Children's Aggressiveness and Prosociality Prediction and Comparison

Ya-Yun Chen[1] and Tai-Jung Chen[2]
[1]Department of Psychology, Virginia Tech
[2]Department of Industrial & Systems Engineering, Virginia Tech

Problem Statement

With appropriate social behaviors, individuals are able to shape friendly social exchanges and to form reciprocal relations groups.

- More prosocial behaviors: a high well-being society
- More aggressive behaviors: harm the whole society

A greater understanding of how humans develop these two sorts of behaviors will offer insight into establishing a family and social safety net, preventing violence and promoting happiness.

Aims

The direct competition, orthogonal or simple dichotomies, between these two social behaviors was explored less.

We aimed to use decision trees to explore the relationship between parents/family/neighborhood, social behavior related neural networks, and children’s social behaviors.

(1) decision trees: prediction models of aggressive behavior and a prediction model of prosocial behavior;
(2) compare the similarity of these prediction models
(3) used ensemble and logistic regression models to **improve the predictability** of aggressive model.

Data Description

The present study will be carried out using data retrieved from the Adolescent Brain Cognitive Development (ABCD) Data Repository (https://abcdstudy.org), which is tracking 11,874 individuals aged 9-11 years from 21 data collection sites around the United States.

In sum, we used the neighborhood-parent problems-family (N-P-F) factors and brain connectivities to predict the child's aggressive tendency and prosociality

Data Pre-processing

---

(1) Concatenate all dataset with inner join method.

(2) remove NaN, 777 (refuse to answer) and 999 (don't know).

(3) Aggregate the negative social behavior scores and positive social behavior scores respectively.

(4) Use domain knowledge to select the target cortical network to subcortical region connectivity.

Data Exploration

We used data exploration techniques to implement attribute selection and labeling.

The correlation analyses were applied to select potential attributes.

The descriptive statistics were utilized to group high and low aggressive groups and prosocial groups respectively.

Data Exploration

Attribute selection:

Originally there were 55 N-P-F attributes
Inclusion criteria: |r| >= 0.1.

- 26 N-P-F factors survived for predicting aggressive tendency
- 15 N-P-F factors survived for parent-report prosociality
- 2 N-P-F factors survived for child-report prosociality

Data Exploration

Labeling:

Aggressive tendency:

- score = minimum score (-2.02) as the non-aggressive group (label = 0, N = 4534)
- score >= 75 percentile (0.35-33) as the aggressive group (label = 1, N = 2903).

Prosocial tendency:

- score <= 3 as the low prosocial group (label = 0, N = 1368 for parent report and N = 1302 for child report)
- score = 6 as the high prosocial group (label = 1, N = 7328 for parent report and N = 5242 for child report).
- For the imbalanced data in prosociality, bootstraps with N = 1500 for each group were applied.

Model Building

We used the attributes survived in correlation analysis in aggressive tendency and barin connectivities to build the aggressive model and prosocial models.

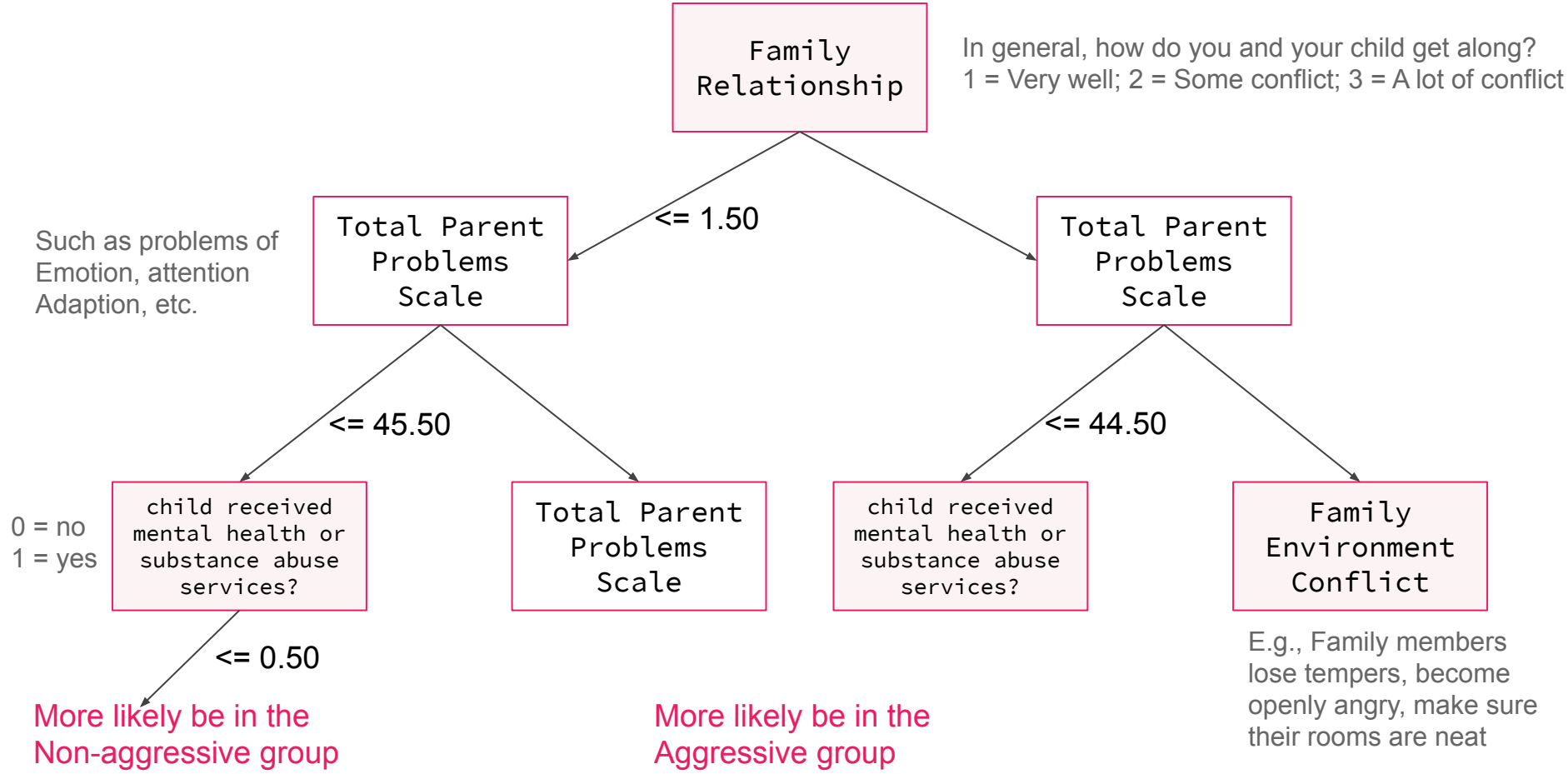
If these two sorts of social behaviors are **simple dichotomy**, the prediction **accuracy and the structure of the decision trees should be similar**.

In contrast, if they are **orthogonal**, the prediction **accuracy of the procial model would be around the chance level** or the s**tructure of the decision trees are different from the aggressive model.**

Model Evaluation

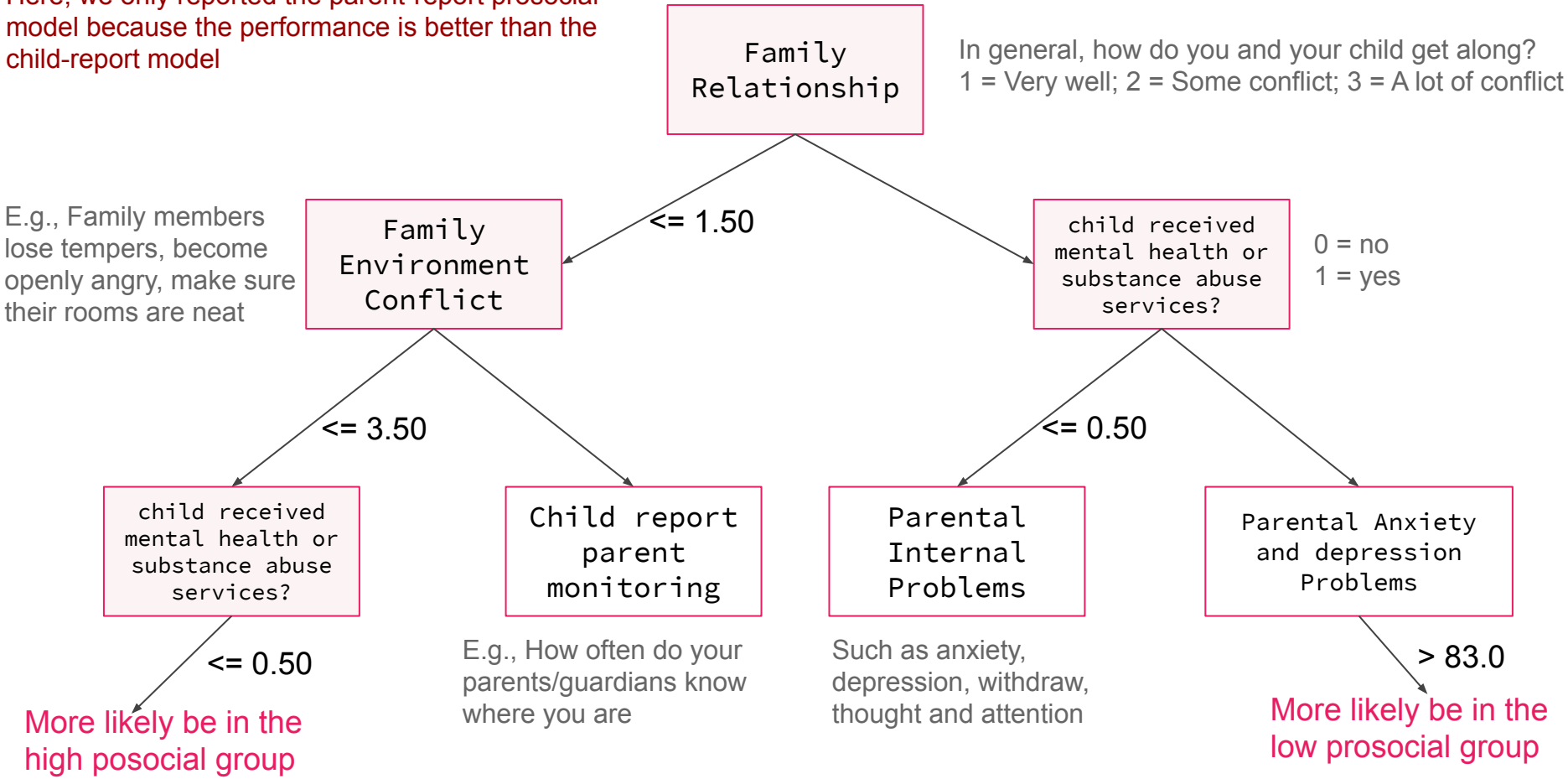
- **Aggressive model predicted by the N-P-F factors.**
  The accuracy = 0.76, f1 score = 0.69, recall = 0.70, precision = 0.68, Kappa = 0.49.
- **Aggressive model predicted by the brain connectivities.**
  The accuracy = 0.58, f1 score = 0.47, recall = 0.47, precision = 0.46, Kappa = 0.12.
- **Parent-report prosocial model predicted by the N-P-F factors.**
  The accuracy = 0.64, f1 score = 0.65, recall = 0.63, precision = 0.67, Kappa = 0.29.
- **Parent-report prosocial model predicted by the brain connectivities.**
  The accuracy = 0.50, f1 score = 0.51, recall = 0.51, precision = 0.50, Kappa = 0.01.
- **Child-report prosocial model predicted by the N-P-F factors.**
  The accuracy = 0.61, f1 score = 0.63, recall = 0.52, precision = 0.64, Kappa = 0.22.
- **Child-report prosocial model predicted by the brain connectivities.**
  The accuracy = 0.52, f1 score = 0.53, recall = 0.54, precision = 0.52, Kappa = 0.04.

Aggressive model predicted by the N-P-F factors.

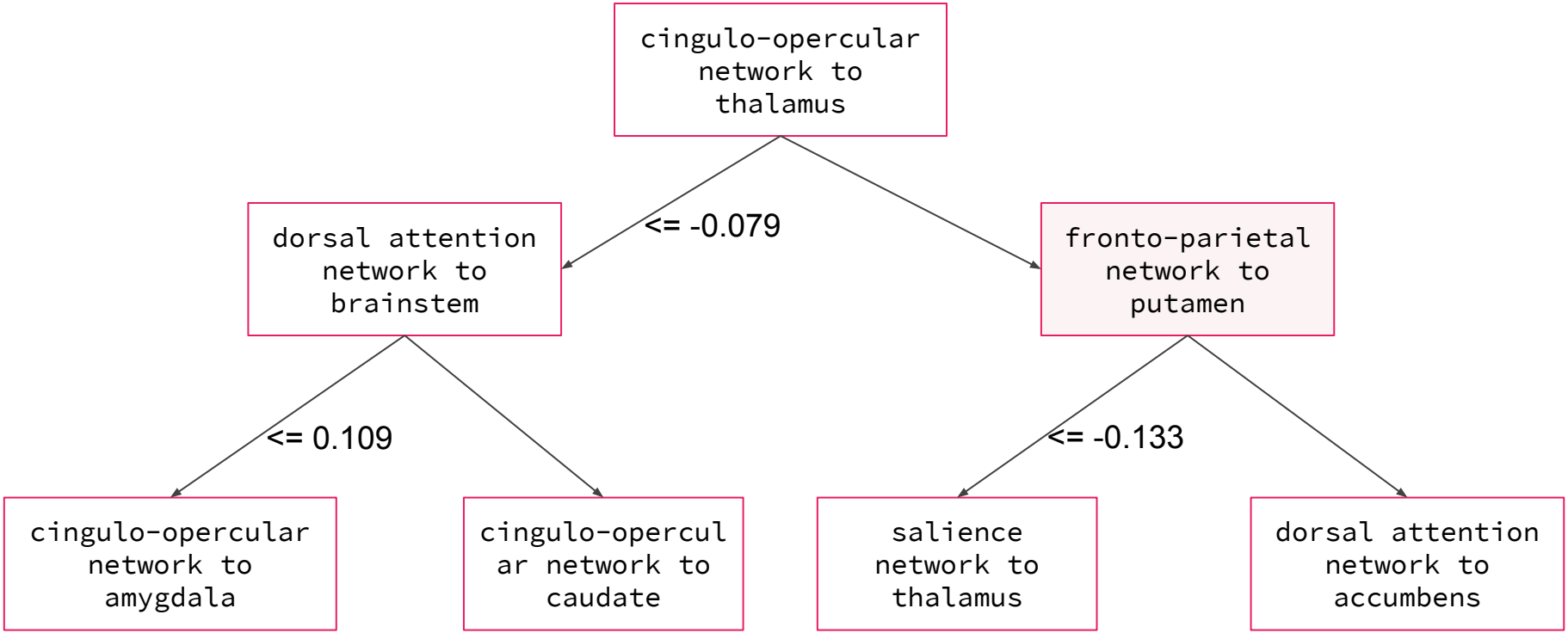


Prosocial model predicted by the N-P-F factors.

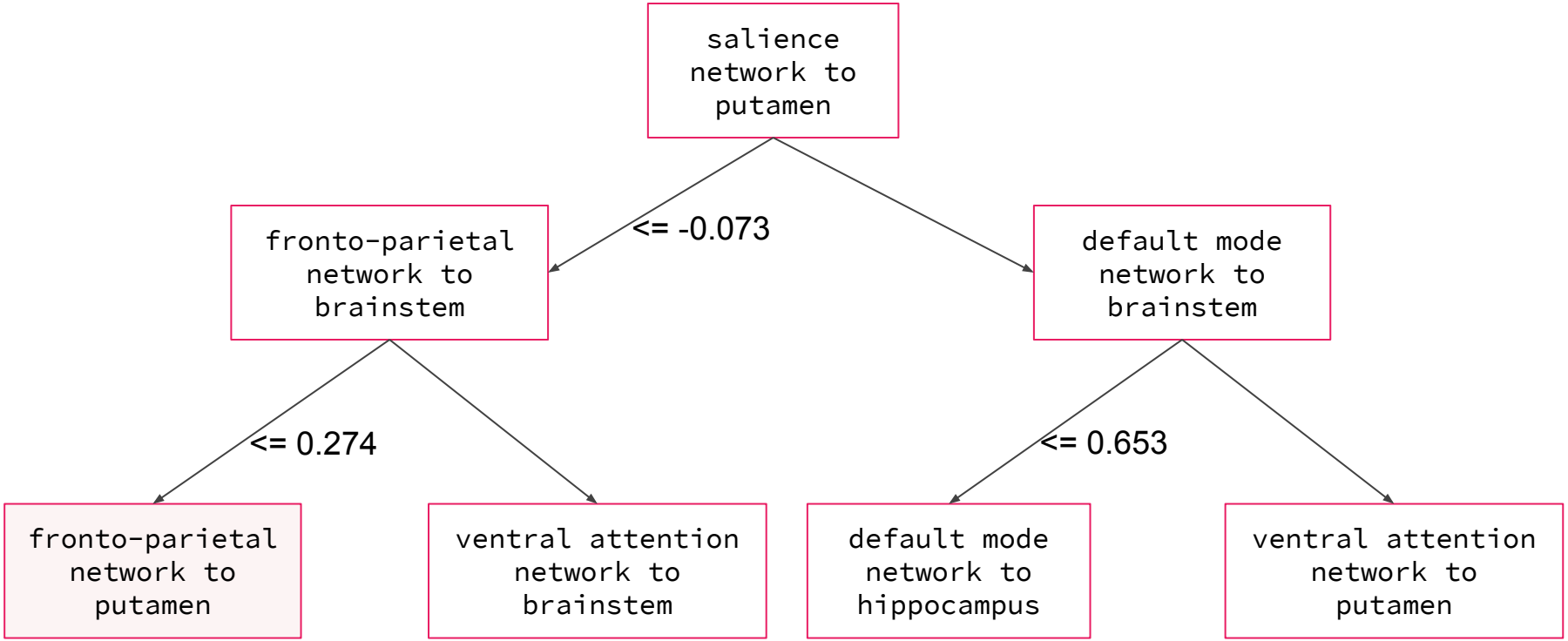
Here, we only reported the parent-report prosocial model because the performance is better than the child-report model



Aggressive model predicted by the brain connectivities.

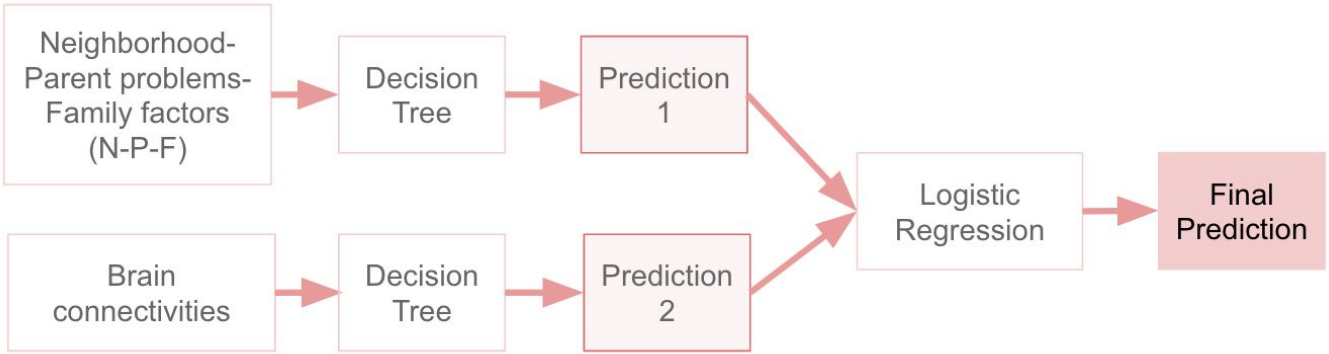


Prosocial model predicted by the brain connectivities.



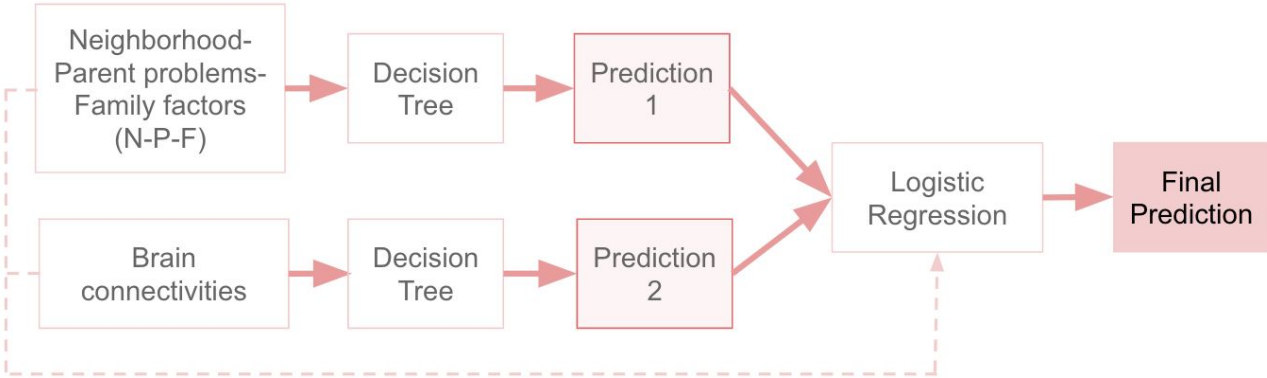
Predictability Improvement (1):

- The best accuracy and kappa of the decision trees of aggressive model were .76 and .49 respectively.
- We tried to use an ensemble model to improve the performance.
- Unfortunately, the accuracy and kappa of this ensemble model **were decreased slightly to .75 and .46** respectively.



Predictability Improvement (2):

- The best accuracy and kappa of the decision trees of aggressive model were .76 and .49 respectively.
- Tried to feed more attributes in the ensemble model as the second attempt to improve the performance.
- The accuracy and kappa of this second attempt model were **improved slightly to .78 and .53** respectively.



Predictability Improvement (3):

- The best accuracy and kappa of the decision trees of aggressive model were .76 and .49 respectively.
- Since the N-P-F factors and brain data improved the performance of the models, we tried to feed these two categories only without the prediction results from the decision trees.
- The accuracy and kappa of this logistic regression model were **significantly improved to .82 and .60** respectively.
- A further examination found that the improvement were from the N-P-F factors (accuracy = .83; kappa = .64).

Real-world Insights

- **The family relationship (conflict between parent and child)** plays a pivotal role in predicting children's aggressive and prosocial tendencies, supported by both decision tree and logistic regression model.
- **Other important factors** shared by both aggressive and prosocial tendency predictions are: **whether the child ever received mental health or substance abuse services**, **family environment conflict**, and **the frontoparietal network to putamen connectivity.**
- The current findings suggest that the relationship between aggressive and prosociality tendencies is not orthogonal nor simple dichotomies.

Lessons learned

Ya-Yun Chen:

1. If have more time, we may have a more accomplish plan to do the analysis. In the current work, we tries different models to improve the performance. However, a better way was to tune the model to make it better from the same structure.
2. Current work only simply compare the first three layers of the trees. It should be algorithm to make the comparison more comprehensively.

Tai-Jung Chen:

3. Clear up a lot of details from what we learned in the lectures when implementing these concepts. In practice, we usually have to write some individual code to be flexible to the scenarios.
4. Communications with my teammate are also a precious takeaway. Since Ya-Yun is more familiar in the domain knowledge of this project, and I am more familiar with coding in Python, we discussed a lot about the interpretation of the outcomes and how we can verify some hypothesis through the model.
5. We might want to try more algorithms and dig deeper to connect the theories with the outcomes, if we are given the opportunity again.