAMD), retinal vein occlusion (RVO), and diabetic macular edema (DME) treated according to a treat-and-extend regimen (TER).

Design: Retrospective cohort study.

Participants: Three hundred seventy-seven eyes (340 patients) with nAMD and 333 eyes (285 patients) with RVO or DME treated with anti–vascular endothelial growth factor agents (VEGF) according to a predefined TER from 2014 through 2018.

Methods: Eyes were grouped by disease into low, moderate, and high treatment demands, defined by the average treatment interval (low, >10 weeks; high, ≤5 weeks; moderate, remaining eyes). Two random forest models were trained to predict the probability of the long-term treatment demand of a new patient. Both models use morphological features automatically extracted from the OCT volumes at baseline and after 2 consecutive visits, as well as patient demographic information. Evaluation of the models included a 10-fold cross-validation ensuring that no patient was present in both the training set (nAMD, approximately 339; RVO and DME, approximately 300) and test set (nAMD, approximately 38; RVO and DME, approximately 33).

Main Outcome Measures: Mean area under the receiver operating characteristic curve (AUC) of both models; contribution to the prediction and statistical significance of the input features.

Results: Based on the first 3 visits, it was possible to predict low and high treatment demand in nAMD eyes and in RVO and DME eyes with similar accuracy. The distribution of low, high, and moderate demanders was 127, 42, and 208, respectively, for nAMD and 61, 50, and 222, respectively, for RVO and DME. The nAMD-trained models yielded mean AUCs of 0.79 and 0.79 over the 10-fold crossovers for low and high demand, respectively. Models for RVO and DME showed similar results, with a mean AUC of 0.76 and 0.78 for low and high demand, respectively. Even more importantly, this study revealed that it is possible to predict low demand reasonably well at the first visit, before the first injection.

Conclusions: Machine learning classifiers can predict treatment demand and may assist in establishing patient-specific treatment plans in the near future. Ophthalmology Retina 2021;5:604–624 © 2021 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Supplemental material available at www.opthalmologyretina.org.

See Editorial on page 601.
long-term would help to plan, set patients expectations, and find more efficient and precise treatment solutions.

Machine learning models have been studied to help individualize anti-VEGF therapy for chronic eye diseases. The common approach is to predict at the early stage the long-term treatment response and demand of the patient, respectively. Machine learning models aim to predict whether a patient will have a high or low treatment demand and will show a rather good or poor response. The definition of respective terms plays a critical role here. Rasti et al used a procedural definition, leveraging the number of injections prescribed to ensure the stability of the disease. Functional criteria such as best-corrected visual acuity (BCVA) improvement are used sparsely.

Rasti et al predicted the response of DME patients to anti-VEGF medication after an every-3-month injection treatment by analyzing OCT volume features at baseline visits. The definition of the responder classes is based on the change in retinal thickness relative to the third injection. Specifically, patients are divided into 2 groups: a responsive group with a reduction of 10% of the retinal thickness at the fourth visit and a nonresponsive group for the remaining patients. The method relies on a deep learning strategy and was trained and tested over a retrospective cohort of 127 patients with DME, with 49 B-scans for each OCT volume. The proposed method reached a mean area under the receiver operating characteristic curve (AUC) of 0.866 to classify responsive and nonresponsive patients. However, the method lacks interpretability on the features extracted and used for the decision making at the B-scan level. A validation, whether the image information used by the method is coherent to the image information used by the clinicians, is not possible. Another limitation is that the responder definition does not rely on long-term information, but instead relies only on information at the baseline and the fourth visit.

Alternatively, the definition of low and high demand as no more than 5 injections and at least 16 injections over a 24-month period, respectively, also has been proposed. Remaining patients then are defined as having moderate demand, and the method used classifies patients according to information from the first 3 visits. Information used included morphological features from the OCT volume, BCVA, and demographic data. Two random forest models were trained to classify low demand and another for high demand over a cohort of 317 eyes with nAMD collected during a 2-year clinical trial. Both classifiers obtained an AUC of 0.7 and 0.77, respectively.

Importantly, current works have been trained and tested only using clinical trial data and considered a single pathologic feature. Such clinical trial data in practice may differ from real-world patient data in 2 aspects. First, in clinical trials, visit schedules generally are well followed by patients (some exclusion criteria on the schedule deviation). Second, the image quality of OCT scans is checked more carefully for further evaluations. In retrospective data, visit schedule can present some significant deviations, which impacts the labeling and makes the prediction more difficult. Along with this, image quality generally is lower, which may harm the extraction of reliable morphological features.

In this work, we aimed to observe the practical feasibility of such a prediction tool for routinely collected retrospective clinical cohorts considering 3 different pathologic features (nAMD, DME, and ME-related RVO). Specifically, we wished to observe the feasibility of predicting the long-term demand of anti-VEGF medication at the early stage of a 1-year TER in a routine clinical setting. For each pathologic group, we trained 2 random forest classifiers for identifying low and high demand and analyzed in detail their performance and the consistency of the most important features used with those leveraged by clinicians.

### Methods

#### Study Cohort and Data

This was a retrospective study with a total of 710 eyes (625 patients) treated for at least 1 year with anti-VEGF therapy according to a predefined TER from 2014 through 2018 at the University Hospital of Bern, Bern, Switzerland. Three hundred seventy-seven eyes with nAMD (340 patients), 155 eyes with ME-related RVO (150 patients), and 178 eyes with DME (135 patients) were included. Henceforth, we consider ME-related RVO and DME in a single pathologic group and refer to it as the retinal vascular diseases group, with 333 eyes (285 patients). The inclusion criteria were the following: (1) patients with nAMD and retinal vascular diseases, (2) starting and following a TER with either ranibizumab or aflibercept for at least 1 year (365 days) between January 1, 2014, and December 31, 2018. The initial visits were included in the study. Both eyes from the same patient were assessed independently.

Patient data such as sex and age, as well as the OCT volume data of the enrolled eyes acquired at each visit, were included. Macular 6 × 6-mm OCT volumes were obtained with a Spectralis SD-OCT imaging system (Heidelberg Engineering, Inc., Heidelberg, Germany) using a 49 B-scans acquisition protocol and a resolution of 496 × 512 pixels per B-scan. The average volume dimensions over the entire cohort in terms of length (49 B-scans), width (512 px), depth (496 px) was 5.90 mm × 5.75 mm × 1.92 mm. Follow-up function was set at the initial OCT scan. We report that we did not have access to the image quality grades, and thus did not exclude patients with poor-quality or ungradable OCT volumes. This study was approved by the ethics committee of the University of Bern (KEK no.: 2019-00285) and followed the tenets of the Declaration of Helsinki. Informed consent was waved because of the retrospective design of the study.

#### Treat-and-Extend Regimen and Demand Definitions

The TER is illustrated in Figure 1. Briefly, 4 weeks after the first intravitreal injection, a second injection is given. Depending on the disease activity, that is, the presence or absence of subretinal fluid (SRF) or intraretinal fluid (IRF) and sub–retinal pigment epithelium fluid seen on the OCT, the treatment interval for the next injection is adjusted. As such, pigment epithelial detachment (PED)—and especially an increase in PED size—is a criterion for reducing the treatment interval. The treatment interval is extended by 2 weeks if stable conditions or inactivity are present or is shortened by 1 week if signs of activity are present, with a minimum treatment interval of 4 weeks. When the treatment interval is shortened, the next treatment interval must not be extended for the next 6 months, but may be shortened if activity remains at any
After 6 months, the treatment interval can be extended again.2,7–9 For nAMD, we established an exit criterion that is met when the maximum interval of 16 weeks is reached and is maintained for 3 consecutive injections. In the case of retinal vascular diseases, no fixed exit criterion was established. After an interval of 16 weeks, the treatment interval for these patients may be extended by 4 weeks instead of 2 weeks, and potential treatment cessation is discussed individually with the patient if the treatment interval has been extended by more than 28 weeks without disease recurrences.

The treatment demand being either low, high, or moderate was defined by the average treatment interval during the first 1-year treatment period. It reflects the number of required injections to obtain disease stability or inactivity throughout 1 year. Low treatment demand was defined by an average treatment interval of 10 weeks or more, high treatment demand was defined by an average treatment interval of 5 weeks or less, and all remaining eyes were considered to have moderate treatment demand.

It is particularly critical to avoid a true-positive prediction label of high demand being classified as low or moderate demand to avoid potential undertreatment, and therefore potential severe vision loss. Therefore, we opted to use an upper bound definition for high demand of 4 weeks with a margin of 1 week because of fluctuations in treatment schedules. Low demand in turn reflects a patient cohort showing disease control with a small number of injections and an increasingly longer treatment interval. Moderate demand was used as a third class to consider eyes with medium demand.

Low and High Predictor

The proposed method relies on a machine learning technique to predict the long-term demand for anti-VEGF treatment using information from the initial visits. In this study, we defined the initiation and early stage of treatment as the first 3 visits. Further, we investigated predictions made from only the first baseline visit, as well as from the baseline visit including 1 or 2 follow-up visits.

To predict low and high demand, we design 2 binary classifiers: low versus others (LvO), which predicts the probability of an eye showing low demand, and high versus others (HvO), which predicts the probability of an eye showing high demand. Both classifiers (LvO and HvO) are trained similarly and vary only in terms of the training labels used: the LvO classifier (respectively, HvO) is trained by labeling low (respectively, high) demand as the positive group, and high and moderate (respectively, low and moderate) demand as the negative group.

We use random forest plots13 as the classifiers for both LvO and HvO, respectively. In both cases, we use morphological features extracted from the OCT volumes as well as additional patient data (age at each visit and sex of the patient), which we describe in the following sections. We chose this type of classifier because it is known to work well when trained on relatively small datasets and to provide interpretability of the prediction process. To construct compact and meaningful input morphological features, we use existing algorithms of retina and layer segmentation and biomarker presence detection. In addition, we also can extract interpretable decision information from random forest plots. This is in line with previous work that also successfully applied random forest plots with satisfactory accuracy in a PRN regimen for patients with nAMD enrolled in a clinical trial.

Our random forest classifiers used 1000 trees, with a maximum tree depth of 100. To train and evaluate both classifiers, we used Python 3.7 and its open-source package Scikit-learn version 0.20.3 for the random forest implementation. To reduce the variance of the model, we used bootstrapping when building the trees. Each classifier used all the following features.

1. Segmentation-based morphological features. Using the segmentation algorithm,10–13 retinal and choroidal layers, IRF, SRF, and PED segmentations were computed on all OCT volumes. A description of this algorithm is given in Appendix A (available at www.ophthalmologyretina.org) and an illustration of the algorithm outcome is given in Figure 2. Features from these segmentations were computed similar to Bogunovic et al, whereby mean thickness of the inner retina, outer nuclear layer, retinal pigment epithelium layer computed at the volume level and at 13 regions specified by the Early Treatment Diabetic Retinopathy Study (ETDRS) grid. Additionally, we computed a novel feature named patient-independent retinal thickness to account for variability of thickness found in healthy retinas. In total, we thus compute 4 (groups of layers) × 1 (mean thickness) × 13 (ETDRS regions) + 3 (fluids) × 2 (en face area map and volume) × 13 (ETDRS regions) + 1 (patient-independent retinal thickness) = 131 segmentation-based morphological features per OCT volume.

Further computation can be found in Appendix B (available at www.ophthalmologyretina.org). Note that segmentations were not manually corrected.

2. Biomarker-based morphological features. We automatically computed the likelihood of occurrence of several retinal biomarkers using a state-of-the-art method.13 Specifically, for each B-scan in a volume, 10 biomarkers are estimated. These include healthy B-scan (defined by the absence of all the herein listed 10 biomarkers), SRF, IRF, hyporeflective foci, drusen, reticular pseudodrusen, epiretinal membrane, geographic atrophy, outer retinal atrophy, and fibrovascular PED. From these 10 biomarkers, we constructed 2 types of detection-based features: (1) the number of B-scans in which the presence probability, for a given biomarker, is superior or equal to a threshold of 0.75 and (2) the maximum probability (between 0 and 1) across all B-scans for a given biomarker. Although the first feature accounts for biomarker quantities, the second feature carries the confidence of the detections. Thus, we have 2 × 10 = 20 detection-based morphological features per OCT volume (Appendix C, available at www.ophthalmologyretina.org).

3. Nonmorphological features. These include the sex and the age (in years) of the patient at each used visit.

4. Differential features. To characterize progression in morphological features over different visits, we estimated the relative and absolute changes. This was computed by differences in the different morphological features mentioned above. Adding these differential features for the first 3 visits led to: (131 (segmentation-based) + 20 (detection-based)) × 3 (V01, V02 and V03) + 604 (differential features) + 1 (sex) + 3 (age at each visit) = 1061 features. When considering the baseline and the subsequent visit, the total number of features used decreases to 607, and when considering only the baseline visit, this number goes to 153.

Evaluation of the Predictive Model

First, the performance of the classifiers was measured quantitatively using the AUC. This is motivated by the fact that a classifier model built from random forest analysis provides a probability of belonging to one class and not the estimated class itself of a sample. This is also the reason why we do not rely on the accuracy metric, because its use implies fixing a sensitivity and a specificity value that requires considering multiple factors that go beyond the scope of this work. Specifically, we performed a 10-fold cross-validation using 90% of the data for training and 10% for validation of the classifiers, respectively. For each fold, we ensured that no patient was present in
both the training and the test set. We computed the mean AUC over all 10 folds. Potentially to set an operating point for both classifiers, we reported the prediction for both classifiers and the results of the joint predictions using the combination rules given in Figure 3. We further computed the quadratic weighted $\kappa$ value to assess the performance of the classification.

We also evaluated the relationship between the features and the treatment demander classes by considering both the data and the model (feature importance in the model) and separately considering only the data (statistical significance of the features). To assess the feature importance of the model, we computed the contribution of each input feature in the classifier training and the final prediction. Specifically, we considered the relative importance score of all features. This was quantified during the classifier training phase by measuring the normalized total reduction of the Gini impurity (1 — the sum of squares of success probabilities of the 2 classes) brought by each feature: the features that most contribute to the performance of the prediction are the most important. From the relative importance score, we extracted the top 3 and analyzed the distribution of their values with respect to the demand class. To assess their statistical significance in relation to the prediction, we selected a set of the main features and used a Wilcoxon rank-sum test and the Holm-Bonferroni$^{15}$ correction to account for multiple comparisons.

Finally, we computed the Lvo and HvO classifiers at 3 time points using the baseline information only (V01), the baseline information and that of the next consecutive visit (V02), and the information of baseline and 2 consecutive visits (V03),

Figure 1. Schematic procedure of the treat-and-extend protocol used in the University Hospital of Bern, Bern, Switzerland. IVT = intravitreal injection; nAMD = neovascular age-related macular degeneration; Ret. Vasc. = retinal vascular.

Figure 2. Illustration of the segmentation results on 1 B-scan using the segmentation algorithm.$^{10–12}$ A, B-scan. B, B-scan including overlaid segmentation. C, Segmentation of the B-scan. GCL = ganglion cell layer; IPL = inner plexiform layer; NFL = nerve fiber layer; ONL = outer nuclear layer; OPL = outer plexiform layer; PR = photoreceptor layer; RPE = retinal pigment epithelium; SRF = subretinal fluid.
respectively. The classifier performances were put into perspective of the amount of necessary external inputs, that is, the number of visits. This is specifically important for the application of this algorithm in clinical practice.

### Results

#### Cohort Description

Our study cohort consisted of 710 eyes distributed over 2 pathologic groups as follows: 377 nAMD eyes (340 patients) and 333 eyes with retinal vascular diseases (285 patients). In the nAMD cohort, we identified 127 eyes, 42 eyes, and 208 eyes with low, high, and moderate treatment demand, respectively. In the retinal vascular diseases cohort, we identified 61 eyes, 50 eyes, and 222 eyes with low, high, and moderate demand, respectively. Table 1 provides the details of the cohort characteristics.

#### Classification Performance

For the 2 pathologic groups, Figure 4 shows the mean receiver operating characteristic curve over the 10 folds for the 2 classifiers trained after the third visit. We obtained a mean AUC of 0.79 for both nAMD classifiers. For eyes with retinal vascular diseases, the LvO classifier reached a mean AUC of 0.76 and the HvO classifier reached a mean AUC of 0.78. Figure 5 shows the impact on classification performance when only 1 and 2 visits are considered, respectively. For the HvO classifier, information from more visits induced better prediction performances with an increase of approximately 0.1 of the mean AUC. The mean AUC of the LvO classifier remained stable regardless of number of visits.

We observed a small performance decrease in the LvO classifier for nAMD and retinal vascular diseases as a function of the visit number (Fig 5). Errors in feature computations (e.g., oversegmentations and undersegmentations of fluid) in images of low quality were found to be an important contributor to why some eyes correctly classified at V01 were incorrectly classified after V02, respectively. Figure 6 exemplifies the errors in the estimation of the total retinal thickness at V02. The central B-scan displayed in Figure 6A illustrates that such an error may originate from noisy, low-quality OCT scans. Figure 6A, B shows that some retinal regions were segmented incorrectly as fluids, leading the decision trees of the LvO classifier to incorrectly vote for high demand. Nevertheless, we cannot conclude that these errors are specifically and generally related to V02 or V03. This phenomenon may be specific to our cohort study and may be verified with different cohort data.

#### Specific Contribution of Input Features

Figures 7 and 8 give the top 15 most important features for the classifiers trained on all 3 visits over the 10 folds and for the 2 pathologic groups. In both figures, we also display the relative importance score of these 15 features for the classifiers trained after the second visit (classifier at V02) and at baseline (classifier at V01). Figures 9 and 10 depict the contribution of all features after the third visit using 3 different perspectives. Figures 11 and 12 show, for both pathologic groups, the boxplots of the top 3 feature values for the classifiers at the 3 time points.

Table 2 lists the set of features that are used for statistical significance. These features correspond to 84 nondifferential segmentation- and biomarker-based features. Table 3 and 4 report their corrected P values (with Holm-Bonferroni correction and \( \alpha = 0.05 \)) for the 2 pathologic groups.

For the nAMD patients, significant differences were found for all features for the first 3 visits between low and high treatment demands, except for the features related to IRF at V03 and 1 IRF feature from V02 (\( P > 0.07 \)). For the low and moderate demand classes, the most important, statistically significant features are extracted at the baseline visit. In patients with retinal vascular diseases, strong significant differences were found between low and high demand for the features related to IRF and retinal

### Table 1. Characteristics of the Included Eyes

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. of Eyes</th>
<th>Age (Mean ± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>nAMD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low demand</td>
<td>79</td>
<td>48</td>
</tr>
<tr>
<td>High demand</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Moderate demand</td>
<td>118</td>
<td>90</td>
</tr>
<tr>
<td>Subtotal</td>
<td>218</td>
<td>159</td>
</tr>
<tr>
<td>Retinal vascular diseases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low demand</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>High demand</td>
<td>79</td>
<td>33</td>
</tr>
<tr>
<td>Moderate demand</td>
<td>79</td>
<td>143</td>
</tr>
<tr>
<td>Subtotal</td>
<td>123</td>
<td>210</td>
</tr>
</tbody>
</table>

nAMD = neovascular age-related macular degeneration.
Classifiers for nAMD patients

Classifiers for patients with ret. vasc. diseases

Figure 4. Mean receiver operating characteristic (ROC) curves of both classifiers and for the 2 pathologic groups, neovascular age-related macular degeneration (nAMD) and retinal vascular (ret. vasc.) diseases. Standard deviation (SD) corresponds to the SD of the area under the ROC curve (AUC) values obtained from the 10-fold cross-validation (CV). **Left column**, Performance of the classifier low versus others (LvO). **Right column**, Performance of the classifier high versus others (HvO).

thickness \( P < 0.0007 \) for the first 3 visits. Between low and moderate demand, almost every feature related to IRF and the retinal thickness showed a statistically significant impact at the 3 time points. Similar to the nAMD cohort, the features provided by the second and the third visits are significantly different between moderate and high demand.
Adapting the operating point by maximizing the specificity led to 99.6% specificity and 25% sensitivity for the LvO classifier in the nAMD cohort and to 100% specificity and 33.3% sensitivity in the retinal vascular diseases cohort. For the HvO classifier, jointly maximizing the specificity and sensitivity yielded 81.2% specificity and 75% sensitivity for the HvO classifier in the nAMD cohort and 42.9% specificity and 80% sensitivity in the retinal vascular diseases cohort, respectively. Figure 13 reports the distribution of the predicted classes for both classifiers. For the HvO classifier, we noted that for eyes with nAMD and retinal vascular diseases, the number of overtreated eyes—and thus with low and moderate treatment demand—that were classified as having high demand was 22.2% and 40.8%, respectively. Figure 14 shows for each fold the receiver operating characteristic curves of both classifiers for both pathologic groups.

Discussion

Results Analysis

Our experimental results show that a machine learning technique can predict the 1-year treatment demand using a TER with a reasonably good AUC of between 0.76 and 0.79 for patients with nAMD and retinal vascular diseases using real-life retrospective patient data. Interestingly, classifying high treatment demand over low and moderate demand, respectively, seems to be a more difficult task. The information from all 3 visits is necessary to achieve a similar performance (Fig 5). However, to predict low treatment demand, only the information from the baseline visit is necessary for both disease groups. Thus, features from V01 seem sufficient to classify low demand, whereas additional features from V02 and V03 are necessary to classify high demand (see Figs 7A, 8A, 9A, C, and 10A, C). We also can conclude that a reasonably good prediction of high demand can be obtained after the second visit (V02).

Our analysis of the feature contributions to the final prediction, as reported in Figures 9A–D and 10A–D, enables comparison of the process of decision making of the classifiers with the TER. For the classifiers trained and tested on patients with nAMD, the most informative features for the classifier were derived from SRF, IRF, and total retinal thickness, whereas for the classifiers trained and tested on patients with retinal vascular diseases, the most informative features were derived from IRF and total retinal thickness. This is consistent with previous clinical observations and previous artificial intelligence-based assessments of these patient populations, in which the presence of SRF in nAMD was identified overall as a predictive marker for low treatment demand and superior functional outcome, whereas the presence of IRF was identified as a marker for an inferior treatment outcome. For retinal vascular diseases, IRF and retinal layer thicknesses are vastly more important than other features to distinguish the 3 demander classes, which is consistent with clinical observations. Subretinal fluid seems to be less important for the classifiers with this pathologic group. Our analysis also underlined that the differential features (see Figs 9C, D and 10C, D) are surprisingly not so discriminative in general. This also has been reported with the PRN regimen for patients with nAMD. An exception is notable for the LvO classifier for the time element V02 to V01 with respect to SRF in patients with nAMD and IRF for patients with retinal vascular diseases, respectively (Figs 7A and 8A). These findings nevertheless are consistent with TER clinical practices.

Neovascular Age-Related Macular Degeneration

Interestingly (Fig 7), the maximum probability of SRF (outlined in the biomarker detector) at the baseline visit (V01) is the most important feature for LvO and HvO in
Figure 6. Eyes with inconsistent predicted responses over time because of noisy OCT acquisition. Representation of the input features from the OCT scans for the 3 first visits of 2 eyes, (A) and (B). The first row shows the central B-scan and the second through fifth rows show the area en face maps of the total retinal thickness (TRT), subretinal fluid (SRF), intraretinal fluid (IRF), and pigment epithelial detachment (PED). The sixth row shows the presence probability of 10 biomarkers across the OCT volume. For each eye, its pathologic features, its sequence of interval, its true response, and the prediction from the classifier low versus others (LvO) at visit 1 (V01), visit 2 (V02), and visit 3 (V03) also are indicated. ERM = epiretinal membrane; FPED = fibrovascular pigment epithelial detachment; GA = geographic atrophy; HF = hyperreflective foci; nAMD = neovascular age-related macular degeneration; ORA = outer retinal atrophy; RPD = reticular pseudodrusen.
nAMD at the 3 time points. For identifying low treatment demand, only the baseline visit is necessary to gather a significant number of the most discriminative features leading to a similar prediction performance as with adding additional information from the next 2 consecutive visits. For identifying high demand, information from the V03 and V02 seems to be significantly important, as Figure 8B, D demonstrates. This shows that the HvO classifier tends to use the latest visit information available to detect high demand.

Figure 7 shows that, in general, the differential features are surprisingly not discriminative and that patient data, that is, sex and age, are not discriminative at all. To discriminate low from high and moderate demand in nAMD, the features from the biomarker detector are more informative than the ones from the segmentation method. However, to discriminate high from low and moderate demand, the presence of SRF at V01 and PED at V02 and V03 are the most useful. Using Figure 9E, F, we can assess that the central 3-mm region is very informative for detecting a low demand, whereas almost
all ETDRS regions are important for discriminating high from low and moderate demand. In Figure 11, we note that the maximum probability of SRF at the baseline (V01_SRF) and at the baseline and second visit (V02_SRF) shows a clear difference between the low and the 2 other demand classes. Putting these data into clinical perspective, we may refer to previous post hoc analyses that outlined that patients with SRF needed less intensive treatment and its presence seems predictive for superior visual function gain.18,19 However, also the dynamics of SRF may play a crucial role. Although small, stable amounts of SRF do not seem to influence morphological and functional outcomes,22,23 and larger amounts of SRF at baseline and new occurrence of SRF seemed unfavorable and are a sign of disease activity.18,24 Although PED per se is not considered to be a retreatment indication for TER, we noted in Figures 7B and 9D that some features related to PED at V02 and V03 contributed significantly to the final prediction of the HvO classifier. This suggests that, when the spread on the en face area map or the volume of PED increases, or both, clinicians shorten the treatment interval. This is consistent with previous observation that PED is associated with higher treatment need and baseline PED with consecutive development of IRF is associated with higher treatment...
Figure 9. Sum of relative importance scores for both neovascular age-related macular degeneration classifiers at visit 3 (V03) over all 10 folds and considering 3 different perspectives. The relative importance scores reflect the relative contribution of each input feature in the final prediction of the classifier. A, B, Relative importance scores as a function of the origin of the input feature to the classifier, that is, if the feature was constructed from the segmentation method or the biomarker detector or if the feature was nonmorphological (age or sex of the patient). C, D, Relative importance scores as function of the type of input morphological features to the classifier, that is, retinal layers or fluids. E, F, Relative importance scores as function of the location of segmentation-based morphological features using the Early Treatment Diabetic Retinopathy Study (ETDRS) grid. To measure this contribution, we summed the relative importance scores across all 10 folds. The segmentation-based morphological features (SEG) comprise 131 features per input OCT volume (i.e., time point), which measure retinal layers and fluids through their thickness and en face area map or volume, respectively, at different regions of the retina. The biomarker-based morphological features (DET) comprise 20 features per input OCT volume (i.e., time point), which measure the presence of 10 biomarkers over the entire volume. HvO = high versus others; IRF = intraretinal fluid; LvO = low versus others; PED = pigment epithelial detachment; SRF = subretinal fluid; V01 = visit 1; V02 = visit 2; w.r.t. = with regard to.
Figure 10. Sum of relative importance scores for both classifiers at visit 3 (V03) for retinal vascular diseases over all 10 folds and considering 3 different perspectives. The relative importance scores reflect the relative contribution of each input feature in the final prediction of the classifier. A, B, Relative importance scores as function of the origin of the input feature to the classifier, that is, if the feature was constructed from the segmentation method or the biomarker detector or if the feature was nonmorphological (age or sex of the patient). C, D, As a function of the type of input morphological features to the classifier, that is, retinal layers or fluids. E, F, As function of the location of segmentation-based morphological features using the Early Treatment Diabetic Retinopathy Study (ETDRS) grid. To measure this contribution, we summed the relative importance scores across all 10 folds. The segmentation-based morphological features (SEG) comprise 131 features per input OCT volume (i.e., time point) that measure retinal layers and fluids through their thickness and en face area map or volume, respectively, at different regions of the retina. The biomarker-based morphological features (DET) comprise 20 features per input OCT volume (i.e., time point) that measure the presence of 10 biomarkers over the entire volume. HvO = high versus others; IRF = intraretinal fluid; LvO = low versus others; PED = pigment epithelial detachment; SRF = subretinal fluid; V01 = visit 1; V02 = visit 2; w.r.t. = with regard to.
demand and inferior BCVA improvement. Also, PED is associated with more frequent disease recurrences.\cite{2,25,26}

Such an explanation sounds reasonable because, in practice, clinicians may consider, with a certain degree of discretion, PED to be a retreatment criterion when its size is too important. Therefore, an increase in PED size currently is considered to be disease activity in nAMD, which requires intensification of treatment.\cite{27}

Figure 11. Box-and-whisker plots showing the value distribution of the top 3 most important features for the classifiers low versus others (LxO) and high versus others (HxO) for patients with neovascular age-related macular degeneration across the 3 time points. For each graph, the upper title (e.g., top left, V01_SRF - LxO & HxO) indicates the name of the feature (V01_SRF) and the classifier type (LxO and HxO) for which it is on the top 3 most important features. Note that some of these features are present in the top 3 of both classifiers and at different time points. FPED = fibrovascular pigment epithelial detachment; IRF = intraretinal fluid; PED = pigment epithelial detachment; RT = retina thickness; SRF = subretinal fluid; V01, V02, V03 = visit 1, 2, and 3, respectively.
Classifiers for patients with ret. vasc. diseases at the three time points

Figure 12. Box-and-whisker plots showing the value distribution of the top 3 most important features for the classifiers low versus others (LvO) and high versus others (HvO) for patients with retinal vascular diseases across the 3 time points. For each graph, the upper title (e.g., top left, V01_IRF - LvO) indicates the name of the feature (V01_IRF) and the classifier type (HvO) for which it is on the top 3 most important features. Note that some of these features are present in the top 3 of both classifiers and at different time points. FPED = fibrovascular pigment epithelial detachment; IRF = intraretinal fluid; PED = pigment epithelial detachment; RT = retina thickness; SRF = subretinal fluid; V01, V02, V03 = visit 1, 2, and 3, respectively.
We used a Wilcoxon rank-sum test (with a significance value of \( \alpha = 0.05 \)) to compare the two classes. All the other 60 tested features were not statistically significant between the two classes.

**Retinal Vascular Diseases**

In particular, all IRF-related features seem important in retinal vascular diseases for the prediction of low and high treatment demand, respectively (Fig 8). Also, retinal thickness parameters seem to be informative for the classifier (Fig 10C). Consistently with nAMD, the LvO classifiers rely predominantly on features gathered at baseline, whereas the HvO classifiers rely on the latest image data. Figure 10A–D reveals that, for both classifiers, the segmentation-based morphological features of IRF clearly are the most important ones. However, although the various features of IRF at baseline (V01) were very discriminative to detect low demand, the second visit (V02) provides the most informative feature material to distinguish high from low and moderate demand. The distribution and location of important features are important (Fig 10E, F). To classify high demand, the central 1-mm region is particularly the most informative region, whereas most of the retina is important to classify low demand. In Figure 12, the top 3 features related to IRF show notable differences between low and high demand. This is the case for the maximum presence probability (V01_IRF) and the fluid volumes in some sections of the 3-mm and 6-mm ETDRS region.

We also analyzed the statistical significance of the features (Tables 3 and 4) used in the classifiers and the contributions of these features in the final prediction (Figs 7–12). These results support that the construction of the classifiers (via the tree growth of the random forest technique) follows the clinician indications for extending and reducing the interval between anti-VEGF injections. For instance, in Figures 11 and 12, the value distribution of the top 3 most important features among the 3 demand classes points out that differences in median and spread are consistent with the results of the statistical significance analysis.

<table>
<thead>
<tr>
<th>Time Points</th>
<th>Feature Type</th>
<th>Feature Description</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>V01, V02, V03</td>
<td>Segmentation</td>
<td>Retinal thickness, central 6/3/1 mm</td>
<td>9</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Segmentation</td>
<td>SRF, area and volume central 6/3/1 mm</td>
<td>18</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Segmentation</td>
<td>IRF, area and volume central 6/3/1 mm</td>
<td>18</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Segmentation</td>
<td>PED, area and volume central 6/3/1 mm</td>
<td>18</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Segmentation</td>
<td>Patient-independent retinal thickness</td>
<td>3</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Biomarker</td>
<td>SRF type 1 and 2</td>
<td>6</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Biomarker</td>
<td>IRF type 1 and 2</td>
<td>6</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Biomarker</td>
<td>PED type 1 and 2</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>84</td>
</tr>
</tbody>
</table>

IRF = intraretinal fluid; PED = pigment epithelial detachment; SRF = subretinal fluid; V01, V02, V03 = visit 1, 2, and 3, respectively.

### Table 3. Results of the Statistical Significance Test for a Subset of Features for the 3 Demand Classes among the Patients with Neovascular Age-Related Macular Degeneration

<table>
<thead>
<tr>
<th>Between low and high demand classes</th>
<th>Feature Type</th>
<th>Feature Description</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V02, V03</td>
<td>Biomarker</td>
<td>IRF type 2</td>
<td>&gt; 0.14</td>
</tr>
<tr>
<td>V03</td>
<td>Biomarker</td>
<td>IRF type 1</td>
<td>&gt; 0.15</td>
</tr>
<tr>
<td>All the other 75 tested features were statistically significantly different between the 2 classes</td>
<td></td>
<td></td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Between low and moderate demand classes</th>
<th>Feature Type</th>
<th>Feature Description</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V01</td>
<td>Segmentation</td>
<td>Retinal thickness, central 6/3/1 mm</td>
<td>&lt; 3.8e-6</td>
</tr>
<tr>
<td>V01</td>
<td>Segmentation</td>
<td>Patient-independent retinal thickness</td>
<td>&lt; 2.2e-5</td>
</tr>
<tr>
<td>V01, V02</td>
<td>Segmentation</td>
<td>IRF and SRF, area and volume central 6/3/1 mm</td>
<td>&lt; 6.0e-4</td>
</tr>
<tr>
<td>V01</td>
<td>Biomarker</td>
<td>SRF type 1 and 2</td>
<td>&lt; 2.9e-8</td>
</tr>
<tr>
<td>V01</td>
<td>Biomarker</td>
<td>IRF type 1 and 2</td>
<td>&lt; 5.4e-6</td>
</tr>
<tr>
<td>V02</td>
<td>Segmentation</td>
<td>IRF, area central 3 mm and volume central 6/3 mm</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>V03</td>
<td>Segmentation</td>
<td>SRF, volume central 6 mm</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>All the other 60 tested features were not statistically significantly different between the 2 classes</td>
<td></td>
<td></td>
<td>&gt; 0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Between moderate and high demand classes</th>
<th>Feature Type</th>
<th>Feature Description</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V01</td>
<td>Segmentation</td>
<td>PED, volume central 3/1 mm</td>
<td>&lt; 0.038</td>
</tr>
<tr>
<td>V01, V02</td>
<td>Biomarker</td>
<td>SRF type 1</td>
<td>&lt; 0.029</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Biomarker</td>
<td>SRF type 2</td>
<td>&lt; 0.013</td>
</tr>
<tr>
<td>V01, V02, V03</td>
<td>Biomarker</td>
<td>PED type 2</td>
<td>&lt; 0.0044</td>
</tr>
<tr>
<td>V02, V03</td>
<td>Segmentation</td>
<td>Retinal thickness, central 6/3 mm</td>
<td>&lt; 0.02</td>
</tr>
<tr>
<td>V02, V03</td>
<td>Segmentation</td>
<td>PED, area central 6/3 mm and volume central 6/3/1 mm</td>
<td>&lt; 0.013</td>
</tr>
<tr>
<td>All the other 60 tested features were not statistically significantly different between the 2 classes</td>
<td></td>
<td></td>
<td>&gt; 0.05</td>
</tr>
</tbody>
</table>

IRF = intraretinal fluid; PED = pigment epithelial detachment; SRF = subretinal fluid; V01, V02, V03 = visit 1, 2, and 3, respectively.

We used a Wilcoxon rank-sum test (with a significance value of \( \alpha = 0.05 \)) and the Holm-Bonferroni\(^1\) correction to mitigate the effect of multiple comparisons.
A core element of this work that has a certain clinical impact is the definition of high and low treatment demand. No consensus exists regarding its definition or, more generally, regarding the definition of good or poor response. Treatment demand also may change over time over the course of the disease, as Iacono et al. reported in 2018 that the label of, for example, ranibizumab supports TER not only in age-related macular degeneration, also in DME and RVO. Also, the European Medicines Agency reported in 2018 that the label of, for example, ranibizumab supports TER not only in age-related macular degeneration, but also in DME and RVO. Interestingly, when looking at the feature importance, we found similar features for the prediction in a TER compared with other studies that evaluated feature importance and treatment prediction in a PRN regimen, and so it could mean that the exact treatment regimen used is not as crucial, although this would need to be validated in future studies. The second point that weakens, at first sight, the generalizability of our work is the absence of 3 loading doses in the TER we used. We underline here 4 main reasons that support that our algorithm remaining relevant when a loading phase is used and our data is significant for the clinical and scientific community. First, not all physicians using a TER are actually implementing a loading dose of 3 injections. Second, we acknowledge the exact treatment regimen used is not as crucial, although this would need to be validated in future studies. Nevertheless, in our TER, the third dose would be given at strict 4-week intervals is also a frequently used strategy. Third, the loading dose targets a maximum initial treatment response, which is always achieved in a TER because consensus exists for administering a maximum initial treatment response, which is always achieved in a TER because consensus exists for administering a loading dose. In our TER, we administer 2 loading doses in the TER we used. We underline here 4 main reasons that support that our algorithm remaining relevant when a loading phase is used and our data is significant for the clinical and scientific community. First, not all physicians using a TER are actually implementing a loading dose of 3 injections. Second, we acknowledge that variation exists among practitioners in the number of given loading doses. In our TER, we administer 2 loading doses before considering extension, but 3 loading doses at strict 4-week intervals is also a frequently used strategy. Nevertheless, in our TER, the third dose would be given at either 4- or 6-week intervals, and this minimal difference may not have a significant impact on the overall prediction results. In future, external testing of our prediction algorithms in cohorts using a 3-loading dose TER may help to show generalizability. Third, the loading dose targets a maximum initial treatment response, which is always achieved in a TER because consensus exists for administering monthly injections until disease inactivity is achieved. Also, several large prospective trials in which a loading phase was not mandatory achieved similar results as trials.
using a loading dose.\textsuperscript{16,40,41} Fourth, a recent publication\textsuperscript{42} observed that a loading dose is not necessary to achieve results in a real-life setting, findings that are comparable with the outcome of clinical trials.

The clinical implication and relevance of these findings are manifold. Being able to tell at baseline, before starting treatment, whether the patient will respond very well or rather poorly is of great interest because patients usually want to know their prognosis. It also may be used to adapt the TER in terms of treatment extension intervals with faster and longer TER intervals in case of low demand. However, additional experiments and clinical considerations remain necessary to assess the best integration strategy of such an algorithm in the clinical routine and to avoid undesired phenomena such as the self-fulfilling prophecy (high demand can be predicted as moderate or low demand, and then its treatment interval is extended, defining automatically this demand as moderate or low). Assessing the best time point to use this prediction and defining its importance in the decision-making process will be investigated in the future.

**Study Limitations**

This study has 6 main limitations. First, even if the cohort size is large enough to reach significant conclusions with the machine learning technique we used, a similar study with a larger cohort in terms of number of patients and overall treatment time would be beneficial. A consequent limitation can be that patients with DME and ME-related RVO were considered as a single group. Although both diseases belong to the retinal vascular disease spectrum, their specific behavior and demand for treatment can differ and were not captured and examined separately in our study.

Second, the classifier performance depends strongly on the performance of the semantic segmentation and detection methods and on the image quality of the B-scans. For
instance, we observed that, in very few B-scans, the automated segmentation method can misinterpret between SRF and PED and that some corrupted B-scans (noisy acquisition, intermittent offset at the B-scan level, etc.) lead to incorrect or incomplete semantic segmentation and detection through the whole volume. In Figure 6A, the OCT acquired at V02 illustrates that noisy B-scans can lead to retina layers being segmented or classified as fluids.
However, these corruptions or incorrect segment classifications are minimized thanks to 2 elements of our method. The first element is the construction of the input features, which relies on the aggregation of the segmentation and detection results through various operators (area, volume, mean value, difference value) and the 13 ETDRS regions. This allows eradication of the errors at the B-scans level as long as they are not dominant in 1 of the 13 ETDRS regions. The second element is the joint use of 2 different algorithms to extract the morphological features.

Third, another type of noise is present in this real-world cohort data: noisy treatment intervals. This comes from the fact that discrepancies exist between the prescribed treatment intervals and the actual ones because of visit rescheduling or other unknown real-world factors. This may impact the association of some patients with the treatment demand groups. However, this impact cannot be evaluated with the current data. Note that the treatment intervals we used to define treatment demand are not the prescribed ones, but the true ones.

Fourth, BCVA, additional demographic data, and data from other image methods such as fluorescein angiography were not included as in a study using a PRN regimen, because these data were not available for our retrospective study. However, in the same study, the analysis of the feature importance revealed that data had medium importance (for BVCA) or were not important (for the demographic data and fluorescein angiography) in the decision-making process. The moderate contribution of BVCA can be surprising; however, it underlines the impact of the treatment demander definition. In fact, if the demander classes were defined by a functional improvement of BCVA, then BCVA would be the most important feature. Given that the morphological OCT data and not BCVA are used to determine disease activity and treatment demand in clinical practice, it seems reasonable that the algorithm focuses on morphological data.

Fifth, as Table 1 reports, patients involved in the cohort generally are elderly, which prevents us from drawing conclusions regarding the generalization of our algorithm to younger patients. However, we note that the mean ages of our cohorts are representative for the 2 pathologic groups: 78 years for the nAMD group and 66 years for the retinal vascular disease group. The mean age for the multiple nAMD groups studied by Heier et al was 76 years. The mean age for the multiple DME groups studied by Korobelnik et al was 63 years. Brown et al reported a mean age of 66 years for patients with RVO.

Sixth, as in the study by Bogunović et al this study assumed that both eyes of the same patient are independent. In our study, for each fold, no patient was present in both the training and the test set. It is reasonable initially to assess the capabilities of machine learning with respect to the prediction task this way. However, in practice, a correlation exists between both eyes of a patient, and in the study cohort, we observed that a significant number of patients are treated for both eyes simultaneously.

Conclusions and Future Work

This study showed that it is possible to predict the individual long-term demand of anti-VEGF therapy in a TER. Low treatment demand can be identified with reasonable accuracy already at baseline before the start of treatment, which underlines the potential to improve individualization of treatment regimens at a very early stage. We were able to report promising prediction performance using real patient data and for 2 major disease entities, nAMD and retinal vascular disease. Using a random forest strategy allowed us to interpret the decision-making process of the trained classifiers, which is a strong asset for introducing such tools into clinical practice. Along with other relevant biomarkers and other categories or subcategories of pathologic features, future research will focus on other demand definitions based on morphological changes, that is, changes of SRF, IRF, and PED, and the prediction of the average treatment interval to gain further insights into the individual demand of patients.

Footnotes and Disclosures

Originally received: December 10, 2020.
Final revision: May 1, 2021.
Accepted: May 3, 2021.

1 AIMI, ARTORG Center, University of Bern, Bern, Switzerland.
2 Department for Ophthalmology, Inselspital, University Hospital, University of Bern, Bern, Switzerland.
3 RetinAI Medical AG, Bern, Switzerland.

Disclosure(s): All authors have completed and submitted the ICMJE disclosures form.
The author(s) have made the following disclosure(s): M.R.M.: Consultant - Zeiss, Novartis, Luminitera, Allergan, RetinAI Medical AG, Isarna; Financial support - Bayer, Novartis; Equity owner - RetinAI Medical AG T.K.: Financial support - Innoussais M.S.Z.: Consultant - Bayer, Novartis, Roche; Financial support - Bayer, Lecturer - Novartis; Equity owner - Novartis S.D.Z.: Employee and Equity owner - RetinAI Medical AG A.M.: Employee and Equity owner - RetinAI Medical AG S.W.: Consultant - Bayer, Novartis, Roche, Chengdu Kanghong Biotechnology, Zeiss, Heidelberg Engineering; Nonfinancial support - Zeiss R.S.: Financial support - Optona-Update; Equity owner - RetinAI Medical AG Supported in part by EUROSTARS, programme co-funded by Eureka and the European Union Horizon 2020 (grant no.: E!12712).

HUMAN SUBJECTS: Human subjects were included in this study. The human ethics committees at the University of Bern approved the study. All research adhered to the tenets of the Declaration of Helsinki. Informed consent was waved because of the retrospective design of the study.
No animal subjects were included in this study.

Author Contributions:
Conception and design: Gallardo, Munk, Wolf, Sznitman
Analysis and interpretation: Gallardo, Munk, Wolf, Sznitman
Data collection: Gallardo, Kurmann, De Zanet, Mosinska, Karagöz, Zinkernagel, Wolf

Obtained funding: Sznitman

Overall responsibility: Gallardo, Munk, Sznitman

Abbreviations and Acronyms:

AUC = area under the receiver operating characteristic curve;
BCVA = best-corrected visual acuity; DME = diabetic macular edema;
ETDRS = Early Treatment Diabetic Retinopathy Study; fl = fluid versus others; IRF = intraretinal fluid; LvO = low versus others; ME = macular edema; nAMD = neovascular age-related macular degeneration; PED = pigment epithelial detachment; PRN = pro re nata; RVO = retinal vein occlusion; SRF = subretinal fluid; TER = treat-and-extend regimen; VEGF = vascular endothelial growth factor; V01, V02, V03 = visit 1, 2, and 3, respectively.

Keywords:

Anti-VEGF therapy, Machine learning, OCT, Treatment demand.

Correspondence:

Mathias Gallardo, PhD, AIMI, ARTORG Center, University of Bern, Muttenstrasse 50, 3008 Bern, Switzerland. E-mail: Mathias.Gallardo@gmail.com.

References


**Pictures & Perspectives**

**Coats-like Exudative Vitreoretinopathy in NMNAT1 Leber Congenital Amaurosis**

A now 6-year-old girl with NMNAT1 Leber congenital amaurosis (p.Arg66Trp/p.Arg237Cys) was known to have macular coloboma-like degeneration and early retinal vessel attenuation in both eyes at the age of 6 months (A). Yellowish subretinal exudation with tel-angiectasia (arrow) was noted only in the right eye at 6 years of age, suggestive of Coats-like exudative vitreoretinopathy (B). A 1.5-disc size endophytic mass (arrow) without calcification was noted on ultrasonography (C). Coats-like exudative vitreoretinopathy is known to appear mainly in retinal dystrophy caused by CRB1, CRX, CEP290, PRPF8, and RPGR mutations, but it can occur in NMNAT1 Leber congenital amaurosis.

JINU HAN, MD
TAE YOUNG KIM, MD
MIN KIM, MD, PhD

Institute of Vision Research, Department of Ophthalmology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea