The Opioid Epidemic, the rapid increase in the use and addiction of prescription or non-prescription opioids, has concerning impacts on the health, economy, and welfare of society. Opioid addiction mostly impacts the middle class, significantly impairing the work force, and is the leading cause of death for Americans under the age of 50, two thirds of which are caused by opioids. The U.S. spends $700 billion annually, $78.5 billion of which is due to opioids. The ability to predict the risk of opioid dependence for an individual patient could help decrease the risk of that patient suffering from addiction to prescribed analgesic drugs. Through the use of SSADDA and genetic data, previous studies were able to identify specific genes that can increase the risk of addiction. In my project, I proposed a machine learning methods, which will predict whether or not a person should have these genes by finding patterns in their phenotypes and genotypes. I started by collecting the specific data I needed from files containing the subject’s experience with drug abuse, medical history, and environment growth up. I extracted specific variables and combined them into one file. I recorded that data into a dataset of variables (1s and 0s) and dummy variables, while removing variables that contain too many NAs (>1300), making the variable useless. With more time, I would have used care to analyze the data and produce meaningful results.

Expected Results

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Opiate Dependence</th>
<th>Cocaine Dependence</th>
<th>Alcohol Dependence</th>
<th>Tobacco Dependence</th>
<th>Marijuana Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>5189</td>
<td>4892</td>
<td>1797</td>
<td>1872</td>
<td>2334</td>
</tr>
<tr>
<td>EA</td>
<td>291</td>
<td>291</td>
<td>291</td>
<td>291</td>
<td>291</td>
</tr>
</tbody>
</table>

Due to a lack of time, I was unable to perform the machine learning and produce results for my project. However, I can make predictions on what my results might have been based off a scientific paper about “Enhancing Genetic Prediction of Substance Dependence Using Personality Trait Data”.

In that experiment, 1,873 AA and 1,487 EA test subjects were used. These subjects were administered the same SSADDA test as in my project, as well as the Revised Neo Personality Inventory (NEO PI-R) and were genotyped. Then, several tests were conducted with the genomic data to determine if including the subject’s personality data improved the accuracy of the test. The ROC curves (See Figure 1) show an improved performance overall when genetic data was tested with the personality data, as opposed to testing the genetic data alone. However, the personality data could not predict addiction on its own, because people’s personality traits can vary highly regardless of genetics. There is a “high genetic variability of personality traits”.

Within the different addictive substances tested, cocaine addiction worked the best; the personality data drastically improving the results with the genetic data (see Figure 1). However, the adding the personality data with the genetic data actually worsened the accuracy for opioid addiction (see Figure 2).

In my project, I think the machine learning would have produced more accurate results than this test because I used medical history data as opposed to personality data, which can be very subjective. Regarding my abstract, when testing opioid dependence, I may not have received the results I was expecting. But since I tested other drugs as well, if the results might have been produced that show the drug environmental data and psychiatric data can improve the accuracy of genome-wide association tests.

To the left (Figure 1), the black line shows the test run with genetic data alone. The three other colored lines are tests done with personality data at different percentages. The blue line, which is 50% personality data added, is a much stronger curve and results shown validated.

Below (Figure 2), three lines with personality data added, and the top line showing significantly different results compared to the genetic data alone.

These figures are based on Ali Z. Jiwani, Ryan Koesterer, David Van Lin, and Lily M. Rajaipakse in their paper “Enhancing Genetic Prediction of Substance Dependence Using Personality Trait Data”.

Discussion

Machine Learning is a subfield of computer science, and a novel approach to analyzing non-linear relationships in data. The machine learning method is basically a combination of pattern learning methods and artificial intelligence methods, through a complex system of mathematics optimization and data mining.

With large, complex data sets with many variables and thousands of elements, the machine learning method can be extremely beneficial and producing astounding results. In my project, I have added the machine learning method to a model with the genetic data was tested with the personality data, as opposed to testing the genetic data on their own. I also added the personality data to the machine learning method because it does not require a linear combination of predictor variables, yet, it is still able to produce a pattern. Where as, it would be impossible to use a normal linear regression method to create any sort of pattern and produce results.

Conclusions

With more time on my project, I could’ve run the machine learning and produced results. From this point, it would be easy to continue this project. Future directions would include programming the machine learning, analyzing its results, and allow the computational machine learning process to make predictions on genotypes based on the subject’s substance-dependence, environmental factors, and psychiatric data and medical history. Although the expected results point to the fact that the personality data is more than a small percentage of the genome associated with the introductions of personality data, it still showed promising results for the other substances. Through my project and machine learning, early diagnoses of addiction can be made for a vast number of different substances, and can lead to early treatment and preventative measures to keep people safe and help to decrease the drug abuse epidemic in the U.S.