A IMPORTANT NOTATIONS

We list the important notations in Table 6.

Table 6: Important notations

Notation	Definition
T	The number of visit
V	The CFP sequence
g	The genotype vector
e^g	The embedding vector of g
v_t	The CFP image in the t^{th} visit
$e_t^{c,a}$	The CFP feature generated by attention module
p_i	The CFP feature vector in patient pool
$e_t^{c,p}$	The aggregate features
$e_t^{c,d}$	The abnormal features generated by CA
D	The sociodemographic sequence
d_t	The sociodemographic vector in the t^{th} visit
e_t^d	The embedding vector of d_t
v_t^p	The AMD progression vector before the t^{th} visit
e_t^p	The embedding vector of v_t^p
y_t^c	The predicted AMD stage probability in the t^{th} visit
$\hat{y}_t^c$	The ground truth of AMD stage in the t^{th} visit
y_t	The predicted late AMD probability after t^{th} visit
$\hat{y}_t$	The late AMD ground truth after t^{th} visit
W_*, b_*	The learnable parameters
Δ_t	The elapsed time between $t-1^{th}$ and t^{th} visit
x_t	The combination of multi-modal features in t^{th} visit
C_t	The memory vector of T-LSTM in t^{th} visit
h_t	The hidden state of T-LSTM in t^{th} visit

B HYPER-PARAMETER OPTIMIZATION

There is a hyper-parameter λ in Eq. (14). Note that we only jointly train the late AMD prediction model and auto-encoder when subtyping CFP sequences. We use the clustering evaluation metrics CHI and DBI to select the value of λ. As Table 7 shown, when $0.3 \leq \lambda \leq 0.9$, clustering performance is not sensitive to λ. In our experiment, we set $\lambda = 0.5$.

Table 7: Hyper-parameter optimization for λ in Eq. (14).

λ	0	0.1	0.3	0.5	0.9	1
CHI ↑	0.57	1.56	1.80	**1.82**	**1.82**	1.75
DBI ↓	2.35	0.89	**0.74**	0.75	0.76	0.88

C K SELECTION FOR K-MEANS.

We try to use different K for k-means when clustering the CFP sequences. As shown in Figure 6, when $K = 3$, we have the best DBI value for CFP clustering. It is also the elbow point for CHI. Thus we cluster the CFP sequences into 3 subphenotypes.

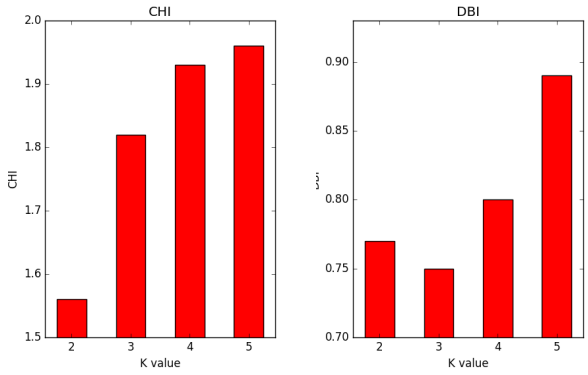

Figure 6: CHI and DBI across different K for k-means to cluster the CFP sequences. When $K = 3$, we have the best DBI value for CFP clustering. It is also the elbow point for CHI.

D LATE AMD RATE IN VARIOUS PATIENT GROUPS

Patients with different demographics (e.g., gender and age) have different risks of progressing to late AMD stages. Table 8 displays the late AMD rates in various patients groups. Higher age (e.g., age > 80) and smoking history have positive correlations with late AMD rates. Due to the existence of the imbalanced late AMD distribution, machine learning methods might learn the bias and discrimination from the data. Thus we propose to use contrastive attention to remove the common features in patient groups and force the model to learn to focus on the abnormalities on fundus images.

Table 8: Late AMD rate in different patient groups.

Sociodemographic	Value	Late AMD rate
Gender	Female	15.0%
	Male	13.8%
Smoking history	No	13.1%
	Yes	16.7%
Age	< 70	11.2%
	70-80	14.7%
	> 80	19.8%

E POSITIVE/NEGATIVE SAMPLE DISTRIBUTION

Table 9 displays the positive/negative sample distribution for late AMD detection and prediction settings.

F GENETIC MARKER VISUALIZATION

We further analyze the alternative allele distribution of 52 AMD-associated genetic markers across various subphenotypes. Figure 7 displays the alternative allele distribution of AMD-associated

Table 9: Positive/Negative sample distribution for late AMD prediction.

Methods	Late AMD Detection	n-year Late AMD Prediction					
		1-year	2-year	3-year	4-year	5-year	All-year
#. of positive visit samples	9,207	753	1,558	2,240	2,861	3,390	4,713
#. of negative visit samples	57,262	37,375	32,371	27,497	23,020	17,498	33,415
Positive rate	16.07%	1.97%	4.59%	7.53%	11.05%	16.22%	12.36%

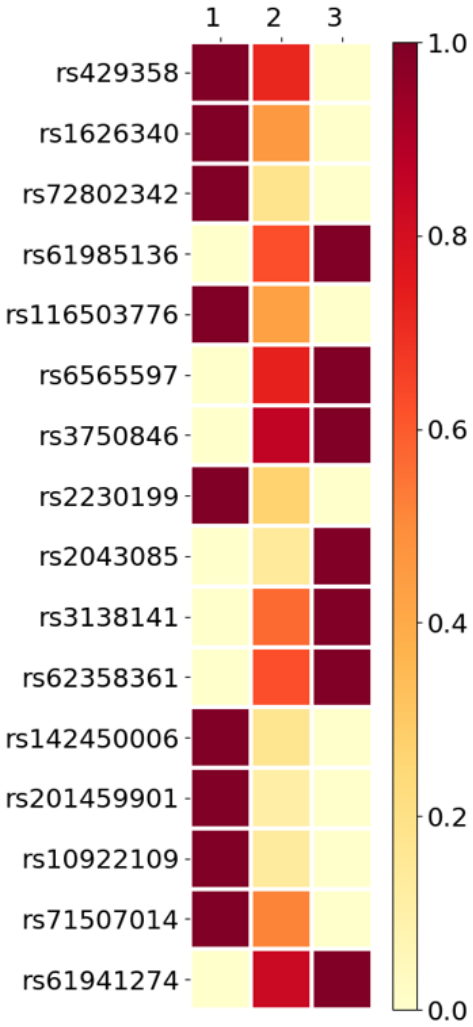

Figure 7: The distribution of AMD-associated genetic markers' alternative allele across the three subphenotypes. The subphenotypes are 1: subphenotype I; 2: subphenotype II; 3: subphenotype III. Because the alternative allele rates of different genetic markers vary a lot, we normalize the rates when visualizing the distribution.

genetic markers with significant distribution difference (P-value < 0.05) across 3 subphenotypes.