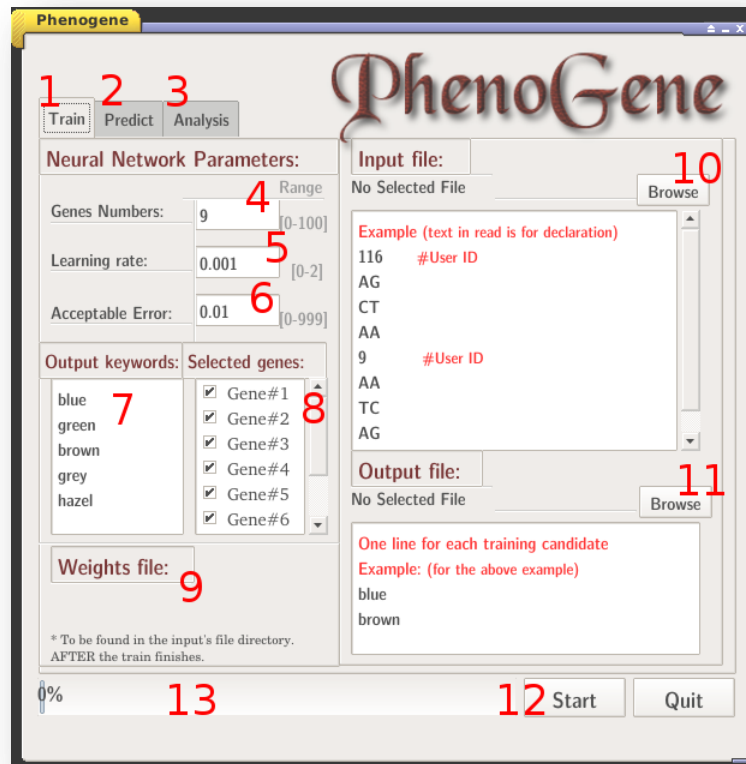


APPENDIX A | USER MANUAL

TRAIN TAB



1. Training tab
2. Predicting tab
3. Analysis tab
4. Number of input genes, i.e. how many gene SNPs do you have for each user?
5. Learning rate: This is usually a number between 0.3-0.7, if the dataset is too small choose even a smaller number. This number is directly proportional to your dataset size.
6. The least error you'd be tolerant about. Usually if this is the first time you're training the network you won't know this value, choose a small number or even zero then. The network will then continue to train until maximum iterations number is reached.
7. Phenotypic traits you're predicting. These keywords should match with your expected_output_file keywords.
8. Select which genes to train. This is mainly used for studying the effect of excluding a gene or more on the prediction and error rates.
9. The weights file will be produced upon successful training. You can find it in the same directory as the input and expected files you browsed earlier.

Mainly, the weights file describes the relation between the input and the output, and it will be used later on for actual prediction on new data.

10. Browse your input file. Choose one that follows the correct format specified below:

Your data file is a text file that has many individual records. For each individual the record starts with an integer number, his ID, follows the 2-letters gene SNPs of that individual each in a line. Then the second ID for the second individual follows his gene SNPs and so on so forth.

Example:

```
1
AA
AT
2
CC
CG
```

11. Browse your expected file. Choose one that follow the correct format specified below:

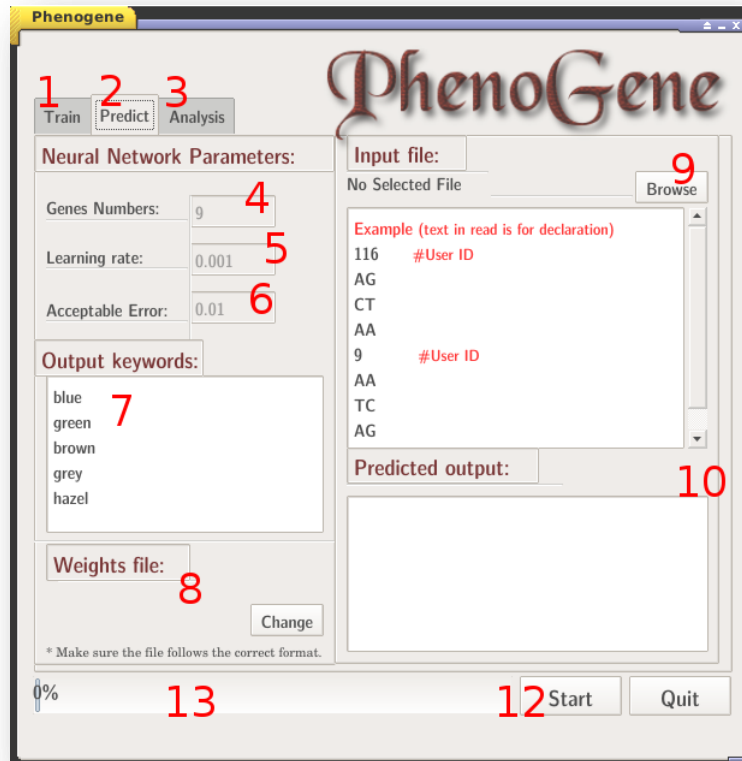
Your data file is a text file that has many individual records. For each individual there's exactly one line. The line may have one keyword or more than one keyword separated by a comma.

Example:

```
blue
green
```

12. Press this button when you have filled all the data.
13. Progress bar, this is updated each time the network makes a progress.

PREDICT TAB



1. training tab
2. Predicting tab
3. Analysis tab
- 4, 5, 6 and 7 are not editable in the predicting tab. If you really want to edit these parameters then go to the training tab and edit, the fields will be automatically synchronized then.
8. Choose the weights file produced earlier in the training phase.
9. Browse input file, it has the same format as the training tab specified earlier.
10. This is the predicted output. Another output file with the same data is produced in the same directory as the input file.

ANALYSIS TAB

PhenoGene

1 Train 2 Predict 3 Analysis

[Warning: This is an expert mode tab. Do not edit if you don't know what you're doing!]

Number of hidden nodes: 7 4 [0-100]

Momentum: 0.00000 5 [0-2]

Maximum iterations: 10000 6 [0-10000000]

Error Analysis

Error Percentage: 7 77.0833% Learning Rate: 15 0.0001

Number of mismatches: 8 37 Accepted Error: 16 0.01

Iterations used: 9 10000

Least actual error reached: 10 28.4555

Number of input nodes: 11 2

Number of hidden nodes: 12 7

Number of output nodes: 13 5

Momentum: 14 0

17 Import configurations

18 Export configurations

19 Produce PDF report

Quit

This is an expert mode tab

4. Number of hidden nodes is usually a number between input and output nodes number.
5. Momentum is used to accelerate the learning when the network is too slow, or the dataset is too big.
6. Maximum iterations number is reached when the acceptable error isn't reached during the training phase.

7-16 error measurements to be produced after the training has finished.

- 17.Import all the configurations saved earlier.
- 18.Save all the current configurations to a file so you can restore them later.
- 19.Produce a PDF report of the current results. The .pdf is to be found on the input's directory. Make sure no previous PDF's is written in the same directory as this will overwrite any previous reports.