

Binghamton University

The Open Repository @ Binghamton (The ORB)

Research Days Posters Spring 2020

Division of Research

2020

Drug Usage in Heart Disease as a Risk factors for Mental Illness using Machine Learning

Jay Sivakumar

Binghamton University--SUNY

Follow this and additional works at: https://orb.binghamton.edu/research_days_posters_spring2020

Recommended Citation

Sivakumar, Jay, "Drug Usage in Heart Disease as a Risk factors for Mental Illness using Machine Learning" (2020). *Research Days Posters Spring 2020*. 84.

https://orb.binghamton.edu/research_days_posters_spring2020/84

This Book is brought to you for free and open access by the Division of Research at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Research Days Posters Spring 2020 by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

Drug Usage in Heart Disease as a Risk factors for Mental Illness using Machine Learning

Jayanth Sivakumar¹, Lina Begdache², Daehan Won¹

¹Department of Systems Science and Industrial Engineering, Watson School of Engineering and Applied Science

²Health and Wellness Studies Department, Decker School of Nursing

Abstract

Patients with heart disease among mid-life and old age are at a higher risk for mental illness like depression, anxiety or schizophrenia. In this research, we investigate the role of cardiovascular disease as a risk factor for mental illness. Long term use of antibiotics, along with gender, age, BMI, hypertension and the usage of drugs like clarithromycin, Z-pak, and supplements such as folate, vitamin B6 are all a risk factors for predicting mental illness.

Objective

- ❑ This research focuses on identifying the mental illnesses for patients with heart disease using machine learning
- ❑ The research also focuses on identifying the effects of drug/supplement usage as a risk factor on mental illnesses

Background

- ❑ Coronary Heart Disease (CHD) and mental illness are one of the major reasons of mortality and morbidity among general population
- ❑ Person with Cardiovascular Disease (CVD) has more depression than general population [1]
- ❑ On the other hand, person with more depression are likely to develop CVD and has higher mortality than general population [2]
- ❑ Research show that mood and anxiety disorders are one of the major symptoms of causing cardiovascular disease [3,4]
- ❑ The goal of the following research is to predict Depression, Anxiety and Schizophrenia using this dataset

Dataset

- ❑ **Dataset was obtained from United Health Services.**
- ❑ The Dataset contains 68647 rows and 29 attributes
- ❑ 67% of the patients has no disease
- ❑ 33% of the patients has one or more mental health disease
- ❑ Attributes of dataset are as follows:
 - Gender, age, InsulinDependentDiabetes, Hypertension, Osteoarthritis, CancerMalignant, Obesity, Depression, Anxiety, Schizophrenia, CongenitalDiseaseOfHeart, HeartFailure, CerebrovascularDisease, Atherosclerosis, CoronaryArteryDisease, NutritionDeficiency, ElevatedCRP, ElevatedESR, LongTermUseOfAntibiotics, BMI, LAB, LabValue, E_Mycin, Clarithromycin, Z_Pak, Folate, VitB6, CoQ, Omega3FishOil
- ❑ **Disease** is a computed column
- ❑ **If** depression is yes, or anxiety is yes, or schizophrenia is yes, or all the mental illnesses are a yes **then Disease = Yes**
- ❑ **Class imbalance**
 - 15% depression cases
 - 22% anxiety cases
 - 0.38% schizophrenia cases
- ❑ Patients with more than one illness is very less (≈ 0.30%)

Method

- ❑ 97% of LAB and LabValue are missing data – removed from the analysis
- ❑ Compute Disease for each row
- ❑ 70% of the dataset is used for 5-fold CV for training the model and 30% of the dataset is used for testing the prediction model
- ❑ Methods applied:
 - **Method 1:** Whole dataset is processed without sampling
 - **Method 2:** One-hot encoded targets without sampling
 - **Method 3:** Under sampled dataset
 - Under sampling: The total sample size of non-mental illness cases are matched to number of total mental illness case
 - **Method 4:** Under sampled dataset split to Male/Female
 - **Method 5:** Deep learning method with 1 hidden layer (16 neurons)
- ❑ Classification Algorithms:
 - Random Forest
 - Decision Tree Classifier
 - XGBoost (Extreme Gradient Boosting)
 - LightGBM
 - Deep Learning
 - Deep learning method with 1 hidden layer (16 neurons)
 - 1 Input layer (24 nodes), 1 Hidden Layer(16 nodes), Dropout layer to prevent
 - overfitting (Early stopping criterion), 1 output layer

Metrics Used

- ❑ **Accuracy (Goal)**
 - Ratio of correctly classified illness/non-illness cases to total rows in the dataset
 - Higher Accuracy = Better Performance
- ❑ **Precision**
 - How many of the non-mental illness case labeled are actual non-illness?
- ❑ **Recall/Sensitivity**
 - How many non-mental illness cases were predicted to be non-illness
- ❑ **F1 Score**
 - Average of precision and recall
- ❑ **Specificity (Goal)**
 - Off all mental illness cases, how many did we correctly predict as mental illness
- ❑ **AUROC (Area Under the Receiver Operating Curve) (Goal)**
 - Tells us how well can the model differentiate between mental and non-mental illness cases
 - The higher the score, the model can predict and approximate well on new samples

Results

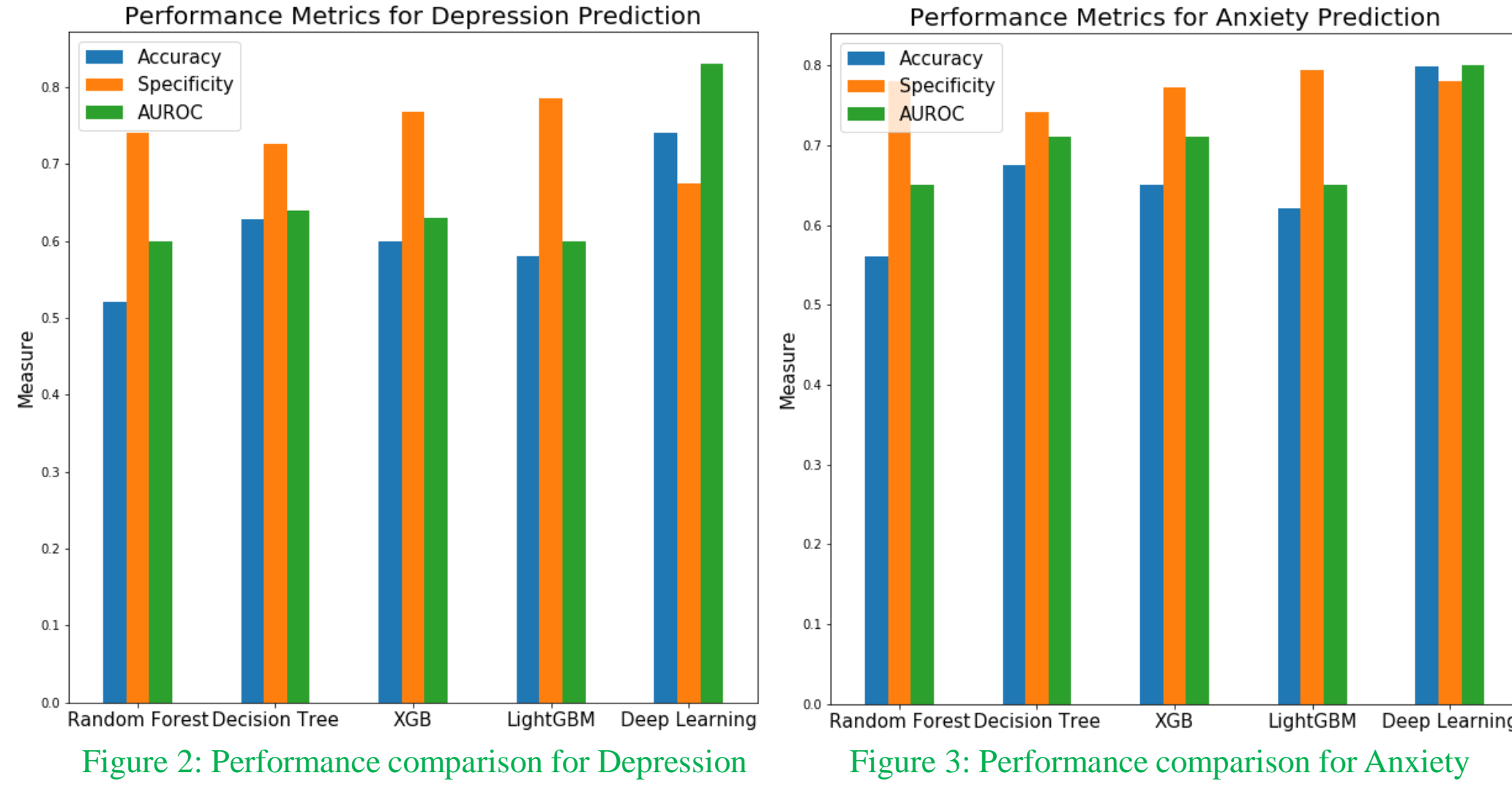


Figure 2: Performance comparison for Depression

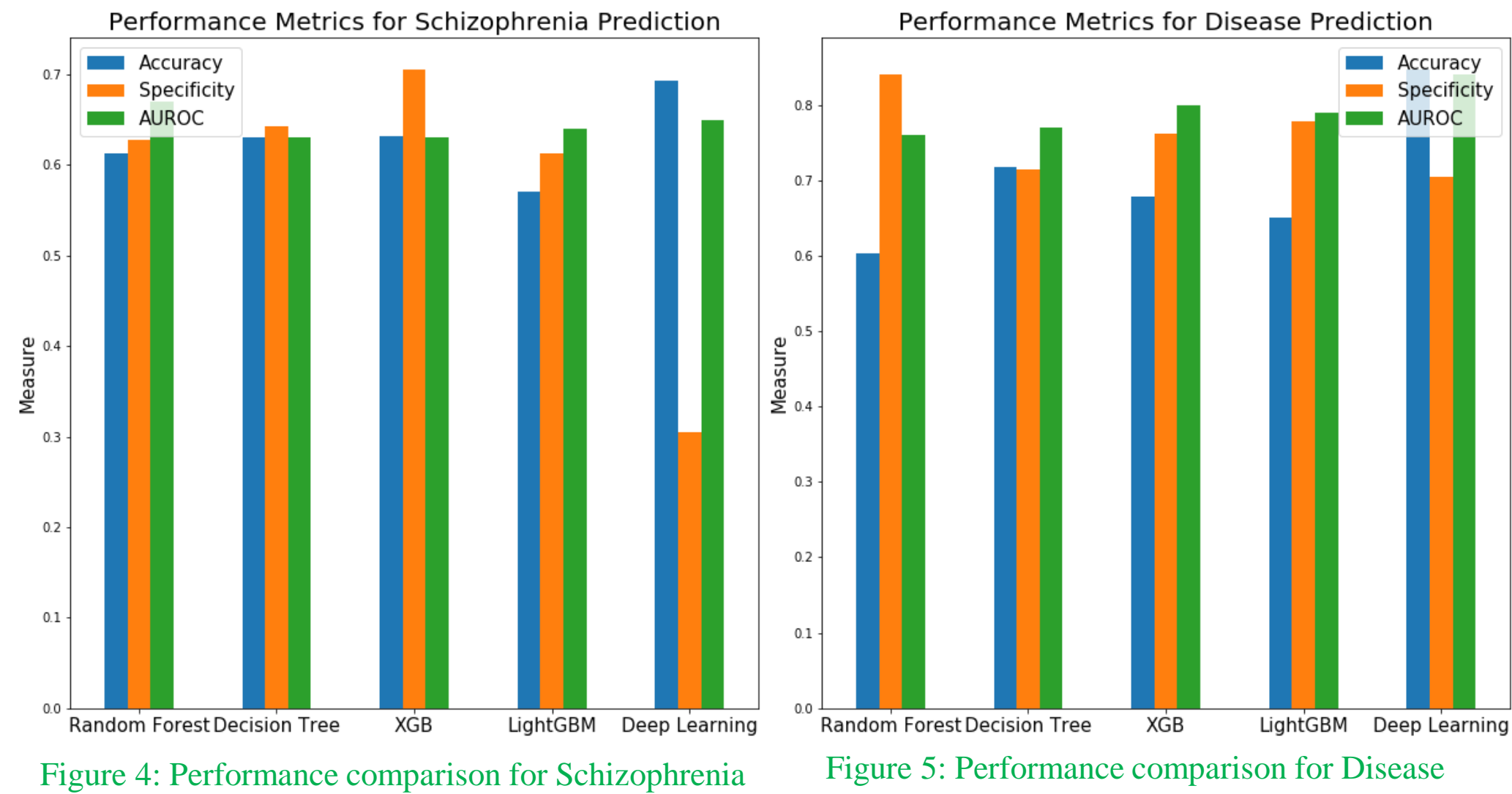


Figure 3: Performance comparison for Anxiety

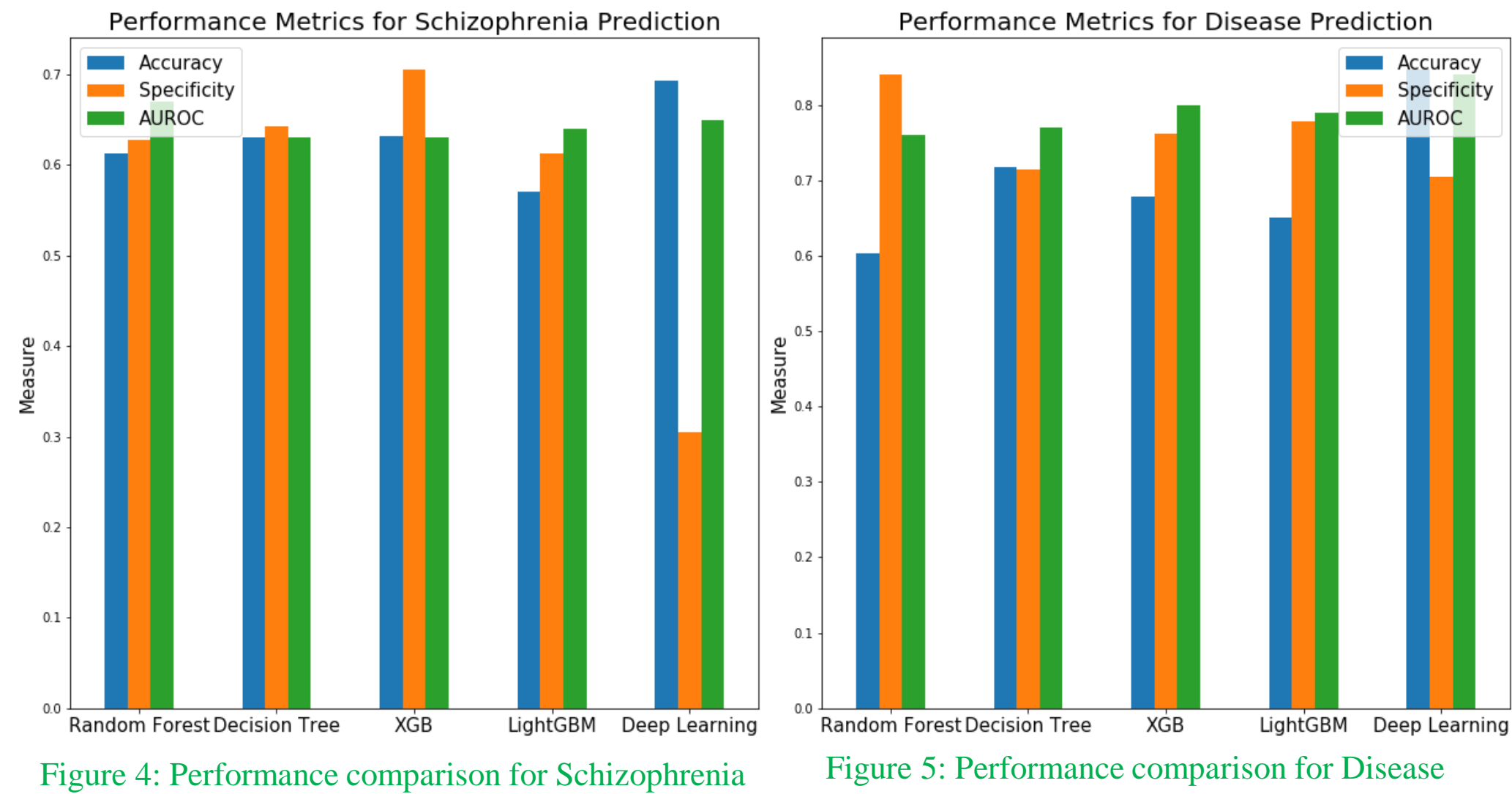


Figure 4: Performance comparison for Schizophrenia

- ❑ **Best model** with high accuracy, specificity and AUROC is selected
- ❑ Models that perform poorly in any of these measures are removed from best performing model
- ❑ **Deep learning** model showed high accuracy, specificity and AUROC – best performing model for **Depression, Anxiety and Disease**
- ❑ **XGB's** performance is better for **Schizophrenia**

Variable Selection

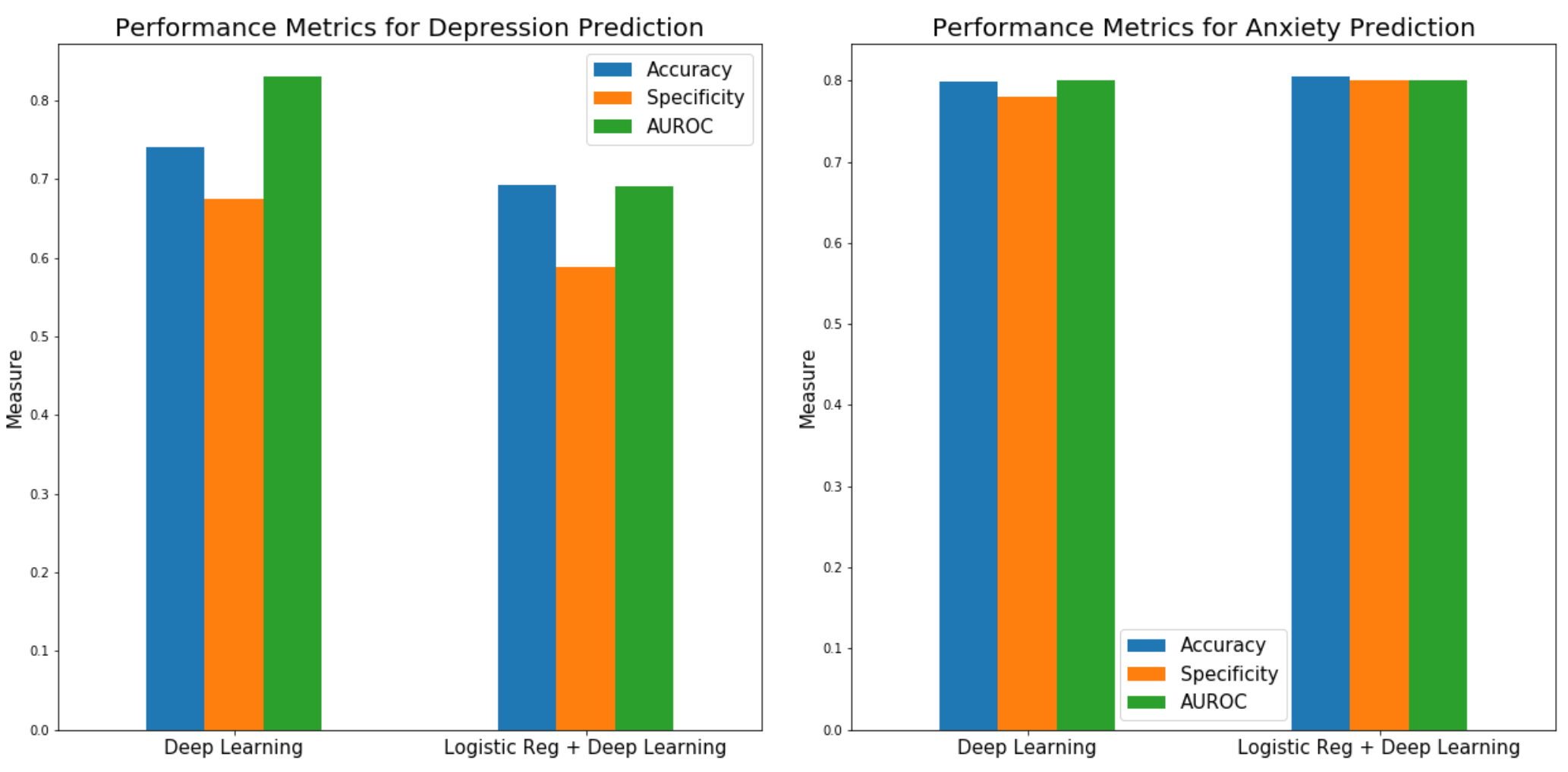


Figure 6: Performance comparison after variable selection for Depression

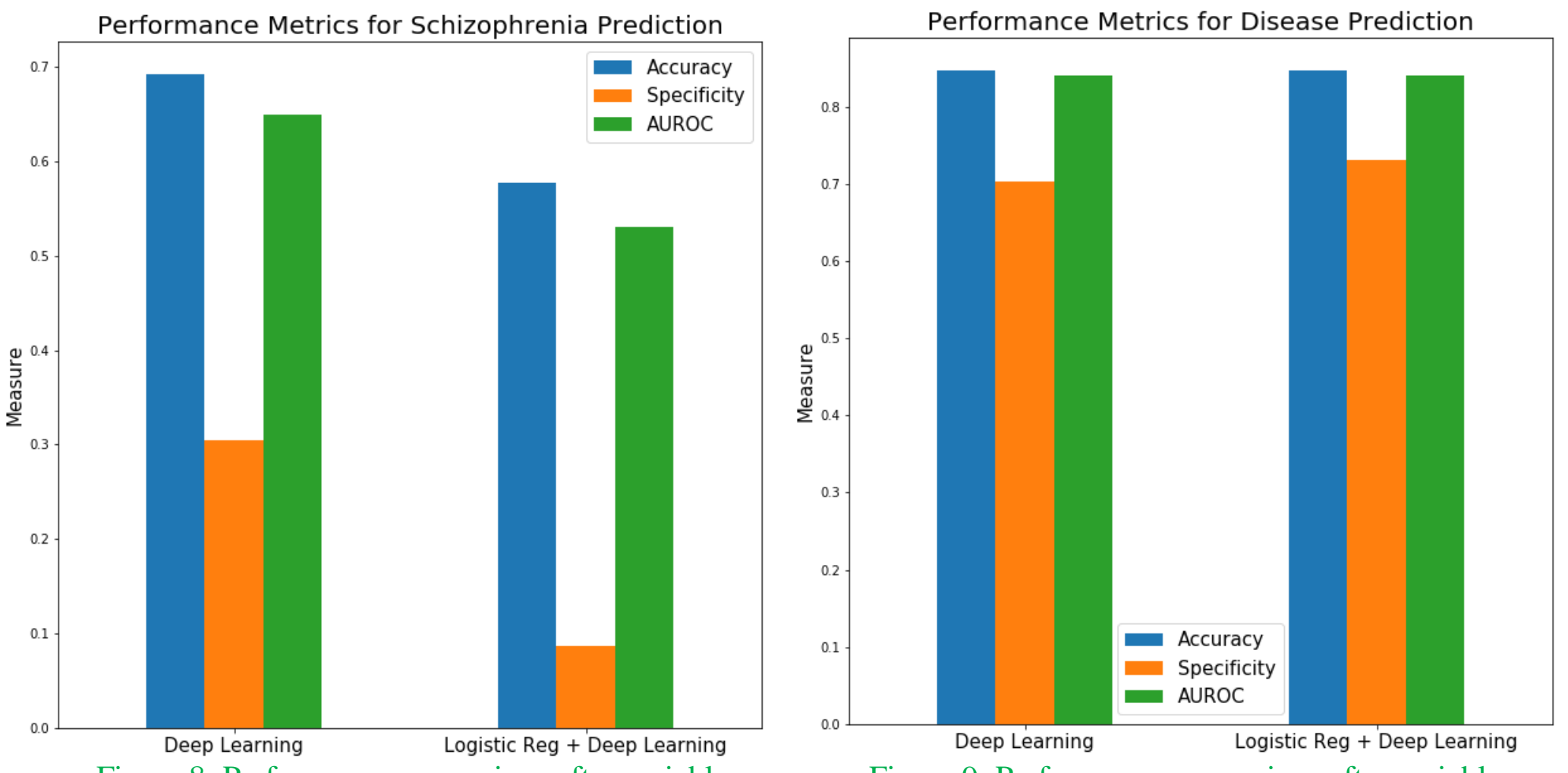


Figure 7: Performance comparison after variable selection for Anxiety

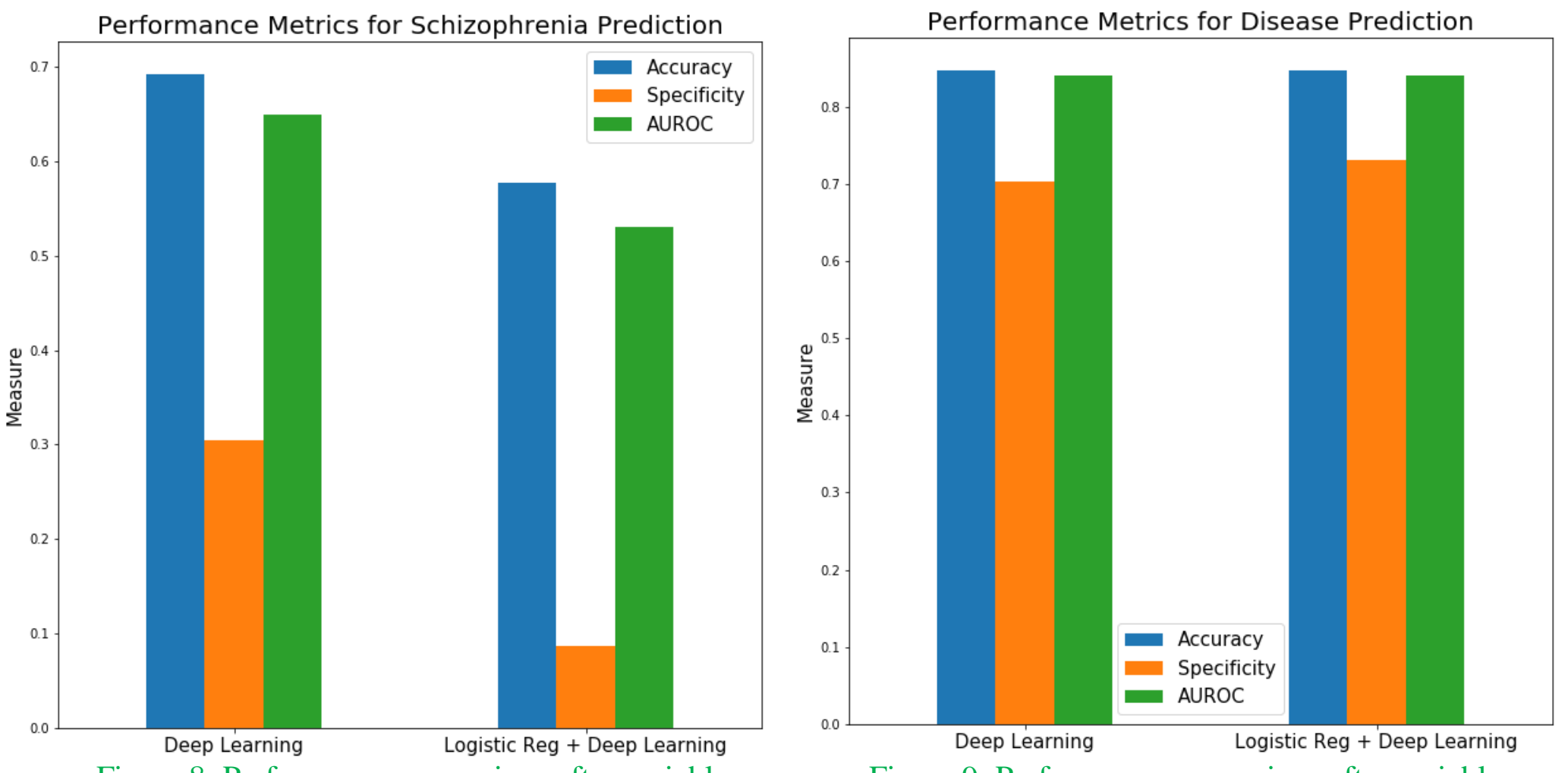


Figure 8: Performance comparison after variable selection for Schizophrenia



Figure 9: Performance comparison after variable selection for Disease

- ❑ Using logistic regression, statistically significant ($p<0.05$) variables are selected and trained using deep learning (Fig 6-9)
- ❑ **18 variables** of 26 predictors are enough to predict Anxiety and Disease
- ❑ **Highlighted** represent common features in all the mental illnesses
- | <u>Variables for Depression</u> | <u>Variables for Schizophrenia</u> |
|---------------------------------|------------------------------------|
| ➤ age | ➤ Gender |
| ➤ Hypertension | ➤ age |
| ➤ Osteoarthritis | ➤ Hypertension |
| ➤ CancerMalignant | ➤ Osteoarthritis |
| ➤ HeartFailure | ➤ Obesity |
| ➤ CoronaryArteryDisease | ➤ HeartFailure |
| ➤ ElevatedCRP | ➤ CoronaryArteryDisease |
| ➤ LongTermUseOfAntibiotics | ➤ BMI |
| ➤ Clarithromycin | ➤ Z_Pak |
| ➤ Z_Pak | |
| ➤ Folate | |
| ➤ VitB6 | |

- | <u>Variables for Anxiety</u> | <u>Variables for Disease</u> |
|------------------------------|------------------------------|
| ➤ Gender | ➤ Gender |
| ➤ age | ➤ age |
| ➤ InsulinDependentDiabetes | ➤ InsulinDependentDiabetes |
| ➤ Hypertension | ➤ Hypertension |
| ➤ Osteoarthritis | ➤ Osteoarthritis |
| ➤ CancerMalignant | ➤ CancerMalignant |
| ➤ Obesity | ➤ Obesity |
| ➤ CongenitalDiseaseOfHeart | ➤ CongenitalDiseaseOfHeart |
| ➤ HeartFailure | ➤ HeartFailure |
| ➤ Atherosclerosis | ➤ Atherosclerosis |
| ➤ CoronaryArteryDisease | ➤ CoronaryArteryDisease |
| ➤ LongTermUseOfAntibiotics | ➤ LongTermUseOfAntibiotics |
| ➤ BMI | ➤ BMI |
| ➤ E_Mycin | ➤ E_Mycin |
| ➤ Clarithromycin | ➤ Clarithromycin |
| ➤ Z_Pak | ➤ Z_Pak |
| ➤ Folate | ➤ Folate |
| ➤ VitB6 | ➤ VitB6 |

- ❑ These variables **may increase or decrease** the mental illnesses that are yet to be investigated

Conclusion

- ❑ **Risk Factors for mental illness:** Long term use of antibiotics, along with gender, age, BMI, hypertension and the usage of drugs like clarithromycin, Z-pak, folate, vitamin B6
- ❑ Using the selected features from the previous section, the below mentioned accuracy, specificity and AUROC was achieved for each mental illness
- ❑ **Depression** has 74% accuracy, 67% specificity and 83% AUROC using deep learning
- ❑ **Anxiety** has 80% accuracy, specificity and AUROC using logistic regression + deep learning
- ❑ **Schizophrenia** has 63% accuracy, 70% specificity and 63% AUROC using XGB
- ❑ **Disease** has 84% accuracy, 73% specificity and 84% AUROC using logistic regression + deep learning

Reference

- De Hert, M., Detraux, J., & Vancampfort, D. (2018). The intriguing relationship between coronary heart disease and mental disorders. *Dialogues in clinical neuroscience*, 20(1), 31.
- Hare, D. L., Toukhsati, S. R., Johansson, P., & Jaarsma, T. (2014). Depression and cardiovascular disease: a clinical review. *European heart journal*, 35(21), 1365-1372.
- Musselman, D. L., Evans, D. L., & Nemeroff, C. B. (1998). The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Archives of general psychiatry*, 55(7), 580-592.
- Kawachi, I., Sparrow, D., Vokonas, P. S., & Weiss, S. T. (1994). Symptoms of anxiety and risk of coronary heart disease. The Normative Aging Study. *Circulation*, 90(5), 2225-2229.