Deep Learning with GWAS to Predict AMD Progression

Tao Sun1,2, Wei Chen1,2 and Ying Ding3

1Department of Biostatistics, University of Pittsburgh
2Department of Pediatrics, Children’s Hospital of Pittsburgh of UPMC

Abstract

To establish an accurate survival prediction model for Age-related Macular Degeneration (AMD) progression, we develop a novel framework, which builds deep neural networks on time-to-event outcomes to effectively extract features from the wealth of GWAS data. Using data from two large randomized clinical trials on AMD progression, Age-Related Eye Disease Study (AREDS) and AREDS2, we develop and evaluate three machine/deep-learning-based survival models to predict the risk of progression to late-AMD given the patient’s clinical and genetic profiles. We establish the up-to-date most accurate survival prediction model for AMD progression. The results provide valuable insights to early prevention and tailored intervention for AMD patients.

Age-related Macular Degeneration

- AMD: an eye disease and a leading cause of blindness in elders.
- A progressive disorder leading to blindness at the late-AMD stage.
- Several Genome-wide Association Studies (GWAS) have found AMD progression is significantly associated with age, smoke, and genetic variants (SNPs) [1-2].
- Objective: develop an accurate survival prediction model for AMD progression using GWAS.

Survival Prediction Models

LASSO: minimize the negative log-partial likelihood function with \( \ell_1 \) penalty (under Cox model)

\[
-\frac{1}{N} \sum_{i=1}^{N} \left( \eta_i \beta - \log \sum_{j \in R_i} e^{\eta_i \beta} \right) + \lambda ||\beta||_1
\]

- Data for subject \( i \): \( \{ Y_i, \delta_i, Z_i \} \)
- \( Y_i \): observed time,
- \( \delta_i \): event status,
- \( Z_i \): covariates,
- \( R_i \): at risk set at time \( Y_i \),
- \( \beta \): parameter of interest.
- Only account for linear covariate effects.

Deep Neural Network (DNN): input layer of covariates, hidden layers and output layer (i.e., risk score)

Implementation and Evaluation

- LASSO and RSF: standard methods in glmnet and randomForestSRC R packages.
- We develop a novel survival DNN framework in R: Deep Learning with GWAS to Predict AMD Progression.

Conclusions

- We establish the up-to-date most accurate survival prediction model for AMD progression.
- We demonstrate DNN’s strong predictive power and capacity in learning complex structures (with simulations).
- We also implement a novel predictor importance algorithm for interpreting the DNN survival model.

References


Acknowledgements

This work is supported by the CTSI QuMP Pilot Grant from the University of Pittsburgh.

Table 1: Top risk factors of AMD progression

<table>
<thead>
<tr>
<th>Types</th>
<th>Risk factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical age</td>
<td>smoke, education level</td>
</tr>
<tr>
<td>Genetic 663 SNPs (GWAS on AMD progression)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Prediction results

<table>
<thead>
<tr>
<th></th>
<th>c-index (sd)</th>
<th>4-year-BrS (sd)</th>
<th>10-year-BrS (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRS</td>
<td>74.1 (2.4)</td>
<td>0.113 (0.005)</td>
<td>0.151 (0.005)</td>
</tr>
<tr>
<td>LASSO</td>
<td>74.4 (1.3)</td>
<td>0.112 (0.004)</td>
<td>0.146 (0.006)</td>
</tr>
<tr>
<td>RSF</td>
<td>70.1 (1.8)</td>
<td>0.119 (0.004)</td>
<td>0.170 (0.006)</td>
</tr>
<tr>
<td>DNN</td>
<td>76.1 (2.9)</td>
<td>0.113 (0.006)</td>
<td>0.136 (0.011)</td>
</tr>
</tbody>
</table>

Figure 1: AMD progression

Figure 2: An example neural network

Figure 3: Neural network optimization

Figure 4: Random Survival Forest

Figure 5: Left: Time-dependent Brier scores; Right: Time-dependent AUC

Figure 6: Left: Predictor importance; Right: Distinct subgroups

Application to AMD Progression

- The event of interest is the onset of late-AMD.
- Right-censoring rate is 73%.
- Predictors include age, smoke, education and 663 top SNPs[6] (p < 1 x 10^-5).

Figure 6: Left: Predictor importance; Right: Distinct subgroups