# Applying Machine Learning to Predict ε3ε4 Gene based on Wayfinding of Sea Hero Quest Game and ACE Cognitive Exam

Lee Yu-Hsiang, Master of Computer Science Information Management, National Changshua University of Education
Shao-Wei Yan, Ph.D., Department of Information Engineering, National Cheng Kung University
Pei-Fen Wu Ph.D., Department of Design, National Yunlin University of Science and Technology

The Apo-E gene is located on the 19th pair of human chromosomes. People with an ε4 (ε3ε4) type gene have a 3-5 times increased chance of getting Alzheimer's disease (AD), while the normal human gene is ε3ε3.

"Sea Hero Quest" (SHQ) is a game that can detect Alzheimer's disease early. According to Dr. Gillian Coughlan's paper(Coughlan, 2019), it was found that the average way-finding distance of ε3ε3players is lower than ε3ε4 at the way-finding game scenes specially in Level 6 and Level 8 , and the routes are also relatively concentrated than ε3ε4.

Addenbrooke's Cognitive Exam (ACE) is an extended cognitive screening technique that including five cognitions: attention, memory, language, language fluency, and visuospatial skills. It had often been used to evaluate AD.

The Rey Complex Figure Test (RCFT, Proposed by Meyers ,1995) consists of a replication test and one or more recall tests . The experiment requires candidates to draw a copy of the graphics presented on the table in front of them. It includes a 3-minute recall trial and a 30-minute recall trial, or an identification test. It was used to detect brain disorders such as Alzheimer's disease, traumatic brain injury, schizophrenia, obsessive-compulsive disorder, Huntington's disease, bipolar disorder, and epilepsy.

This research hopes to combine the SHQ L6, L8 levels way-finding distance , ACE five cognitive items, RCFT test and age to build a neural network model by using SPCN5.0 (Ye ,2009), This work doesn't measure genes directly, but can predict the possibility of getting ε3ε4 gene which may increase chance of getting AD.

Through Dr. Gillian Coughlan's paper (Coughlan, 2019) 59 volunteers' experimental data. We divide these 59 people into 51 people for training and 8 people for verification. And we use "age sorting" to make 5 different combinations of total 59 data by using stratified sampling method as following table:

| Combination | I | II | III | IV | V |
|---|---|---|---|---|---|
| Numbering To do Verification | 1、14、30、34、38、42、44、52 | 2、11、22、25、31、34、49、57 | 6、14、18、22、31、39、41、54 | 1、11、30、34、38、42、44、52 | 2、14、19、26、33、35、41、51 |

The combination table of " Age sorting" and "Stratified Sampling"

Then Input of each combination is persons' age, L6, L8 ACE-total and RCFT, Each combination has 8 verification data and the rest 51 is as the training data. We train the model by SPCN5.0 BPN simulation program, then we get 5 verification results.

| Variation | L6+L8+RCFT AGE + ACE | L6+L8+ AGE + ACE | L6+L8+ AGE+RCFT | L6+L8+ ACE+RCFT | L6+L8+ AGE | L6+L8+ RCFT | L6+L8+ ACE |
|---|---|---|---|---|---|---|---|
| Combination1 | 62.5% | 62.5% | 62.5% | 62.5% | 75% | 62.5% | 75% |
| Combination2 | 100% | 100% | 62.5% | 87.5% | 75% | 75% | 100% |
| Combination3 | 62.5% | 75% | 62.5% | 87.5% | 62.5% | 75% | 100% |
| Combination4 | 25% | 62.5% | 50% | 50% | 75% | 50% | 75% |
| Combination5 | 75% | 100% | 75% | 87.5% | 50% | 100% | 87.5% |
| Average | 65% | 80% | 62.5% | 75% | 67.5% | 72.5% | 87.5% |

Five parameters is < 75%

Four parameters, Two groups≥ 75% (L6,L8, ACE,AGE) , (L6,L8,ACE,RCFT)

Three parameter, One group ≥ 75% (L6,L8,ACE)

Interestingly, if variation groups ≥ 75% , they have L6,L8,ACE factors together!

We go on doing BPN simulation for Prediction., then we get 5 prediction results.

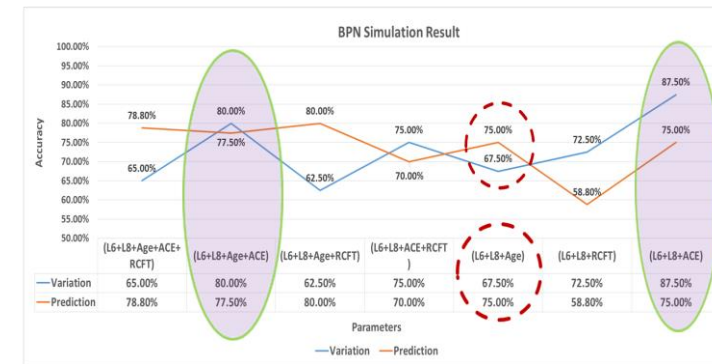| Prediction | L6+L8+RCFT AGE + ACE | L6+L8+ AGE + ACE | L6+L8+ AGE+RCFT | L6+L8+ ACE+RCFT | L6+L8+ AGE | L6+L8+ RCFT | L6+L8+ ACE |
|---|---|---|---|---|---|---|---|
| Combination1 | 81.3% | 81.3% | 75% | 75% | 75% | 68.8% | 75% |
| Combination2 | 81.3% | 75% | 87.5% | 68.8% | 81.3% | 56.3% | 75% |
| Combination3 | 68.8% | 75% | 81.3% | 62.5% | 75% | 37.5% | 75% |
| Combination4 | 75% | 81.3% | 75% | 75% | 75% | 68.8% | 75% |
| Combination5 | 87.5% | 75% | 81.3% | 68.8% | 68.8% | 62.5% | 75% |
| Average | 78.8% | 77.5% | 80% | 70% | 75% | 58.8% | 75% |

(L6,L8,ACE,RCFT,AGE) ≥ 75%

Four parameters, Two groups≥ 75% (L6,L8, ACE,AGE) , (L6,L8,AGE,RCFT)

Three-parameter, Two groups≥ 75% (L6,L8,AGE), (L6,L8,ACE)

Similarly , if prediction groups ≥75%, they have both L6 and L8.

The RCFT seems effect is minor both at variation and prediction!

Now we may look closely on "Line Chart".



| | BPN Simulation Result | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameters | (L6+L8+Age+ACE+ RCFT) | (L6+L8+Age+ACE) | (L6+L8+Age+RCFT) | (L6+L8+ACE+RCFT) | (L6+L8+Age) | (L6+L8+RCFT) | (L6+L8+ACE) |
| Variation | 65.00% | 80.00% | 62.50% | 75.00% | 67.50% | 72.50% | 87.50% |
| Prediction | 78.80% | 77.50% | 80.00% | 70.00% | 75.00% | 58.80% | 75.00% |

There are only two groups that both variation≥75% and the prediction also ≥75% .

(L6+L8+Age+ACE): Variation Average Value=80.0% Prediction Average Value=77.5%

(L6+L8+ACE}: Variation Average Value=87.5% Prediction Average Value=75 %

The L6, L8 and ACE still are the major factors in BPN simulation ( but without RCFT) !

As for Age, we can see the red dotted line of three parameters (L6+L8+Age). It's worse at variation (67.5%) but good at prediction (75%). We can hypothesis age is not an important factor in BPN. All we need may be just L6,L8 and ACE-total in BPN.

Conclusion: Combine the Sea Hero Quest L6, L8 levels way-finding distance and ACE five cognitive items, Applying Machine Learning to Predict ε3ε4 Gene is a potential good way to predict.

國立彰化師範大學
National Changhua University of Education