

Correlating the personality and diseases using genomes

Gowtham Velmurugan
2015503517

Swetha B
2015503049

Girinathaprasad B
2015503013

*Under the Guidance of
Dr. P. Jayashree*

INTRODUCTION

Personality traits are the relatively enduring patterns of thoughts, feelings and behaviors that reflect the tendency to respond in certain ways under certain circumstances. Twin and family studies have showed that personality traits are moderately heritable, and can predict various lifetime outcomes, including psychopathology. Predict the Personality Trait of the person based on his genome analysis to forecast the disease that could occur to the individual. Personality Traits were measured with the NEO Personality Inventory, which was designed to characterize the five-factor model of the 'Big Five Dimensions' of Personality. With the development of behavioral genetics, personality was established an inheritable trait. On average female adults were more neurotic and agreeable than male adults. The results tended to support previous research in that a negative relationship was found between neuroticism and cancer, and a positive relationship was found between neuroticism, stress-related symptoms, and cardiovascular symptoms.

LITERATURE SURVEY

- 1) Pathway analysis of genome-wide association datasets of personality traits, H.-N. Kim, B.-H. Kim, J. Cho, S. Ryu, H. Shin, J. Sung, C. Shin, N. H. Cho, Y. A. Sung, B.-O. Choi and H.-L. Kim. Analyzing correlation between genome and traits based on 1042 pathways. Analyzing of interactions of gene set with the personality traits.
- 2) Genome-wide association study of the five-factor model of personality in young Korean women. Han-Na Kim, Seung-Ju Roh, Yeon Ah Sung, Hye Won Chung, Jong-Young Lee, Juhee Cho Hocheol Shin and Hyung-Lae Kim. Performed a GWA study of 2,053,685 SNPs from 1089 women and analyzed the SNPs that relate to the personality trait. It Showed the key genes associated with the personality traits.
- 3) Genome-wide association scan for five major dimensions of personality. A Terracciano, S Sanna, M Uda, B Deiana, G Usala, F Busonero, A Maschio, M Scally, N Patriciu¹, W-M Chen³. Recruited subjects, Personality assessment, Genotyping and imputation, GWA analysis on the data with traits. The Merit is Provide the P values and Z values associated with each trait.

BACKGROUND STUDY

CONVOLUTION NEURAL NETWORK:

Convolutional Neural Networks are very similar to ordinary Neural Networks from the previous chapter: they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. SVM/SoftMax) on the last (fully-connected) layer and all the tips/tricks we developed for learning regular Neural Networks still apply. A Conv Net architecture is in the simplest case a list of Layers that transform the image volume into an output volume (e.g. holding the class scores). There are a few distinct types of Layers (e.g. CONV/FC/RELU/POOL are by far the most popular). Each Layer accepts an input 3D volume and transforms it to an output 3D volume through a differentiable function. Each Layer may or may not have parameters (e.g. CONV/FC do, RELU/POOL don't). Each Layer may or may not have additional hyperparameters (e.g. CONV/FC/POOL do, RELU doesn't)

CONVOLUTION NET ARCHITECTURE:

LENET: The first successful applications of Convolutional Networks were developed by Yann LeCun in 1990's. Of these, the best known is the LeNet architecture that was used to read zip codes, digits, etc.

ALEXNET: The first work that popularized Convolutional Networks in Computer Vision was the AlexNet, developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton. The AlexNet was submitted to the ImageNet ILSVRC challenge in 2012 and significantly outperformed the second runner-up (top 5 error of 16% compared to runner-up with 26% error). The Network had a very similar architecture to LeNet, but was deeper, bigger, and featured Convolutional Layers stacked on top of each other (previously it was common to only have a single CONV layer always immediately followed by a POOL layer).

ZF NET: The ILSVRC 2013 winner was a Convolutional Network from Matthew Zeiler and Rob Fergus. It became known as the ZF Net (short for Zeiler & Fergus Net). It was an improvement on AlexNet by tweaking the architecture hyperparameters, in particular by expanding the size of the middle convolutional layers and making the stride and filter size on the first layer smaller.

GOOGLENET: The ILSVRC 2014 winner was a Convolutional Network from Szegedy et al. from Google. Its main contribution was the development of an *Inception Module* that dramatically

reduced the number of parameters in the network (4M, compared to AlexNet with 60M). Additionally, this paper uses Average Pooling instead of Fully Connected layers at the top of the Conv Net, eliminating a large amount of parameters that do not seem to matter much. There are also several follow up versions to the GoogLeNet, most recently Inception-v4.

VGGNET: The runner-up in ILSVRC 2014 was the network from Karen Simonyan and Andrew Zisserman that became known as the VGGNet. Its main contribution was in showing that the depth of the network is a critical component for good performance. Their final best network contains 16 CONV/FC layers and, appealingly, features an extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end. Their pretrained model is available for plug and play use in Caffe. A downside of the VGGNet is that it is more expensive to evaluate and uses a lot more memory and parameters (140M). Most of these parameters are in the first fully connected layer, and it was since found that these FC layers can be removed with no performance downgrade, significantly reducing the number of necessary parameters.

RESNET: Residual Network developed by Kaiming He et al. was the winner of ILSVRC 2015. It features special *skip connections* and a heavy use of batch normalization. The architecture is also missing fully connected layers at the end of the network. The reader is also referred to Kaiming's presentation (video, slides), and some recent experiments that reproduce these networks in Torch. ResNets are currently by far state of the art Convolutional Neural Network models and are the default choice for using ConvNets in practice (as of May 10, 2016). In particular, also see more recent developments that tweak the original architecture from Kaiming He et al. Identity Mappings in Deep Residual Networks (published March 2016).

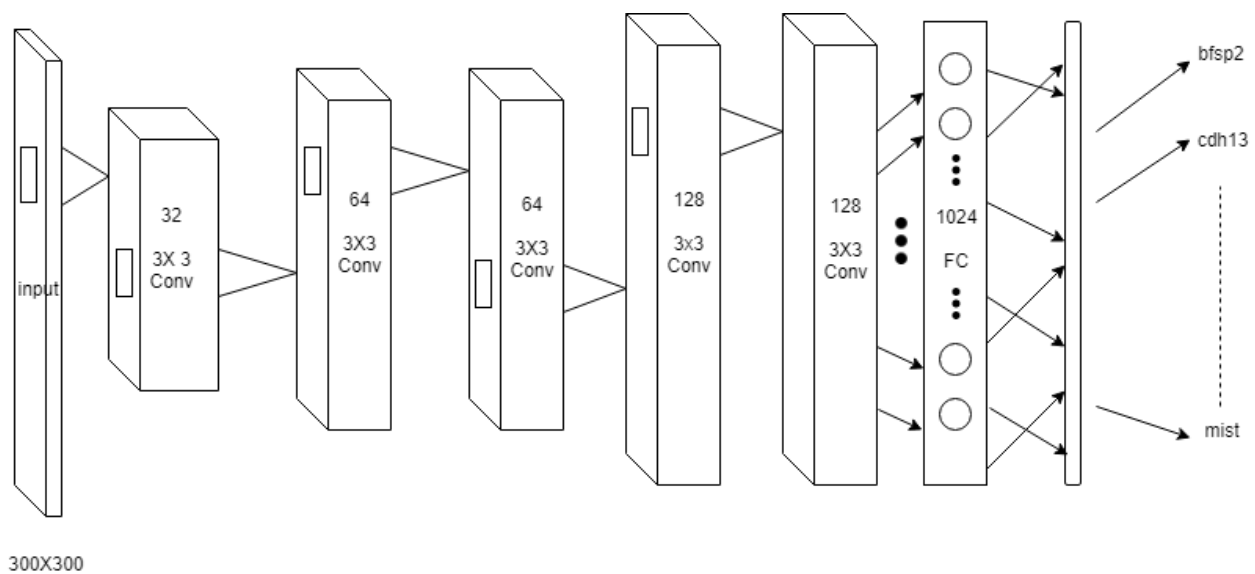
PROPOSED WORK

As we know that the Personality traits are the relatively enduring patterns of thoughts, feelings and behaviors that reflect the tendency to respond in certain ways under certain circumstances. Twin and family studies have showed that personality traits are moderately heritable, and can predict various lifetime outcomes, including psychopathology. We can analyze the genotype and the single nucleotide polymorphism associated with a person and which can be used to solve the problem of identifying the personality trait of the person. The Main ideology of the project is that finding personality trait and finding the diseases associated with the genotype and SNP which was published earlier research works which was mentioned in the references. Parse the Genome sequence for processing in our case we have represented the genomic sequences into the image with using five colors as its that is Guanine (G) -> Green, Thymine (T) -> Blue, Cytosine(C) -> Black, Adenine(A) -> Red, Gaps -> White. Convert all the gene sequences that are related with the diseases and personality and form the image dataset which is going to form as the dataset for the gene type prediction and SNP prediction associated with the sequences. The gene prediction and SNP prediction problem can be solved

with the help of Convolution neural network here in our case we have used the VGG Net which is mentioned in the Background Study. Then by using the correlation which have already mentioned in the reference paper [2]. We should make a fuzzy inference model or any machine learning model (under discussion phase) that is going to give apt predictions. Then using the result obtained we should find the correlation between the personality trait and diseases with genomes as its base.

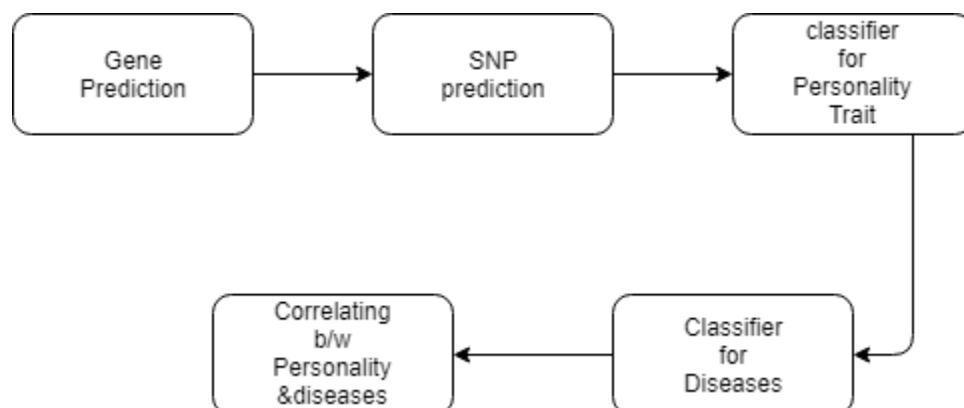
MODEL ARCHITECTURE

Convolutional Neural Net Architecture for Genotype Prediction



MODULES INVOLVED

The Following Flow Represents the Modules that are involved in this project.

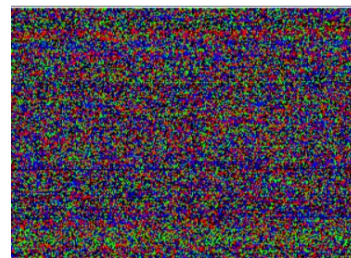


RESULT

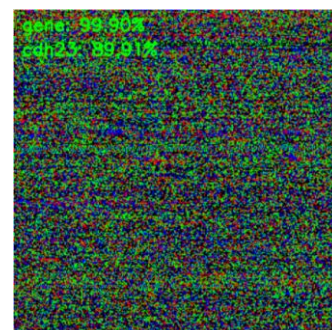
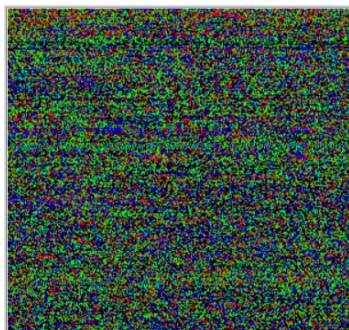
As of Now module one is Successfully completed. Here Comes the final outcomes of the Module one. The Model predicts with an accuracy of 92.86% and inference time is 2 seconds (approx.) with a model size of 11,000 Kilobyte.

SEQUENCE TO IMAGE CONVERSION:

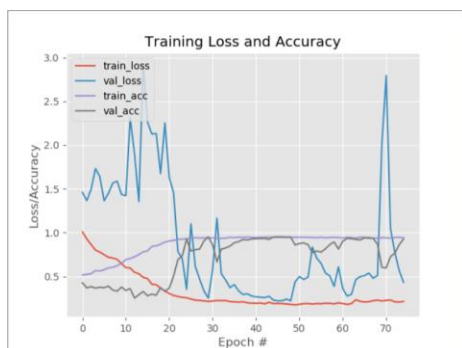
```
>NC_000003.12:c133491146-133426402 Homo sapiens chromosome 3, GRCh38.p12 Primary Assembly
TGAACCCATGGCTGCGGGCTGTGCTGAGGTCAATGTCAGGCCCTGAAGGACTCGCTGAGCATGACAATCA
AGTGAAGAGGGCTCTTGAATTGCACCTTGCCCTGAGGGGGGCCCTTGCTCTGCCACCTTGGTCTCTGC
TCAGGCTGCTGGGGCTGAGGCCCTGAGGCTGAGGATGGGGAGAGAGCCTTCAGTATGTGGGCCGCCCTCT
GTCAGAACAGTGTGTTGCTGCTCTCTGCTATCTACGTGAGTAGCTCTCAATCTGCTGATAATGTGGGTA
TCTTTGCTCTCTACAGTTAGCTGTGAGTTGATGGCTTCAAAGCCAGTCTCTCAATAGCTTCTGCTCTCA
GGGAGGGAGCTCCATGCTTGAAGAGGGGCCATTGCTGCTGCTGGTGGGGCTAAATCAGGCCCTGGCTGGG
TCATCGGCACCGGTGATAGCAGGGCAGGCTCCGCACAGTGTGGGCTCTCAGCACAGAGGGTGGCAC
CGTGGCATCCACCTGGGCCACACTGTATCTGGGGGAAGAACTCCATGGCAATGTTTCATTATACCTGGC
CCTGCTGGCTCATCCAGGTGGTCACTATGACACATCCTTACCACACTGACATTGCTGCTTTAGCCACC
AACATGCGACCAACCAACACAGATCACCTTGGTCAGCCAGCAACCTCCACCTCCCAACAGTCAATGTC
TATCAGTCAAAATTCCTAGAATCAATAAGTATAAAAGAACTTAACCTCTCACTTAAAAAATTTTG
CTTCTGCTGCTGATGAAATGTTAATATTGTGAATCTGAGTGAAGGGTATAGGCAATCTTCAACT
ATTCTTTCAACTGCTCTGTAAGCCTGAAATATGTTAAAAATAAAATTCAAAAACAAAGCCAAAAACCT
TGTACCTTAAGGGTTACTTGAGTTAGTGGTTCACAGCAACAGATACTCAATCTGGTTGTGTACTTC
TGAAAAATTAAGCTTCATATCTTTTCTTCATGGCTCTTCTGCTGTAAGTACAACCTGAATCTCATGTT
```



GENE PREDICTION:



PLOTTED GRAPH FOR PREDICTION:



REFERENCES

- [1] A Terracciano, S Sanna, M Uda, B Deiana, G Usala, F Busonero, A Maschio, M Scally. *Genome-Wide Association Scan for Five Major Dimensions of Personality*.2013
- [2] Han-Na Kim¹, Seung-Ju Roh¹, Yeon Ah Sung², Hye Won Chung³, Jong-Young Lee⁴, Juhee Cho⁵. *Genome-Wide Association Study of The Five-Factor Model of Personality In Young Korean Women*.2010
- [3] H.-N. Kim, B.-H. Kim, J. Cho, S. Ryu. *Pathway Analysis of Genome-Wide Association Datasets Of Personality traits*.2015.
- [4] S. Sanchez-Roige, J. C. Gray, J. MacKillop, C.-H. Chen, A. A. Palme. *The genetics of human personality*.2017.
- [5] A. Narayanana, E.C. Keedwella, J. Gamalielssonb, S. Tatinenia. *Single-layer artificial neural networks for gene expression analysis*.2003.