Multitask Deep Learning for Joint Detection of Necrotizing Viral and Noninfectious Retinitis From Common Blood and Serology Test Data

Kai Tzu-iunn Ong,[1] Taeyoon Kwon,[1] Harok Jang,[1] Min Kim,[2] Christopher Seungkyu Lee,[3] Suk Ho Byeon,[3] Sung Soo Kim,[3] Jinyoung Yeo,[1] and Eun Young Choi[2]

[1]Department of Artificial Intelligence, Yonsei University College of Computing, Seoul, Republic of Korea
[2]Department of Ophthalmology, Institute of Vision Research, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
[3]Department of Ophthalmology, Institute of Vision Research, Severance Eye Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

Correspondence: Eun Young Choi, Department of Ophthalmology, Institute of Vision Research, Gangnam Severance Hospital, Yonsei University College of Medicine, 211, Eonjuro, Gangnam-gu, Seoul 06229, Republic of Korea; eychoi@yuhs.ac.
Jinyoung Yeo, Department of Artificial Intelligence, College of Computing, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea; jinyeo@yonsei.ac.kr.

KTO and TK are the co-first authors.

Received: June 15, 2023
Accepted: January 9, 2024
Published: February 2, 2024

Citation: Ong KT, Kwon T, Jang H, et al. Multitask deep learning for joint detection of necrotizing viral and noninfectious retinitis from common blood and serology test data. *Invest Ophthalmol Vis Sci.* 2024;65(2):5. https://doi.org/10.1167/iovs.65.2.5

Purpose. Necrotizing viral retinitis is a serious eye infection that requires immediate treatment to prevent permanent vision loss. Uncertain clinical suspicion can result in delayed diagnosis, inappropriate administration of corticosteroids, or repeated intraocular sampling. To quickly and accurately distinguish between viral and noninfectious retinitis, we aimed to develop deep learning (DL) models solely using noninvasive blood test data.

Methods. This cross-sectional study trained DL models using common blood and serology test data from 3080 patients (noninfectious uveitis of the posterior segment [NIU-PS] = 2858, acute retinal necrosis [ARN] = 66, cytomegalovirus [CMV], retinitis = 156). Following the development of separate base DL models for ARN and CMV retinitis, multitask learning (MTL) was employed to enable simultaneous discrimination. Advanced MTL models incorporating adversarial training were used to enhance DL feature extraction from the small, imbalanced data. We evaluated model performance, disease-specific important features, and the causal relationship between DL features and detection results.

Results. The presented models all achieved excellent detection performances, with the adversarial MTL model achieving the highest receiver operating characteristic curves (0.932 for ARN and 0.982 for CMV retinitis). Significant features for ARN detection included varicella-zoster virus (VZV) immunoglobulin M (IgM), herpes simplex virus immunoglobulin G, and neutrophil count, while for CMV retinitis, they encompassed VZV IgM, CMV IgM, and lymphocyte count. The adversarial MTL model exhibited substantial changes in detection outcomes when the key features were contaminated, indicating stronger causality between DL features and detection results.

Conclusions. The adversarial MTL model, using blood test data, may serve as a reliable adjunct for the expedited diagnosis of ARN, CMV retinitis, and NIU-PS simultaneously in real clinical settings.

Keywords: necrotizing viral retinitis, acute retinal necrosis, cytomegalovirus retinitis, common blood test, viral serology, deep learning, multitask learning, adversarial training

Viral retinitis is a rare ocular infection caused by the Herpesviridae family.[1] The disease spectrum is determined by host immunity and causative virus. Acute retinal necrosis (ARN) is predominantly caused by the varicella-zoster virus (VZV) following the herpes simplex virus (HSV), and immunocompetence is a common feature.[2] At the same time, cytomegalovirus (CMV) retinitis is associated with immunocompromise.[3] Both types of necrotizing viral retinitis share clinical features with full-thickness necrotizing due to occlusive vasculitis with varying degrees of panuveitis. Early diagnosis and appropriate antiviral treatment can produce a relatively good visual prognosis.[4] However, delayed diagnosis might lead to severe complications. In patients with ARN, retinal detachment occurs in 70% of cases, and up to 50% have severe visual impairment <20/200.[5] One-third of patients with CMV retinitis have vision loss due to severe macular complications.[6,7]

ARN[8] and CMV retinitis[9] are clinically diagnosed through a physical examination using a slit lamp and indirect ophthalmoscopy, supported by fundus photography, optical coherence tomography, and fluorescein angiography (left column, Fig. 1). Early diagnosis of necrotizing viral retinitis is difficult because ocular manifestations may overlap with noninfectious uveitis of the posterior segment (NIU-PS).[10] Differential diagnosis can become even more challenging

Copyright 2024 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from iovs.arvojournals.org on 04/25/2024

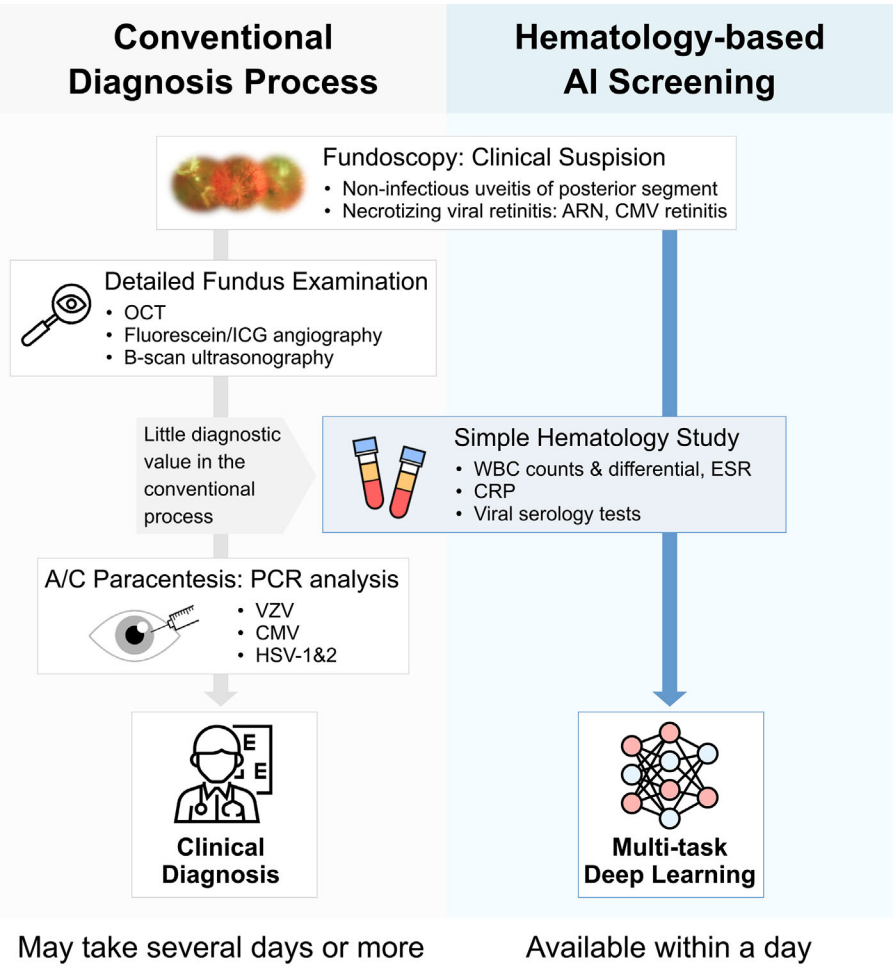


Figure 1. A schematic comparison between the conventional process for diagnosing necrotizing viral retinitis and a proposed artificial intelligence (AI) framework based on blood tests. We hypothesized that necrotizing viral retinitis could be detected early through a cost-effective AI framework, without detailed ophthalmologic examination or invasive PCR testing. A/C, anterior chamber; HSV-1, HSV type 1; HSV-2, HSV type 2; ICG, indocyanine green; OCT, optical coherence tomography.

as the disease progresses if atypical fundus findings are observed or significant vitreous inflammation obscures the fundus.[11] Differentiating between these conditions is crucial as the treatment for necrotizing viral retinitis requires antiviral drugs, while the treatment for NIU-PS generally involves immunosuppressants. Administering corticosteroids before specific antiviral coverage can accelerate the progression of retinal necrosis and increase the risk of permanent vision loss.

Polymerase chain reaction (PCR) analysis has enhanced diagnostic accuracy by detecting viral DNA in intraocular fluids (left column, Fig. 1).[12] However, in primary clinics, both obtaining specimens and performing PCR tests can be difficult. Anterior chamber paracentesis for aqueous humor sampling is invasive[13] and usually only tests one viral etiology at a time.[14] Vitreous specimens can be examined for multiple etiologies, but vitrectomy surgery is often required.[15] On the other hand, common blood and serology tests are easy to perform and require less advanced equipment, but their value in detecting necrotizing viral retinitis has been limited. While white blood cell (WBC) counts can infer a patient's immune status, nonspecific inflammation markers and viral serology provide little diagnostic information. Nevertheless, because common blood and serology tests are more accessible than intraocular fluids, inspired by clinlabomics,[16] which address the clinical application of easily accessible data, we presume that developing methods to differentiate necrotizing viral retinitis based on them could help streamline clinical decision-making, avoid diagnostic delays, and discover their underexplored diagnostic values in ophthalmology.

Recently, deep learning (DL) techniques have garnered attention in the field of ophthalmology. For instance, deep neural networks (DNNs) with multiple layers have been implemented in many studies to automatically extract relevant features from fundus images for the detection of various retinal diseases without requiring manual feature engineering.[17–19] Multitask learning (MTL) is used to resolve the lack of training data by sharing features across multiple target tasks,[20–23] and adversarial training is also used to address the limited availability of data as well as the interpretability and reliability of detection models.[24] Although there has been recent success using MTL to improve DL feature extraction in common retinal diseases,[24,25] few DL-based approaches have been applied to differentiate rare retinal diseases such as viral retinitis. This could partially be due to the limited

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

size of data and the imbalance of normal versus disease data sets.

Motivated by this, we present several detection models that can accurately differentiate necrotizing viral retinitis and NIU-PS using common blood and serology test data via different DL techniques. Our models range from base multilayer perceptron (MLP)–based models that target ARN and CMV retinitis separately to various MTL models that may be able to simultaneously discriminate the three target diseases in situations where the availability of data is limited. After developing the models, we conducted a three-step evaluation process. First, we assessed the detection performance of the models. Second, we explored model interpretability by analyzing the significant input features for detecting ARN and CMV retinitis. Last, we evaluated the reliability of the models by quantifying their ability to reflect the causal relationship between predictions and input features. Our ultimate goal was to verify the accuracy and reliability of the most advanced MTL model, which used adversarial training (right column, Fig. 1), for simultaneously screening the three diseases using only blood test data.

Methods

Data Set and Participants

This retrospective, cross-sectional study was conducted based on patient medical records from two tertiary referral hospitals in South Korea (Severance Eye Hospital and Yongin Severance Hospital) between November 2006 and December 2022. This study was conducted under the Declaration of Helsinki, and the institutional review board approved the protocol of Yonsei University Medical Center (approval number: 2022-1128-001). The requirement for informed consent was waived as the data retrospectively used in this study were anonymized beforehand.

We extracted the electronic medical data of patients who were diagnosed with noninfectious posterior uveitis, ARN, or CMV retinitis. ARN was clinically diagnosed based on the criteria published by the American Uveitis Society[26] with the identification of the pathogen virus such as VZV and HSV. The diagnosis of CMV retinitis is based on clinical features, evidence of immune compromise, and the detection of DNA of CMV using real-time PCR analysis.[27] NIU-PS was diagnosed when the results of multiple clinical examinations and laboratory tests did not meet the aforementioned criteria for necrotizing viral retinitis.

Data regarding the demographic characteristics (age and sex) and the results of laboratory examination of blood were collected on a patient's first visit. Hematologic examinations collected included WBC counts and differential, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and viral antibody test. The CRP level was measured by immunoturbidimetry using Cobas 8000 (Roche, Mannheim, Germany). For WBC measurements, an automated cell counter (ADVIA 2120 Hematology System; Siemens, Eschborn, Germany) was used. Quantitative and qualitative determination of specific immunoglobulin G (IgG) and immunoglobulin M (IgM) antibodies to the pathogenic virus was performed based on immunoassay methods, such as chemiluminescence immunoassay (CLIA) and enzyme-linked fluorescent assay (ELFA). Antibodies against VZV, HSV type 1, and HSV type 2 were detected using the LIAISON XL Analyzer (DiaSorin S.p.A., Saluggia, Italy) employing the CLIA method. VIDAS (BioMérieux, Lyon, France) was used as an automated ELFA for the detection of anti-CMV IgG and IgM antibodies. As input data, antibody concentration was used for quantitative tests for input data, and cutoff index values were used for qualitative tests. The results of the real-time PCR assay for causative viral agents of aqueous humor or vitreous specimens were not included in the data set.

Statistical Analysis

Data statistics of demographic and hematologic data collected at the time of subject diagnosis before preprocessing can be found in Supplementary Table S1. Continuous variables were presented as mean ± SD, and comparisons between groups were performed using one-way analysis of variance. Categorical variables were presented as numbers with percentages of the group total, and comparisons between groups were analyzed using either Fisher exact test or the χ^2 test. A P value <0.05 was considered statistically significant. Logistic regression analysis using the enter method was conducted to assess the associations between the two target diagnoses and clinical variables. Additionally, intercorrelation analysis was performed to examine the correlation between each variable. All statistical analyses were performed on a patient basis using SciPy,[28] NumPy,[29] and statsmodels.[30]

Data Preprocessing

For our research purposes, we performed preprocessing on the collected data using the following steps: imputation of missing values, normalization of input data, relabeling, data split, and oversampling. Additional information regarding each data preprocessing step can be found in Supplementary Method S1.

Task Definition

We formulated two classification tasks: ARN detection and CMV retinitis detection, where the model learns to classify (1) ARN from CMV retinitis and NIU-PS and (2) CMV retinitis from ARN and NIU-PS, respectively.

Development of DL Models

With the collected common blood and serology test data of patients, we developed two models using DNNs for ARN detection and CMV retinitis detection, respectively. As shown in Figure 2A, the DNNs are constructed of MLP and have two parts: an embedding block (EMB) that extracts features from inputs and a classification block that converts extracted features into probabilities for its prediction (these models are denoted as "base DNN models" in this article).

To prepare for the potential unavailability of data in real clinical settings and to address the limitation of only being able to test one viral etiology at a time through aqueous humor sampling,[14] we implemented MTL to train models on two tasks simultaneously. In MTL, features of different diseases are shared, which allows better model performance with limited data and the joint detection of ARN and CMV retinitis. Here, we developed two MTL models with different feature-sharing schemes: a fully shared MTL model, where the features of two tasks are completely shared (Fig. 2B), and a shared-private MTL model, which divides common and disease-specific features into shared and private feature

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

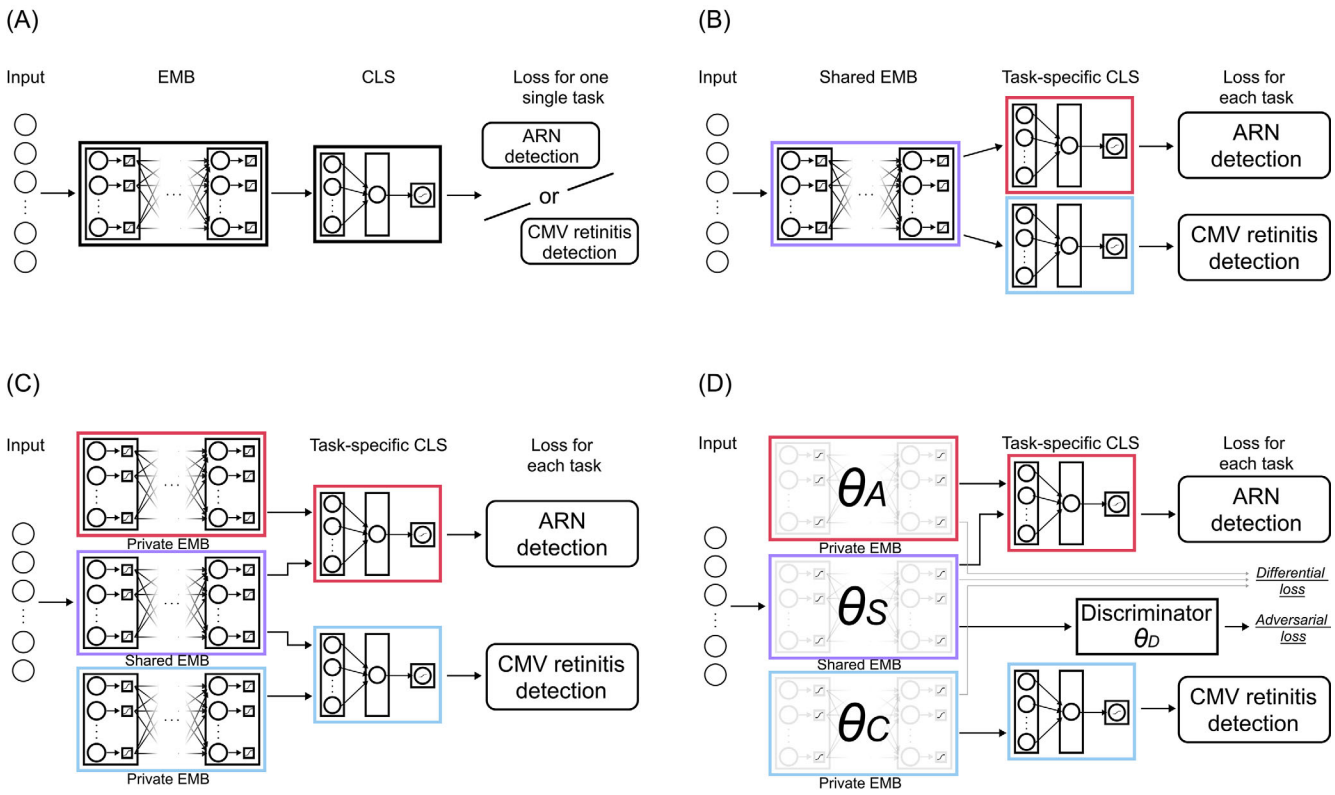


Figure 2. Model architectures: (**A**) base deep neural network model, (**B**) fully shared MTL, (**C**) shared-private MTL, and (**D**) Adversarial MTL. CLS, classification block.

spaces to tackle situations where a specific marker has a differently strong association with ARN and CMV retinitis (Fig. 2C). During training, these two models learn to simultaneously predict the ground-truth labels for ARN detection y_A and CMV retinitis detection y_C by minimizing the loss function $\mathcal{L}_{Task}$, which is the weighted sum of the binary cross-entropy losses $\mathcal{L}_{BCE}$ between the predictions $\hat{y}_A$ and $\hat{y}_C$ and true distribution y_A and y_C:

$$\mathcal{L}_{Task} = \lambda \cdot \mathcal{L}_{BCE}\left(\hat{y}_A, y_A\right) + (1-\lambda) \cdot \mathcal{L}_{BCE}\left(\hat{y}_C, y_C\right)$$

where λ is the preference weight.

Gaining the trust of medical professionals in DL models goes beyond accessing their detection performances. As studied in the preprint by Holzinger and colleagues,[31] it is equally important to ensure that the model's predictions align with human intuition. To achieve this, we incorporated adversarial training into MTL and developed an adversarial MTL model[32] (Fig. 2D) in pursuit of a clearer separation between common and disease-specific features. Our hypothesis is that a more precise separation of features allows the model to better make predictions based on its inputs, resulting in a stronger causal relationship between input features and predictions, which is beneficial for enhancing the reliability of the model. During training, a discriminator works adversarially against the shared embedding block (shared EMB) while using the output of the shared EMB to predict the corresponding disease of the given patient data (ARN, CMV retinitis, or NIU-PS). The main idea is that a properly trained shared EMB should output purer common features (i.e., not extract disease-specific features), such that the discriminator cannot use them to predict the disease correctly. For that, we apply an adversarial loss $\mathcal{L}_{Adv}$ to facilitate purer common features that suppress the discriminator:

$$\mathcal{L}_{Adv} = \min_{\theta_S}(\max_{\theta_D} y_D \log\left[\textbf{Discriminator}\left(h_S, \theta_D\right)\right])$$

where **Discriminator**$(\cdot, \theta_D)$ is the discriminator's predicted probability distribution of the three diseases, and y_D is the true distribution. We further apply a differential loss $\mathcal{L}_{Diff}$, described as orthogonality constraints,[33] to reduce the similarity between the common and disease-specific feature by promoting shared and private embedding blocks to extract nonoverlapped features:

$$\mathcal{L}_{Diff} = \|h_A{}^{\top} h_S\|_F^2 + \|h_C{}^{\top} h_S\|_F^2$$

where $\|\cdot\|_F^2$ is the sum of the squares of all the matrix entries, also known as the squared Frobenius norm, and h_S, h_A, and h_C are the output of embedding blocks for features across diseases, specific to ARN detection and specific to CMV retinitis detection, respectively. Putting these all together, the final loss function $\mathcal{L}_{AMTL}$ for the adversarial MTL model is computed as follows:

$$\mathcal{L}_{AMTL} = \mathcal{L}_{Task} + \tau \mathcal{L}_{Adv} + \gamma \mathcal{L}_{Diff}$$

where τ and γ are preference weights.

The training process, settings of weights, detailed explanations, and the backbone code of all proposed models are provided in Supplementary Method S2.

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

Experiments

Our experiments are guided by the following three research questions (RQs). RQ1: How is the performance of our models in detecting necrotizing viral retinitis? RQ2: How do we interpret the model's working mechanism? RQ3: How reliable is the model prediction?

To evaluate and compare model performance (RQ1) on ARN and CMV retinitis detection, we applied two evaluation metrics: (1) the area under the receiver operating characteristic curve (AUROC) with a random classifier as the baseline (AUROC = 0.50) and (2) the area under the precision-recall curve (AUPRC) with the incidence of diseases in our data set (0.017 for ARN and 0.102 for CMV retinitis) as the baseline.

To analyze model interpretability (RQ2), we first implemented the SHaply Additive exPlanation (SHAP) method[34] to investigate the important input features that affect model predictions in ARN detection and CMV retinitis detection. SHAP is a mathematical method that explains machine learning models' predictions by calculating each feature's contribution to the prediction, providing us with interpretations of the model prediction. SHAP has been widely used in the medical field to understand how a prediction was reached.[35,36] Details on SHAP are provided in Supplementary Appendix S1. After SHAP, we present the two-way partial dependence plots of the important features from SHAP to demonstrate how they are correlated with the diagnosis.

To evaluate the reliability of the models (RQ3), we conducted a counterfactual inference test to investigate the causal relationship between models' predictions and inputs. For that, we measured how different the model predictions are when input features change for the quantification of model causality, following Ong and colleagues.[37] Specifically, taking an original sample x from the test split, we added random Gaussian noise to important features identified via the SHAP analysis to acquire a counterfactual sample $\bar{x}$. We then measured the absolute difference between the predicted probabilities $P(x)$ and $P(\bar{x})$, when given x and $\bar{x}$, respectively. More details on the counterfactual inference test are provided in Supplementary Appendix S2.

Results

The study excluded 6979 patient cases with missing data for each variable of interest. A total of 3080 patient cases are included in this study, which contains 2858 patient cases diagnosed with NIU-PS, 66 patient cases with ARN, and 156 patient cases with CMV retinitis. The demographics and blood test results are summarized in Supplementary Table S1. The results of intercorrelation analysis for each clinical variable are provided in Supplementary Figure S1.

Detection Performance (RQ1)

Figure 3A represents evaluation results in terms of ROC curves and AUROC. We observed that all our models could detect necrotizing retinitis with high performances using common blood and serology test data. Especially, the adversarial MTL model performed best in ARN detection (AUROC = 0.932; 95% confidence interval [CI], 0.846–1.000). As for the CMV retinitis detection, the adversarial MTL model and the fully shared MTL model rank first together with the same AUROC of 0.982 (95% CI, 0.965–1.000 and 95% CI, 0.965–1.000, respectively). Results for AUPRC are presented in Figure 3B. Similarly, all models achieved high detection performances while the adversarial MTL model ranks first on both tasks, with AUPRCs of 0.253 for ARN detection and 0.807 for CMV retinitis detection. Additionally, we developed a classic logistic regression model for both diagnostic tasks using the same set of clinical variables, and the detailed results are presented in Supplementary Table S2. In both tasks, the logistic regression model demonstrated comparable AUROC values to the base DNN model (P value by DeLong's test = 0.808 for ARN detection, 0.532 for CMV retinitis detection).

Model Interpretability (RQ2)

We analyzed model interpretability by investigating important input features that drive the model prediction. We chose the adversarial MTL and fully shared MTL model for the analysis as they have achieved the highest performance in both or either of the tasks. The SHAP results of the adversarial MTL are presented in Figures 4A and 4B (for fully shared MTL, see Supplementary Figure S2). Details on SHAP, including how to read the results, are in Supplementary Appendix S1.

In ARN detection, for adversarial MTL, the top six important features are VZV IgM (17%), HSV IgG (14%), Neutrophil count(#) (11%), WBC(#) (9%), ESR (9%), and Monocyte(#) (9%). These six features account for 69% of the model prediction regarding feature importance. The proportional correlation group for ARN detection includes VZV IgM, HSV IgG, Neutrophil(#), WBC(#), CMV IgG, VZV IgG, CMV IgM, and HSV IgM. The inversely proportional correlation group has ESR, Monocyte(#), CRP, Lymphocyte differential(%), and Monocyte(%). The above features are listed in the order of feature importance.

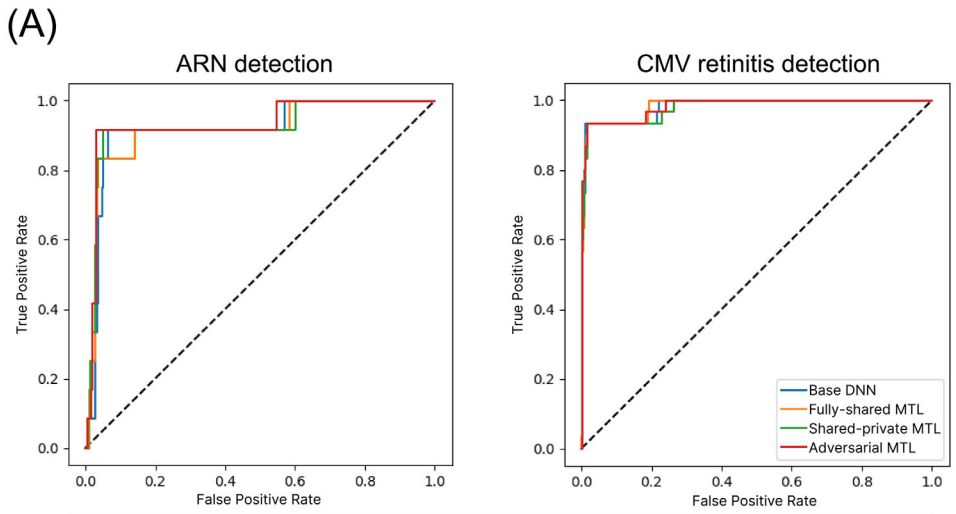
In CMV retinitis detection, the top six important features with regard to CMV retinitis detection are VZV IgM (17%), CMV IgM (13%), Lymphocyte(#) (10%), Lymphocyte differential(%) (13%), CRP (9%), and Monocyte(#) (8%), summing up to 70% of the feature importance. The proportional correlation group for CMV retinitis detection contains CMV IgM, Lymphocyte differential(%), CRP, Monocyte(#), Neutrophil(%), ESR, CMV IgG, VZV IgG, and Monocyte(%). The inversely proportional correlation group has VZV IgM, Lymphocyte(#), Neutrophil(#), and HSV IgG. The above features for CMV retinitis detection are also listed in the order of feature importance. Figures 4C and 4D present the two-way partial dependence plot of the top two important features (by SHAP) for adversarial MTL in ARN and CMV retinitis detection, respectively, using our test split. The results align with the SHAP results. In the context of ARN detection, lower values of VZV IgM and HSV IgG correspond to a decreased probability of ARN, as predicted by the model. In the case of CMV retinitis detection, a higher VZV IgM value and lower CMV IgM values are associated with a reduced probability of CMV retinitis.

Model Reliability (RQ3)

We present the results of the counterfactual inference test in Figure 5. We conducted the test on the adversarial MTL and fully shared MTL models as they achieved the highest performance in both or either of the detection tasks. We applied random Gaussian noise to the values of important features identified by SHAP to obtain counterfactual samples for each detection task. For ARN detection, we modified the

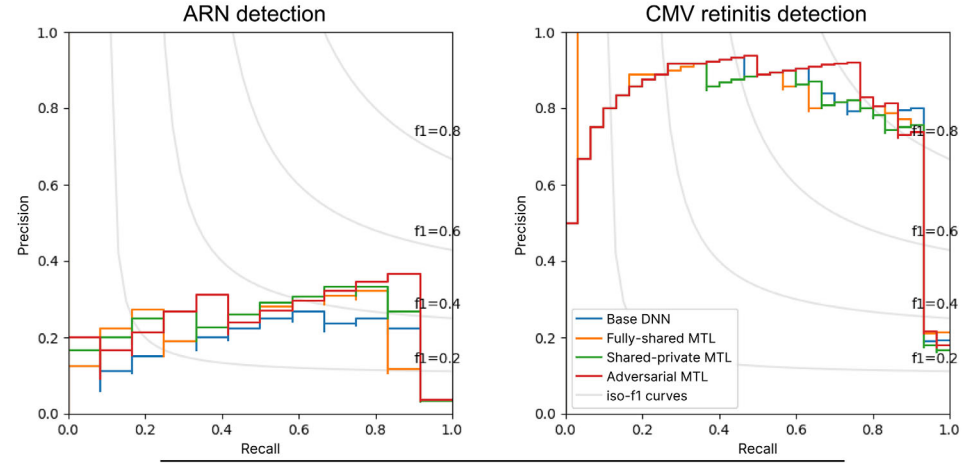
Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024



AUROC (95%CI), *P* Value

Models	ARN detection	CMV retinitis detection
Base DNN (Ref.)	0.919 (0.830, 1.0), -	0.981 (0.961, 1.0), -
Fully-shared MTL	0.918 (0.826, 1.0), 0.948	0.982 (0.965, 1.0), 0.378
Shared-private MTL	0.926 (0.831, 1.0), 0.243	0.976 (0.956, 1.0), 0.153
Adversarial MTL	0.932 (0.846, 1.0), 0.012	0.982 (0.962, 1.0), 0.674



AUPRC

Models	ARN detection	CMV retinitis detection
Base DNN	0.118	0.795
Fully-shared MTL	0.222	0.801
Shared-private MTL	0.245	0.781
Adversarial MTL	0.253	0.807
Prevalence (Ref.)	0.017	0.102

Figure 3. Performance evaluation: (**A**) AUROC comparison of MTL models against the base DNN model on both ARN detection and CMV retinitis detection. (**B**) AUPRC of base DNN and MTL models for ARN and CMV retinitis detection presenting against disease prevalence.

VZV IgM and HSV IgG values, and for CMV retinitis detection, we modified the CMV IgM values.

In the bin charts, the first bin indicates the number of the original-counterfactual sample pairs yielding similar diagnoses where the differences between predicted probabilities are less than 2.5%. The lower the first bin is (i.e., fewer pairs with small *DIF*), the better the model reflects the changes of input features in its prediction. We observed that when given

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

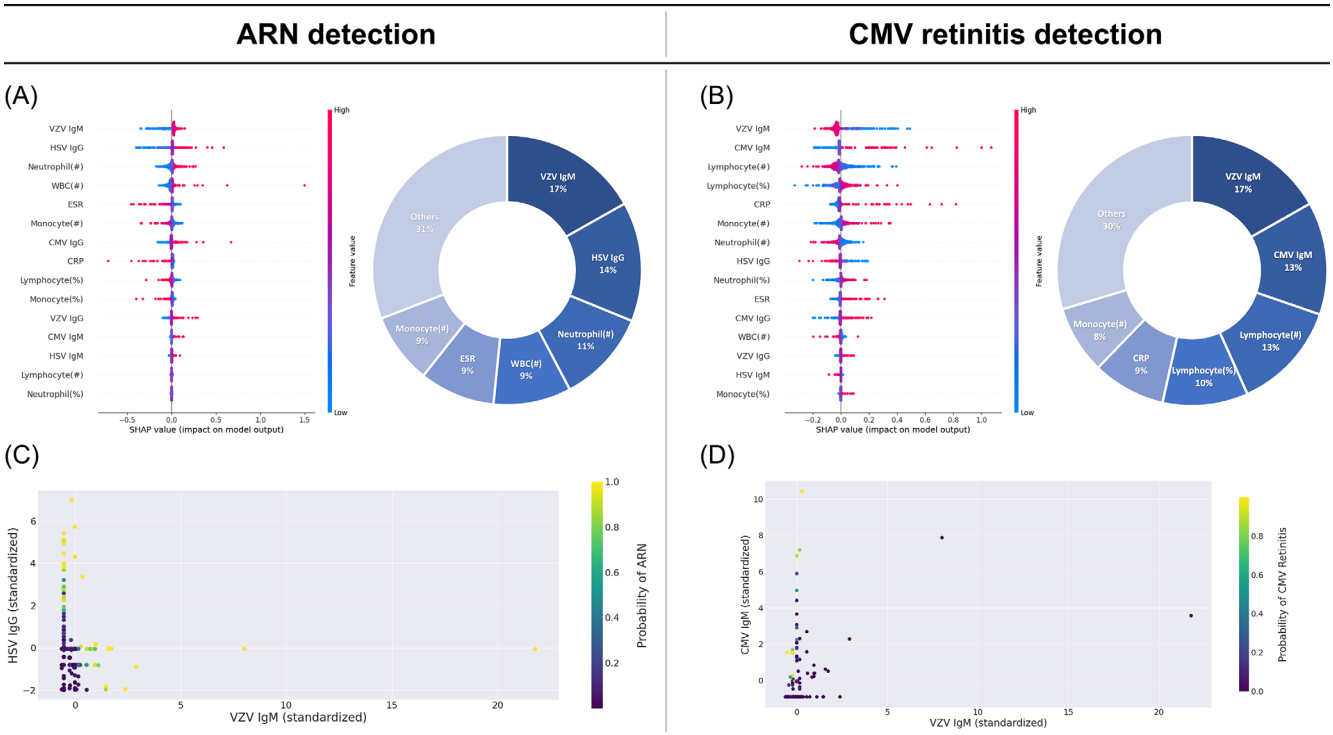


Figure 4. The SHAP results (**A**, **B**) and two-way partial dependence plots (**C**, **D**) of the adversarial multitask learning model: the SHAP results are presented by the beeswarm plots and the pie charts. In the beeswarm plot, the x-axis refers to each feature's impact on model prediction according to its SHAP value, whose absolute value quantifies "how much" the impact is, and the positive/negative value indicates that the feature increases/decreases the predicted probability for the target disease. The color of each dot represents the magnitude of the value of the corresponding input feature, from *red* (high) to *blue* (low). The pie chart visualizes the ranking of feature importance with regard to the model prediction. It is derived from the SHAP values presented in the beeswarm plot. (**A**) The SHAP result for ARN detection. (**B**) The SHAP result for CMV retinitis detection. (**C**) The two-way partial dependence plot for ARN detection using the top two important features obtained with SHAP. (**D**) The two-way partial dependence plot for CMV retinitis detection using the top two important features obtained with SHAP.

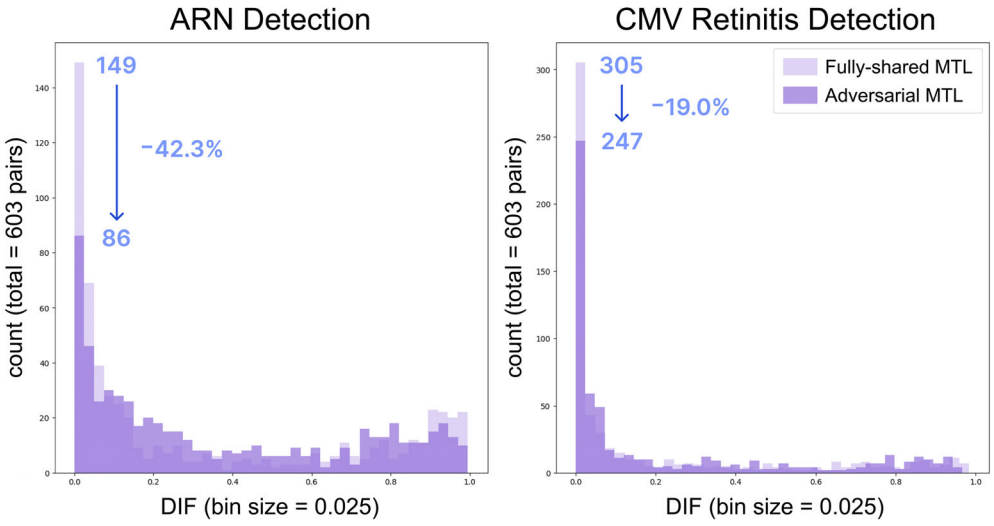


Figure 5. Comparison of counterfactual inference test results for the fully shared and adversarial MTL models on two tasks.

original-counterfactual sample pairs, the adversarial MTL model had a lower count of instances yielding similar predictions (the first bins: $0 \leq DIF < 0.025$) compared to the fully shared MTL model in both ARN detection (42.3% decrease) and CMV retinitis detection (19.0% decrease). These results quantified how the adversarial MTL model demonstrated a stronger causal relationship between input features and model predictions than the fully shared MTL model.

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

Discussion

In this multicenter cross-sectional study, we empirically showed that we could use common blood and serology test data to develop diverse models that accurately differentiate necrotizing viral retinitis from NIU-PS. Of those models, the adversarial MTL model can simultaneously detect ARN and CMV retinitis with the best performance in both tasks in all metrics. Our hematology-based approach is novel, as our models achieved high performances in differentiating viral retinitis without conventional diagnostic procedures such as fundus examination and PCR analysis. This makes screening safer and more accessible in primary care settings while reducing required time, cost, and equipment. By minimizing diagnostic delay, we expect our approach to improve patient outcomes.

The base DNN models could accurately detect one of two rare diseases (ARN or CMV retinitis) from the other and NIU-PS. This indicates that the base DNN models were able to extract valuable diagnostic values for diagnosing these diseases from the blood test data. However, the amount of the collected data was small for joint detection of three diseases at the same time, and DL-based detection models tend to yield suboptimal performances or overfitting in such situations.[38] To address these issues, we employed MTL to train the models on two diagnosis tasks to learn shared features across diseases. Such information sharing encourages the model to learn representations that are generalizable to all target tasks, rather than memorizing task- or class-specific noises and overfitting to individual tasks. Additionally, we introduce adversarial training into the shared-private MTL model (i.e., the adversarial MTL model) to enhance the separation of common and disease-specific features, resulting in further performance gains.

A common remedy to overfitting when developing diagnosis models is transfer learning, which involves pretraining a model with available large data and then fine-tuning it with small data of target diseases. For instance, Zhou et al.[39] recently proposed RETFound, a foundation model pretrained with large-scale (1.6 million) retinal images, which can be used to develop models for diagnosing eye diseases by further fine-tuning with small patient data. However, in the absence of large-scale data or tasks for pretraining, our adversarial MTL has proved to be an efficient method to overcome the limitations posed by small data sets. If a pretrained foundation model appropriate for our tasks becomes available, it would be valuable to investigate the performance improvement achieved by transfer learning with adversarial MTL for fine-tuning.

The unknown diagnostic value of hematology data can be inferred from findings on the important features of each diagnostic task. Antiviral IgG and IgM, the conventional markers of immunity and recent infection,[40] did not help diagnose necrotizing viral retinitis much. Although reduced cellular immunity is a well-known risk factor for ocular viral infections,[41] WBC differential has not been used explicitly to diagnose necrotizing viral retinitis. In the order of importance, the adversarial MTL model found VZV IgM, HSV IgG, Neutrophil(#), and WBC(#) to be important for a positive diagnosis of ARN; ESR and Monocyte(#) are significant for a negative diagnosis of ARN; CMV IgM, Lymphocyte(%), CRP, and Monocyte(#) are important for a positive diagnosis of CMV retinitis; and VZV IgM and Lymphocyte(#) are significant for a negative diagnosis of CMV retinitis. Although no pronounced correlation was found between antiviral antibody concentrations and cutoff index values in each virus pathogen, it is noteworthy that cutoff index values were informative in determining necrotizing viral retinitis. Furthermore, these findings highlight MTL's superiority in developing diagnostic models compared to existing tree-based algorithms. The latter often necessitate manual feature selection, involving processes such as feature engineering. These procedures are time-consuming and heavily reliant on prior knowledge of each feature's diagnostic value.

The counterfactual inference test allowed us to quantify the causality between important input features and model predictions. The results showed that the adversarial MTL model had fewer original-counterfactual sample pairs yielding similar predictions when important input features were altered, demonstrating that the adversarial MTL model better reflects the causal relationship between inputs and predictions than fully shared MTL. Such model causality matches the human thinking process: "Changes in important features lead to a different diagnosis." As elaborated in the preprint by Holzinger and colleagues,[31] this property is essential for enhancing medical professionals' trust in DL models. We hence could confirm the reliability of our adversarial MTL model.

This study has limitations. First, our models do not address other retinal infections caused by viruses (e.g., Epstein–Barr virus retinitis) and other pathogens (e.g., tuberculosis). However, they can distinguish the two most sight-threatening infectious retinitis (either separately or simultaneously) from noninfectious uveitis. Second, while our MTL diagnostic model has shown promise in achieving high accuracy and reliability using our institutional data, the general applicability of the model in clinical practice has not been confirmed. Qualitative serology test results may have varying cutoff index values due to different testing equipment and reagents. To adapt our model to various clinical settings, appropriate calibration of the cutoff values may be required through further multi-institutional research. Once validated, our MTL model can be offered through a user interface for clinical use. Although public data for validation were absent for our diagnostic task, the prospective accumulation of real clinical data through an open user interface holds the potential for robust external validation. Fine-tuning model parameters may be necessary for consistent performance when applied to calibrated data from other institutions. Last, as fundus photographs were not incorporated into our model, we could not verify the impact of it on model performance. We acknowledge the potential for improved accuracy with fundus photographs but also recognize the challenges posed by cases with nonspecific or obscured fundus findings and the modeling of multimodal input features (numeric blood and serology test results and images). Future research could explore the integration of fundus findings into our models and the associated implications for diagnostic performance.

In conclusion, this study manifests the previously unexplored diagnostic values of common blood and serology test data in the differential diagnosis of necrotizing viral retinitis. We presented and compared diverse ways to develop detection models with such data. Of those models, the adversarial MTL model showed the best performance in simultaneously detecting ARN and CMV retinitis as well as demonstrated its reliability. We presume clinicians can implement our adversarial MTL model to achieve a faster diagnosis, appropriate treatment, and improved visual outcomes. It can offer standardized and advanced medical services for areas with limited access to tertiary medical care. We also expect our

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

framework can be applied to other medical fields for the screening of rare diseases and the discovery of previously unknown markers.

Acknowledgments

The authors thank Eunjin Kim (Yonsei University College of Medicine) and Hyunseo Lee (Yonsei University College of Medicine) for obtaining ethical approval.

Sponsored by Samsung Research Funding Center of Samsung Electronics (Project Number: SRFC-TF2103-01) and by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (NO. 2022-0-00077, AI Technology Development for Commonsense Extraction, Reasoning, and Inference from Heterogeneous Data).

All data relevant to the study, except for the participant data, are included in the article or uploaded as supplementary materials. Questions regarding this should be directed to the corresponding author: Eun Young Choi (eychoi@yuhs.ac).

Disclosure: **K.T. Ong**, None; **T. Kwon**, None; **H. Jang**, None; **M. Kim**, None; **C.S. Lee**, None; **S.H. Byeon**, None; **S.S. Kim**, None; **J. Yeo**, None; **E.Y. Choi**, None

References

1. Lee JH, Agarwal A, Mahendradas P, et al. Viral posterior uveitis. *Surv Ophthalmol*. 2017;62(4):404–445.
2. Schoenberger SD, Kim SJ, Thorne JE, et al. Diagnosis and treatment of acute retinal necrosis: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2017;124(3):382–392.
3. Hennis HL, Scott AA, Apple DJ. Cytomegalovirus retinitis. *Surv Ophthalmol*. 1989;34(3):193–203.
4. Baltinas J, Lightman S, Tomkins-Netzer O. Comparing treatment of acute retinal necrosis with either oral valacyclovir or intravenous acyclovir. *Am J Ophthalmol*. 2018;188:173–180.
5. Sims JL, Yeoh J, Stawell RJ. Acute retinal necrosis: a case series with clinical features and treatment outcomes. *Clin Exp Ophthalmol*. 2009;37(5):473–477.
6. Schneider EW, Elner SG, Van Kuijk FJ, et al. Chronic retinal necrosis: cytomegalovirus necrotizing retinitis associated with panretinal vasculopathy in non-HIV patients. *Retina*. 2013;33(9):1791–1799.
7. Kempen JH, Jabs DA, Wilson LA, Dunn JP, West SK, Tonascia JA. Risk of vision loss in patients with cytomegalovirus retinitis and acquired immunodeficiency syndrome. *Arch Ophthalmol*. 2003;121(4):466–476.
8. Standardization of Uveitis Nomenclature (SUN) Working Group. Classification criteria for acute retinal necrosis syndrome. *Am J Ophthalmol*. 2021;228:237–244.
9. Davis JL. Differential diagnosis of CMV retinitis. *Ocul Immunol Inflamm*. 1999;7(3–4):159–166.
10. Balansard B. Necrotising retinopathies simulating acute retinal necrosis syndrome. *Br J Ophthalmol*. 2005;89(1):96–101.
11. Davis J. Diagnostic dilemmas in retinitis and endophthalmitis. *Eye*. 2012;26(2):194–201.
12. Priya K, Madhavan HN, Malathi J. Use of uniplex polymerase chain reaction & evaluation of multiplex PCR in the rapid diagnosis of viral retinitis. *Indian J Med Res*. 2003;117:205–210.
13. Anwar Z, Galor A, Albini TA, Miller D, Perez V, Davis JL. The diagnostic utility of anterior chamber paracentesis with polymerase chain reaction in anterior uveitis. *Am J Ophthalmol*. 2013;155(5):781–786.
14. Short GA, Margolis TP, Kuppermann BD, Irvine AR, Martin DF, Chandler D. A polymerase chain reaction-based assay for diagnosing varicella-zoster virus retinitis in patients with acquired immunodeficiency syndrome. *Am J Ophthalmol*. 1997;123(2):157–164.
15. Zhao X, Xia S, Chen Y. Role of diagnostic pars plana vitrectomy in determining the etiology of uveitis initially unknown. *Retina*. 2020;40(2):359–369.
16. Wen X, Leng P, Wang J, et al. Clinlabomics: leveraging clinical laboratory data by data mining strategies. *BMC bioinform*. 2022;23(1):1–20.
17. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136(7):803–810.
18. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962–969.
19. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*. 2018;125(9):1410–1420.
20. Zhang Y, Yang Q. An overview of multi-task learning. *Natl Sci Rev*. 2018;5(1):30–43.
21. Thung KH, Wee CY. A brief review on multi-task learning. *Multimed Tools Appl*. 2018;77:29705–29725.
22. Caruana R. Multitask learning. *Mach Learn*. 1997;28(1):41–75.
23. Zhang Y, Yang Q. A survey on multi-task learning. *IEEE Trans Knowl Data Eng*. 2021;34(12):5586–5609.
24. Ju L, Wang X, Zhao X, et al. Synergic adversarial label learning for grading retinal diseases via knowledge distillation and multi-task learning. *IEEE J Biomed Health Inform*. 2021;25(10):3709–3720.
25. Wang X, Ju L, Zhao X, Ge Z. Retinal abnormalities recognition using regional multitask learning. *International Conference on Medical Image Computing and Computer Assisted Intervention; Oct 13–17, 2019*, https://link.springer.com/chapter/10.1007/978-3-030-32239-7_4. Accessed November 23, 2022.
26. Holland GN. Standard diagnostic criteria for the acute retinal necrosis syndrome. *Am J Ophthalmol*. 1994;117(5):663–666.
27. Standardization of Uveitis Nomenclature (SUN) Working Group. Classification criteria for cytomegalovirus retinitis. *Am J Ophthalmology*. 2021;228:245–254.
28. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17:261–272.
29. Harris CR, Millman KJ, Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357–362.
30. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. *Proc 9th Python Sci Conf* 2010;57:10–25080.
31. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv.1712.09923. 2017.
32. Liu P, Qiu X, Huang X. Adversarial multi-task learning for text classification. *Proc 55th Annu Meeting Assoc Computational Linguistics*. 2017;1:1–10.
33. Bousmalis K, Trigeorgis G, Silberman N, Krishnan D, Erhan D. Domain separation networks. Advances in neural information processing systems; December 5–10, 2016, https://proceedings.neurips.cc/paper_files/paper/2016/file/45fbc6d3e05ebd93369ce542e8f2322d-Paper.pdf. Accessed September 1, 2022.

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024

34. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems; December 4–9, 2017, https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf. Accessed October 3, 2022.
35. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584.
36. Su PY, Wei YC, Luo H, et al. Machine learning models for predicting influential factors of early outcomes in acute ischemic stroke: registry-based study. *JMIR Med Inform*. 2022;10(3):e32508.
37. Ong KT, Kim H, Kim M, et al. Evidence-empowered transfer learning for Alzheimer's disease. IEEE International Symposium on Biomedical Imaging (ISBI); April 18–21, 2023, https://arxiv.org/abs/2303.01105. Accessed May 1, 2024.
38. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759.
39. Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156–163.
40. Hangartner L, Zinkernagel RM, Hengartner H. Antiviral antibody responses: the two extremes of a wide spectrum. *Nat Rev Immunol*. 2006;6(3):231–243.
41. Guex-Crosier Y, Rochat C, Herbort CP. Necrotizing herpetic retinopathies a spectrum of herpes virus-induced diseases determined by the immune state of the host. *Ocul Immunol Inflamm*. 1997;5(4):259–266.

Investigative Ophthalmology & Visual Science

Downloaded from iovs.arvojournals.org on 04/25/2024