Supplementary materials

Vision Transformer-based Interpretable Metabolic Syndrome Classification Using Retinal Images

Tae-Kwan Lee[1*], So Yeon Kim[1,2*], Hyuk Jin Choi[3,4], Eun Kyung Choe[5,6§], Kyung-Ah Sohn[1,2§]

[1]Department of Artificial Intelligence, Ajou University, Suwon, South Korea

[2]Department of Software and Computer Engineering, Ajou University, Suwon, South Korea

[3]Department of Ophthalmology, Seoul National University College of Medicine, Seoul, South Korea

[4]Department of Ophthalmology, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, South Korea

[5]Department of Surgery, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, South Korea

[6]Department of Surgery, Seoul National University College of Medicine, Seoul, South Korea

{qwan96, jebi1771, kasohn}@ajou.ac.kr, choe523@gmail.com, docchoi@hanmail.net

[*]These authors contributed equally to this work

[§]Corresponding authors

Contents

Supplementary Table 1. Detailed characteristics of the participants.

	Total set (N=3000)	Train/validation set (N=2400)	Test set (N=600)
Smoking status			
Non-smoker	1103 (42.2%)	910 (43.3%)	193 (37.9%)
Ex-smoker	966 (37.0%)	759 (36.1%)	207 (40.7%)
Current Smoker	544 (20.8%)	435 (20.7%)	109 (21.4%)
Alcohol consumption			
Regular (< 14 cups/week)	1752 (68.1%)	1405 (68.4%)	347 (67.0%)
Heavy (≥ 140 cups/week)	819 (31.9%)	648 (31.6%)	171 (33.0%)
Hypertension medication			
No	2332 (77.7%)	1883 (78.5%)	449 (74.8%)
Yes	668 (22.3%)	517 (21.5%)	151 (25.2%)
Diabetes medication			
No	2831 (94.4%)	2257 (94.0%)	574 (95.7%)
Yes	169 (5.6%)	143 (6.0%)	26 (4.3%)
Dyslipidemia medication			
No	2648 (88.3%)	2115 (88.1%)	533 (88.8%)
Yes	352 (11.7%)	285 (11.9%)	67 (11.2%)
Height	167.7 ± 8.1	167.7 ± 8.0	167.8 ± 8.3
Weight	68.9 ± 12.8	68.8 ± 12.7	69.3 ± 13.4
Waist circumference (cm)	86.3 ± 9.4	86.3 ± 9.4	86.5 ± 9.4
Glucose	104.0 ± 21.5	103.9 ± 21.5	104.3 ± 21.7
Triglyceride(TG)	137.3 ± 91.1	137.1 ± 91.8	138.0 ± 88.6
High density lipoprotein(HDL)	51.1 ± 12.2	51.1 ± 12.2	50.9 ± 12.2
HBA1C	5.8 ± 0.7	5.8 ± 0.7	5.8 ± 0.7
Exercise			
None	2305 (76.8%)	1860 (77.5%)	445 (74.2%)
Mild	542 (18.1%)	421 (17.5%)	121 (20.2%)
Aggressive	153 (5.1%)	119 (5.0%)	34 (5.7%)
Hemoglobin(HB)	14.6 ± 1.4	14.6 ± 1.4	14.7 ± 1.4
CALCIUM	9.3 ± 0.3	9.3 ± 0.3	9.3 ± 0.3
PHOSPHORUS	3.5 ± 0.5	3.5 ± 0.5	3.5 ± 0.5
Blood urea nitrogen(BUN)	14.2 ± 3.6	14.2 ± 3.7	14.3 ± 3.5
URIC_ACID	5.7 ± 1.5	5.7 ± 1.5	5.8 ± 1.5
CREATININE	0.8 ± 0.2	0.8 ± 0.2	0.8 ± 0.2
Total cholesterol	194.6 ± 34.7	194.5 ± 34.7	194.9 ± 34.6
Protein	7.3 ± 0.4	7.3 ± 0.4	7.3 ± 0.4
Albumin	4.6 ± 0.3	4.6 ± 0.3	4.6 ± 0.3
Total bilirubin	1.0 ± 0.4	1.0 ± 0.4	1.0 ± 0.4
Alkaline phosphatase	54.9 ± 15.0	54.8 ± 14.9	55.1 ± 15.3
GOT	25.0 ± 12.3	24.8 ± 12.0	25.7 ± 13.5
GPT	26.9 ± 19.1	26.7 ± 18.8	27.9 ± 20.3
GGT	42.9 ± 51.7	42.9 ± 52.5	42.9 ± 48.6
Low density lipoprotein(LDL) cholesterol	122.0 ± 30.1	121.9 ± 30.2	122.4 ± 29.9

Supplementary Table 2. Performance comparison of each deep learning model and pretrained data

Model	Image size	Precision	Sensitivity	Specificity	F1 score	Accuracy	AUC
ResNet50	224×224	0.61345 (±0.0292)	0.64060 (±0.0903)	0.59901 (±0.0934)	0.62460 (±0.0374)	0.61967 (±0.0232)	0.66357 (±0.0226)
	512×512	0.62426 (±0.0262)	0.64631 (±0.1304)	0.61225 (±0.1040)	0.63034 (±0.0608)	0.62917 (±0.0201)	0.68354 (±0.0138)
ResNet101	224×224	0.60664 (±0.0203)	0.65772 (±0.0419)	0.57781 (±0.0563)	0.63051 (±0.0148)	0.61750 (±0.0144)	0.66166 (±0.0217)
	512×512	0.60446 (±0.0230)	0.70235 (±0.0662)	0.54371 (±0.0866)	0.64837 (±0.0169)	0.62250 (±0.0142)	0.67793 (±0.0089)
InceptionV3	299×299	0.59217 (±0.0260)	0.69329 (±0.0809)	0.52616 (±0.0928)	0.63707 (±0.0289)	0.60917 (±0.0209)	0.65300 (±0.0311)
	512×512	0.60575 (±0.0110)	0.71812 (±0.0630)	0.53775 (±0.0603)	0.65625 (±0.0209)	0.62733 (±0.0049)	0.67962 (±0.0112)
DenseNet169	224×224	0.60644 (±0.0095)	0.64497 (±0.1082)	0.58642 (±0.0751)	0.62271 (±0.0490)	0.61550 (±0.0183)	0.66077 (±0.0068)
	512×512	0.60039 (±0.0388)	0.67383 (±0.0699)	0.55629 (±0.0703)	0.63411 (±0.0427)	0.61467 (±0.0403)	0.66122 (±0.0458)
RETFound (ImageNet)	224×224	0.59858 (±0.0104)	0.71745 (±0.0474)	0.52450 (±0.0480)	0.65209 (±0.0163)	0.62033 (±0.0065)	0.65303 (±0.0067)
	512×512	0.61061 (±0.0227)	**0.75772** (±0.0303)	0.52185 (±0.0572)	0.67587 (±0.0013)	0.63900 (±0.0208)	0.69094 (±0.0130)
RETFound (Retinal)	224×224	**0.66093** (±0.0210)	0.73691 (±0.0147)	0.62616 (±0.0403)	**0.69666** (±0.0095)	**0.68117** (±0.0160)	**0.74014** (±0.0081)
	512×512	0.65397 (±0.0237)	0.71074 (±0.0404)	**0.62715** (±0.0582)	0.68044 (±0.0095)	0.66867 (±0.0120)	0.72188 (±0.0147)

* The mean performances ± 95% confidence intervals, obtained after a 5-fold cross-validation, are presented.
* The original images were resized to 350x350 and then center-cropped to 224x224. For the 512x512 images, the original images were resized to 775x775 and then center-cropped to 512x512.
* The above results are the results of learning using the retinal image of one side of the participant, assuming that one retinal image is one participant. The left and right retinal images of one participant were set to co-exist in one of the train, validation, and test datasets.
* In the table above, RETFound(ImageNet) uses parameters pretrained with ImageNet-21k, and RETFound(Retinal) uses parameters pretrained with MEH-MIDAS and Kaggle EyePACS.

Supplementary Table 3. Performance comparison of machine learning models for classifiers

Model	Precision	Sensitivity	Specificity	F1-score	Accuracy	AUC
DT	0.67840 (±0.0182)	0.70537 (±0.0178)	0.66954 (±0.0316)	0.69146 (±0.0103)	0.68733 (±0.0131)	0.68745 (±0.0130)
RF	0.70846 (±0.0181)	0.70738 (±0.0201)	0.71258 (±0.0226)	0.70784 (±0.0159)	0.71000 (±0.0157)	0.78293 (±0.0106)
XGBoost	0.76659 (±0.0069)	0.77584 (±0.0108)	**0.76689** (±0.0080)	0.77117 (±0.0076)	0.77133 (±0.0071)	0.85048 (±0.0073)
SVM	0.75681 (±0.0052)	0.80000 (±0.0163)	0.74636 (±0.0062)	0.77778 (±0.0095)	0.77300 (±0.0081)	0.85754 (±0.0068)
LR	**0.77184** (±0.0066)	**0.80134** (±0.0130)	0.76623 (±0.0095)	**0.78628** (±0.0072)	**0.78367** (±0.0064)	**0.87249** (±0.0056)

* DT : Decision Tree, RF : Random Forest, SVM : Support Vector Machine, LR : Logistic Regression
* The mean performances ± 95% confidence intervals, obtained after a 5-fold cross-validation, are presented.
* Features used for learning are age, gender, BMI, and retinal image features.

Supplementary Figure 1. Classification results by BMI distribution of normal and metabolic syndrome participants in each fold via cross validation, categorized into four groups. Each group is the result of classifying normal and metabolic syndrome using either only the clinical features of each fold or a combination of features (age, gender, BMI, and retinal image features) for the test dataset, and is distributed according to obese (BMI ≥ 25 kg/m²) and non-obese (BMI < 25 kg/m²). For participants in Group A, both clinical features alone and combined features correctly classified participants as normal or metabolic syndrome. For participants in Group B, both clinical features alone and combined features misclassified participants' condition. For participants in Group C correctly classified participants as normal or having metabolic syndrome when only clinical features were used, but incorrectly classified them when combined features were used. In contrast, participants in Group D were correctly classified as normal or metabolic syndrome when the combined features were used, but incorrectly classified when only the clinical features were used. This demonstrates that incorporating retinal image features for both obese participants with normal and non-obese participants with metabolic syndrome across all folds can improve the result of classifying the patient's condition.