

Predictive Models in Mental Health: From Diagnosis to Treatment

Sandy H Huang¹, Paea LePendur¹, Srinivasan V Iyer¹, Ming Tai-Seale², David Carrell³,
Nigam H Shah¹

¹Stanford University, Stanford, CA ²Palo Alto Medical Foundation, Palo Alto, CA

³Group Health Research Institute, Seattle, WA

Abstract

We investigate the potential of using electronic health record (EHR) data for improving healthcare for mental illness. Our focus is depression, a prevalent disorder that is difficult to diagnose and treat. We developed and evaluated models that use EHR data for predicting the diagnosis of depression and levels of severity, as well as identifying moderators for personalized treatment. The models were trained and evaluated on two datasets: a set of 35,000 patients selected from a database of 1.2 million patients from the Palo Alto Medical Foundation (PAMF) and a set of patients treated for depression from the Group Health Research Institute (GHRI). Using both structured and unstructured EHR data as features, our models performed with good sensitivity and specificity for predicting onset and severity. Baseline severity was the strongest prognostic indicator for treatment response, and more work is required to identify prescriptive characteristics for personalizing treatment.

Introduction: Depression is a prevalent disorder affecting about 14% of individuals worldwide. Despite its prevalence, diagnosing and treating depression is a challenge; primary care physicians identify only about 50% of true depression cases [1]. The growing amount of data available from EHR systems has proven useful in building predictive models for other disorders, and there is a need for such models for depression. In addition, very few of these efforts utilize free-text derived features. We evaluate the effectiveness of utilizing both structured and unstructured EHR data to predict whether patients will be diagnosed with depression, to predict the severity of depression in patients, and to personalize treatment for patients [2].

Methods: We created a cohort of 5,000 depressed patients and 30,000 non-depressed patients from PAMF, matched on age and visit history, and trained a model (penalized logistic regression) to predict a diagnosis of depression up to 12 months prior to the actual diagnosis. We also trained two models using a dataset from GHRI of 7,000 patients treated for depression, that have been scored using the PHQ-9 both at treatment and after a 90-day follow-up. The first model (penalized logistic regression) predicts *minimal* (PHQ score of 0-4) vs. *severe* (PHQ 20-27) depression, and the other model (logistic regression) predicts improvement on follow-up based on moderator effects of patient characteristics on treatment modality (medication vs. psychotherapy). The features of all models include gender, age, average number of visits per year, ICD-9 codes, and disease and drug ingredient terms extracted from the clinical text. Each model is trained on a randomly selected 80% of the patients and tested on the remaining 20%.

Results: Our model for early detection achieves an area under the receiver operating characteristic curve (AUC) of 0.82 – 0.85 for 12 months and 6 months prior to diagnosis, with over 65% sensitivity at a specificity of 80%. Top features include fatigue, anxiety, insomnia, and gender. We achieve an average AUC of 0.73 for severity level and 0.70 for treatment effectiveness (Table 1). The strongest *prognostic* characteristic [2] is the baseline PHQ score. Possible *prescriptive* characteristics include dyschesia, Raynaud syndrome, and marital life events.

Discussion: The use of EHR data may inform the timely diagnosis and treatment of depression. The accuracy of our predictive model for diagnosis rivals that of primary care physicians – 50% sensitivity at a specificity of 80% [1]. A limitation of our work, however, is that misdiagnosis by primary care physicians impacts the quality of our gold standard. If baseline severity is a strong prognostic indicator, then models that correctly estimate severity and onset could be used to pool datasets that do not assign PHQ-9 scores to all patients. More work is required to match and control patients to eliminate possible confounders, given the observational nature of our datasets. Future work will involve developing and extending other types of models and pooling additional feature-rich datasets.

References

1. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;374:609-19.
2. Simon GE, Perlis RH. Personalized Medicine for Depression: Can we Match Patients with Treatments? *American Psychiatric Association*. 2010 Dec;167(12):1445-55.

Table 1 AUC's of the predictive models

Model	AUC
Early diagnosis	0.82 – 0.85
Severity level	0.72 – 0.73
Effectiveness of treatment	0.68 – 0.76