Project 3 GeNIe

Before we start, we will explain the features of my model:

The age attribute indicates the subject's age at the time of recruitment, while sex is encoded as '0' for male and '1' for female, highlighting potential gender influences on disease progression.

Test\_time represents the duration since recruitment, with the integer part denoting the number of days. This helps in tracking disease progression over time.

Two crucial clinical measures included are motor\_UPDRS and total\_UPDRS, both of which are linearly interpolated. The motor\_UPDRS score assesses motor function, evaluating symptoms such as tremors and rigidity, whereas the total\_UPDRS score encompasses both motor and non-motor aspects of the disease.

The dataset also contains several measures of variation in fundamental frequency, known as jitter measures. Jitter(%) quantifies frequency stability as a percentage, Jitter(Abs) measures the average absolute difference between consecutive periods in microseconds, Jitter

focuses on short-term perturbations by averaging differences over three periods, Jitter

averages over five periods, and Jitter

highlights fine-grained changes by calculating differences between consecutive periods.

Similarly, there are measures of variation in amplitude, known as shimmer measures. Shimmer represents amplitude stability, Shimmer(dB) provides this measure in decibels, Shimmer

averages amplitude differences over three periods, Shimmer

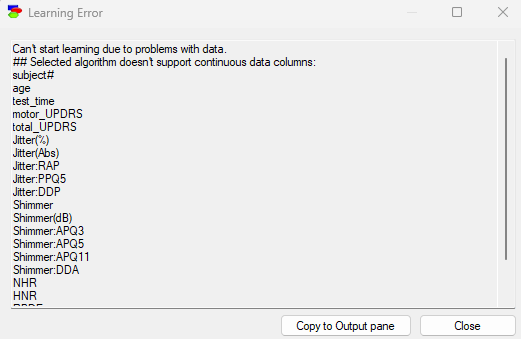
over five periods, and Shimmer

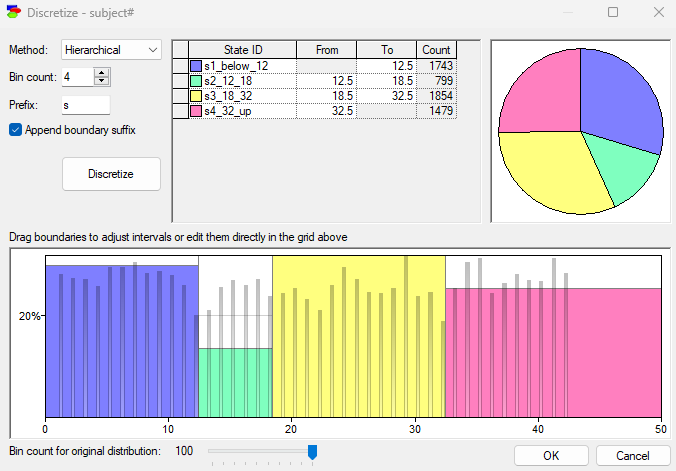
over eleven periods. Shimmer

highlights fine details by calculating differences of differences between consecutive amplitudes.

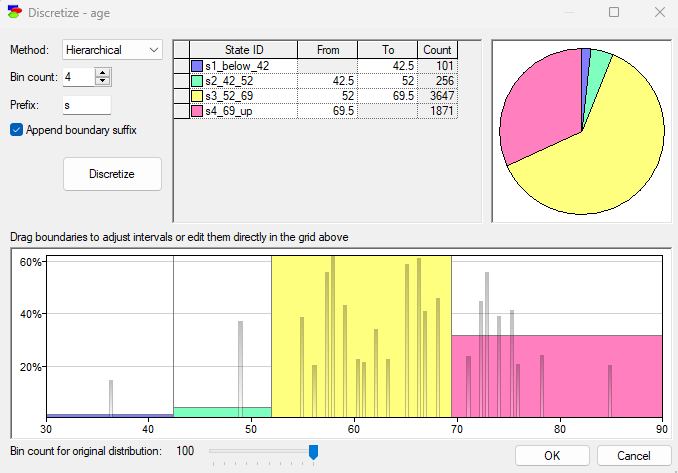
Two measures assess the ratio of noise to tonal components in the voice: NHR (Noise-to-Harmonics Ratio) and HNR (Harmonics-to-Noise Ratio), where higher NHR indicates more noise, and higher HNR indicates clearer, more tonal voice.

RPDE (Recurrence Period Density Entropy) is a nonlinear dynamical complexity measure indicating signal predictability and complexity. DFA (Detrended Fluctuation Analysis) measures the fractal scaling properties of a time series, assessing its long-range correlations. Lastly, PPE (Pitch Period Entropy) evaluates the complexity of pitch variations, with higher values indicating more complex and less stable pitch variations.

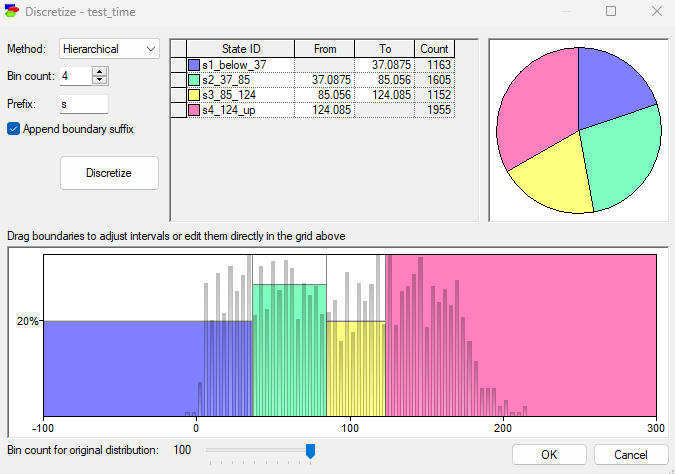
First step is loading the data, but since its all continuous data I got this error.  
  
  
so we have to discretize them:

Subject#:  


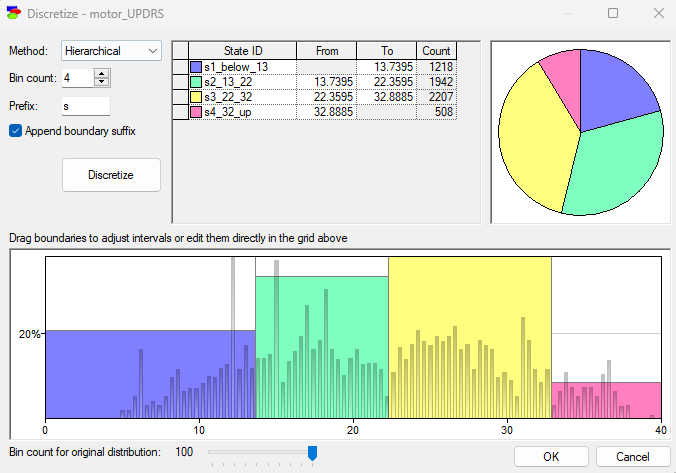
Age:



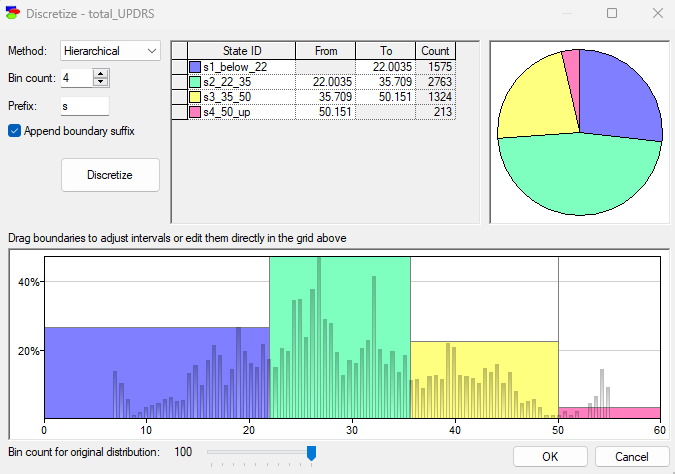
Test\_time:



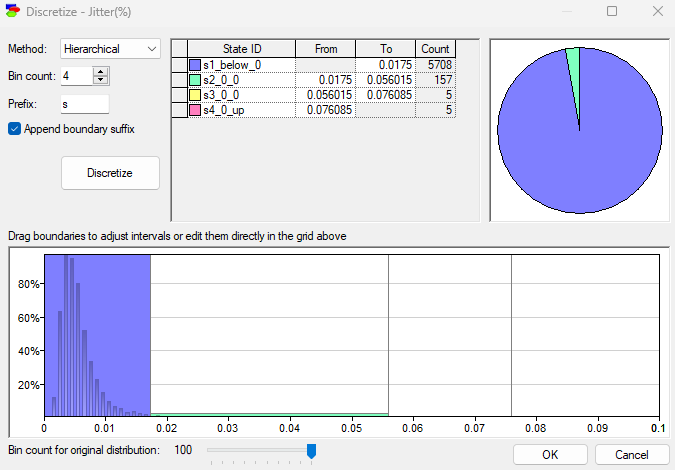
Motor\_UPDRS:



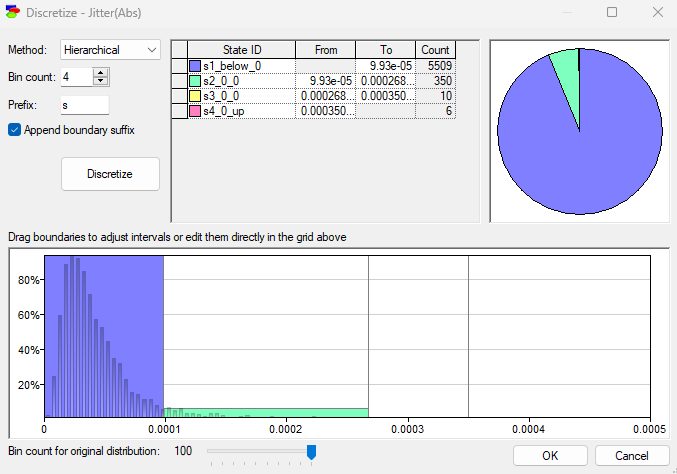
Total\_UPDRS:



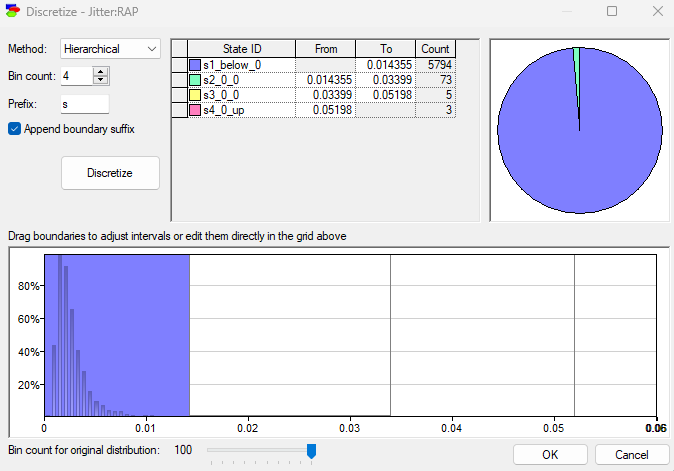
Jitter(%):



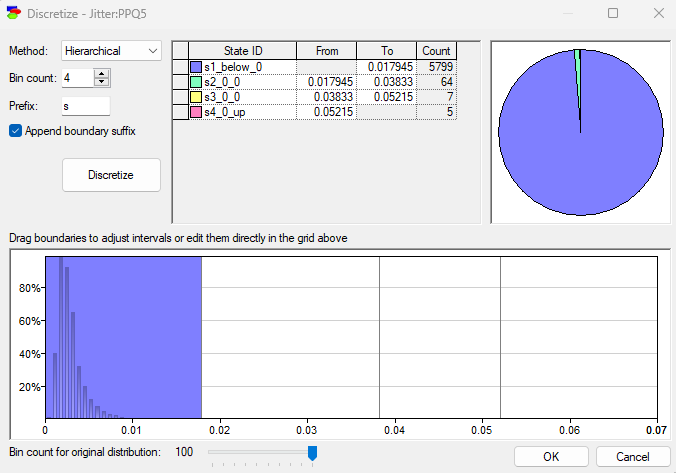
Jitter(Abs):



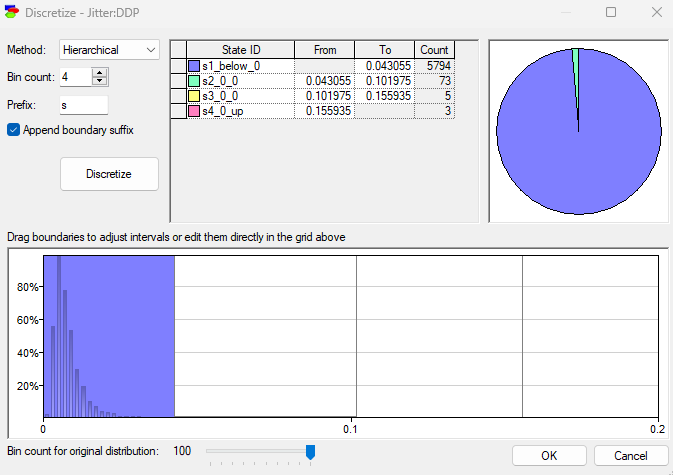
Jitter:RAP:



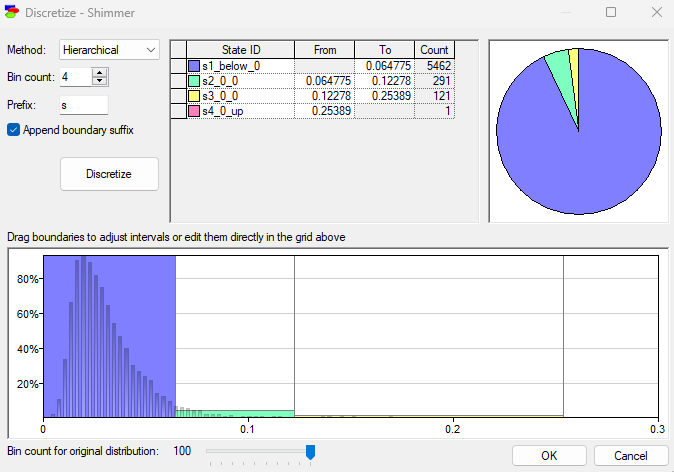
Jitter:PPQ5:



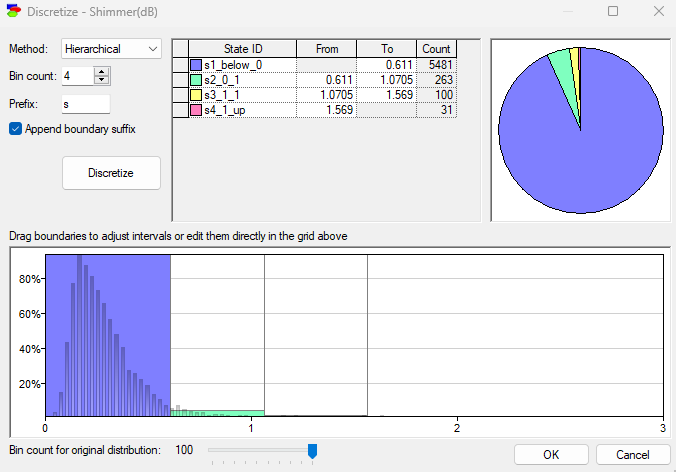
Jitter:DDP:



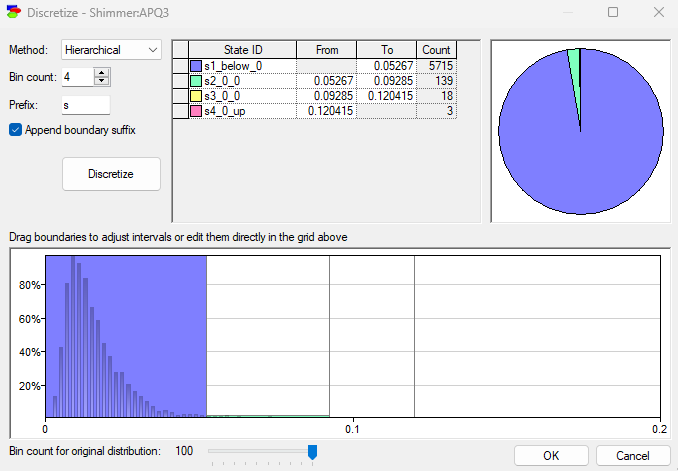
Shimmer:



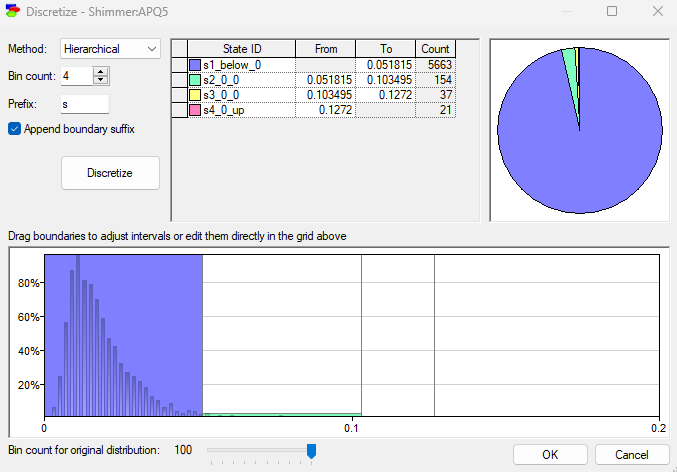
Shimmer(dB):



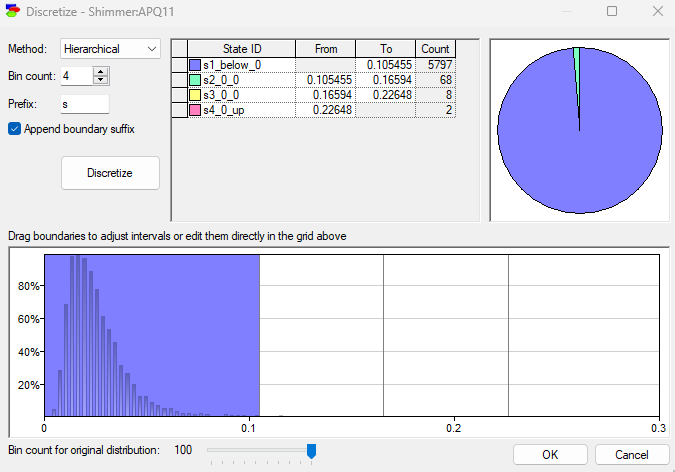
Shimmer:APQ3:



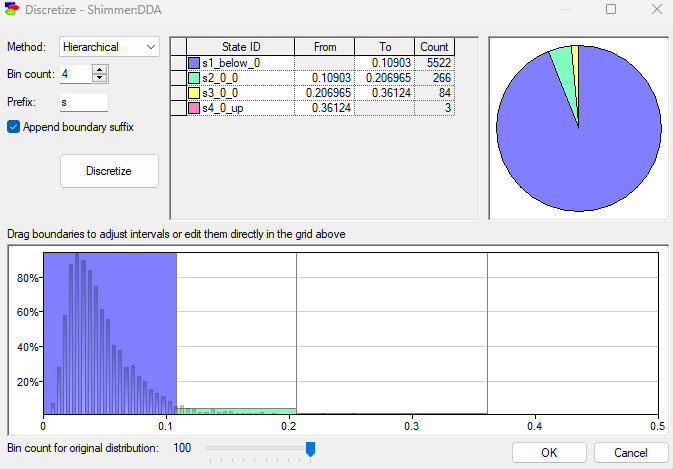
Shimmer:APQ5:



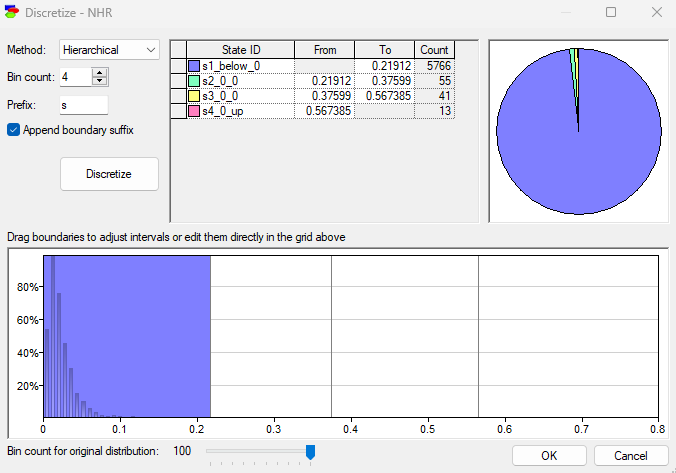
Shimmer:APQ11:



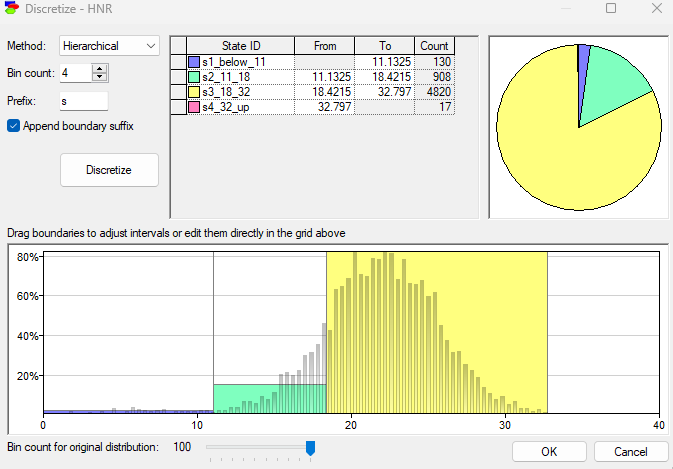
Shimmer:DDA:



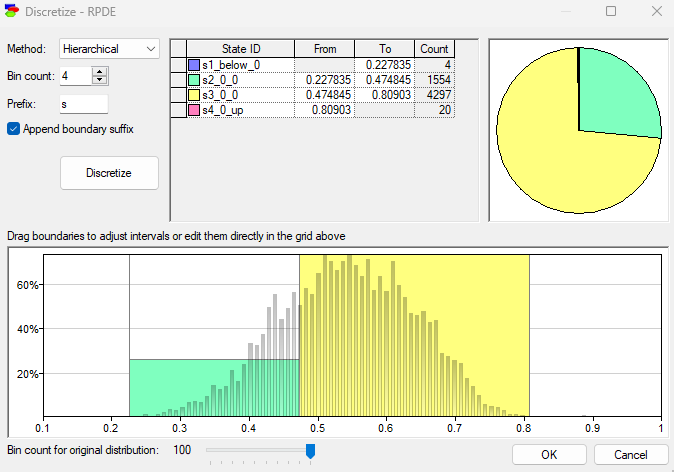
NHR:



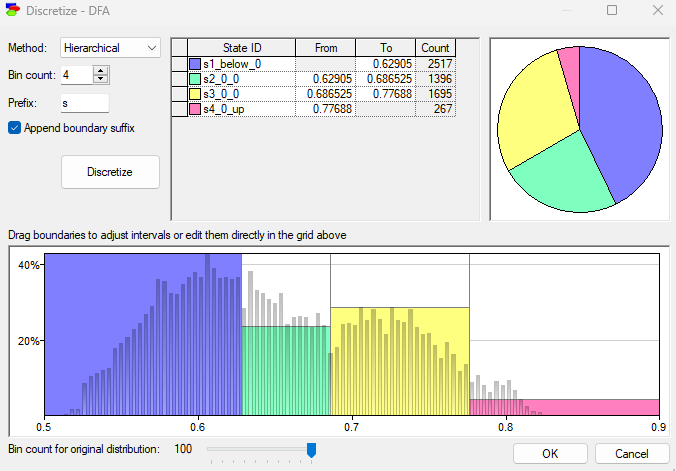
HNR:



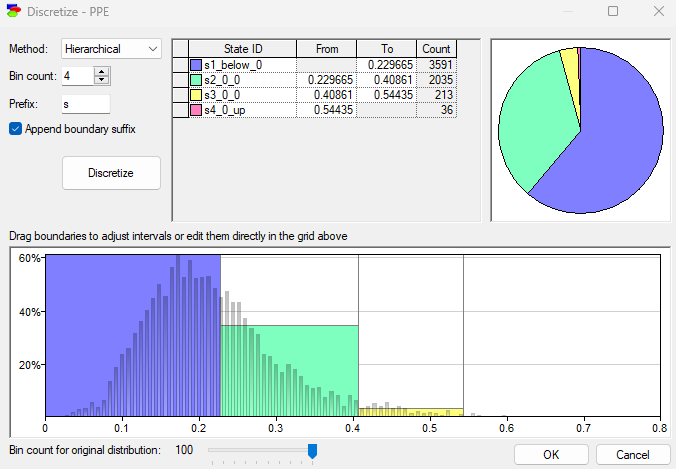
RPDE:

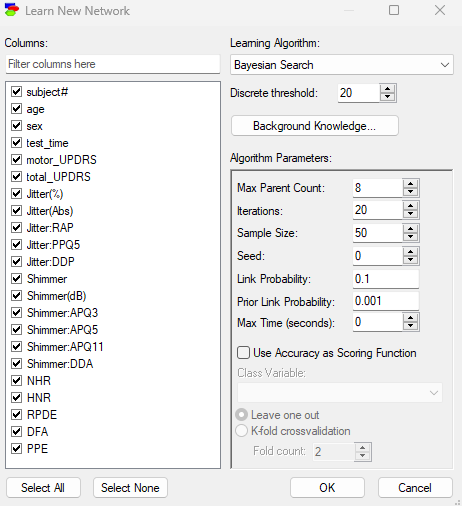


DFA:

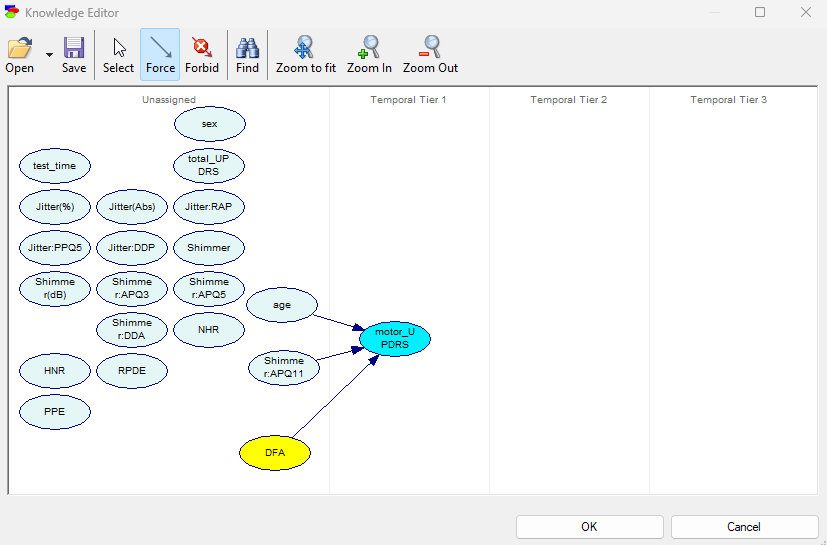


PPE:

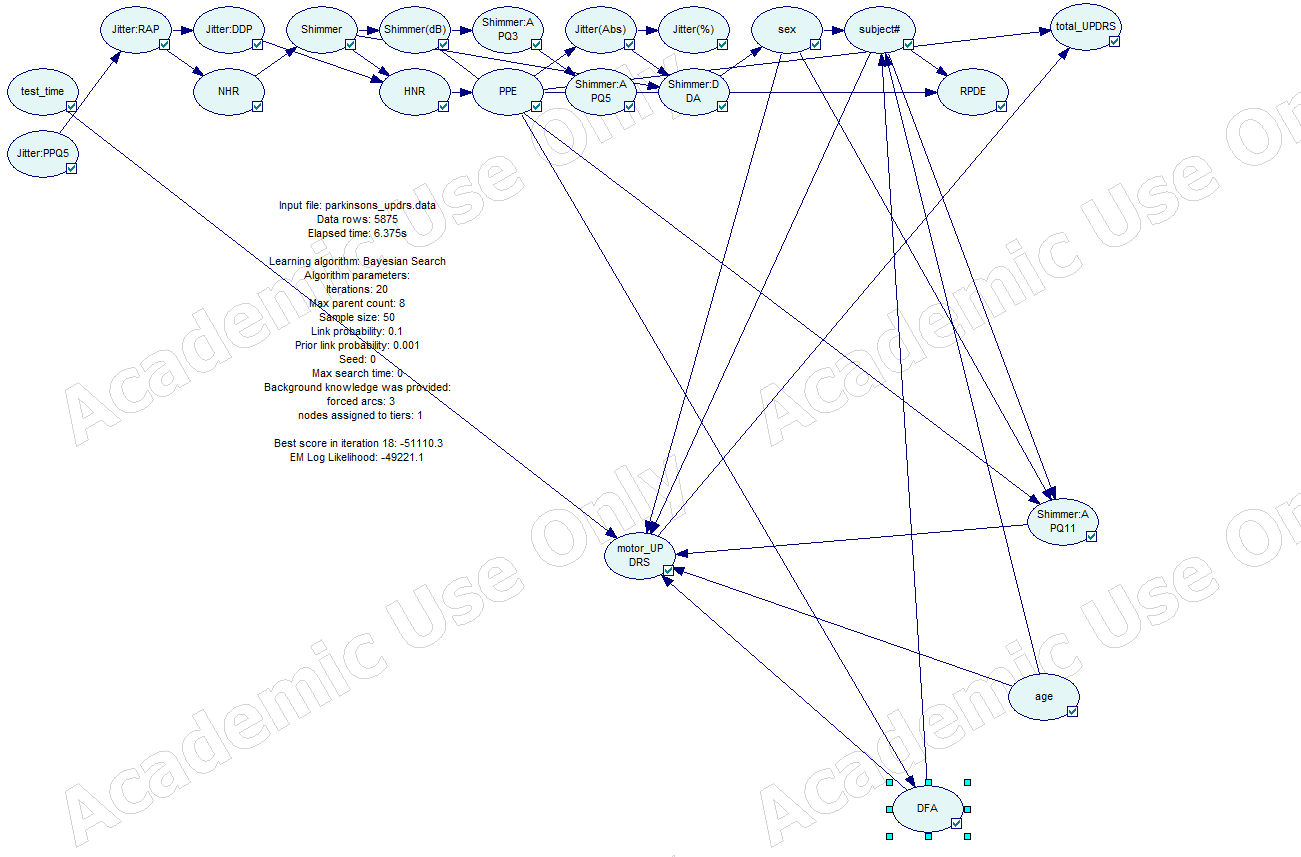


Next step is to learn the network:  


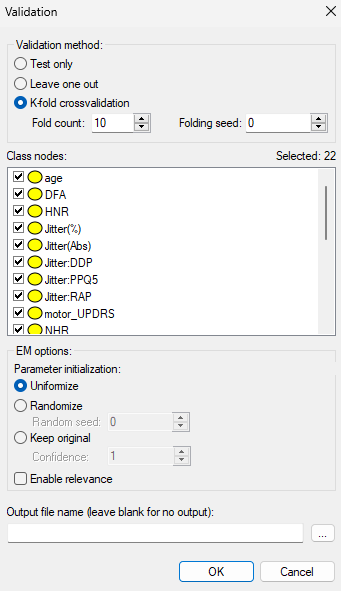
We enforce 3 direct link which had the highest IV value during project 2 (age/Shimmer:APQ11/DFA)



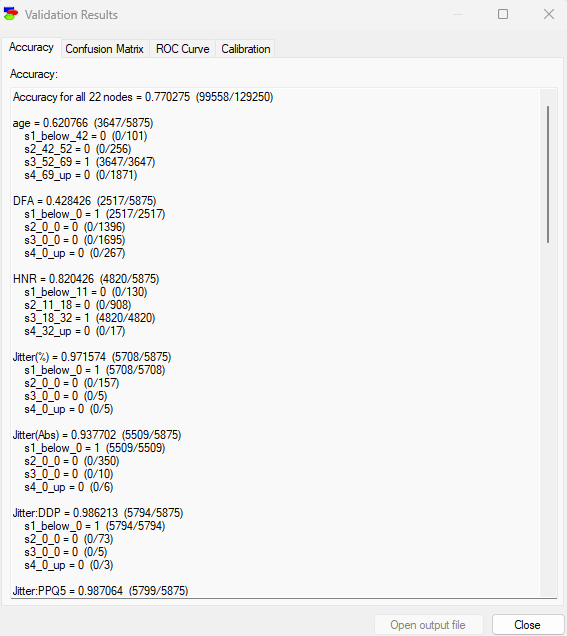
After that, the network will be generated as follow:



Next will be using CrossValidation, which devides the data set into K parts of equal size, trains the network on K-1 parts and uses the last part for testing purposes. (in this case 9 training, and 1 testing parts)

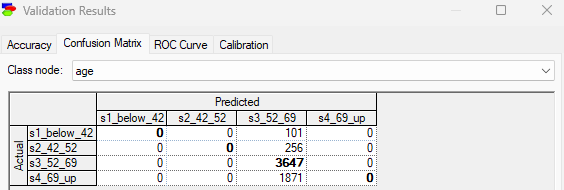


Scoring results:

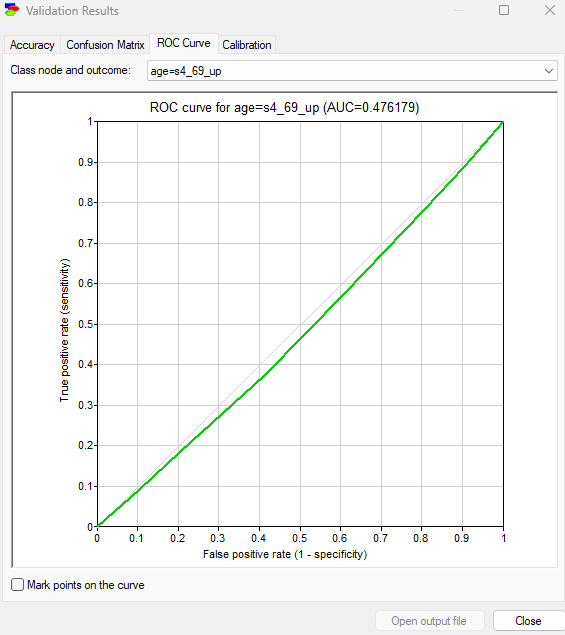


Confusion Matrix:

The confustion matrix shows us the number of correct and incorrect classified records.



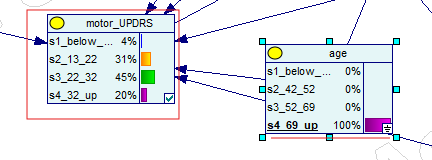
ROC Curve for variable “age”:

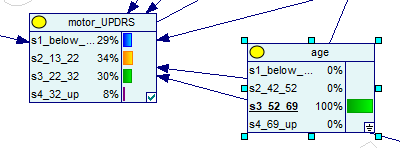


## Sensitivity analysis:

Case example 1:

Setting the age attribute to 69+, we can see the result obtained in our data model, that showed the severity of the parkinsons (20% above 32, 45% between 22 and 32).

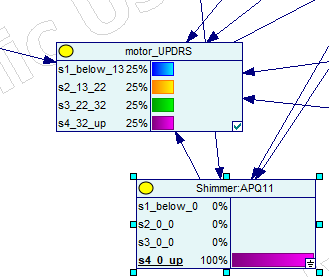
  
  
then we set age to be between 52\_69 to 100%



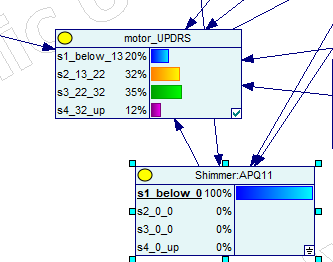
We can see the result, the data model was updated to show us, that the severity of the parkinsons went down on the high end (8% above 32, 30% between 22 and 32).

Case example 2:

Setting the Shimmer:APQ11 attribute to the highest level, we can see the result obtained in our data model, that showed the severity of the parkinsons (25% above 32, 25% between 22 and 32).



then we set Shimmer:APQ11 to the lowest level



We can see the result, the data model was updated to show us, that the severity of the parkinsons went down on the high end (12% above 32).

References used:  
-<https://support.bayesfusion.com/docs/GeNIe.pdf>

-https://arxiv.org/abs/1512.02406  
-[1512.02406 (arxiv.org)](https://arxiv.org/pdf/1512.02406)