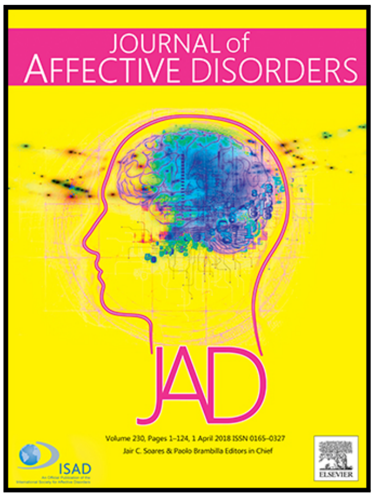
Accepted Manuscript



Identifying Depression in the National Health and Nutrition Examination Survey Data using a Deep Learning Algorithm

Jihoon Oh , Kyongsik Yun , Uri Maoz , Tae-Suk Kim , Jeong-Ho Chae

PII: S0165-0327(19)30441-0
DOI: https://doi.org/10.1016/j.jad.2019.06.034
Reference: JAD 10879

To appear in: *Journal of Affective Disorders*

Received date: 19 February 2019
Revised date: 30 April 2019
Accepted date: 29 June 2019

Please cite this article as: Jihoon Oh , Kyongsik Yun , Uri Maoz , Tae-Suk Kim , Jeong-Ho Chae , Identifying Depression in the National Health and Nutrition Examination Survey Data using a Deep Learning Algorithm, *Journal of Affective Disorders* (2019), doi: https://doi.org/10.1016/j.jad.2019.06.034

Highlights

- Estimating epidemiological contributors to depression and predicting the prevalence of depression are still challenging.
- We aimed to estimate factors affecting depression in National Health and Nutrition Examination Survey (NHANES) datasets using deep learning and machine learning algorithms.
- Deep-learning achieved a high performance for identifying depression on the NHANES datasets of both the United States and South Korea.
- Trained deep-learning and machine learning algorithms are useful for estimating the prevalence of depression.

Identifying Depression in the National Health and Nutrition Examination Survey Data using a Deep Learning Algorithm

Jihoon Oh, MD[1*], Kyongsik Yun, PhD[2,3*], Uri Maoz, PhD[2,4,5,6], Tae-Suk Kim, MD, PhD[1], Jeong-Ho Chae, MD, PhD[1†]

1. Department of Psychiatry, Seoul St. Mary's Hospital, The Catholic University of Korea, College of Medicine, Seoul, Korea
2. Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA
3. Bio-Inspired Technologies and Systems, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
4. Computational Neuroscience, Health and Behavioral Sciences and Brain Institute, Chapman University, Orange, CA, 92866, USA
5. Institute for Interdisciplinary Brain and Behavioral Sciences, Chapman University, Orange, CA, 92866, USA
6. Department of Anesthesiology, School of Medicine, University of California Los Angeles, Los Angeles, CA, 90095, USA

[*] **These authors contributed equally to this study**

[†] **Corresponding author at:** Department of Psychiatry, Seoul St. Mary's Hospital, The Catholic University of Korea, College of Medicine, 222 Banpo-Daero, Seocho-Gu, Seoul, 06591, Korea

Tel.: +82 2 2258 6083 Fax: +82 2 594 3870

E-mail: alberto@catholic.ac.kr

Abstract

Background: As depression is the leading cause of disability worldwide, large-scale surveys have been conducted to establish the occurrence and risk factors of depression. However, accurately estimating epidemiological factors leading up to depression has remained challenging. Deep-learning algorithms can be applied to assess the factors leading up to prevalence and clinical manifestations of depression.

Methods: Customized deep-neural-network and machine-learning classifiers were assessed using survey data from 19,725 participants from the NHANES database (from 1999 through 2014) and 4,949 from the South Korea NHANES (K-NHANES) database in 2014.

Results: A deep-learning algorithm showed area under the receiver operating characteristic curve (AUCs) of 0.91 and 0.89 for detecting depression in NHANES and K-NHANES, respectively. The deep-learning algorithm trained with serial datasets (NHANES, from 1999 to 2012), predicted the prevalence of depression in the following two years of data (NHANES, 2013 and 2014) with an AUC of 0.92. Machine learning classifiers trained with NHANES could further predict depression in K-NHANES. There, logistic regression had the highest performance (AUC, 0.77) followed by deep learning algorithm (AUC, 0.74).

Conclusions: Deep neural-networks managed to identify depression well from other health and demographic factors in both the NHANES and K-NHANES datasets. The deep-learning algorithm was also able to predict depression relatively well on new data set—cross temporally and cross nationally. Further research can delineate the clinical implications of machine learning and deep learning in detecting disease prevalence and progress as well as other risk factors for depression and other mental illnesses.

Keywords: machine learning; depression; National Health and Nutrition Examination Survey; deep learning

Introduction

At least 1 in 23 people in the world suffer from depression, with rates as high as 1 in 13 for some demographics (World Health Organization, 2017). This has made depression the leading cause of disability worldwide and a major contributor to the overall global burden of disease (Smith, 2014). Large-scale national surveys have therefore been conducted to identify prevalence and risk factors for depression. In the United States, there are several nation-wide surveys that measure the occurrence of depression in adolescents (Avenevoli et al., 2015) and in the general population (Center for Behavioral Health Statistics and Quality, 2017). Based on these survey data, a number of previous studies identified associations between demographic (Weissman et al., 1996), social (Vallance et al., 2011) and biological factors (Ford and Erlinger, 2004) and depression.

Conventional machine-learning methods like multivariate logistic regression contributed to the localization of clinical manifestations of depression in these survey data. Recent machine-learning algorithms have been successfully able to correlate depression with co-morbid afflictions and predict its occurrence (Van Loo et al., 2014). For example, logistic regression well predicted treatment-resistance depression in the STAR*D (Sequenced Treatment Alternatives to Relieve Depression) cohort trial (Perlis, 2013). Similarly, a machine-learning-based approach predicted treatment outcome in depression in cross-clinical trials (Chekroud et al., 2016). Machine-learning boosted regression analysis also found some biomarkers related to depression in the NHANES (National Health and Nutrition Examination Survey) dataset (Dipnall et al., 2016) and machine-learning models also predicted the persistence and severity of depression with baseline self-reports (Kessler et al., 2016). However, to the best of our knowledge, there has been no systematic estimation of the factors related to depression within and across large datasets using deep-learning algorithms.

Deep-learning algorithms have recently made important contributions to the detection of certain disease (e.g. diabetic retinopathy) (Gulshan et al., 2016) from medical images. In psychiatry, deep-learning algorithms are introduced to detect depression using EEG signals (Acharya et al., 2018) and from video information such as facial appearance and images (Zhou et al., 2018; Zhu et al., 2018). Multi-modal approaches—using video, audio, and text streams—also successfully recognized patients with depression (Yang et al., 2017). These studies suggest that deep-learning could be a useful technique for the detection of psychiatric illnesses, based on visual and other information.

Note that the term "machine-learning" is often used as a parent concept of the term "deep-learning". However, here we distinguish the two terms. We use the term "machine-learning" to refer to conventional machine-learning algorithms (e.g., logistic regression, support-vector machine) while the term "deep-learning" refers only to deep neural-network models. Deep learning typically outperforms conventional machine-learning approaches for large datasets (LeCun et al., 2015). When dealing with multi-dimensional images, the feature representation in each layer of the neural network is rather clear (i.e., edges in the 1st, object parts in the 2nd and objects in the 3rd layers) (LeCun et al., 2015). However, when basing deep learning on textual and numeric aspects of clinical history and questionnaires, the inner workings of the algorithms tend to be opaque, rendering deep-learning networks a type of "black box" (Barak-Corren et al., 2017). Thus, in psychiatry, where numeric and textual data prevail, conventional machine learning classifiers (like Bayesian models) have generally been favored over deep learning (Barak-Corren et al., 2017).

We therefore focused our investigation on how deep-learning algorithms assess the epidemiological, demographic, life-style and other factors leading up to depression in large survey data sets. We further compared the performance and inner-workings of deep-learning algorithms with several conventional machine-learning algorithms (e.g. support vector machine, and logistic regression) on the same data sets. Our aim was to assess the utility of machine learning and deep learning in deciphering relevant risk factors for depression in two nation-wide survey data sets.

Method

Data sets

National Health and Nutrition Examination Survey datasets of the United States (NHANES) and South Korea (K-NHANES) were used to train the deep-learning algorithms and other machine learning classifiers. NHANES is a nationwide survey to assess the health and nutritional status of the general population in the United States. It consists of demographic, dietary, and other questionnaire data as well as on a medical examination and various laboratory tests. Having begun in the 1960s, it samples approximately 5,000 people a year using a multi-stage stratification design (Center for

Disease Control and Prevention, 2014). All the NHANES data, except the pediatric survey information, are in the public domain and are available on the website of the National Center for Health Statistics (https://www.cdc.gov/nchs/nhanes).

In South Korea, K-NHANES data has been collected since 1998, with a complex, multi-stage stratification sample design for the entire South Korean population. This nation-wide survey is being carried out by the Korean Centers for Disease Control and Prevention, and targets individuals starting the aged of 1. Two stages of stratified clustering—consisting of primary sampling units and households—were applied to the data collected from the Population and Housing Census in Korea (Kim et al., 2014). The extracted samples had their own weights, and were representative of the health and nutritional status of the general population. Similar to NHANES, K-NHANES is composed of demographic variables, health questionnaires, medical examination, and a nutritional survey (Kim et al., 2014).

In our analysis of the NHANES data set we included all the variables from all the categories. We also combined the datasets from 1999 through 2014, based on the sequential numbers assigned to each participant. The dataset initially included 83,731 participants and 2,864 variables. To mitigate the effects of multi-collinearity, we deleted duplicate variables that represented the same information in different units (e.g., cholesterols in laboratory data, mg/dl [code: LBXSCH] and mmol/dl [code: LBDSCHSI]). Then we removed all qualitative variables as well as a variable that directly expressed the status of depression ("How long have you suffered from depression, anxiety or emotional problems?" [code: PFD069D from 1999 to 2000, PFD069DG from 2000 to 2008]). The value '9', '99', and '999' represent the "Don't know" answer for continuous variables and were therefore treated as missing values. Finally, we exclude all variables with more than 10%, of the samples missing. Thus, 157 variables were used for training the algorithms.

In the K-NHANES data set, a variable named Patient Health Questionnaire 9 (PHQ-9) was only added in 2014. We therefore focused on the one-year dataset of K-NHANES which surveyed 7,550 participants and contained 652 variables. As the laboratory data of K-NHANES used only one type of unit, there were no duplicate variables that had to be removed to avoid multi-collinearity. Similarly to NHANES, we removed 7 variables which could directly imply depression, such as "Have you ever been diagnosed with depression?" (code: DF2_dg) or "How depressed and anxious are you?" (code: LQ_5EQL). And, as with the NHANES data set, numerical values representing no-response were treated as missing values for continuous variables.

After selecting the variables according to above criteria, we calculated the proportion of missing values in each variable and included only variables that had less than 25% missing values. This is because incomplete data and the proportion of missing items tends to affect the prediction power of machine-learning classifiers (Williams et al., 2007). We were then left with 157 of 2,864 variables in NHANES and 314 of 652 variables in K-NHANES (excluding the sample index numbers and the depression outcome variables) were used to train the deep-learning and machine-learning classifiers (lists of all the selected variables are presented in Supplementary Table 1 and Supplementary Table 2).

Also, not all participants were assessed for depression in both datasets. Hence, only records of those who had replied to the depression screening questionnaires were included for further analysis. In NHANES, 28,280 of 83,731 participants and in K-NHANES, 4,949 of 7,550 participants were evaluated for depression using reliable psychiatric scales (Figure 1). We therefore ended up with a dataset of 28,280 participants with 157 variables for NHANES and with 4,949 participants with 314 variables for K-NHANES. These datasets were used to train and validate the deep-learning algorithms and machine-learning classifiers.

Evaluation of Depression

NHANES and K-NHANES used a validated screening tool for depression to test for depression in the general population. NHANES used the automated version of the World Health Organization Composite International Diagnostic Interview, version 2.1 (CIDI-Auto 2.1) between 1999 and 2004. CIDI-Auto 2.1 was designed to assess mental disorders and is especially suitable for large populations (Andrews and Peters, 1998). We used the depression score among panic disorder, generalized anxiety disorder, and depressive disorder diagnostic modules of CIDI-Auto 2.1 to create positive or negative diagnoses of depression (code: CIDDSCOR), which we used as the label of this dataset. In this period, a total of 2,216 individuals were assessed for depression and 148 (6.7%) had a positive diagnosis (Figure 1).

From 2005 to 2014 another validated screening tool for depression, PHQ-9, replaced the CIDI-Auto 2.1 for NHANES. The PHQ-9 consists of 9 questionnaires, which are based on the diagnostic criteria of depression from the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) (Kroenke and Spitzer, 2002). It is a reliable and valid measurement in screening depression (Kroenke et al., 2001). We chose 10 to be our threshold for the diagnosis of depression, as this threshold had reliable sensitivity and specificity for detecting major depressive disorders (Kroenke et al., 2001). In the 2005-2014 period, 26,064 participants were assessed for depression and 2,094 were diagnosed with depression (8.0%).

In a cross-sectional study of K-NHANES, the standardized Korean version of PHQ-9 was used to assess depression (Choi et al., 2007). Among 7,550 individuals who participated in the survey, 4,949 were assessed for depression and 344 (7.0%) had a PHQ-9 total score of 10 or more.

To compare the prevalence of each predictor between depression and non-depression groups, we performed univariate logistic regression with the binary status of depression as the outcome variable and each predictor as an independent variable. Collection and pre-processing of the survey datasets and conventional univariate logistic regression analysis were carried out using SPSS for Mac version 18.0 (SPSS, Chicago, Illinois).

Development and Validation of Algorithms

For the deep-learning algorithm, we devised a dense-layer, feed-forward, neural-network model. To optimize the number of nodes and layers, we tried combinations of 10, 100, 500, 1000, and 1500 nodes per layer with 1, 2, 3, 4, 5, or 6 layers. All the neurons had sigmoid activation functions except for the output layer, which consisted of softmax neurons. The network was trained with scaled conjugate gradient backpropagation. Network parameters were initialized using Gaussian distributed random numbers with a mean of 0 and a standard deviation of 1. For each of these combinations of the number of nodes and the number of layers we computed the area under the receiver operating characteristic curve (AUC) with 10-fold cross validation. The 5-layer network with 500 nodes per layer produced the maximal AUC.

The NHANES network therefore had 157 input nodes and 28,280 samples on which to run. The K-HANES network had 314 input nodes and ran on 4,949 samples. Both networks were trained to classify depression in the participants. The training and evaluation of the deep-learning algorithm were carried out using TensorFlow (Google Inc.) (Abadi et al., 2016).

To compare classification performance between deep-learning and machine-learning algorithms, we ran 5 different commonly used machine-learning algorithms: decision trees, logistic regression, support-vector machine, K-nearest neighbor, and Ensemble classifiers. Decision-tree learning refers to the algorithm that generates a set of prediction rules based on deciding a threshold on a variable, one variable at a time (Quinlan, 1986). Logistic-regression learning is based on binominal classification using the logistic-regression function (Dreiseitl and Ohno-Machado, 2002). Support-vector machines use probabilistic, binary, linear classifiers to construct a set of hyperplanes with maximal margins, potentially in higher dimensional space, nominally, using kernels (Suykens and Vandewalle, 1999). K-nearest neighbor is a method of predicting new data from the majority or plurality of information among the k-closest neighbors of existing data (Cover and Hart, 1967). In each category, there were 1 to 6 sub-classifiers and all of them were tested on the training set (Supplementary Table 4).

We ran 10-fold cross-validation for all algorithms and datasets to validate the performance of each classifier and to avoid overfitting (Liu Ling, 2009). These 10 sub-datasets were then also used to compare the performance between deep-learning and machine-learning algorithms. For example, a model was trained with 9 of 10 sub-datasets and the prediction performance of algorithm (AUC values) in the remaining dataset was measured. Then, an one-way ANOVA and post-hoc t tests were carried out to judge whether the deep-learning algorithm significantly outperformed each of the machine-learning algorithms (Supplementary Table 6).

To compare the classification performance of the algorithms between the US and Korean NHANES datasets, we extracted the variables that completely match between the two datasets. Overall 41 of the 157 variables in NHANES and 316 variables in K-NHANES were identical in the text of the questionnaire and the response items. Of these, we excluded the item of PHQ-9 total score, which was used as an outcome indicator. The remaining 40 variables are presented in Supplementary Table 3.

We also examined how the performance of deep-learning and logistic regression changed as the number of predictors varied. Using the NHANES dataset, after extracting subsamples with 99 and 49 predictors from 157 ones, we tested how classification performance varied when each algorithm was trained using each subsample (Supplementary Figure 1). Classification Learner Application in MATLAB R2017a (Mathworks, Natick, Massachusetts) was addressed for model training and for the analysis of the results of the machine-learning classifiers.

Contribution-Ranking Analysis of Variables

In addition to the above, we carried out contribution-ranking analysis to investigate and interpret the inner-mechanism of the selected deep-learning model as well as that of the conventional machine-learning classifiers. Among the conventional machine-learning classifiers, logistic regression was chosen as the comparison target it is widely used in clinical studies. Another reason for choosing logistic regression was that its t-statistic values can be used to determine the contribution of its variables. For deep-learning, we used the cross-entropy measure as the comparison statistic. It reflects the accuracy and the confidence in the learning performance of the neural networks (Le et al., 2011; Oh et al., 2017).

Results

NHANES and K-NHANES Data Sets Characteristics

How the participants were selected as well as the number of participants with depression in the NHANES and K-NHANES datasets are shown in Figure 1. In NHANES, the original data set included 83,731 individuals, who participated in the survey from 1999 to 2014. Among them, it was impossible to infer depression from 55,451, because they were not asked about depression (being under the age of 19) or they did not complete the depression-screening questionnaires. Thus, the remaining 28,280 participants were chosen for further analysis and the overall prevalence of depression in this population was 7.9% (2,242 [148 from year 1999 to 2004 and 2,094 from year 2005 to 2014] of 28,280).

In K-NHANES, which was a one-year survey conducted in 2014, 7,550 individuals were enrolled and 4,949 were assessed for depression. The prevalence of depression in the K-NHANES data set was 6.95% (344 of 4,949) when using the same criteria as NHANES (PHQ-9 total score of 10 or more). The total number of predictors (or features) used in training the machine-learning classifiers were 157 for the NHANES dataset and 314 for the K-NHANES dataset (see Methods for details).

Identification of Depression in NHANES and K-NHANES Data Sets

Figure 2 summarizes the performance of deep learning and conventional machine-learning classifiers for identifying depression in NHANES and K-NHANES. It depicts the ROC curves resulting from 10-fold cross-validation. Deep learning detected depression with an AUC of 0.91 on NHANES and an 0.89 on K-NHANES. Although deep learning resulted in the highest accuracy, there were not large differences between its performance and those of some of the other, more-conventional machine-learning classifiers. In NHANES, linear support vector machine (SVM) and logistic regression had only slightly worse results (AUC of 0.89 for both), while complex tree had the worst performance (AUC of 0.82). There were significant differences between the classification performance of the algorithms (one-way ANOVA; $F = 38.581$; $p < 0.001$). Further, post-hoc analysis revealed that deep-learning was superior to coarse KNN ($p < 0.001$) and to complex tree algorithms ($p < 0.001$). But its performance was not significantly better than linear SVM ($p = 0.222$) and logistic regression ($p = 0.347$) (see Supplementary Table 6 for details).

Deep learning was also the most accurate for K-NHANES, with an AUC of 0.89, followed by boosted tree and linear SVM (AUC of 0.86 and 0.85, respectively). Here, the coarse KNN classifier exhibited the worst performance (AUC of 0.78). In K-NHANES, the classification performance between algorithms was statistically significantly different (one-way ANOVA; $F = 28.361$; $p < 0.001$). Post-hoc analysis showed that deep-learning had higher performance than linear SVM ($p = 0.004$), logistic regression ($p < 0.001$), and coarse KNN ($p < 0.001$). Though its classification performance was not significantly better than boosted trees ($p = 0.983$).

Prediction of Depression in Cross-temporal and Cross-National Modalities

We wanted to further assess the ability of deep-learning and other machine-learning algorithms to predict the occurrence of depression across time as well as across national datasets. For our cross-temporal analysis, classifiers were trained on 14 years of NHANES (1999 -2012) and their prediction performance was then tested on newer NHANES data from 2013-2014. Deep learning achieved an AUC of 0.92 (Figure 3A), with linear SVM and logistic regression tying for second (with an AUC of 0.80**)**. Coarse KNN trailed behind (AUC of 0.77) and the complex tree classifier did even worse (AUC of 0.72).

In the cross-national validation test, we aimed to detect depression in one country, with the classifier trained on another country's survey data. All the variables that were common to NHANES and K-NHANES were chosen—41 in total (Supplementary Table 3). We then trained the various classifiers on the NHANES dataset and tested for depression on the K-NHANES data set. All classifiers achieved relatively similar accuracies in this analysis. Logistic regression reached the highest accuracy, with an AUC of 0.77 (Figure 3B), followed by deep learning and the ensemble subspace discriminant classifier (both with an AUC of 0.74). Coarse KNN did a bit worse (AUC of 0.72). Prediction of depression in the NHANES dataset with models trained on the K-NHANES dataset did not show reliable performance (Supplementary Table 5). This is likely due to the relatively small number of samples in the K-NHANES dataset with respect to the NHANES one.

Contribution of Predictors in Deep Learning and Logistic Regression Algorithms

We analyzed the contribution of the various variables used to estimate depression with deep learning and logistic regression (Figure 4). Variables tended to have similar contributions for deep learning as measured by their cross entropy. For the NHANES dataset, the cross entropy was 25.488 ± 0.002 (mean ± standard deviation, here and below), resulting in a ratio of standard deviation to mean of 8×10^{-5} (Figure 4A). For K-NHANES, the cross entropy was 0.170 ± 0.008, and the standard-deviation-to-mean ratio was 0.05 (Figure 4C).

In logistic regression, in contrast, the variables had more varied contributions. The absolute value of the t-statistics across all variables was 1.06 ± 1.03 in NHANES and 0.92 ± 0.85 in K-NHANES. In NHANES, 20 of 157 variables (12.7%) achieved statistical significance ($p < 0.05$, uncorrected); in K-NHANES, 24 of 314 variables 7.6%) were statistically significant. The standard-deviation-to-mean ratio was 0.97 in NHANES and 0.92 in K-NHANES. These are much higher values than the ratios for deep learning. These results suggest that deep learning used most of the variables for identifying depression whereas logistic regression used only a subset of the variables. For the logistic-regression analysis, the variable contributing most to identifying depression was one related to the subjective feeling of health in both NHANES and K-NHANES (Table 1). Univariate logistic regression analysis also showed that the most contributing variables in estimating depression are statistically significant (NHANES; Odds Ratio [OR] = 1.108, $p < 0.001$; K-NHANES; OR = 3.306, $p < 0.001$), and these findings suggest that logistic regression could detect variables with higher odds in patients with depression than without.

Discussion

In this study, we showed that a deep-learning algorithm can well decode the occurrence of depression from other health-related and demographic factors in large, survey datasets. Deep learning further significantly outperformed conventional machine-learning classifiers for the K-NHANES dataset. It also outperformed all conventional machine-learning classifiers for the NHANES dataset, though it did so significantly for two of the four conventional classifiers. Importantly, deep learning also demonstrated predictive ability over novel datasets—both across time and across national surveys. This hints at its potential future clinical implications.

Previous studies have attempted to estimate the prevalence and the severity of depression with automated algorithms. In one study, van Loo et al. found that data mining techniques could classify subtypes of major depressive disorder according to the long-term disease course using the World Health Organization’s World Mental Health Surveys dataset (Van Loo et al., 2014). Following these findings, a more recent study showed that machine learning algorithms could predict the course and severity of depression with an AUC of 0.71-0.76 (Kessler et al., 2016). Our dataset did not include information from which we could predict the severity or prognosis of depression.

Instead, it estimated the occurrence of self-reported depression based on other health and demographic factors in the general population with a relatively high degree of accuracy. This suggests that machine learning algorithms can be used to classify risk groups for depression from general survey data and might further point to key contributing factors for depression in the general population.

Machine learning algorithms have yielded notable prediction accuracies in other fields of medicine (Gulshan et al., 2016). Nevertheless, it was suggested that the inner-workings many machine-learning algorithms—and especially deep neural-networks—are unclear. So, applying them in clinical settings may be difficult (Barak-Corren et al., 2017). We found that deep learning generally used a combination of many features to achieve its classification accuracy, whereas logistic regression tended to use a smaller subset of the features (Fig. 4.) Hence, to look further into individual factors that machine learning found to influence depression, we used several more conventional machine-learning approaches—and in particular logistic regression—and compared their results with deep learning (Table 1 and Supplementary Table 1). Interestingly, our analysis of ranked contributing-factors for logistic regression showed that the algorithm relied on clinically well-known risk factors—such as amount of physical activity and a degree of daily stress—to identify depression in our model (Table 1). In NHANES, the questionnaire item 'Number of days mental health was not good' was ranked top among contributing factors for logistic regression. This is to be expected, as certain screening tools for depression includes this specific question (Kroenke et al., 2009). Similar results were found for K-NHANES. 'Subjective health status' and 'The degree of stress in daily life' were highest and second highest ranked.

Other highly ranked contributing factors for depressing using logistic regression over NHANES were 'How many times urinate in night?', 'How often have urinary leakage', and 'Urinate before reaching the toilet' (ranked 5th, 6th and 11th respectively, Table 1). Univariate logistic regression analysis also showed that these items had significantly higher odds in patients with depression than who had no depression (Odds ratio [OR]= 1.314, $p < 0.001$; OR = 1.400, $p < 0.001$, respectively). Idiopathic urinary incontinence is known to be strongly related to depression by altered serotonergic function (Zorn et al., 1999). And the overall prevalence of urinary incontinence in women was 38% in NHANES data from 1999 to 2000 (Anger et al., 2006). Our results might have therefore revealed a key factor affecting depression in the US general population.

Similarly, in K-NHANES, a quarter of the top 20 ranked features appeared related to Chronic Obstructive Pulmonary Disease (COPD) (Table 1; ranked 4th, 8th, 15th, 17th). Among these items, the item of 'Cough experience for 3 consecutive months' and 'Diagnosis of Asthma' had significantly higher odds in patient with depression than were not (OR = 3.230, $p < 0.001$; OR = 2.856, $p < 0.001$, respectively). These results too go hand in hand with epidemiological findings in South Korea: most COPD patients there suffer from depression, a much higher rate than healthy controls (Ryu et al., 2010). These results suggest that conventional machine-learning classifiers, especially logistic regression, can identify depression utilizing localized disease characteristics.

As for comparing the performance of deep learning and conventional machine-learning techniques, deep-learning best detected depression in both NHANES and K-NHANES. But, while deep-learning was significantly superior to all the conventional machine learning algorithms we tried over K-NHANES, its accuracy was not significantly different from linear SVM and logistic regression over NHANES (Supplementary Table 6). One reason for this might be the different number of samples and predictors in the two datasets. The number of samples in NHANES was more than 5.5 times larger than in K-NHANES, while number of variables in the former was less than half in the latter (NHANES vs. K-NHANES: number of samples 28,280 vs. 4,949, number of variables 157 vs. 316, respectively). Hence the samples to features ratio for NHANES was more than 11 times that of K-NHANES. This ratio is a well-known, crucial factor affecting the performance of deep neural networks (Subana and Samarasinghe, 2016). Ideally, the ratio should be as small as possible, meaning that we want as few features and as many samples as possible to build a robust prediction and avoid overfitting. In K-NHANES, we speculate the homogenous demographics of Korean population helped improve the performance.

We should also note that our results showed that deep-learning can be useful for data that does not have a visual component. Our deep-learning algorithm well decoded depression from numerical data, without visually representing any images. This ability of neural networks has already been demonstrated in other studies. Yang et al. reported that text-only information, acquired from transcription of conversation with patients suffering from depression, let them detected the disease well using deep learning (Yang et al., 2017). Another example is a deep neural-network trained on EEG signals that could classify depressive patients from normal ones (Acharya et al., 2018). Along the lines of this literature, we proposed that numerical data based on survey questionnaires and laboratory tests could also be used to detect certain psychiatric illnesses.

While only a limited number of key parameters determined much of the classification performance in logistic regression, deep learning used roughly all the parameters for classification (Figure 4). This suggests that, while the two algorithms perform similarly on novel datasets (Figure 3), the internal workings of the classification algorithms were vastly different. The deep-learning architecture we used was highly nonlinear (as it was 5 layers deep with sigmoid activation functions), and highly-complex (2,500 parameters) compared to logistic regression. Examining the 41 chosen variables suggests that they did not happen to be ones that were favored by logistic regression for neither the NHANES nor K-NHANES datasets. Instead, additional analysis that we carried out showed that deep learning out-performed logistic regression when the number of variables was larger than 100, but it was logistic regression that did better as the number of trained variables decreased (Supplementary Figure 1). It is likely that, as more features becomes available, the performance of deep learning would further increase over logistic regression (Krizhevsky et al., 2012).

Several important limitations of this study should be mentioned. First, this study viewed depression as a binary variable. So, we could not evaluate the correlation between the various factors and the severity of depression. Second, as both NHANES and K-NHANES were cross-sectional surveys (rather than longitudinal), we could not measure the prognosis of the disease or the future occurrence of depression in the population. Third, the performance of the deep-learning and conventional machine-learning algorithms was clearly reduced cross-nationally (Figure 3B; AUC of 0.74). This may be due to cross-national diversity in datasets or to cultural differences in understanding the questions, in propensity for responding in specific manners, and so on. But the result may also be due to the smaller number of variables used for training (41 variables in the cross-national datasets—26.1% of the variables in NHANES and 13.1% in K-NHANES), as the performance of the neural network largely depends on the number of predictors (Karayiannis N, 1993).

Needless to say, though our model estimated the presence of depression with relatively high accuracy across the population, it could not replace the conventional, individual screening tools for depression (e.g., PHQ-9 or Beck Depression Inventory). Rather, a trained deep-learning algorithm might be used in aggregate—for instance to estimate the regional prevalence of depression in regions where individual mental-health surveys were not run. Additional research into the performance of deep-learning on longitudinal and other datasets and on other mental disorders might reveal more information about the incidence, prevalence, and progression of mental disorders, on comorbidity of such disorders, and on correlations between mental disorders and various demographic, life-style, and other factors.

Contributors

J Oh, K Yun, and J-H Chae conceived the idea and designed the study. J Oh and K Yun organized the data and coded algorithms. J Oh and K Yun drafted the manuscript, J-H Chae, U Maoz and T-S Kim reviewed the data, suggested additional analyses, and revised the manuscript.

Role of the Funding Sources

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HM15C1054). Uri Maoz was funded by the Bial Foundation (grant 388/14).

Conflict of Interests

All authors report no competing interests.

Acknowledgements

None

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., Brain, G., 2016. TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning, in: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). pp. 265–284. https://doi.org/10.1038/nn.3331

Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H., Subha, D.P., 2018. Automated EEG-based screening of depression using deep convolutional neural network. Comput. Methods Programs Biomed. 161, 103–113. https://doi.org/10.1016/j.cmpb.2018.04.012

Andrews, G., Peters, L., 1998. The psychometric properties of the Composite International Diagnostic Interview. Soc Psychiatry Psychiatr Epidemiol 33, 80–88. https://doi.org/10.1007/s001270050026

Anger, J.T., Saigal, C.S., Stothers, L., Thom, D.H., Rodríguez, L. V., Litwin, M.S., 2006. The Prevalence of Urinary Incontinence Among Community Dwelling Men: Results From the National Health and Nutrition Examination Survey. J. Urol. 176, 2103–2108. https://doi.org/10.1016/j.juro.2006.07.029

Avenevoli, S., Swendsen, J., He, J.P., Burstein, M., Merikangas, K.R., 2015. Major Depression in the National Comorbidity Survey–Adolescent Supplement: Prevalence, Correlates, and Treatment. J. Am. Acad. Child Adolesc. Psychiatry 54, 37–44.e2. https://doi.org/10.1016/j.jaac.2014.10.010

Barak-Corren, Y., Castro, V.M., Javitt, S., Hoffnagle, A.G., Dai, Y., Perlis, R.H., Nock, M.K., Smoller, J.W., Reis, B.Y., 2017. Predicting suicidal behavior from longitudinal electronic health records. Am. J. Psychiatry 174, 154–162. https://doi.org/10.1176/appi.ajp.2016.16010077

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression: A machine learning approach. The Lancet Psychiatry 3, 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

Choi, H. S., Choi, J. H., Park, K. H., Joo, K. J., Ga, H., Ko, H. J., & Kim, S.R., 2007. Standardization of the Korean version of Patient Health Questionnaire-9 as a screening instrument for major depressive disorder. J. Korean Acad. Fam. Med. 28, 114–119.

Center for Behavioral Health Statistics and Quality (2017). 2016 National Survey on Drug Use and Health: Detailed Tables. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Centers for Disease Control and Prevention (CDC) (2014). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Cover, T.M., Hart, P.E., 1967. Nearest Neighbor Pattern Classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/TIT.1967.1053964

Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N., Meyer, D., 2016. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. PLoS One 11. https://doi.org/10.1371/journal.pone.0148195

Dreiseitl, S., Ohno-Machado, L., 2002. Logistic regression and artificial neural network classification models: A methodology review. J. Biomed. Inform. 35, 352–359. https://doi.org/10.1016/S1532-0464(03)00034-0

Ford, D.E., Erlinger, T.P., 2004. Depression and C-reactive protein in US adults: data from the Third National Health and Nutrition Examination Survey. Arch. Intern. Med. 164, 1010–4. https://doi.org/10.1001/archinte.164.9.1010

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., Webster, D.R., 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316, 2402. https://doi.org/10.1001/jama.2016.17216

Karayiannis N, V.A., 1993. Artificial Neural Networks: Learning Algorithms, Performance Evaluation, and Applications. Kluwer Academic Publishers, Boston, MA.

Kessler, R.C., van Loo, H.M., Wardenaar, K.J., Bossarte, R.M., Brenner, L.A., Cai, T., Ebert, D.D., Hwang, I., Li, J., de Jonge, P., Nierenberg, A.A., Petukhova, M. V, Rosellini, A.J., Sampson, N.A., Schoevers, R.A., Wilcox, M.A., Zaslavsky, A.M., 2016. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol. Psychiatry 21, 1366–1371. https://doi.org/10.1038/mp.2015.198

Kim, Y., Kim, N.H., Lee, H.K., Song, J.S., Kim, H.C., Choi, S., Chun, C., Khang, Y., Oh, K., Laibson, P., Lemp, M., Meisler, D., Castillo, J. Del, O'Brien, T., Pflugfelder, S., Rolando, M., Schein, O., Seitz, B., Tseng, S., Setten, G. van, Wilson, S., Yiu, S., 2014. The Korea National Health and Nutrition Examination Survey (KNHANES): Current Status and Challenges. Epidemiol. Health 36, e2014002. https://doi.org/10.4178/epih/e2014002

Krizhevsky, A., Sutskever, I., Geoffrey E., H., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst. 25 1097–1105. https://doi.org/10.1109/5.726791

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: A New Depression Diagnostic and Severity Measure. Psychiatr. Ann. 32, 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: Validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114, 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Le, Q. V, Coates, A., Prochnow, B., Ng, A.Y., 2011. On Optimization Methods for Deep Learning. Proc. 28th Int. Conf. Mach. Learn. 265–272. https://doi.org/10.1.1.220.8705

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Oh J, Yun K, Hwang J-H, Chae J-H., 2017. Classification of Suicide Attempts through a Machine Learning Algorithm Based on Multiple Systemic Psychiatric Scales. Front. Psychiatry 8, 192. https://doi.org/10.3389/fpsyt.2017.00192

Perlis, R.H., 2013. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol. Psychiatry 74, 7–14. https://doi.org/10.1016/j.biopsych.2012.12.007

Quinlan, J.R., 1986. Induction of Decision Trees. Mach. Learn. 1, 81–106. https://doi.org/10.1023/A:1022643204877

Refaeilzadeh, Payam, Lei Tang, and H.L., 2009. Encyclopedia of database systems. Springer US, pp. 532–538.

Ryu, Y.J., Chun, E.-M., Lee, J.H., Chang, J.H., 2010. Prevalence of depression and anxiety in outpatients with chronic airway lung disease. Korean J. Intern. Med. 25, 51–57. https://doi.org/10.3904/kjim.2010.25.1.51

Smith, K., 2014. Mental health: A world of depression. Nature 515, 180–181. https://doi.org/10.1038/515180a

Subana, S., Samarasinghe, S., 2016. Artificial Neural Network Modelling. Springer US.

Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9, 293–300. https://doi.org/10.1023/A:1018628609742

Vallance, J.K., Winkler, E.A.H., Gardiner, P.A., Healy, G.N., Lynch, B.M., Owen, N., 2011. Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). Prev. Med. (Baltim). 53, 284–288. https://doi.org/10.1016/j.ypmed.2011.07.013

Van Loo, H.M., Cai, T., Gruber, M.J., Li, J., De Jonge, P., Petukhova, M., Rose, S., Sampson, N.A., Schoevers, R.A., Wardenaar, K.J., Wilcox, M.A., Al-Hamzawi, A.O., Andrade, L.H., Bromet, E.J., Bunting, B., Fayyad, J., Florescu, S.E., Gureje, O., Hu, C., Huang, Y., Levinson, D., Medina-Mora, M.E., Nakane, Y., Posada-Villa, J., Scott, K.M., Xavier, M., Zarkov, Z., Kessler, R.C., 2014. Major depressive disorder subtypes to predict long-term course. Depress. Anxiety 31, 765–777. https://doi.org/10.1002/da.22233

Weissman, M.M., Bland, R.C., Canino, G.J., Faravelli, C., Greenwald, S., Hwu, H.G., Joyce, P.R., Karam, E.G., Lee, C.K., Lellouch, J., Lépine, J.P., Newman, S.C., M, R.-S., Wells, J.E., Wickramaratne, P.J., Wittchen, H., Yeh, E.K., 1996. Cross-national epidemiology of major depression and bipolar disorder. JAMA 276, 293–299. https://doi.org/10.1001/jama.1996.03540040037030

Williams, D., Liao, X., Xue, Y., Carin, L., Krishnapuram, B., 2007. On classification with incomplete data. IEEE Trans. Pattern Anal. Mach. Intell. 29, 427–436. https://doi.org/10.1109/TPAMI.2007.52

World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.

Yang, L., Jiang, D., Xia, X., Pei, E., Oveneke, M.C., Sahli, H., 2017. Multimodal Measurement of Depression Using Deep Learning Models. pp. 53–59. https://doi.org/10.1145/3133944.3133948

Zhou, X., Jin, K., Shang, Y., Guo, G., 2018. Visually Interpretable Representation Learning for Depression Recognition from Facial Images. IEEE Trans. Affect. Comput. https://doi.org/10.1109/TAFFC.2018.2828819

Zhu, Y., Shang, Y., Shao, Z., Guo, G., 2018. Automated Depression Diagnosis Based on Deep Networks to Encode Facial Appearance and Dynamics. IEEE Trans. Affect. Comput. 9, 578–584. https://doi.org/10.1109/TAFFC.2017.2650899

Zorn, B.H., Montgomery, H., Pieper, K., Gray, M., Steers, W.D., 1999. Urinary incontinence and depression. J. Urol. 162, 82–4. https://doi.org/10.1097/00005392-199907000-00020

Figure Legends

Figure 1. Participants Selection and Prevalence of Depression in NHANES and K-NHANES

(A) Data collection and participant selection in the NHANES data set (1999-2014). (B) The corresponding numbers to A in the K-NHANES dataset.

Figure 2. Area Under Curve (AUC) for Identifying Depression in NHANES and K-NHANES

Performance of deep learning and conventional machine learning classifiers trained with NHANES 1999 to 2014 data sets (A) and K-NHANES 2014 data set (B).

Figure 3. Estimation of Depression in NHANES and Cross-National Estimation of Depression

Performance of various machine learning algorithms when predicting depression across time and across national surveys. (A) Predicting depression in the last two years of available data on NHANES (2013-2014) with various machine learning algorithms trained on the previous 14 years of NHANES data (1999-2012). (B) Predicting depression in K-NHANES (2014) with various machine-learning algorithms trained on 16 years of NHANES data (1999-2014).

Figure 4. Contribution of Predictors for Deep Learning and Logistic Regression over the NHANES and K-NHANES datasets

The figure depicts the distribution of cross entropy values over all variables in NHANES (1999-2014) in (A) and K-NHANES (2014) in (C) for deep learning. (Note the inset in (C) that covers a smaller range of cross entropy values.) Cross entropy represents the contribution of each variable in estimating depression during the training process of the deep-learning algorithm. It also shows the distribution of t-statistics across all variables in NHANES (B) and K-NHANES (D) for logistic regression. The horizontal red lines designate the threshold of statistical significance (above the red line denotes $p < 0.05$; uncorrected).

Figure 1. Case Selection of NHANES (1999-2014) and K-NHANES (2014) Datasets

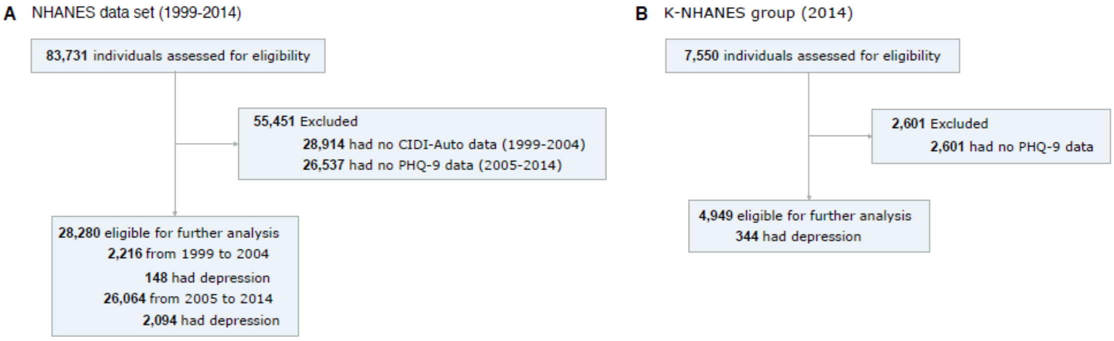


Figure 2. Estimation of Depression in NHANES (1999-2014) and K-NHANES (2014) via Machine Learning Algorithms

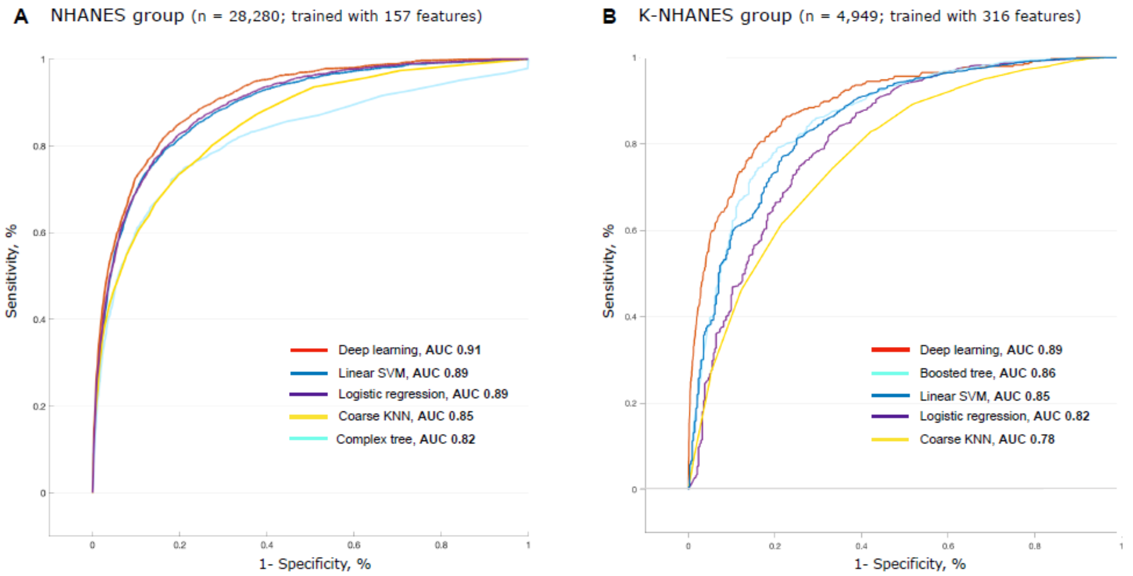


Figure 3A. Prediction of Depression in NHANES 2013-2014 Data Using a Model Trained with NHANES 1999-2012 Data

Figure 3B. Estimation of Depression in K-NHANES 2014 Using a Model Trained with NHANES 1999-2014

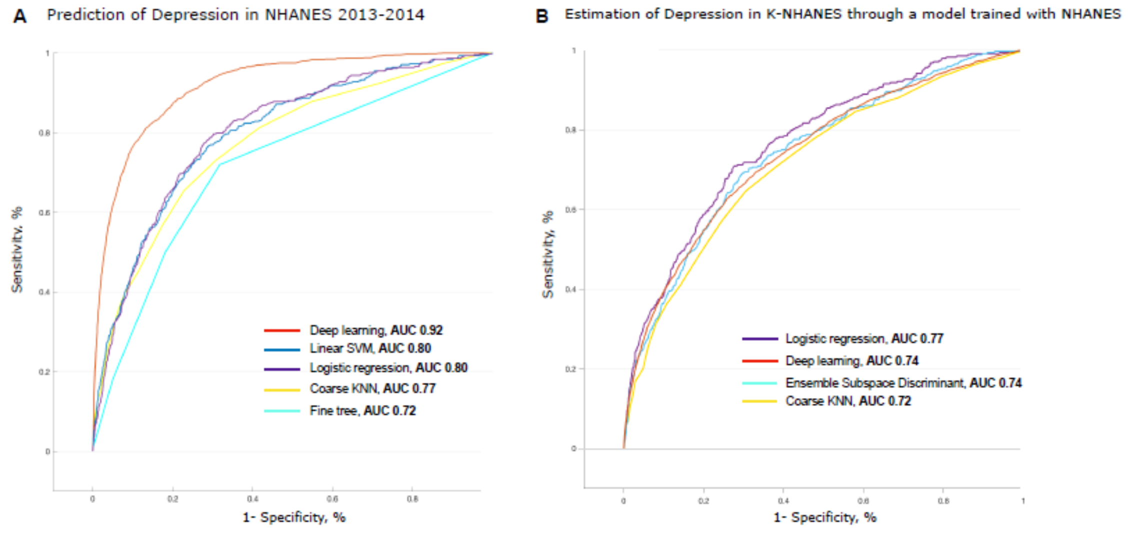


Figure 4. Contribution of Variables in Estimating Depression with Deep learning and Logistic Regression Algorithms

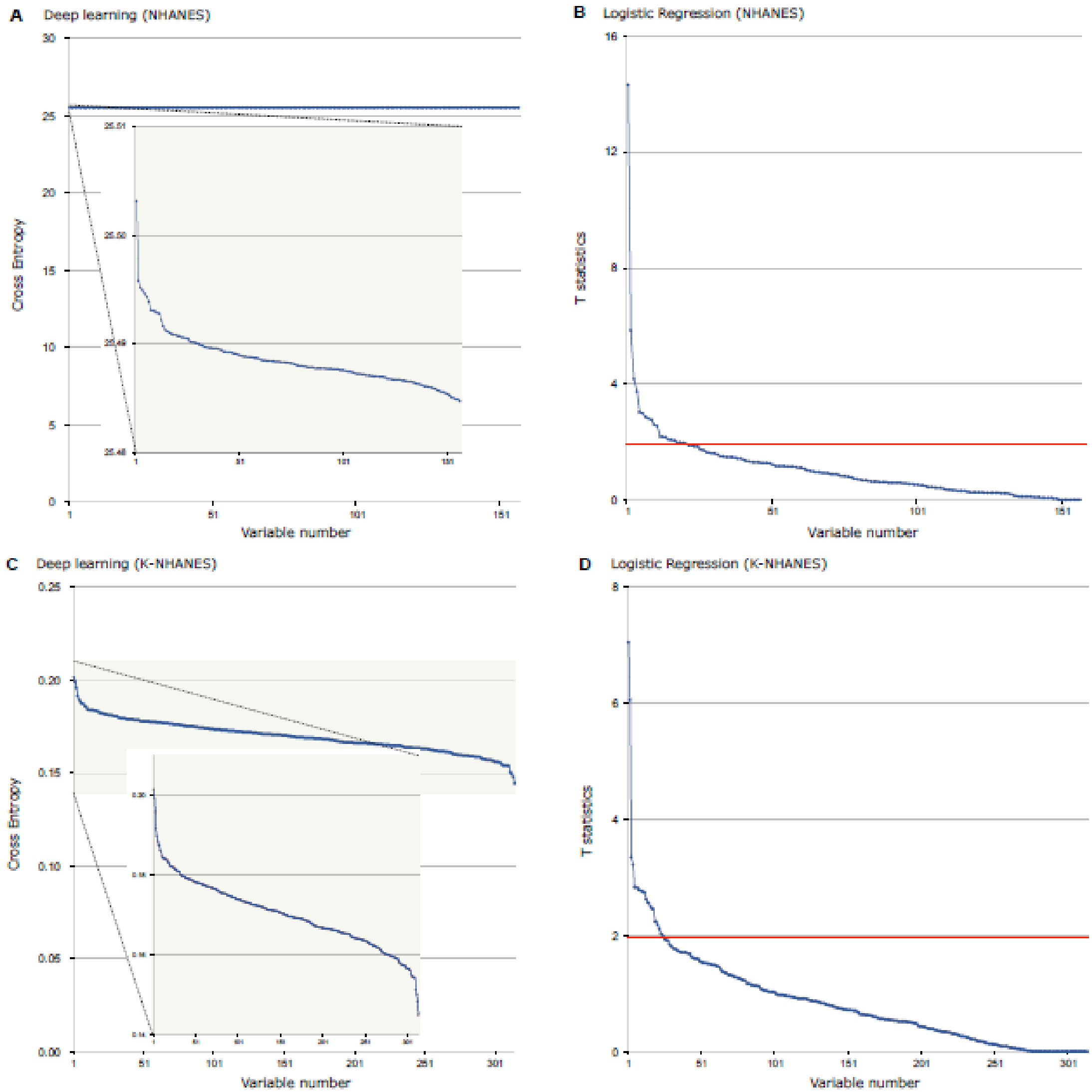


Table 1. Top 20 ranked variables used to estimate depression in NHANES and K-NHANES data sets through deep learning and logistic regression algorit

Predictor Ranking	Code	Label	Cross Entropy	Predictor Ranking	Code	Label	*T* statistics	*P* value
NHANES								
Deep Learning				Logistic Regression				
1	PFQ054	Need special equipment to walk	25.503	1	HSQ480	no. of days mental health was not good	29.409	<.001
2	LBXGH	Glycohemoglobin:(%)	25.496	2	HSQ490	Inactive days due to phys./mental hlth	10.763	<.001
3	IMQ020	Received hepatitis B 3 dose series	25.495	3	HSD010	General health condition	9.911	<.001
4	PFQ057	Experience confusion/memory problems	25.495	4	HUQ090	Seen mental health professional /past yr	-7.751	<.001
5	FSD151	HH Emergency food received	25.495	5	KIQ480	How many times urinate in night?	7.583	<.001
6	MCQ300A	Close relative had heart attack?	25.494	6	KIQ005	How often have urinary leakage	6.191	<.001
7	BPXSY3	Systolic: Blood pres (3rd rdg) mm Hg	25.494	7	DBQ700	How healthy is the diet	5.522	<.001
8	PFQ090	Require special healthcare equipment	25.493	8	SMAQUEX2	Questionnaire Mode Flag	-5.135	<.001
9	WHD140	Self-reported greatest weight(pounds)	25.493	9	PFQ059	Physical, mental, emotional limitations	-5.010	<.001
10	MCQ300C	Close relative had diabetes?	25.493	10	URXCRS	Creatinine, urine (umol/L)	4.838	<.001
11	HSQ571	SP donated blood in past 12 months?	25.493	11	KIQ044	Urinated before reaching the toilet	-4.777	<.001
12	URXCRS	Creatinine, urine (umol/L)	25.493	12	PFQ057	Experience confusion/memory problems	-4.594	<.001
13	MCQ140	Trouble seeing even with glass/contacts	25.492	13	HUQ010	General health condition	4.085	<.001
14	HSQ520	SP have flu, pneumonia, ear infection?	25.492	14	HSQ510	SP have stomach or intestinal illness?	-3.963	<.001
15	LBXBCD	Cadmium (ug/L)	25.491	15	SMQ020	Smoked at least 100 cigarettes in life	-3.689	<.001
16	LBDEONO	Eosinophils number	25.491	16	RIDAGEYR	Age at Screening Adjudicated - Recode	-3.588	<.001
17	FSDAD	Adult food security category	25.491	17	RIAGENDR	Gender	3.439	.001
18	BPQ150D	Had cigarettes in the past 30 minutes?	25.491	18	WHQ040	Like to weigh more, less or same	-3.397	.001
19	BMXHT	Standing Height (cm)	25.491	19	SMD410	Does anyone smoke in the home	-3.242	.001
20	MCQ160L	Ever told you had any liver condition	25.491	20	DIQ050	Taking insulin now	-3.059	.002
K-NHANES								
Deep Learning				Logistic Regression				
1	HE_PFTag	Age when diagnosed COPD	.201	1	DJ4_pt	Current treatment of Asthma	7.046	<.001
2	HE_fev1p	FEV1	.199	2	DI3_ag	Age when diagnosed as cerebral stroke	-6.046	<.001
3	HE_Unitr	Nitrate	.196	3	BE3_81	Middle-strength physical activity	-3.340	.001
4	Y_MTM_D2	Breastfeeding period	.192	4	HE_PFTtr	Treatment of COPD	3.217	.001
5	HE_HCHOL	Hypercholesterolemia	.190	5	DC6_pt	Treatment of lung cancer	-2.830	.005
6	DC1_dg	Diagnosis of gatric cancer	.188	6	HE_DMfh3	Diagnosis of diabetes mellitus (siblings)	2.822	.005
7	HE_DMfh1	Diagnosis of diabetes mellitus (father)	.187	7	E_NWT	Average near-field working time per day	2.822	.005
8	sex	sex	.187	8	HE_cough1	Cough experience for 3 consecutive montł	2.783	.005
9	HE_dbp1	Diastolic blood pressure (1st)	.186	9	HE_STRfh3	Diagnosis of cerebral stroke (siblings)	-2.775	.006
10	HE_DMfh2	Diagnosis of diabetes mellitus (mother)	.186	10	LQ4_01	Reason for limiation of activity: fracture, jc	2.758	.006
11	HE_fev1fvc	FEV1/FVC	.184	11	L_LN_FQ	Number of lunches	2.755	.006
12	BE3_75	High-strength physical activity	.184	12	BM8	Speaking problems	-2.735	.006
13	T_Q_OMDG	Diagnosis of otitis media	.184	13	LK_LB_CO	Knowing nutritional labelling	2.621	.009
14	DI3_dg	Diagnosis of cerebral stroke	.184	14	BD2	Age when start to drinking	2.568	.010
15	wt_pft	Pulmonary function test (weight)	.184	15	HE_fev1fvc	FEV1/FVC	-2.552	.011
16	HE_dbp3	Diastolic blood pressure (3rd)	.184	16	HE_HBsAg	Hepatitis B surface antigen	2.488	.013
17	BM13	Experience of dental damage	.184	17	HE_PFTdr	Diagnosis of COPD	-2.455	.014
18	BH9_11	Flu vaccine status	.183	18	wt_tot	Questionnaire, screening, nutrition (weigh	-2.454	.014
19	HE_fvcp	FVC	.183	19	DK8_pr	Presence of Hepatitis B	-2.231	.026
20	E_DO	Automatic refractometer test not possible	.182	20	LQ2_ab	Stay out of work for one month	-2.230	.026