Research Article



Open Access

Development and Validation of Predictive Models for Depression Using PHQ-9 Data

Jonathan C. Huang

22 Whitewoods Lane, Malvern, PA, United States jonathanhuang19@gmail.com

Abstract: Depression, the leading cause of suicide worldwide, is a serious, widespread, and growing mental health disorder that has now been labeled a global health epidemic. The Patient Health Questionnaire-9 (PHQ-9), a depression-screener questionnaire, has emerged as an effective diagnostic tool globally. Using U.S. PHQ-9 patient response data and corresponding demographic data from 2013-2014 and 2015-2016, this study conducts a comprehensive big data analysis of the response data to develop and validate predictive models for depression probability. Age at screening, gender, race/ethnicity, education level, and body weight were proposed as factors correlated with depression. Two models were constructed using RStudio to explore these correlations: a logistic regression model, and an artificial neural network. The logistic regression predictive model performed better than the artificial neural network in an unfamiliar dataset, whereas the opposite was true in a familiar dataset. Both models supported that the proposed factors are indeed significantly correlated with depression. The logistic regression model indicated that females and those with weight problems are more likely to have depression, and that the likelihood of depression increases with age, decreases with higher education levels, and varies by race. The artificial neural network indicated that age, the Asian race, some college education, and weight problems are the most significant factors affecting depression probability, in that order. Based on these results, populations most at-risk for depression are identified and appropriate measures should be taken to combat depression.

Keywords: depression, logistic regression, artificial neural network, PHQ-9, big data, correlogram, ROC

INTRODUCTION

A sudden suicide oftentimes deeply shocks loved ones, local communities, and the larger global community. However, the crucial underpinning of the suicide is frequently overlooked: about ninety percent of people who commit suicide suffer from a mental illness at the time of their death, the most prominent of which is depression (Kahn, 2016). Depression, also known as major depressive disorder, is a serious mental health disorder that negatively affects how an individual feels, thinks, and acts, leading to a variety of emotional and physical problems that range from mild to severe (NIH, 2018). Currently, depression statistics indicate that it is an extremely serious and growing health issue, to the extent that it has been labeled an "epidemic" (Rottenberg, 2014). In any given year, depression affects an estimated one in fifteen adults, and an estimated one in six people will experience depression at some time in their life (APA, 2017). Depression, in addition to being common, is also a growing trend. The World Health Organization (WHO, 2017) states, "Depression is the leading cause of ill health and disability worldwide. More than 300 million people are now living with depression, an increase of more than 18% between 2005 and 2015." In light of depression's widespread, serious, and growing state, societies around the world have prioritized the identification of the risk factors of depression, and the possible treatment and prevention of the disorder.

Several risk factors for depression currently identified include but are not limited to differences and variance in brain biochemistry, genetics, personality, and environmental factors (NIH, 2018). Many previous studies have been conducted to investigate the correlation between depression and various factors. Mirowsky et al.

(1992) analyzed the relationship between depression and age and concluded that depression was shown to reach its lowest level in the middle ages and its highest level in adults aged 80 or older. Gender and depression is another pertinent relationship: a study conducted by the American Psychological Association found that after age fifteen, females are approximately twice as likely to be depressed than males. (Nolen-Hoeksema and Girgus, 1994). Another study conducted by Munce et al. (2018) affirmed this age/depression relationship, stating that in 131, 535 adults, the prevalence of depression in women was almost twice that of men. Roxburgh (2009) also investigated the gender and depression relationship and found that black and white men are "significantly" less depressed than black and white women. A distinct dynamic also exists between race/ethnicity and depression. A Canadian study found that English Canadians had better mental health than Jewish Canadians, worse mental health than East Asians, South Asians, Chinese Canadians, and black Canadians, and similar mental health as all other racial/ethnic groupings (Wu et al., 2003). Variance in likelihood of depression was also investigated by Riolo et al. (2005), and found that prevalence of depression differed significantly by racial/ethnic group, with the highest prevalence in White participants. They also concluded that Mexican American and White individuals had significantly earlier onset of depression compared to African American individuals. Socioeconomic standing and a factor closely tied to it, education level, have also been clearly linked to depression. In the same study, Riolo et al. (2005) also found that people living in poverty were 1.5 times more likely to have depression, and after controlling for poverty, the same study found that lack of education remained a significant risk factor. Moreover, a study conducted by de Wit et al. (2009) explored the relationship between depression and the physical health measure body mass index. They found a very significant U-shaped association between BMI categories (underweight, normal, overweight and obesity) and depression, that is, depression is much more frequent in people with abnormally low or high BMIs.

With regards to treatment, the majority of patients respond well to either medication or therapy, but the crux of the issue is that one in every two people with depression go undiagnosed and thus untreated. Untreated depression is found to get progressively worse over time (WebMD, 2017). Therefore, it is imperative to diagnose depression at the earliest stage possible. While many diagnostic methods have been developed, one particularly effective method is depression questionnaires. These are designed to screen for depression by asking the patient to answer important questions, and based on their results, the level of severity of depression can be diagnosed. Early depression questionnaires administered in 1995 and 1997 ranged from 2-28 questions (Whooley, 1997), and each case-finding instrument was "available to help primary care clinicians identify patients with major depression" (Mulrow et al., 1995).

One specific questionnaire with nine questions, the Patient Health Questionnaire-9 (PHQ-9), was developed by doctors from Columbia University in 1999 (Joo et al., 1999). After a validity test was conducted for the questionnaire, it was concluded that "In addition to making criteria-based diagnoses of depressive disorders, the PHQ-9 is also a reliable and valid measure of depression severity. These characteristics plus its brevity make the PHQ-9 a useful clinical and research tool (Kroenke, et al., 2001). The PHQ-9 was also endorsed by the National Institute for Health and Clinical Excellence "in measuring baseline depression severity and responsiveness to treatment" (Smarr and Keefer, 2011). The emergence of PHQ-9 as a very reliable and effective depression screener can be attributed to its numerous advantages, such as ability to facilitate diagnosis of depression, ability to assess symptom severity, ease of administration, and ability to effectively be administered to a wide variety of populations (Univ. of Washington, 2018). As previously mentioned, the PHQ-9 is not only able to screen for depression, but also depression severity. A major upside of the questionnaire is its ease of administration, including the fact that it is free, and its brevity: being only nine questions, it can be completely administered in minutes and is comparatively shorter than other depression measures, such as the Beck Depression Inventory-II (Kung et al., 2012). The PHQ-9 can also be reliability administered in ways other than in person by a clinician; it can be self-administered (Univ. of Washington, 2018), or even administered over the telephone (Pinto-Meza et al., 2005). Furthermore, the PHQ-9 has proven to be very effective in a range of populations varying by both age and ethnicity. It has been administered to adolescents as well as adults, and is used on a global scale since it is available for free and in thirty languages other than English (APA, 2018). The PHQ-9 has become the

most common depression measure in the United Kingdom's National Health Service (Kroenke, et al., 2010) and has been an extremely valuable tool in screening and treating depression in many other regions of the world including Nigeria (Adewuya et al., 2006), South Korea (Han et al., 2007), and Honduras (Wulsin et al., 2002).

Currently, little study has been conducted analyzing the data from patient responses to the PHQ-9 in the United States administered by the Centers for Disease Control and Prevention (CDC)'s National Health and Nutrition Examination Survey (NHANES). In the wake of the current depression epidemic in the United States, and utilizing the relevance of the PHQ-9 in relation to depression, this study aims to be the first to conduct an extensive big data analysis of the PHQ-9 responses. Analyzing the latest data of the PHQ-9, the objective of this study is to broaden societal understanding of depression as a whole by investigating potential new trends and identifying the factors most correlated with depression, and based on these results, to inform the government and general public for the purpose of improving societal health.

After literary review and preliminary background research, the hypothesis of this study is "The probability that a patient is classified by the PHQ-9 with clinically relevant depression is heavily correlated with the following factors: age at screening, gender, race/ethnicity, education level, and weight."

DATA AND METHODS

Data

The two latest datasets (2013-2014 and 2015-2016) of the PHQ-9 from the CDC's NHANES were combined and used in this study. NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the United States. Specifically, the overall dataset was obtained by first merging the the PHQ-9 dataset and the demographics dataset for each of the two surveys, then combining both surveys. All data is located at https://www.cdc.gov/nchs/nhanes/about_nhanes.htm. The PHQ-9 is shown in Table 1.

PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)						
Over the <u>last 2 weeks</u> , how often have you been bothered by any of the following problems? (Use " "" to indicate your answer)	Not at all	Several days	More than half the days	Nearly every day		
1. Little interest or pleasure in doing things	0	1	2	3		
2. Feeling down, depressed, or hopeless	0	1	2	3		
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3		
4. Feeling tired or having little energy	0	1	2	3		
5. Poor appetite or overeating	0	1	2	3		
 Feeling bad about yourself — or that you are a failure or have let yourself or your family down 	0	1	2	3		
 Trouble concentrating on things, such as reading the newspaper or watching television 	0	1	2	3		
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3		
 Thoughts that you would be better off dead or of hurting yourself in some way 	ο	1	2	3		

Table1. The Patient Health Questionnaire-9

Methods

Using RStudio version 1.1.456, the PHQ-9 patient response dataset was analyzed in this study by employing three different tools, namely, a correlogram, a logistic regression model, and an artificial neural network.

American Research Journal of Psychiatry

Correlogram

A correlogram is a graphical representation of the cells of a matrix of correlations. It displays the pattern of correlations in terms of their signs and magnitudes using visual thinning and correlation-based variable ordering. The cells of the matrix are shaded or colored to show the correlation value. The positive correlations are shown in blue, while the negative correlations are shown in red. Furthermore, the darker the hue, the greater the magnitude of the correlation. The correlogram of this study is a visual depiction of the relationship between each of the variables in this study. The most important correlations are between the independent variables and the dependent variable.

Logistic Regression Model

Logistic regression is one of a category of statistical models called generalized linear models. It provides a way to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a combination of these. Typically, the dependent variable is dichotomous, whereas the independent variables can be either categorical or continuous.

The logistic regression model can be expressed with the following formula:

$$\ln(P/P-1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

where P is the probability of being depressed, β_0 is a constant (i.e., the intercept or "reference" or base from which the probabilities are calculated), and β_1 through β_n are the regression coefficients and X_1 through X_n are the independent variables listed in table 1. For simplicity, the item on the left side of the equation is often referred to as "the logit". The interpretation of the coefficients describes the impact of the independent variables on the natural logarithm of the odds, rather than directly on the probability P.

Artificial Neural Network

An artificial neural network (ANN) consists of an interconnected group of artificial neurons based on biological neural networks. Processing information using a connectionist approach to computation, ANNs are used to provide a visual interpretation of the connection weights among neurons. In more practical terms, ANNs are non-linear statistical data modeling tools. They can be used to model complex functional relationships between inputs and outputs or to find patterns in data, therefore are a useful tool to harvest information from datasets in the process known as data mining. A package called "neuralnet" in R was used to conduct neural network analysis. The package neuralnet focuses on multi-layer perceptrons using neurons and synapses, which are well applicable when modeling functional relationships. The input neuron layer consists of all covariates in separate neurons and the output layer consists of the response variables. The layers in between are referred to as hidden layers, as they are not directly observable. Input layer and hidden layers include a constant neuron relating to intercept synapses. ANNs are adaptive and fitted to the data by learning algorithms during a training process. This study will specifically visualize the ANN through a neural interpretation diagram, and will also utilize Garson's algorithm to produce a bar plot depicting the relative importance of each of the independent variables in relation to the dependent variable (Olden and Jackson, 2002).

Model Evaluation

Discrimination was used to evaluate the quality of the logistic regression and artificial neural network models (Dreiseitla and Ohno-Machadob, 2002). To provide an impartial estimate of the discrimination of a model, these values have to be calculated from a dataset which is not used in the model building process. Normally, a portion of the original dataset, called the test or validation set, is set aside for this purpose. For small datasets, however, there may not be enough data items for both training and testing. In this case, a dataset is divided into n pieces, n-1 pieces are used for training, and the last piece becomes the test set. For this study, the overall dataset was large enough to be split evenly into a training sample and a testing sample, and the process of cross validation was conducted using these two samples.

The discriminatory ability refers to the capacity of the model to separate cases from non-cases, with 1.0 and 0.5 meaning perfect and random discrimination, respectively. In this study, logistic regression and ANN discriminatory ability was determined using receiver operating characteristic (ROC) curve analysis. An ROC curve is defined by false positive rate (FPR) and true positive rate (TPR) on x and y axes respectively, which depicts relative trade-offs between true positive and false positive. Since TPR is equivalent to sensitivity and FPR is equal to 1 – specificity, the ROC graph is often called the sensitivity vs (1 – specificity) plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space. The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). As a result, point (0,1) is also called a perfect classification. A random guess would give a point along a diagonal line (i.e., line of no-discrimination) from the lower left corner to the upper right corner. ROC curves are commonly used to summarize the diagnostic accuracy of risk models, and more importantly, to evaluate the improvements made to such models that are gained from adding more risk factors. For each new risk factor added, sensitivity, specificity and accuracy are recalculated and compared.

Variables

The outcome or dependent variable is clinically relevant depression (CRD) positive or negative based on the PHQ-9, which was used to measure signs and symptoms of depression. PHQ-9 total scores>=10 are defined as CRD. CRD was coded in the data using variable name called CRD (1, yes; 0, no). In order to analyze the impact of age at screening, gender, race, education level and body weight on depression, the PHQ-9 dataset was merged with demographic data, which contains a wide range of information about the patients who completed the PHQ-9. The information related to age, gender, race, education level and body weight were identified as the independent variables in this study. The independent variables are listed in Table 2, together with the variables used as the reference group, namely, race6 (i.e., other races), edu5 (i.e., college graduate or above), normal weight, age twenty, and male, which is an implicit reference variable for gender.

Variable Name	Column Name in Data File	Description
gender	riagendr	Gender: 1 male; 2 female
age	ridageyr	Age in years at screening
race1	Ridreth3 (=1)	Mexican-American
race2	Ridreth3 (=2)	Other Hispanic
race3	Ridreth3 (=3)	Non-Hispanic White
race4	Ridreth3 (=4)	Non-Hispanic Black
race5	Ridreth3 (=6)	Non-Hispanic Asian
race6	Ridreth3 (=7)	Other Race (including Multi-Racial)
edu1	Dmdeduc2 (=1)	Less than 9th grade
edu2	Dmdeduc2 (=2)	9-11th grade (including 12th grade with no diploma)
edu3	Dmdeduc2 (=3)	High school graduate/GED or equivalent
edu4	Dmdeduc2 (=4)	Some college or AA degree
edu5	Dmdeduc2 (=5)	College graduate or above
overweight	Whq030 (=1)	Overweight
underweight	Whq030 (=2)	Underweight
normal weight	Whq030 (=3)	Normal weight

Table2. Independent variables

RESULTS

In total, the data for 11659 patients were obtained from the merged dataset, from which only 9942 patients' information is complete and used as the overall sample in this study. The age range for the 9942 patients was from 20 to 80 years old, with average age of 49.5 years old. In order to perform model validation, the 9942 overall sample dataset is randomly divided into to two datasets, namely, training sample and testing sample. After

experimentation, the optimal division was found to be an equal split. So, both the training sample and testing sample have 4971 patients. The characteristics of the overall, training, and testing samples are summarized in in Table 3. It is worthy to note that the partition of the overall sample into training and testing samples was sufficiently random as both have almost the same cases of CRD with just a 4 case difference.

Variables		Training		Testing		Overall	
		Sample		Sample		Sample	
CRD	With CRD	442	8.9%	446	9.0%	888	8.9%
	Without CRD	4529	91.1%	4525	91.0%	9054	91.1%
Gender	Male	2399	48.3%	2422	48.7%	4821	48.5%
	Female	2572	51.7%	2549	51.3%	5121	51.5%
Race/ Ethnicity	Mexican American	814	16.4%	728	14.6%	1542	15.5%
	Other Hispanic	519	10.4%	564	11.3%	1083	10.9%
	Non-Hispanic White	1956	39.3%	1937	39.0%	3893	39.2%
	Non-Hispanic Black	1009	20.3%	1038	20.9%	2047	20.6%
	Non-Hispanic Asian	505	10.2%	533	10.7%	1038	10.4%
	Other Race (Including Multi-Racial)	168	3.4%	171	3.4%	339	3.4%
Education Level	Less than 9th grade	472	9.5%	430	8.7%	902	9.1%
	9-11 grade (Includes 12th grade with no diploma)	651	13.1%	613	12.3%	1264	12.7%
	High school graduate/GED or equivalent	1095	22.0%	1115	22.4%	2210	22.2%
	Some College or AA Degree	1518	30.5%	1541	31.0%	3059	30.8%
	College graduate or above	1235	24.8%	1272	26.0%	2507	25.2%
Weight	Overweight	2604	52.4%	2644	53.2%	5248	52.8%
	Underweight	269	5.4%	287	5.8%	556	5.6%
	Normal weight	2098	42.2%	2040	41.0%	4138	41.6%

Table3. Patient characteristics in training sample, testing sample, and overall sample

Correlogram

Using the entire dataset, the correlations between all variables are analyzed with a correlogram, which depicts a matrix of the correlations, and the result is shown in Figure 1.



Fig1. Correlogram depicting matrix of correlations between variables

American Research Journal of Psychiatry

Logistic Regression Model

The logistic regression model was developed with the method outlined in section 2.2. using the training sample (N=4971). The reference group (i.e. β_0 as discussed previously) was established as male, twenty years of age at screening, race6 (other races including multi-racial), edu5 (college graduate or above), and normal weight. Age is a continuous variable. Malee was coded as gender=0, whereas female was coded as gender=1. The other variables listed in Table 2 were coded as No=0, and Yes=1. The result of the logistic regression model is shown in Table 4. In applying the result to calculate the probability p based on the logistic regression model formula, the intercept estimate is β_0 , whereas estimates for the other variables correspond to β_1 , β_2 , etc. The z value is defined as Estimate/Std. Error. The probability Pr is based on the z-test. The odds ratio, which is defined as Probability / (1 - Probability), was calculated with the logit for each variable.

	Estimate	Std. Error	z value	Pr(> z)	Odds Ratio
(Intercept)	-3.257	0.2867	-11.358	< 0.0001	0.039
gender	0.514	0.107	4.797	< 0.0001	1.672
age	0.186	0.179	1.037	0.300	1.204
race1	-1.130	0.268	-4.218	< 0.0001	0.323
race2	-0.671	0.273	-2.455	< 0.05	0.511
race3	-0.443	0.241	-1.840	0.066	0.642
race4	-0.688	0.255	-2.702	< 0.01	0.503
race5	-1.433	0.340	-4.213	< 0.0001	0.239
edu1	1.639	0.207	7.972	< 0.0001	5.150
edu2	1.390	0.185	7.504	< 0.0001	4.015
edu3	0.721	0.179	4.021	< 0.0001	2.056
edu4	0.505	0.172	2.934	< 0.01	1.657
overweight	0.638	0.118	5.408	< 0.0001	1.893
underweight	1.285	0.192	6.686	< 0.0001	3.615

Table4. Logistic regression result with training sample

Artificial Neural Network

The ANN model for the training sample included all variables listed in Table 2: patient age at screening, gender, race, education level, and body weight. The neural interpretation diagram is plotted in Figure 2 and the bar plot produced by Garson's algorithm is plotted in Figure 3.



American Research Journal of Psychiatry

In Figure 2, the thickness of the lines (axons) joining neurons is proportional to the magnitude of the connection weight, and the shade of the line indicates the direction of the interaction between neurons: black connections are positive, whereas gray connections are negative. After experimentation, the optimal number of hidden layers was found to be five for the data in this study. Due to the fact that the model includes hidden layers, a direct or clear relationship between each independent variable and the outcome (i.e., dependent) variable cannot be discerned. However, the ANN result is sufficient for the training algorithm to converge. Therefore, the model can be used for predictive purposes.

In order to identify the importance of the independent input variables, Garson's algorithm is employed to partition the neural network connection weights (Garson, 1991). The result of applying the algorithm to the training sample is shown in Figure 3. It is important to note that Garson's algorithm uses the absolute values of the connection weights when calculating variable contributions, and therefore does not provide the direction of the relationship between the input and output variables.



Fig3. Bar plot showing the relative importance of each independent variable for predicting CRD based on Garson's algorithm

Model Validation: ROC Curve Sketching

After logistic regression and artificial neural network analysis were conducted in the training sample, the outputs from both models were used to predict the probabilities of CRD in the testing sample (N=4971). The performance of the prediction can be visualised by plotting ROC curves, thus allowing the performance of the built models to be validated. Figure 4 shows the ROC curves for both the logistic regression model and ANN model in the testing sample. Figure 5 shows the ROC curves for both models in the training sample.







Fig5. ROC curves of logistic regression versus artificial neural network in training sample

DISCUSSION

Correlogram Analysis and Interpretation of Results

The visual nature of the correlogram allows it to serve as a quick glance overview of the relationships between all variables in this study. The most useful information that the correlogram provides are the correlations between the dependent variable CRD and the independent variables listed in Table 2. As shown in Figure 1, CRD has a slight positive correlation with all the independent variables except race1 (Mexican American), race5 (Asian) and edu 4 (some college or AA degree), which CRD instead has slight negative correlations to.

Verification of the Binomial Assumption of Logistic Regression

The use of logistic regression is based on the assumption that the binomial distribution is the assumed distribution for the conditional mean of the dichotomous outcome. This assumption implies that the same probability is maintained across the range of predictor values. Though this assumption was not verified or tested, the binomial assumption is known to be robust as long as the data sample is random. According to the CDC, "Each year approximately 7,000 randomly-selected residents across the United States have the opportunity to participate in the latest NHANES." (CDC, 2011)Since the NHANES is a random data sample, this assumption is satisfied.

Logistic Regression Analysis and Interpretation of Results

Of the 4971 patients in the training sample, 442 (8.9%) were classified as having CRD and 4529 (91.1%) were not. According to the logistic regression model result listed in Table 4, at significance level of 0.05, the predictive model of clinically relevant depression is as follows:

 $\label{eq:predicted logit of CRD = -3.257 + 0.514*Gender + 0.186*Age - 1.13*Mexican-American - 0.671*Other Hispanic - 0.443*White - 0.688*Black - 1.433*Asian + 1.639*Less than 9th grade + 1.39*Grades9-11 + 0.721*Highschool Graduate + 0.505*Some College + 0.638*Overweight + 1.285*Underweight$

The coefficients of the parameters were interpreted as follows. At the significance level of 0.05:

- On average, controlling other variables, a female is 67.2% more likely to be classified as having CRD than a male.
- On average, controlling other variables, a person eighty years of age at screening is 20.4% more likely to be classified as having CRD than a person twenty years of age at screening.
- On average, controlling other variables, a person who is Mexican-American is 67.7% less likely to be classified as having CRD than a person who is Other Race (including Multi-Racial).

- On average, controlling other variables, a person who is Other Hispanic is 48.9% less likely to be classified as having CRD than a person who is Other Race (including Multi-Racial).
- On average, controlling other variables, a person who is White is 35.8% less likely to be classified as having CRD than a person who is Other Race (including Multi-Racial).
- On average, controlling other variables, a person who is Black is 49.7% less likely to be classified as having CRD than a person who is Other Race (including Multi-Racial).
- On average, controlling other variables, a person who is Asian is 76.1% less likely to be classified as having CRD than a person who is Other Race (including Multi-Racial).
- On average, controlling other variables, a person with an education level of less than high school is 415% more likely to be classified as having CRD than a person with an education level of college graduate or above.
- On average, controlling other variables, a person with an education level of 9-11th grade (including 12th grade with no diploma) is 301.5% more likely to be classified as having CRD than a person with an education level of college graduate or above.
- On average, controlling other variables, a person with an education level of High school graduate/GED or equivalent is 105.6% more likely to be classified as having CRD than a person with an education level of college graduate or above.
- On average, controlling other variables, a person with an education level of some college or AA degree is 65.7% more likely to be classified as having CRD than a person with an education level of college graduate or above.
- On average, controlling other variables, a person who identifies as overweight is 89.3% more likely to be classified as having CRD than a person who identifies as having normal weight.
- On average, controlling other variables, a person who identifies as underweight is 261.5% more likely to be classified as having CRD than a person who identifies as having normal weight.

It is indicated in the model that being female, being older than twenty years, being Mexican-American, being Other Hispanic, being White, being Black, being Asian, having education levels of less than high school, 9-11th grade (including 12th grade with no diploma), High school graduate/GED or equivalent, some college or AA degree, and identifying as overweight or underweight are significant factors in predicting CRD. In comparison to the established reference group, being Mexican-American, Other Hispanic, White, Black, or Asian may decrease the probability of being classified as having CRD, while being female, older than twenty years, having an education below college graduate or above, and being overweight or underweight may increase the probability of being classified as having CRD.

The logistic regression model reveals broader trends for each demographic category. Firstly, the model indicates that as people age, they are more likely to be classified as having CRD, although not significantly. This trend agrees with the investigation result by Mirowsky and Ross (1992) Secondly, the model indicates that females are significantly more likely to be classified as having CRD, which confirms the result from a study conducted by American Psychological Association (APA, 1994 study). Thirdly, the model indicates that Other Race (including Multi-Racial) is the race/ethnicity most likely to be classified as having CRD, whereas Asian is the least likely race/ethnicity, followed by Mexican American, Black, Other Hispanic and White. This result differs from studies

cited in the introduction (Wu et al., 2003; Riolo et al., 2005). The model also indicates that the lower a person's education level, the more likely they are to be classified as having CRD, supporting a previous study which found the correlation of education and depression (Riolo et al., 2005). Lastly, the model indicates that a person who does not identify as normal weight (ie overweight or underweight) is more likely to be classified as having CRD, supporting the conclusion of a previous depression versus BMI study (de Wit et al., 2009).

Artificial Neural Network Analysis and Interpretation of Results

It is difficult to interpret quantitative results from the neural interpretation diagram. The bar plot produced by Garson's algorithm yields quantitative results shown in Figure 3. The plot indicates that age is the independent variable most significantly correlated with CRD (about 18.7%), followed by race5 (i.e. Asian, about 10.2%), edu4 (i.e. college graduate or above, 9.8%), underweight (9.0%), and overweight (7.1%). The variable importance/ significance of all other variables to CRD is less than 7%. The correlation direction of each of these variables to CRD is not depicted by the plot alone. However, more sense can be made by looking at the neural network results in conjunction with the logistic regression model results. It is worthy to note that different models may produce different results, which is not abnormal since they are based on different theories and starting assumptions.

Model Evaluation: ROC Curve Analyses and Interpretation

In Figure 4, the areas under the ROC curves are 0.66 and 0.63 for the logistic regression model and the artificial neural network model, respectively. This result indicates that in the testing sample, the logistic regression model performs better than the ANN model in predicting CRD. The validation of the models was also conducted in training sample. The areas under the ROC curves in Figure 5 are 0.70 and 0.75 for logistic regression model and ANN model, respectively. The result indicates that in the training sample, the ANN model performs better than the logistic regression model in predicting CRD. Moreover, Figure 4 and Figure 5 in conjunction indicate that both models are better at predicting data samples they are familiar with, and that the models' performance will decrease with brand new data samples.

CONCLUSION

In this study, 9942 patient responses to the Patient Health Questionnaire Depression Screener (PHQ-9) from 2013-2014 and 2015-2016 were analyzed in relation to corresponding demographic information including age at screening, gender, race, education level, and body weight. These factors were used to develop and validate a logistic regression model and an artificial neural network model for predicting the probability of depression. Both models were validated using Receiver Operating Characteristic (ROC) analysis. The areas under the ROC curves were 0.66 for the logistic regression model and 0.63 for the artificial neural network model when new testing data was used, thus favoring the logistic regression model. However, when familiar data was used, the areas under the ROC curves were 0.70 for the logistic regression model and 0.75 for the artificial neural network model, thus favoring the artificial neural network model. Both models are stable and therefore valid for predictive purposes.

The results of the predictive models are very significant. Since specific populations were found to be more at risk for depression such as females, those undereducated, those older, and those with weight problems, society should take measures to increase depression awareness among these populations especially, and should place an increased emphasis on physical health and education in an effort to combat depression.

Further study is recommended to increase the reliability of each model. One action step in the future is to identify additional significant demographic variables and include them in the models to improve model performance.

Additional PHQ-9 data from previous or future years, or from outside the U.S. can also be added into the study to train the models and improve their performance in predicting depression.

REFERENCES

- Adewuya, A. O., Ola, B. A. and Afolabi, O. O. "Validity of the Patient Health Questionnaire (PHQ-9) as a Screening Tool for Depression amongst Nigerian University Students." ScienceDirect, Elsevier - Journal of Affective Disorders Volume 96, Issues 1–2: 89-93, 20 July 2006, www.sciencedirect.com/science/article/ pii/S0165032706002539.
- 2. American Psychiatric Association. What Is Depression? American Psychiatric Association, Jan. 2017, www. psychiatry.org/patients-families/depression/what-is-depression.
- 3. American Psychological Association. "Patient Health Questionnaire (PHQ-9 & PHQ-2)." American Psychological Association, 2018. www.apa.org/pi/about/publications/ caregivers/practice-settings/assessment/tools/patient-health.aspx
- 4. Centers for Disease Control and Prevention. "History." National Health and Nutrition Examination Survey, Centers for Disease Control and Prevention, 14 Nov. 2011, www.cdc.gov/nchs/nhanes/history.htm.
- de Wit, L. M., van Straten, A., van Herten, M, Penninx, B. W.J.H. and Pim Cuijpers, P. "Depression and Body Mass Index, a u-Shaped Association." BMC Public Health, BioMed Central, 13 Jan. 2009, bmcpublichealth. biomedcentral.com/articles/10.1186/1471-2458-9-14.
- 6. Dreiseitla, S. and Ohno-Machadob, L. "Logistic regression and artificial neural network classification models: a methodology review." Journal of Biomedical Informatics 35 (2002) 352–359
- 7. Garson, G.D., 1991. Interpreting neural-network connection weights. Artif. Intell. Expert 6, 47–51.
- Han, C., Jo, S. A., Kwak, J. H., Pae, C. U., Steffens, D., Jo, I. and Park, M. H. "Validation of the Patient Health Questionnaire-9 Korean Version in the Elderly Population: the Ansan Geriatric Study." ScienceDirect, Elsevier - Comprehensive Psychiatry Volume 49, Issue 2: 218-223, 25 Oct. 2007, www.sciencedirect.com/ science/article/pii/S0010440X07001162.
- Joo et al. "Diagnosing Mood Disorders Part 1." Primary Care Depression, Center for Family Medicine New York Presbyterian / Columbia University College of Physicians and Surgeons Director of Predoctoral Education, www.columbia.edu/itc/hs/medical/clerkships/primcare/case/depression/depression01_05. html.
- 10. Kahn, A. "Suicide and Suicidal Behavior: Causes, Symptoms, and Diagnosis." Healthline, Healthline Media, 27 Jan. 2016, www.healthline.com/health/suicide-and-suicidal-behavior.
- 11. Kroenke, K., Spitzer, R. L., Williams, J. B. and Löwe, B. "The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a Systematic Review." ScienceDirect, Elsevier Inc., 2010, https://www.sciencedirect.com/science/article/pii/S0163834310000563?via%3Dihub.
- 12. Kroenke K., Spitzer R. L. and Williams, J. B. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine*, 16(9): 606-613, 2001.
- 13. Mirowsky, J. and Ross, C. E. "Age and Depression." Journal of Health and Social Behavior, vol. 33, no. 3, 1992, pp. 187–205. JSTOR, JSTOR, www.jstor.org/stable/2137349.

- 14. Mulrow, C. D., Williams, J. W., Gerety, M. B., Ramirez, G., Montiel, O. M. and Kerber, C. "Case-Finding Instruments for Depression in Primary Care Settings." Ann Intern Med. 1995;122:913–921.
- Munce, S. E. P., and Donna E. Stewart. "Gender Differences in Depression and Chronic Pain Conditions in a National Epidemiologic Survey." ScienceDirect, Elsevier - Psychosomatics Volume 48, Issue 5: 394-399, 7 Apr. 2011, www.sciencedirect.com/science/article/pii/S0033318207710022.
- 16. National Institute of Mental Health. "Depression." National Institute of Mental Health, U.S. Department of Health and Human Services, Feb. 2018, www.nimh.nih.gov/health/topics/depression/index.shtml.
- 17. Nolen-Hoeksema, S. and Girgus, J. S. "The Emergence of Gender Differences in Depression during Adolescence." APA PsycNET, American Psychological Association Psychological Bulletin, Vol 115(3) 424-443, May 1994.
- Olden, J. D. and Jackson, D. A. "Illuminating the "Black Box": a Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks." Ecological Modeling, Elsevier - Ecological Modelling 154 (2002) 135–150, 4 Mar. 2002, labs.eeb.utoronto.ca/jackson/ecol.%20modelling%20ANN.pdf.
- 19. Pinto Meza, A., Serrano Blanco, A., Peñarrubia, M. T., Blanco, E. and Haro, J. M. "Assessing Depression in Primary Care with the PHQ 9: Can It Be Carried Out over the Telephone?" Journal of General Internal Medicine, 20: 738-742, 2005. https://link.springer.com/article/10.1111%2Fj.1525-1497.2005.0144.x.
- Riolo, S. A., Nguyen, T. A., Greden, J. F. and King, A. K. "Prevalence of Depression by Race/Ethnicity: Findings From the National Health and Nutrition Examination Survey III." American Journal of Public Health Publications, American Journal of Public Health - Vol 95, No. 6, June 2005, ajph.aphapublications.org/doi/ pdf/10.2105/AJPH.2004.047225.
- 21. Rottenberg, J. "Why We Can't Stop the Depression Epidemic." Psychology Today, Sussex Publishers, 16 Sept. 2014, www.psychologytoday.com/us/blog/charting-the-depths/201409/why-we-cant-stop-the-depression-epidemic.
- 22. Roxburgh, S. "Untangling Inequalities: Gender, Race, and Socioeconomic Differences in Depression." Wiley Online Library, Sociological Forum Volume 24, Issue 2: 357-381, 15 Apr. 2009, onlinelibrary.wiley.com/ doi/abs/10.1111/j.1573-7861.2009.01103.x.
- 23. Kung, S., Alarcon, R. D., Williams, M. D., Poppe, K. A., Jo Moore, M. and Frye, M. A. "Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) Depression Measures in an Integrated Mood Disorders Practice." ScienceDirect, Journal of Affective Disorders, 25 Sept. 2012, www. sciencedirect.com/science/article/pii/S0165032712005861.
- 24. Smarr, K. L. and Keefer, A. L. "Measures of depression and depressive symptoms: Beck Depression Inventory II (BDI II), Center for Epidemiologic Studies Depression Scale (CES D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire 9 (PHQ 9)." Arthritis Care Res, 63: S454-S466, 2011.
- 25. University of Washington, Psychiatry & Behavioral Sciences Division of Population Health. "PHQ-9 Depression Scale." Aims Center Advancing Integrated Mental Health Solutions, University of Washington, 2018, aims.uw.edu/resource-library/phq-9-depression-scale.
- 26. WebMD. "Symptoms of Depression." WebMD, 9 Sept. 2017, www.webmd.com/depression/guide/detecting-depression#1.

- 27. Whooley, M. A., Avins, A. L., Miranda, J. and Browner, W. S. (1997), Case Finding Instruments for Depression. Journal of General Internal Medicine, 12: 439-445, 1997.
- 28. World Health Organization. "Depression: Let's Talk." World Health Organization, World Health Organization, 4 July 2017, www.who.int/mental_health/management/depression/en.
- 29. Wu, Zheng, et al. "Race, Ethnicity, and Depression in Canadian Society." Journal of Health and Social Behavior, vol. 44, no. 3, 2003, pp. 426–441. JSTOR, JSTOR, www.jstor.org/stable/1519788.
- Wulsin, L., Somoza, E. and Heck, J. "The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary Care in Honduras." Primary Care Companion to The Journal of Clinical Psychiatry 4.5, 2002: 191–195.

Citation: Jonathan C. Huang. "Development and Validation of Predictive Models for Depression Using PHQ-9 Data". American Research Journal of Psychiatry, 2018; 1(1): 1-14.

Copyright © 2018 Jonathan C. Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

American Research Journal of Psychiatry