Use of Deep Learning to Examine the Association of the Built Environment with Mental Health

1 Abstract

In this paper, we adopt a novel method for the task of predicting the prevalence of mental health issues in a city. By utilizing socioeconomic data, as well as built environmental features derived from satellite imagery, to predict the prevalence of mental health issues in a city, the method provides an alternative to telephone and in-person interviews. Through this method, we demonstrate that built environmental features do correlate with the prevalence of mental health issues. Our multi-modal model achieves a maximum of 0.86 R2 score for Intra-City predictions and a maximum of 0.78 R2 score for Multi-city and Cross-City predictions. Hence, our proposed method demonstrates a reliable, uniform method to predict mental health issues, and could be applied to other cities in the world that are not included in this study.

2 Introduction

According to the National Alliance on Mental Health, 1 in 5 U.S. Adults experience mental illness each year. Despite its prevalence, mental health remains a topic that is often glossed over or unspoken. Currently, mental health evaluations are performed through surveys and in-person interviews (BRFSS). Although these surveys are conducted annually and encompass all 50 states, they require immense time and manpower. These surveys are also self-reported and does not include the reasons causing their issues.

Mental health is a complex topic that is affected by a person's genes, psyche, lifestyle and other environmental factors [1]. The amount of green space a person has, or the person's housing condition might create stress and trigger mental disorders. As increasing evidence has been shown to link the impact of environment to mental health, researchers have decided to quantify this impact of built environmental features [2]. Although numerous published studies have shown the impact of built environmental features on one's mental health, inconsistencies exist across multiple findings due to the lack of a universal measurement tool.

In this project, we hope to create a consistent method to measure physical attributes of living environments and correlate the findings to the prevalence of mental health issues within that neighborhood. In doing so, we hope to be able to uncover physical attributes of the neighborhood that influences the mental health status of the residents. In addition, we also hope to provide an alternative to the labor-intensive data collection process by establishing a uniform method to predict mental health issues.

The outcome of the project can be used to help inform policy makers what type of building structures and recreational space they can propose to change in order to reduce the prevalence of mental health illnesses. In addition, the project could also be used to provide a consistent and reliable data point to organisations that evaluate mental health statuses of the neighborhood as a whole.

3 Related Work

On one hand, previous research on the intersection of mental health and built environmental features have been conducted through cross sectional survey and on-the-ground observations [3,4]. On the other hand, previous attempts to predict mental health status have been limited to the use of social media streams such as Twitter and Facebook [5]. While
paper by Adyasha Maharana and Elaine Okanyene Nsoesie examines the association of the environment with prevalence of obesity. The paper uses a convolutional neural network, which is applied to more than 150,000 satellite images of the built environment, to extract features. It uses adult obesity data from the CDP (Centers for Disease Control and Prevention) 500 cities project and uses regression models to quantify the association between the features and obesity prevalence across census tracts [10].

4 Problem

Inconsistencies exist in the measurement of surrounding environmental features. This leads to varying conclusions on how the surrounding environment affects one's mental health status. Our goal is to create a consistent method to measure physical attributes of living environments and correlate the findings to the prevalence of mental health issues within that neighborhood. With a consistent methodology established, we can then begin exploring whether the surrounding environment is associated with the prevalence of mental health issues, and predict the prevalence of mental health issues in the neighborhood in a uniform manner.

5 Data

5.1 Data on Mental Health Prevalence: 500 cities Project

To conduct an analysis on the relationship between built environment and prevalence of mental health issues, we obtained crude and age-adjusted estimates on the prevalence of mental health issues at the census tract level in 2016. The estimates were obtained from the 500 Cities Project and were derived from the United States Centers for Disease Control and Prevention (CDC)'s Behavioral Risk Factor Surveillance System (BRFSS). BRFSS is a telephone-based health survey system used to collect data regarding health-related risk behaviors of U.S. residents. It is conducted annually and encompasses all 50 states of the U.S.

The formal description of the data points obtained is "Mental health not good for ≥14 days among adults aged ≥18 years".

It is important to note that the data collected from the survey is self-reported and is not a formal medical diagnosis. Due to the subjective nature of the data, the data collected is subjected to limitations such as social desirability bias and a lack of coverage in certain areas.

To test whether mental health issues correlate with the surrounding infrastructure, we've narrowed down the sample size to 4 cities: Los Angeles, Memphis, San Antonio and Bellevue. These cities were selected due to availability of pre-processed satellite imagery. Among the 4 cities, Memphis is one of the top 15 cities with the highest prevalence of mental health issues among the 500 cities. Los Angeles and San Antonio have a moderate prevalence of mental health issues and have data values relatively closer to the mean of all data points from the 500 cities than the other cities in this sampling. On the other hand, Bellevue is one of the top 10 cities with the lowest prevalence of mental health issues among the 500 cities.

The table (1) below describes the pre-processed datasets used in this study.

Table 1. 500 Cities Data Description after Pre-processing

City	**Mental health not good for ≥14 days among adults aged ≥18 years (Mean of all data values of the census tracts within the city)**	**Low Confidence Limit (Mean)**	**High Confidence Limit (Mean)**	**Total (Number of census tracts within city)**
Los Angeles	13.057%	11.78%	14.38%	993
Memphis	16.27%	14.69%	17.96%	178
San Antonio	12.52%	11.38%	13.70%	311
Bellevue	8.32%	7.48%	9.26%	28

5.2 Census Tract Shapefiles

In the process of obtaining satellite imagery, we used census tract shapefiles for the 4 cities to divide each survey location into corresponding polygons. This allows us to obtain multiple satellite images encompassing each census tract.

Below is a series of visualization of the census tract shapefiles and its corresponding data points collected from the 500 Cities Project.

Figure 1: Los Angeles Census Tracts

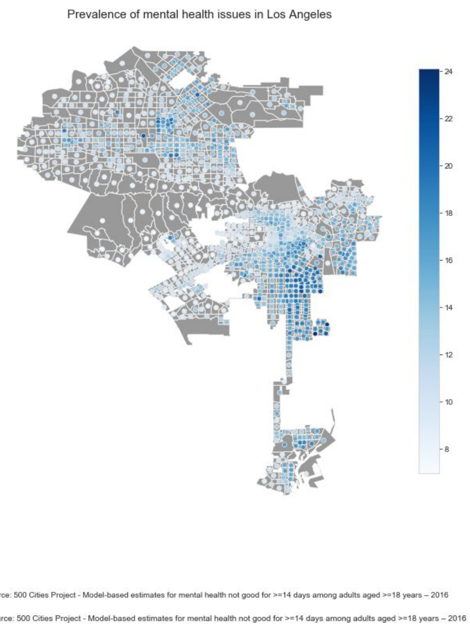

Figure 2: Memphis Census Tracts

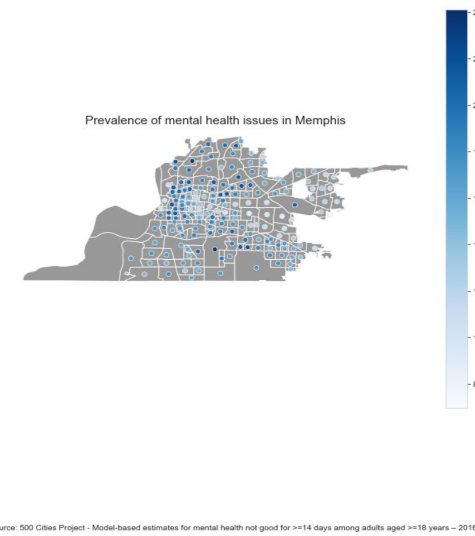

Figure 3: San Antonio Census Tracts

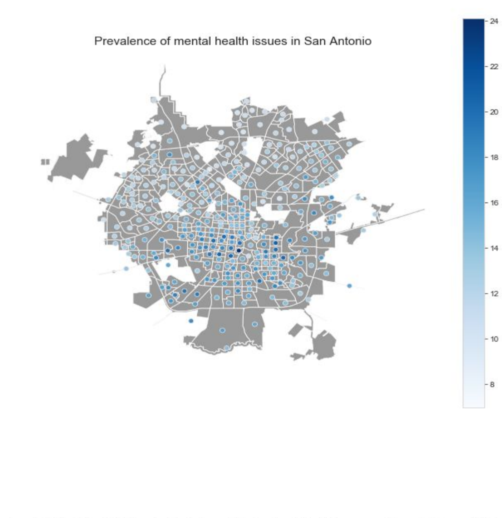

Figure 4: Bellevue Census Tracts

5.3 Satellite Imagery Data

To obtain built environment features, we obtained boundaries for each census tract from shapefiles. For each census tract, we proceeded to obtain all satellite images within the boundaries from the Google Static Maps API. Each image was at a zoom level of 18 and was 400 x 400 pixels. Due to the implementation of the VGG16 CNN, each image was equalized and resized to 224 x 244 pixels for image feature extraction.

The table 2 below describes the satellite image dataset obtained for this paper.

Table 2. Satellite Image Data Description

City	Number of satellite images	Number of census tracts
Memphis	48001	178
San Antonio	99718	311
Bellevue	8931	28
Los Angeles	66527	993

5.4 Socioeconomic Data

To obtain socio-economic features, we obtained data from ACS (American Community Survey) dataset from 2013 to 2017 (5-Year Estimates) in conjunction with data from Social Explorer dataset for the same time period. The datasets represent the average characteristics over the 5-year period of time and are published for all geographic areas including those with a population under 20,000. The columns considered in the dataset were:

- Population Density
- Sex
- Age
- Households by Household Type
- Employment Status for Total Population 16 Years and Over
- Occupation for Employed Civilian Population 16 Years and Over
- Per Capita Income (in 2017 Inflation Adjusted Dollars)
- Gini Index
- Nativity by Citizenship Status

There was a total of 22 columns obtained from the dataset.
The data was collected for 4 states: Washington (Bellevue), California (Los Angeles), Texas (San Antonio), Tennessee (Memphis).

6 Methodology

In this paper, we employed a transfer learning approach to extract surrounding environmental features from the satellite images. Details about convolutional neural networks and the transfer learning approach is described in Section 6.1 and 6.2. After the feature extraction process, the surrounding environmental features for each census tract were then grouped by their corresponding cities and inputted into a regression model to predict the prevalence of mental health issues within the city. Details of the regression model is described in Section 6.2.

6.1 Convolutional Neural Networks

Convolutional Neural Networks are a class of deep neural networks commonly used for image-recognition tasks. Due to the increased availability of data and computing resources, CNNs are now able to perform simple image recognition tasks at a human-level [5]. For instance, a recent paper proposed a parallel-computing methodology that uses an ensemble of convolutional neural networks, and were able to classify handwritten digits with 0.21% error rate [6]. As a result, CNN's are best suited for the task of classifying satellite images.

6.2 Transfer Learning Approach

Although CNN's are promising in the area of image recognition, it requires huge amounts of labelled training data and computing resources. Since there isn't enough computing resources and labelled training data for the task of identifying surrounding environmental features, we've decided to use a pre-trained CNN to perform image feature extraction. This process is called Transfer Learning.

The idea of transfer learning is to apply a pre-existing model architecture with pre-trained weights to a different but similar task [8]. For this paper, we used a pre-existing convolutional neural network, VGG16-CNN, that has been pre-trained on the ImageNet dataset to extract surrounding environmental features. We selected the weights pre-trained on the ImageNet dataset, a widely available dataset with over 15 million labelled images from roughly 22,000 categories [7], due to its high reusability. Since VGG16-CNN has been pre-trained on ImageNet to learn how to interpret images, our goal is to transfer the image-recognition knowledge learnt by the model to interpret satellite images obtained from the Google Statics Map API [9].

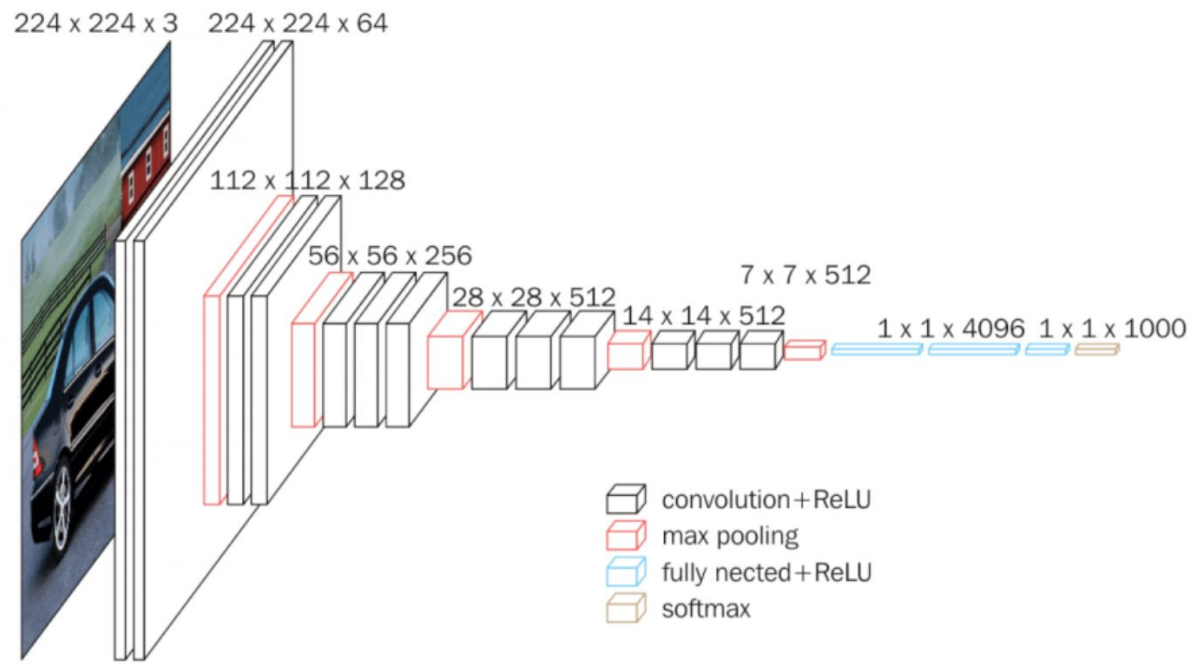

Figure 5: Visualization of VGG16-CNN's architecture

VGG16-CNN consists of 12 convolutional layers, 5 pooling layers, 3 fully-connected layers and a 1000-dimension softmax classifier [11]. To use the network, we pass a resized satellite image of dimension 224x224 as input. The image then goes through a series of convolutional layers that acts as filters to extract high and low-level image features. In each step, the pooling layers allow the network to reduce the parameters of its representation of the image. After the final pooling layer, the image goes through 3 fully-connected layers and a softmax classifier to output a value corresponding to its predicted category.

Since the 1000 categorical values outputted by VGG16-CNN are incompatible with our current task of identifying surrounding environmental features, we removed the prediction layer and used the last fully connected layer 'fc2' as extracted image features instead. As a result, our extracted image feature vectors have 4096 dimensions. These feature vectors were then grouped by city, combined with socioeconomic data, and fed into a regression model to predict the target variables obtained from the 500 cities project.

6.3 Feature Selection

We originally had 22 socioeconomic features, spanning from population density to single parent household percentages. After normalizing all features and implementing a recursive feature selection process, we narrowed down our selected features to 11 features with highest feature importances.

The features displayed below are ranked by their feature importance.

1. Per capita income [Highest]
2. Population density
3. Normalized population of adults aged 35-44
4. Foreign born population
5. Native born population
6. Gini Index (Representing wealth distribution in the neighborhood)
7. Normalized population in labor force
8. Normalized population not in labor force
9. Normalized 25-34 population
10. Normalized 18-24 population
11. Normalized 45-54 population

6.4 Regression Models

Due to the high-dimensionality of the input (4096 + 11 dim), Lasso Regression was used to predict the prevalence of mental health issues within a city based on surrounding environmental features.

Lasso Regression is a regression model that utilizes L1 regularization, a shrinkage and feature selection technique that minimizes prediction errors by reducing the coefficients of some features to zero. This technique allows the model to automatically determine which features are more closely associated to the target variable by looking at features with non-zero coefficients.

In addition to Lasso Regression, we also used Support Vector Regression and Linear Regression as baseline models. Lastly, we also chose Gradient Boosting Regression (GBR) for its ability to learn representations of the data.

Gradient Boosting is an ensemble learning method where the model sequentially builds new trees (weak learners) based on the previous tres's error. In GBR, trees are fitted to residuals/errors and built by choosing features leading to the lowest error. This feature makes GBR a good candidate for most machine learning tasks that require low error rates. It is important to note that GBR is often prone to overfitting and requires hyperparameter tuning for better generalization.

As a result, per each model, Grid Search cross validation was used to find optimized hyperparameters. For the Gradient Boosting Regressor, the maximum depth of each tree was set to 3. All the other hyperparameters were set to default. For Lasso Regression, the alpha was set to 0.1 as recommended by Grid Search Cross Validation. For Support Vector Regression, a 3rd degree rbf kernel was used by default.

6.5 Model evaluation metrics

To evaluate the models, the R2 scores was chosen as the main measure of how the models fared against each other in predicting the prevalence of mental health issues within a city. The R2 score mainly ranges from 0 to 1, and explains how close the actual data is to the fitted regression line. A 0 R2 score indicates that the model fares the same as simply predicting the mean value of the data points, whilst a 1 R2 score indicates that the model predicts the actual data with no error. It is important to note that a negative R2 score may appear when the model is tested on data it has never seen before.

In addition to the R2 score, the Mean Absolute Error, Mean Squared Error and Root Mean Squared Error were used to help evaluate the models.

7 Experiments and Results

To evaluate the performance of our regression models, we first split the set of extracted image features by city and conducted random sampling for fitting and testing purposes. Per each experiment indicated below, 5-fold cross validation was used to evaluate the performance of the regression models based on the measurement indicators listed in Section 6.4. The experiments were tested on an Intra-city, Cross-city and Multi-city basis.

Details and results of our sampling approach can be found below.

7.1 Intracity: Train and test on one city

For all experiments, we used 5-fold cross validation to randomly sample the set of extracted image features for each census tract location in the chosen city. The following tables illustrate the performance of each model in each city.

Note: Measurements are shown for the testing dataset only. R2 score is used.

Los Angeles:

Table 3: Performance of regression models (Intra-city: Train and test on Los Angeles)

	Linear Regression	Lasso Regression	Support Vector Regression	Gradient Boosting Regression
Socioeconomic features only	0.65	0.45	0.79	**0.87**
Image features only	0.06	0.29	0.31	0.38
Multi-modal features	0.67	0.73	0.80	**0.86**

As indicated by Table 3, Gradient Boosting Regressor outperforms Lasso, Support Vector and Linear regression in predicting the prevalence of mental health issues within Los Angeles based on all three types of features (multimodal, socioeconomic and image only).

In addition, the results show that the Gradient Boosting Regression model with multi-modal features outperform all other models with only socioeconomic features or image features. It also fares similarly to the Gradient Boosting Regression model trained on socioeconomic features only, with a 0.01 R2 score difference.

Bellevue:

Table 4: Performance of regression models (Intra-city: Train and test on Bellevue)

	Linear Regression	Lasso Regression	Support Vector Regression	Gradient Boosting Regression
Socioeconomic features only	-0.33	-1.07	-0.36	0.19
Image features only	-2.51	-1.38	-0.49	-1.16
Multi-modal features	-1.85	-0.44	-0.16	***0.27***

Models trained and tested on Bellevue generally have sub-par performance due to the small number of datapoints in the dataset (28). However, we can see that our Gradient Boosting Regression model trained on multi-modal features outperforms all other models and feature types.

San Antonio:

Table 5: Performance of regression models (Intra-city: Train and test on San Antonio)

	Linear Regression	Lasso Regression	Support Vector Regression	Gradient Boosting Regression
Socioeconomic features only	0.65	0.31	0.67	0.73
Image features only	0.24	0.30	0.22	0.27
Multi-modal features	0.46	0.47	0.67	***0.74***

According to table 5, the Gradient Boosting Regression Model trained on Multi-modal features outperforms all other models and feature types.

Memphis:

Table 6: Performance of regression models (Intra-city: Train and test on Memphis)

	Linear Regression	Lasso Regression	Support Vector Regression	Gradient Boosting Regression
Socioeconomic features only	0.62	0.53	0.73	**0.78**
Image features only	0.03	0.22	-0.03	0.17
Multi-modal features	0.19	0.62	0.73	***0.76***

According to table 6, the Gradient Boosting Regression Model trained on Multi-modal features outperforms all other models and feature types. It also fares similar to the Gradient Boosting Regression Model trained on socioeconomic features only, with a 0.02 R2 score difference.

7.2 Cross-city: Train on one city and test on another city

In this experiment, we trained our regression models on all the extracted image features for each census tract location in one city and tested the model on extracted image features belonging to another city.

Cross-city tests are mainly used to see how well a model generalizes across cities with different geographical and socioeconomic features.

Table 7: Cross-city performance of Gradient Boosting Regression model with Multi-modal features (R2 score)

	Los Angeles	San Antonio	Bellevue	Memphis
Los Angeles		0.78	-1.92	0.38
San Antonio	0.77		-1.01	0.33
Bellevue	-1.05	-1.09		-3.33
Memphis	0.38	-0.79	0.38	

Per the cross-city experiments, we resorted to the Gradient Boosting Regression model with multi-modal features due to its consistency in achieving optimal results.

According to table 7, we can see that cities with greater number of datapoints outperform cities with a smaller dataset in the cross-city tests. Bellevue, a city with only 28 census tracts, has the worst performance among all cities. It's subpar performance can not only be attributed to its small dataset, but also the fact that Bellevue is the only city in our dataset with lower prevalence of mental health issues. The same could also be said for Memphis, one of the cities with higher prevalence of mental health issues.

Since Los Angeles and San Antonio have similar prevalences and geography, the model is able to generalize better across these two cities. As a result, the cross-city tests between Los Angeles and San Antonio have the best results among the other cities in the test.

Table 8: Multi-city performance of Gradient Boosting Regression model with Multi-modal features (R2 score)

	Los Angeles	San Antonio	Bellevue	Memphis
Train on all excluding Los Angeles	0.78			
Train on all excluding San Antonio		0.68		
Train on all excluding Bellevue			-1.85	
Train on all excluding Memphis				0.35

According to table 8, we noticed that, again, the model generalizes well for cities with similar features (geographics and socioeconomic data). Among all the results, the model has subpar performance predicting mental health prevalence for Bellevue since it has not been fed data that represents cities with lower prevalence of mental health issues. The same could be said for Memphis, since the city has a higher prevalence of mental health issues than all the other cities.

To summarize the above findings, we discovered that built environmental features derived from satellite images do play a role in the prevalence of mental health issues in the city. While the benefits of the satellite image features were not as significant as we would've hoped, our multi-modal feature set did help the model attain better results than socioeconomic features alone. It is important to note, however, that the models we've developed are only able to generalize to cities with similar geography and prevalence of mental health issues. Nevertheless, the results are promising and indicates the need for future research to also factor in built environmental features in their prediction models.

8 Conclusions

In this paper, we've adopted a novel approach to predict the prevalence of mental health issues in a city using the combination of built environmental features and socioeconomic data. Although results have shown a subpar ability to generalize across cities with varying geographies and prevalence levels, our results are promising for predicting the prevalence of mental health issues in similar cities. To conclude, we believe that, with more work done on model optimization and data selection, our models provides a consistent method to measure the prevalence of mental health within a region.

References

[1] U.S. Department of Health and Human Services. *Mental Health: A Report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health, 1999.

[2] Helbich M. (2018). Mental Health and Environmental Exposures: An Editorial. *International journal of environmental research and public health*, *15*(10), 2207. doi:10.3390/ijerph15102207

[3] Mair C, Roux AVD, Galea SAre neighbourhood characteristics associated with depressive symptoms? A review of evidence*Journal of Epidemiology & Community Health* 2008;62:940-946. 3-1

[4] Galea S, Ahern J, Rudenstine S*, et al*Urban built environment and depression: a multilevel analysis*Journal of Epidemiology & Community Health* 2005;59:822-827.

[5] Hao B., Li L., Li A., Zhu T. (2013) Predicting Mental Health Status on Social Media. In: Rau P.L.P. (eds) Cross-Cultural Design. Cultural Differences in Everyday Life. CCD 2013. Lecture Notes in Computer Science, vol 8024. Springer, Berlin, Heidelberg

[6] Gu, Jiuxiang & Wang, Zhenhua & Kuen, Jason & Ma, Lianyang & Shahroudy, Amir & Shuai, Bing & Liu, Ting & Wang, Xingxing & Wang, Gang. (2015). Recent Advances in Convolutional Neural Networks. Pattern Recognition. 10.1016/j.patcog.2017.10.013.

[7] VGG16 - Convolutional Network for Classification and Detection. (2018, November 21). Retrieved from https://neurohive.io/en/popular-networks/vgg16/.

[8] Transfusion: Understanding Transfer Learning for Medical Imaging

[9] VGG16 - Convolutional Network for Classification and Detection. (2018, November 21). Retrieved from https://neurohive.io/en/popular-networks/vgg16/.

[10] Maharana A, Nsoesie EO. Use of Deep Learning to Examine the Association of the Built Environment With Prevalence of Neighborhood Adult Obesity. JAMA Netw Open. 2018;1(4):e181535. doi:https://doi.org/10.1001/jamanetworkopen.2018.1535

[11] Rizwan, M. (2018, October 5). VGG16 - Implementation Using Keras. Retrieved from https://engmrk.com/vgg16-implementation-using-keras/.