1-1-2019

Symptom Clustering And Decision Tree Analysis Within The Vast-D Randomized Clinical Trial

Yangmin Chen
yangmin.chen@yale.edu

Symptom Clustering and Decision Tree Analysis within the VAST-D Randomized Clinical Trial

Name:	Yangmin Chen
Year Completed:	2019
Year Degree Awarded:	2019
Degree Awarded:	Master of Public Health
Department:	School of Public Health
Advisor/Committee Chair:	Michael Wininger, PhD
Committee Members:	Tassos Kyriakides, PhD

ABSTRACT

As a secondary analysis of the VAST-D clinical trial data, we employed a multi-layered strategy to describe the complicated clinical features of Major Depressive Disorder (MDD) and the heterogeneity among depressive symptoms, using the following analytical approaches:

(1) Cluster analysis was used to transform a large and heterogenous mix of survey questions into a small number of correlated MDD symptom clusters: Four robust and highly-interpretable MDD symptom clusters (core emotional, appetite and weight, sleep disorders, atypical) were identified within the VAST-D trial, consistent with the findings from other relevant studies.

(2) Decision tree analysis was used to identify symptom thresholds with particularly effective discriminability in identifying remitters who were being treated with the three different study medications. Classification trees built for remission using a CART algorithm, were used for each of the three treatments and for the total cohort in the VAST-D study to facilitate:

(a) Generation of practical guidance that could be used to inform decision-making in real clinical settings;

(b) Identification of features for the sub-groups of patients showing low/high responses to each of the three treatments;

(c) Identification of the most important factors for remission through the use of random forests.

Key words: Major Depressive Disorder, symptom cluster, CART, decision tree, random forest, patient subtyping, Biostatistics

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my thesis advisor Dr. Michael Wininger for the continuous support of my thesis, and for his patience, motivation, enthusiasm and immense knowledge. The door to Michael's office was always open and his guidance helped me in all the time of research and writing of this thesis. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

My sincere thanks also go to Dr. Tassos Kyriakides as the second reader of this thesis. I am gratefully indebted to his insightful comments and warm encouragement on this thesis.

Besides my thesis committee, I would like to thank another supervisor of my summer internship, Mr. Gary Johnson, for offering me the internship opportunity and for all the support and guidance along the internship project as well as the thesis.

I would like to express my gratitude and respect to all the experts who were involved in the VAST-D clinical trial for their contribution and efforts in. My research would not have been possible without them.

I would especially like to thank my academic advisor Prof. Hongyu Zhao for guiding and supporting me over my master study. And my sincere thanks go to all the professors who had taught me here at Yale. I am very grateful to all of you.

I would like to thank all my fellow graduate students and friends. You are and always will be my friends.

Finally, I must express my very profound gratitude to my amazing parents for the love, unfailing support and continuous encouragement throughout my years of study. Thank you.

Table of Contents

List of Tables

List of Figures

I. BACKGROUND

A. Major Depressive Disorder

Major depressive disorder (MDD, "depression") espouses a spectrum of different symptoms spread across many axes of daily living, including mood, appetite, fatigue and socialization, each with their own range of severities.[1,2] MDD affects approximately 16% of the U.S. population at some point in their lives, however, only less than one-third of patients achieve remission with their first antidepressant.[3,4] It is believed that a combination of biological, psychological, genetic and social factors are the major causes in the onset of a depressive condition. However, the exact cause and pathophysiology of MDD is not yet understood, and as a result prescription of treatment regimen remains empiric.

Different rating scales have been developed to diagnose MDD and to measure symptom severity. All the commonly-used rating scales assess each of the nine DSM-IV-TR criterion symptom domains[5] (Sleep disturbance, Sad mood, Appetite/weight, Concentration, Self-criticism, Suicidal ideation, Interest, Energy/fatigue, Psychomotor agitation/ retardation) with varying designs. While the extant rating scales have demonstrated great facility in diagnosing MDD, there is great interest in studying the structure and relative importance of each of these surveys, particularly for synergizing plural information sources for the sake of prioritizing treatments.

B. The VAST-D Clinical Trial for treatment of MDD

Given the fact that only less than one-third of patients with MDD respond to their first antidepressant treatment and achieve remission, next-step treatments are in great demand for the large number of unresponsive patients. The VA Augmentation and Switching Treatments for Improving Depression Outcomes (VAST-D) – a multisite randomized, single-blind, parallel-group trial – was conducted to determine the relative effectiveness of three common alternate next-step treatments in patients whose MDD was unresponsive to prior antidepressant treatments. A total of 1,522 veterans (mean age, 54.4 years; men, 1296 [85.2%]) were randomized to one of the three interventions – Switch to a different antidepressant, bupropion (n = 511); augment current treatment with bupropion (n = 506); or augment with an atypical antipsychotic, aripiprazole (n = 505). After a 12-week follow-up period, 28.9% participants in the augment-aripiprazole group, 26.9% in the augment-bupropion group and 22.3% in the switch-to-bupropion group achieved remission respectively. The only significant, although modest, remission comparison was found between the augment-aripiprazole group and the switch-to-bupropion group. Full study design and results of the primary analysis are published elsewhere in detail.[6,7]

C. Thesis Objectives

Objective 1.a: To identify similar symptoms within Major Depressive Disorder by investigating its underlying grouping schemes, i.e. symptom clusters;

Objective 1.b: To compare the grouping schemes of baseline MDD symptoms identified in our study and in other relevant studies: whether the groupings are robust across different studies using different diagnostic tools and study populations.

Objective 2: To construct decision trees for remission for the three different study medications respectively. Based on the decision trees and relevant tree-based analyses, (a) to identify sub-groups of patients showing low or high responsiveness within each study arm, and to pinpoint key features with thresholds distinguishing different sub-groups; (b) to identify the most important factors for remission by random forests.

D. Overview of approach

We first applied hierarchical clustering, an unsupervised machine learning method, to identify baseline symptom clusters. The symptom clusters captured the most important and concise information of depressive symptoms with reduced dimensions. That is, symptom cluster scores were further derived as informative and succinct summary of diverse MDD symptoms, by averaging the scores of multiple symptoms within the same cluster. After combining the newly-derived symptom clusters information with other baseline data, we conducted decision tree analyses for each treatment group and for the study cohort as a whole: (a) The Classification and Regression Tree (CART) model was used to build single decision trees for remission for each treatment respectively. The graphical representation of possible remission status with certain conditions on patient features could be highly intuitive, especially in a clinical context.

(b) Random forest, an ensemble machine learning method that aggregates the results of multiple single decision trees with less overfitting tendency yet less intuitive results, was then applied as a complement to single trees. The *Importance of Variable* indices evaluated by random forest were used to identify the most important factors for remission. Finally, we offered new insights into underlying structures of MDD symptoms and sub-groups of patients or specific symptoms that show high- or low-likelihood of treatment response to anti-depressants in general.

a. Symptom Clustering

Given the clinical diversity of MDD symptoms, it is essential, to investigate underlying coherent clusters of these diverse symptoms. The reduced symptom dimensions would not only provide insight into the underlying nature of the complex disorder, but also assist a great many relevant next-step studies and make the findings more interpretable. Some approximately consistent clusters of depression symptoms were identified in recent studies with different rating scales and different statistical approaches[8-10]. Using data collected in the VAST-D study, we replicated these symptom clusters and evaluated the robustness of the clusters identified across different studies. Our study will be the first to merge data from two different instruments together (*PHQ-9* and *QIDS-C16*). By comparing the results from a merged dataset versus clusters obtained independently from a single instrument, we will provide valuable insight into the results obtained by others, specifically: do the different clusters obtained by various authors reflect their choices of diagnostic tool.

Factor analysis and clustering are two major statistical methods for investigating underlying grouping schemes and relationships of various symptoms. Compared to factor analysis that produces complicated structural relationships between individual symptoms and higher-level groups (called factors in factor analysis), hierarchical clustering produces more concise results where each individual symptom is only assigned to one single cluster. We applied hierarchical clustering in the study for its simplicity of interpreting and visualizing results, given that our main purpose is to investigate the underlying grouping schemes of complicated depressive symptoms.

To summarize, we first applied hierarchical clustering, an unsupervised machine learning method, to examine the baseline symptom clusters. Then we compared our symptom clusters to the ones identified in other studies.

b. Decision Tree Analyses

Due to the complexity of MDD and heterogeneity among depressive symptoms patients, some treatments tend to exhibit differential effectiveness with different patient groups. Many studies have been conducted to show promising evidence of subtypes of MDD based on biological variables or on clinical features[11], indicating potential personalized diagnostics and medication strategies of MDD.[12] However, research results diffuse slowly into clinical practice. In addition, remembering or even evaluating each treatment guideline on an individual symptom level is complicated and inefficient in clinical practice.

Decision trees are a popular supervised machine learning technique, especially useful for classification problems. Tree-based learning algorithms are adept at producing high classification accuracy with very concise representation of gathered knowledge.[13-15] The use of machine learning methods including decision tree techniques to assist in medical diagnosis, decision-making and prediction of medical and health conditions has received substantial attention from researchers in recent years.[16-19] Further, being non-parametric, tree based techniques afford great flexibility (i.e. no constraints on the data type, no assumptions about the space distribution and the classifier structure) compared to other conventional modeling methods.

With an interest in facilitating knowledge translation into clinical practice, our study implemented decision tree analyses to make highly intuitive and easy-to-implement guidelines for individualized treatment selections from the three alternative MDD anti-depressant treatment strategies in the VAST-D trial, and to identify sub-groups of patients or specific symptoms that show high- or low-likelihood of treatment response to each anti-depressant. To the best of our knowledge, ours is the first analytical approach that investigates MDD antidepressant therapy through the use of decision tree techniques.

II. METHODS

The study is a secondary analysis of the VAST-D randomized clinical trial. The primary objective of the VAST-D trial was to evaluate the effectiveness and safety of three treatment approaches for MDD.[6,7]

A. VAST-D: Protocol summary and Study Population

Participants were recruited from 35 VA medical centers. Patients were considered eligible for VAST-D if they had an MDD diagnosis, were non-responsive to at least one course of anti-depressant treatment, and capable and willing to provide informed consent. The criterion for study entry was a score of 16 or more (indicating severe depression) on the Quick Inventory of Depressive Symptomatology-Clinician Rated (QIDS-C16) questionnaire after at least 6 weeks of treatment, or a score of 11 or more (indicating moderately severe depression) after at least 8 weeks of treatment with the three most recent weeks at a stable "optimal" dose.

Patients were randomized to one of three treatments in a 1:1:1 ratio: switch to another antidepressant, namely bupropion sustained release (Sw; Randomized as Treatment A), augmentation of current treatment with bupropion sustained release (AB; Randomized as Treatment B), or augment current treatment with aripiprazole (AA; Randomized as Treatment C).

A total of 1,522 study participants completed the study protocol and their remission status on or before Week 12 was determined. The primary outcome was remission (1=remission, 0= no remission), defined as a QIDS-C16 score of 5 or less at 2 consecutive scheduled follow-up visits during the acute treatment phase.

B. Study Data

In the study, we included all the 1522 participants who completed the study protocol of the VAST-D trial. The outcome measure in our study is the remission status (1=remission, 0=no remission), which is the primary outcome defined in the original VAST-D study. For the variables, we considered available baseline information included demographic information, smoking history and frequency, depression related information, vital signs of health, medication use, psychopathology assessments, adverse effect assessments, quality-of-life assessments, and the PHQ-9 patient Depression Questionnaire. A full description of all variables and instruments included in the study is shown in **Appendix A**.

C. Statistical Analysis

a. Symptom Clustering

PHQ-9 and *QIDS-C16* are the two instruments used in this part of analysis (See **Appendix A** for instrument descriptions). To measure the severity of MDD symptoms, there is a one-to-one correspondence between the nine items of the *PHQ-9* questionnaire and the nine DSM-IV-TR domains of MDD symptoms. Further, there is a one/multiple-to-one correspondence between the sixteen items of the *QIDS-C16* questionnaire and the nine domains. The rating scales are the same for all items within both questionnaires, ranging from 0 (least severity) to 3 (most severity). For the MDD symptom domains (Sleep, Appetite/Weight, Psychomotor) which are each measured by multiple items in the *QIDS-C16* questionnaire, the domain score is the highest score among its related items.

To obtain a more high-resolution dataset of the MDD symptoms, we focused on the 16 MDD symptoms instead of the 9 symptom domains in the analysis. To obtain a more balanced

view of the severity of each symptom, we averaged the corresponding clinician-rated and self-rated symptom scores, of each patient. Based on the averaged scores, hierarchical clustering (*Distance measure: Euclidean distance; Cluster agglomeration method: Ward's method*) was applied to identify the underlying clusters of symptoms.

b. Decision Tree Analyses

1. CART Modeling

Using clinical and demographic data (see **Appendix A**), and the computed symptom cluster scores (average the symptom scores of each symptom cluster), classification and regression trees (CARTs) algorithms were applied to construct decision tree models. For each treatment group and for the total cohort, classification trees utilizing the CART algorithm were intended for the primary outcome (: 1 = remission, 0 = no remission).

A decision tree is a hierarchically organized structure, with each node partitioning the predictor space into disjoint subspaces based on value of a predictor. And same decisions/ predictions are made for all data points on the same predictor subspace.

The decision tree modeling mainly comprises two processes: splitting and pruning. The splitting process produces fully-grown trees utilizing the CART algorithm. The algorithm makes top-down recursive binary division of the predictor space into partitions. Each split is created after considering all the possible splits at each node by examining each predictor in turn. Then to choose the best split so that the resulting child nodes are the "purest", measured by the reduction in an impurity index (*Gini Index*) with respect to the response.

A pruning process follows the partitioning process to prevent potential overfitting issues by trimming the nodes of the tree in a bottom-up fashion: The fully-grown decision trees are further

pruned back based on a cost-complexity algorithm, producing smaller trees with better cross-validation properties.

2. Random Forest

Though highly interpretable, single trees are prone to over-fitting; thus, random forests were also evaluated to give more convincing results. Random forest is a versatile and powerful machine learning technique that mitigates possible overfitting problems of decision trees with robust results, by aggregating the results of a large number of uncorrelated decision trees (*number of bootstrapped trees in the study: 500*) into one final result. As the basic building block of a random forest, each classification tree is created by randomly selecting a pre-specified number of variables from all predictors in each splitting process without pruning (*Number of variables at each split in the study:* 9).

Moreover, as a powerful dimensionality reduction method, random forest is adept at handling large data set with higher dimensionality and identifying the most significant predictors. It generates an index for each predictor variable representing the relative importance of that variable, in terms *of Importance of Variable*. Based on the importance index of each variable, we investigated most important factors (Top 10) for remission, respectively for each treatment group and for the total cohort.

c. Software and packages

We conducted all the analyses in R software. The *"hclust"* package was used for hierarchical clustering for symptom cluster analyses. The *"rpart"* package was used for building CART trees and the *"randomForest"* package was used for random forest analysis for the decision tree analyses.

III. RESULTS AND DISCUSSION

A. Symptom Clustering

We found four highly-interpretable symptom clusters among the sixteen MDD symptoms that measure the nine symptom domains *(Figure 1)*. The four symptom clusters comprise core emotional symptom cluster (bad mood, concentration/decision making, loss of interest/involvement, feelings of worthless/self-outlook, energy/fatigability), appetite and weight symptom cluster (weight/appetite increases/decreases), symptom cluster of sleep disorders (sleep-onset Insomnia, early morning insomnia, mid-nocturnal insomnia, hypersomnia) and atypical symptom cluster (psychomotor agitation/slowing, suicidal ideation).

More importantly, our finding shared a great consistency with the MDD symptom groupings found in other studies, adding value to the confidence of the theory of MDD symptom

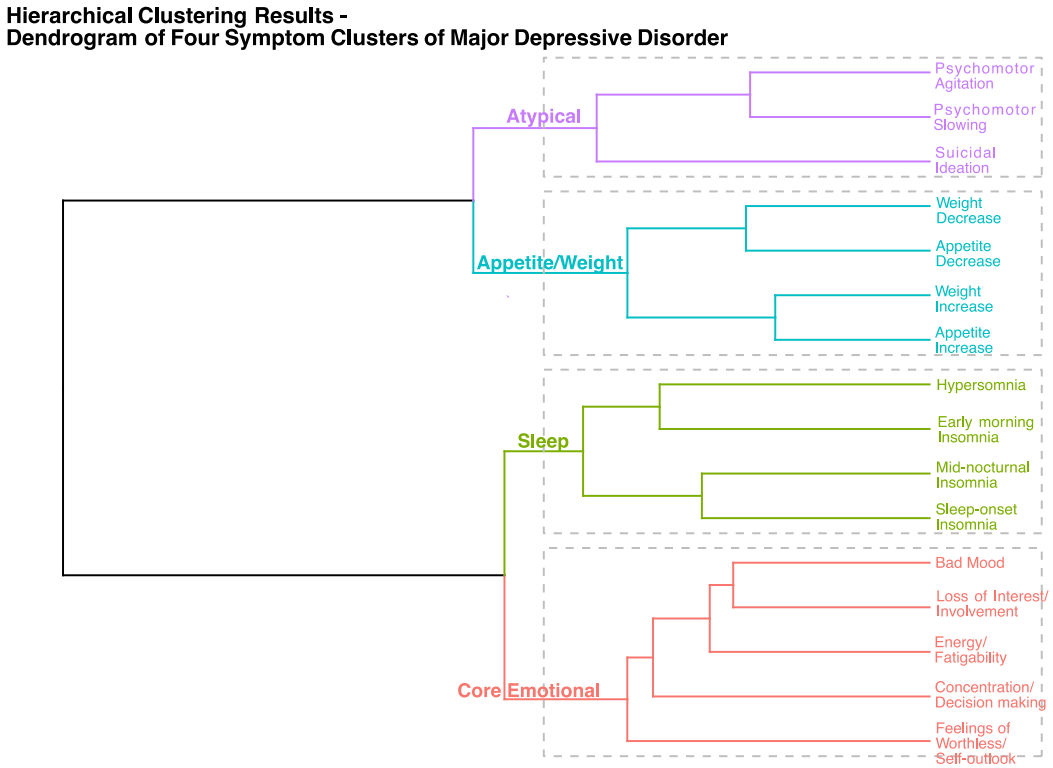


Figure 1 Dendrogram: Four MDD symptom clusters (baseline) identified by hierarchical clustering in the VAST-D study.

clusters (Table 2). Romera[8] performed factor analysis of the Zung self-rating depression scale (ZSDS) and found a clinical interpretable 4-factor structure – a core depressive factor, a cognitive factor, an anxiety factor and a somatic factor – respectively correspond to the core emotional, atypical, sleep and appetite/weight symptom clusters in our study. Li[9] found three meaningful factors reflecting weight/appetite disturbance, general depressive symptoms and sleep disturbance – respectively correspond to the appetite/weight, core emotional, symptom clusters in our study – by means of exploratory and confirmatory factor analysis in a large sample of 6008 depressed Han Chinese women. Chekroud[8] reported three robust symptom clusters - the sleep symptom cluster, core emotional symptom cluster and atypical cluster – validated with three data settings (*QID-SR* scale used in the STAR*D trial, *QID-SR* scale used in the CO-MED trial, *HAM-D* scale used in the STAR*D trial) by hierarchical clustering.

Similar grouping schemes of diverse MDD symptoms were obtained by various studies using different instruments and were reproduced within VAST-D in our study. Therefore, these MDD symptom clusters are suggested to be robust given different choices of instruments or diagnostic tools. Besides uncovering the underlying nature of MDD symptoms, these meaningful symptom clusters could fuel relevant studies by condensing the information and/or reducing the dimension of diverse MDD symptoms. They could also assist clinicians with more concise knowledge of the list of MDD symptoms collected by different questionnaires in the clinical context.

Table 1Summary of symptom clustering in three recent secondary analyses and our study

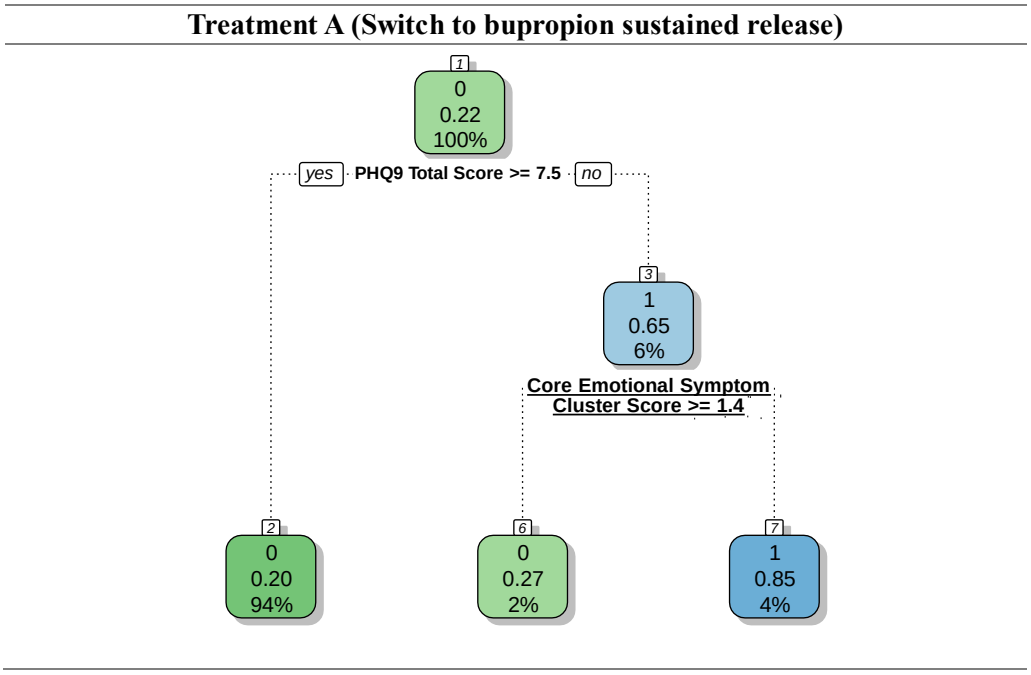
Our study	Romera, 2008[9]	Li, 2013[10]	Chekroud, 2017[8]		
VAST-D	Caballero, 2008[20]	CONVERGE[21]	STAR*D[23-25] and CO-MED[26,27]		
Agglomerative (bottom-up) hierarchical clustering	Unweighted exploratory factor analysis	Exploratory and confirmatory factor analysis	Agglomerative (bottom-up) hierarchical clustering		
PHQ-9 and QIDS-C16	ZSDS[22]	DSM-IV	QIDS-SR in STAR*D	QIDS-SR in CO-MED	HAM-D[28] in STAR*D
Core Emotional ▪ Bad mood ▪ Energy/fatigability ▪ Concentration/decision making ▪ Loss of interest ▪ Feelings of worthlessness/self-outlook	**Core Depressive** ▪ Depressed affect ▪ Crying spells ▪ Decreased libido ▪ Hopelessness ▪ Personal devaluation ▪ Emptiness ▪ Suicidal rumination ▪ Dissatisfaction	**General depressive** ▪ Depressed mood ▪ Anhedonia ▪ Psychomotor retardation ▪ Psychomotor agitation ▪ Loss of energy or fatigue ▪ Feeling of worthlessness	**Core Emotional** ▪ Bad mood ▪ Energy/fatigability ▪ Concentration/ Decision-making ▪ Loss of interest ▪ Feelings of worthlessness/ Self-outlook	**Core Emotional** ▪ Bad mood ▪ Energy/fatigability ▪ Concentration/ Decision-making ▪ Loss of interest ▪ Feelings of worthlessness/ Self-outlook	**Core Emotional** ▪ Somatic anxiety ▪ Psychological anxiety ▪ Guilt and delusions ▪ Sad mood ▪ Loss of interest
Sleep ▪ Mid-nocturnal insomnia ▪ Sleep-onset insomnia ▪ Early morning insomnia ▪ Hypersomnia	**Anxiety** ▪ Sleep disturbance ▪ Psychomotor agitation ▪ Irritability	**Sleep disturbance** ▪ Insomnia ▪ Hypersomnia	**Sleep** ▪ Mid-nocturnal insomnia ▪ Sleep-onset insomnia ▪ Early morning insomnia	**Sleep** ▪ Mid-nocturnal insomnia ▪ Sleep-onset insomnia ▪ Early morning insomnia	**Sleep** ▪ Mid-nocturnal insomnia ▪ Sleep-onset insomnia ▪ Early morning insomnia ▪ Energy/fatigability
Atypical ▪ Psychomotor retardation ▪ Psychomotor agitation ▪ Suicidal ideation	**Cognitive** ▪ Psychomotor retardation ▪ Fatigue ▪ Confusion ▪ Indecisiveness		**Atypical** ▪ Psychomotor retardation ▪ Psychomotor agitation ▪ Suicidal ideation ▪ Hypersomnia	**Atypical** ▪ Psychomotor retardation ▪ Psychomotor agitation ▪ Suicidal ideation ▪ Hypersomnia	**Atypical** ▪ Psychomotor slowing ▪ Psychomotor agitation ▪ Suicide ▪ Reduced libido ▪ Hypochondriasis
Appetite and weight ▪ Loss of appetite ▪ Loss of weight ▪ Increase of appetite ▪ Increase of weight	**Somatic factor** ▪ Decreased appetite ▪ Weight loss ▪ Tachycardia	**Weight/appetite** ▪ Loss of appetite ▪ Loss of weight ▪ Increase of appetite ▪ Increase of weight			

B. Decision Tree Analyses

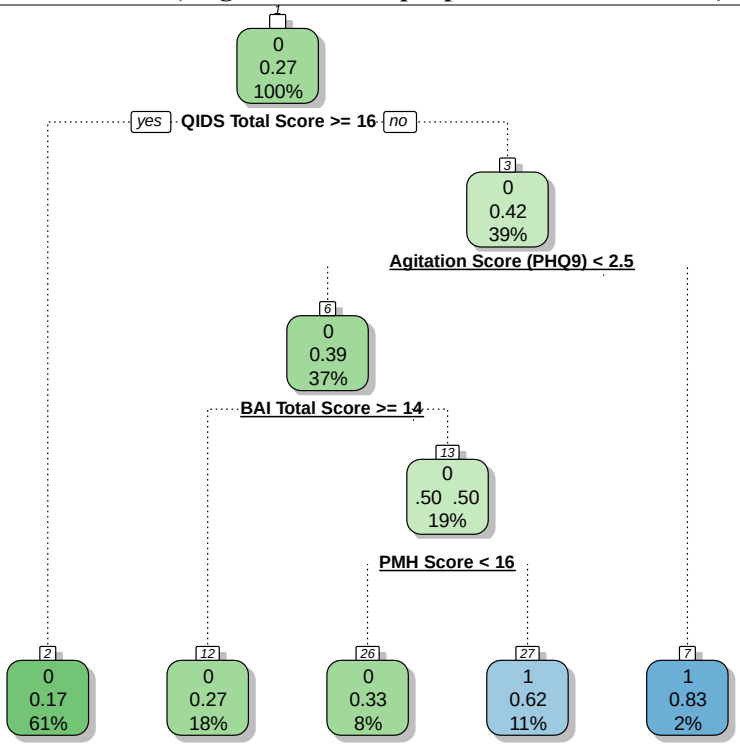
a. CART modeling

Classification decision trees were built based on the CART algorithm, respectively for the three treatment groups and for the total cohort. In each situation, the predicted outcome (remission or not) for a certain patient could be quickly obtained by going through a decision path of some certain simple conditions.

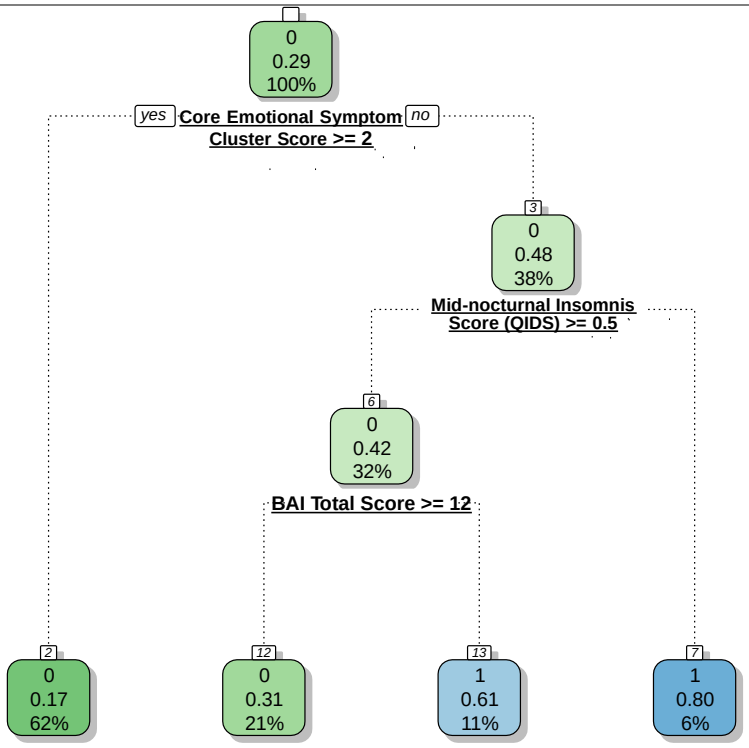
Figure 2 Classification trees for each of the three treatment groups in the VAST-D study (algorithm: CART). Within each node, the predicted outcome is in the first line: 0 – not remit, 1 – remit. The two decimals in the second line are the remission rate. Percentages in the third line is the proportion of participants in this node.



Treatment B (Augment with bupropion sustained release)



Treatment C (Augment with aripiprazole)



1. Personalized treatment selection from the three treatments in VAST-D

Here, we put forward an intuitive and easy to implement guideline for clinicians to make personalized treatment selection from the three MDD treatments in the VAST-D study, and the logic could be extended to other existing treatments if data available:

(1) For a given patient, a clinician could quickly go through the decision tree and record the predicted result (remission or not and/or probability of remission) under each of the treatment scenarios.

(2) Then summarize the predicted results over the three (or more) treatments and make the selection.

For example, a patient is predicted to remit for treatment A while not to remit for treatment B or C. Obviously, treatment A would be the suggested individualized treatment for that patient. Likewise, a treatment would be suggested if its predicted probability of remission is obviously higher than the predicted remission probabilities of other treatments.

Naturally, classifiers derived from group analyses yield probabilistic statements about a deterministic phenomenon: the patient will either remit or not, and the probability can only provide a statement about the likelihood of an individual outcome given a knowledge-based obtained from others. In these situations, the guideline is still valuable to simply and quickly exclude or screen out some treatments with based on the best available information. The final decision could then be made by incorporating clinician expertise and patient preferences.

2. Highly responsive/unresponsive sub-groups of MDD patients

Sub-groups of patients or specific symptoms that show high- or low-likelihood of treatment response to anti-depressants were identified by evaluating the terminal nodes of each classification tree using the VAST-D data.

For a decision tree, the feature space of all terminal nodes is mutually exclusive and collectively exhaustive, and each terminal node represents a group of patients with similar features indicated by its previous nodes. By evaluating the average remission rate and proportion of patients of that terminal node, we distinguished highly responsive/unresponsive sub-groups of MDD patients. Comparisons were based on the results of the primary analysis in the VAST-D study:

The remission rate of treatment A is 22.3%; The remission rate of treatment B is 26.9%; The remission rate of treatment C is 28.9%.

Table 2 Features of highly unresponsive subgroups of MDD patients within each treatment group (There is no highly unresponsive subgroup identified for treatment A)

Highly unresponsive sub-groups of MDD patients *for* Treatment B (Overall remission rate of treatment B: **26.9%**)	
Sub-group b1 (Sub-group remission rate **17%;** Proportion of the treatment group **61%**)	
Conditions	**Interpretation**
QIDS Total Score < 16	Higher levels of clinician-rated depressive symptoms in general
Highly unresponsive sub-groups of MDD patients *for* Treatment C (Overall remission rate of treatment C: **28.9%**)	
Sub-group c1 (Sub-group remission **17%;** Proportion of the treatment group **62%**)	
Conditions	**Interpretation**
Core Emotional Symptom Cluster Score >= 2	Higher levels of core emotional symptoms

Table 3 Features of Highly responsive subgroups of MDD patients within each treatment group

Highly responsive sub-groups of MDD patients *for* Treatment A (Overall remission rate of treatment A: **22.3%**)	
Sub-group A1 (Sub-group remission rate **85%;** Proportion of the treatment group **4%**)	
Conditions	**Interpretation**
PHQ-9 total Score < 7.5 & Core Emotional Symptom Cluster Score < 1.4	Lower levels of patient self-rated depressive symptoms in general & Lower levels of core emotional symptoms
Highly responsive sub-groups of MDD patients *for* Treatment B (Overall remission rate of treatment B: **26.9%**)	
Sub-group B1 (Sub-group Remission rate **83%;** Proportion of the treatment group **2%**)	
Conditions	**Interpretation**
QIDS Total Score < 16 & PHQ-9 Agitation Item Score >= 2.5	Lower levels of clinician-rated depressive symptoms in general & Higher levels of patient-rated psychomotor agitation
Sub-group B2 (Sub-group remission rate **62%;** Proportion of the treatment group **11%**)	
Conditions	**Interpretation**
QIDS Total Score < 16 & PHQ-9 Agitation Item Score < 2.5 & BAI Total Score < 12 & Positive Mental Health Score > 16	Lower levels of clinician-rated depressive symptoms in general & Lower levels of patient-rated psychomotor agitation & Lower levels of self-report anxiety in general & Higher levels of positive mental health
Highly responsive sub-groups of MDD patients *for* Treatment C (Overall remission rate of treatment C: **28.9%**)	
Sub-group C1 (Sub-group remission rate **80%;** Proportion of the treatment group **6%**)	
Conditions	**Interpretation**
Core Emotional Symptom Cluster Score < 2 & QIDS Mid-nocturnal Item Score < 0.5	Lower levels of core emotional symptoms & Lower levels of mid-nocturnal symptoms
Sub-group C2 (Sub-group remission rate **61%;** Proportion of the treatment group **11%**)	
Conditions	**Interpretation**
Core Emotional Symptom Cluster Score < 2 & BAI Total Score < 12 & QIDS Mid-nocturnal Item Score >= 0.5	Lower levels of core emotional symptoms & Lower levels of self-report anxiety in general & Higher levels of mid-nocturnal symptoms

b. Random Forest: Important factors for Remission

According to the *Importance of Variable* indices by random forests, we evaluated ten most important variables in predicting the binary outcome (remission), respectively for each treatment group and for the total cohort. *BAI (Beck Anxiety Inventory) total score, duration of trial* and *BMI* are the three most important factors for remission for treatment A; *BAI total score, duration of trial* and *age* are the three most important factors for remission for treatment B*; QIDS total score, BAI total score* and *core emotional symptom cluster score* are the three most important factors for remission for treatment C; *BAI total score, QIDS total score, BMI* are the three most important factors for remission in general.

An interesting finding is the consistency of significant variables across situations: The four situations have exactly the same ten most important factors for remission with different rankings. The great consistency may implicate most important factors for MDD remission, and fuel related studies concerning a dimensional reduction or model selection process.

In a clinical context, these top factors could assist clinicians to make rough but quick judgement of how probable the remission would occur for a patient in general and/or for each treatment.

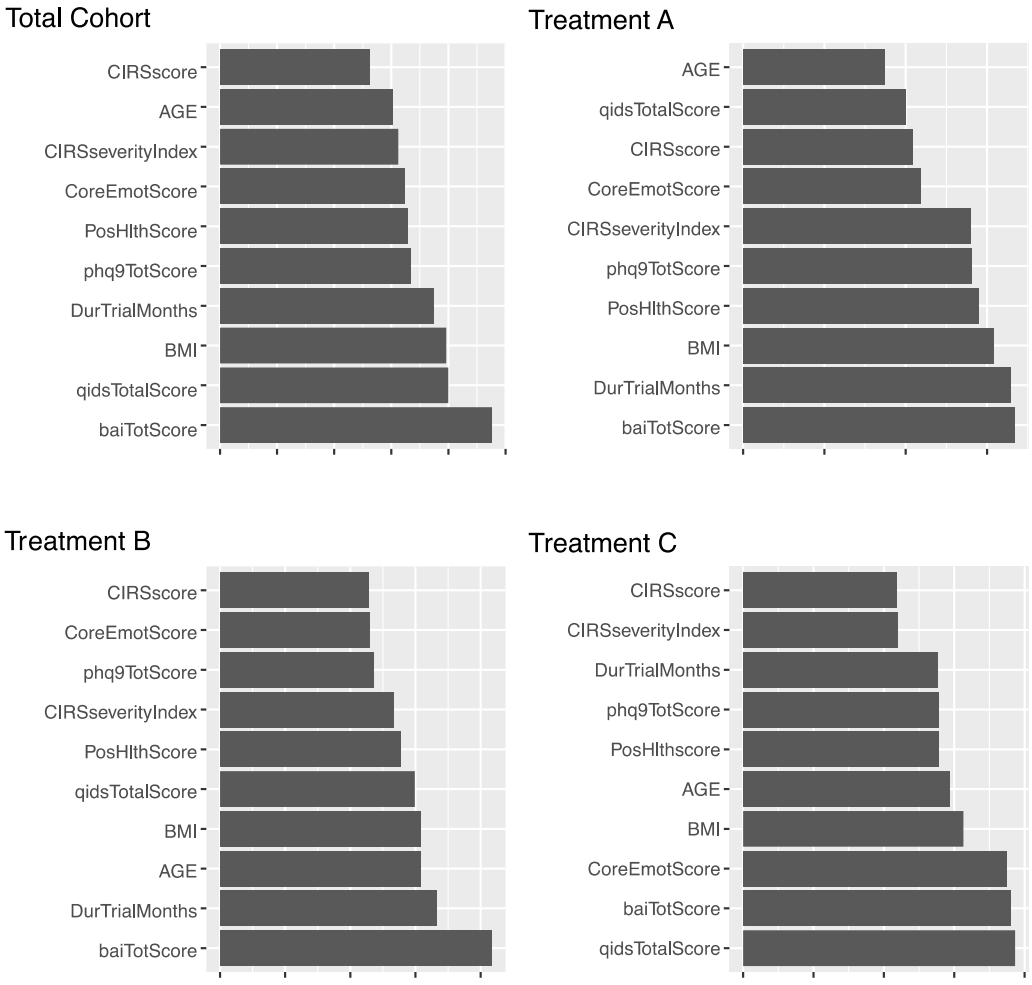


Figure 3 Random Forest Results – Ten most important factors for remission according to the Importance of Variable index

Table 4 Detail information of the ten significant factors for remission.

CoreEmotScore	Score of the core emotional symptom cluster
CIRSscore	Total score of the Cumulative Illness Rating Scale
phq9TotScore	Total score of the Patient Health Questionnaire
CIRSseverity_index	Severity index of the Cumulative Illness Rating Scale
PosHlthscore	Score of the Positive Mental Health instrument
AGE	Age of the patient
BMI	Body Mass Index (WeightLb/(HeightIn*HeightIn))*703
Dur_Trial_Months	Duration of index treatment trial (months)
qidsTotalscore	Total score of the 16-item Quick Inventory of Depressive Symptomatology
baiTotScore	Total Score of the Beck Anxiety Inventory

C. Limitations

The biggest limitation of the study is the inability to generalize the findings: First, potential solutions included in the proposed guidelines for the individualized treatment selection of MDD are limited to the three treatments in the VAST-D trial. However, it is likely that more alternate treatments are considered in a real situation. On the other hand, the guidelines could be easily extended to more solutions by the same framework if data available. Second, decision trees are prone to overfitting. That is, the trees may have better performance in predicting the outcome for the VAST-D trial data than for new data. Although the pruned trees with smaller number of layers in our study could reduce the likelihood of the problem theoretically, the robustness of the tree-based models is expected to be further tested and validated in other relevant studies. Third, older male VA patients predominated in the study participants (mean age, 54.4 years; men, 1296 [85.2%]). Therefore, whether the results could be further generalized to a broader population is unknown. To adjust the results for a broader population, analysis on different study populations using same or similar procedures is suggested. Another concern is that the VAST-D studied patients who had already failed at least one medication, and therefore, the results from our study may not be appropriate for newly diagnosed MDD patients. We expect to apply same or similar procedures proposed in our study on MDD patients with their first treatments. And if our technique could be extended to studies of first-line medication for MDD treatments, it would be a great opportunity to decrease the first-time failure rates.

Improvements could be made in the findings of most important factors for MDD remission identified in our study using random forests: Besides the relative importance of each factor, more informative results could be obtained if the magnitude and direction of influence on remission of each factor were complemented. Integrating the results of a logistic regression or cox regression

model on the same VAST-D study data is suggested in the next step analysis to: 1) provide supplemental information (magnitudes and signs of coefficients) that how the important variables identified by random forest influence the results (remission); 2) compare the significant variables from regression models (may after a model selection procedure) with the important variables from random forests and investigate the consistency and difference. And new insights are expected by comparing our results using random forests within VAST-D study to other studies with similar purposes based on different statistical methods and data setting.

IV. CONCLUSION

In the study, we first applied hierarchical clustering on the baseline data of the VAST-D trial and identified four highly-interpretable clusters for the sixteen MDD symptoms included in the PHQ-9 and QIDS-C16 questionnaires. The four symptom clusters (core emotional, appetite and weight, sleep disorders, atypical) share great consistency with the findings from other studies using other different MDD instruments and/or statistical methods. Therefore, our findings provide new evidence in support of the theory of symptom clusters of MDD by reproducing the grouping schemes of MDD symptoms in the PHQ-9 and QIDS-C16 instruments.

In support of knowledge translation and clinical application via MDD personalized treatment selection, we are the first to implement decision tree analytical techniques in this field and to propose an easy-to-implement and dependable guideline in a clinical context. Compared to most regression models for similar study objectives that produce significance levels and coefficients for each covariates, decision tree models are easy to understand and interpret for people with or without statistical background, and their graphical display could be easily interpreted and adopted in the clinical setting. By quickly going through the three decision trees

generated in the study based on the CART algorithm, an individualized treatment suggestion among the three alternative anti-depressants could be obtained for each MDD patient. Of course, the final treatment decision should consider other important factors including clinicians' experience and judgement and significant findings from other studies. Subgroups of patients with similar features that have specific high/low response to each treatment within the VAST-D trial were also evaluated from the decision trees.

Finally, we screened out ten most important factors from a great many factors for MDD remission using random forests. Together with the findings of the four robust symptom clusters, valuable insights into the underlying structure of complicated MDD symptoms and other related features are proposed that could assist in clinical management of patients.

V. Appendix A

Following patient-level data were included in the analysis:

- Patient study number, randomized group assignment and participating site number, duration of index treatment trial in months, outcomes (=1 remission, =0 not remission);
- Demographics: age, sex, marital status, education level, employment, race, ethnics; health status (BMI, alcohol/drug use history and frequency, total score and severity index of the *CIRS*);
- Psychopathology assessments include total scores of the *QIDS-C16, PHQ-9, PMH and BAI* instruments; indicators of recurrent mania or depressant, and PTSD; the *CGI-Severity* index;
- MDD symptom scores and nine DSM-IV-TR criterion symptom domains scores:

- For the *PHQ-9* instrument, the nine symptom scores (score of each item) are the nine corresponding symptom domain scores.
- For the *QIDS-C16* instrument, there are 16 MDD symptom scores (scores of the 16 items) and 9 symptom domain scores: If there is a one-to-one correspondence between the item and the MDD symptom domain, then the domain score is the item score, if there is a multiple-to-one correspondence between the items and the symptom domain, then the domain score is the highest score of the items of that domain)

Table for Appendix A. Brief description of measures used in the study

Instrument		Description
QIDS-C16[28]	16-item Quick Inventory of Depressive Symptomatology	A 16-item clinician-rated depression scale that adopts a 4-point scale: 0 – 3, higher scores indicate higher degree of severity of that depression symptom during the past 7 days. The 16 items have a one/multiple-to-one correspondence to the nine DSM-IV symptom criterion domains: Sleep disturbance domain - Initial, middle, and late insomnia or hypersomnia (Q1-Q4), Sad mood domain (Q5), Appetite/ weight domain - Decrease/increase in appetite/weight (Q6-Q9), Concentration domain (Q10), Self-criticism domain (Q11), Suicidal ideation domain (Q12), Interest domain (Q13), Energy/fatigue domain (Q14), Psychomotor domain – Psychomotor agitation/retardation (Q15-16). The score of each domain is the highest score of the items within that domain. And the total score of the instrument is the sum of all the nine domain scores, indicating the severity of depression.
PHQ-9[29]	Patient Health Questionnaire	A 9-item self-report depression scale that adopts a 4-point scale (0 – 3, higher scores indicate higher frequencies of being bothered by that item during the last 2 weeks), where each item corresponds to each of the nine DSM-IV-TR criterion symptom domains. The total score is the sum of the 9 items. Higher scores indicate higher degree of severity of MDD.

PMH[30]	Positive Mental Health Instrument	A 47-item instrument included six subscales: general coping (9 items), emotional support (7 items), spirituality (7 items), interpersonal skills (9 items), personal growth and autonomy (10 items), and global affect (5 items). Higher scores indicate higher PMH.
BAI[31]	Beck Anxiety Inventory	A 21-item self-report anxiety scale adopts a 4-point scale (0 = not at all, 1 = mildly, 2 = moderately, 3 = severely). The total score is the sum of the 21 items (0-21 = low anxiety, 22-35 = moderate anxiety, 36 or above = potentially concerning levels of anxiety). Higher scores indicate higher self-report measure of anxiety.
CGI-Severity[32]	Clinical Global Impressions Severity Index	A one-item clinician-rated index that evaluates the severity of psychopathology from 1(least severe) to 7 (most severe).
CIRS[33]	Cumulative Illness Rating Scale	One of the commonly used tools to measure comorbidity that measures the chronic medical illness burden with the severity of chronic diseases considered. Higher score indicates higher severity.

VI. REFERENCE

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. American Psychiatric Association; 2013. doi:10.1176/appi.books. 9780890425596
2. Angst J, Merikangas K. The depressive spectrum: diagnostic classification and course. J Affect Disord. 1997;45(1-2):31-39; discussion 39-40.
3. Kessler RC, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R) JAMA. 2003;289:3095–3105
4. Trivedi MH, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006;163:28–40.
5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994
6. Mohamed S, Johnson GR, Chen P, et al. Effect of Antidepressant Switching vs Augmentation on Remission Among Patients With Major Depressive Disorder Unresponsive to Antidepressant Treatment: The VAST-D Randomized Clinical Trial. JAMA. 2017;318(2):132. doi:10.1001/jama.2017.8036
7. Mohamed S, Johnson GR, Vertrees JE, et al. The VA augmentation and switching treatments for improving depression outcomes (VAST-D) study: Rationale and design considerations. *Psychiatry Res*. 2015;229(3):760-770. doi:10.1016/j.psychres.2015.08.005

8. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the Efficacy and Predictability of Antidepressant Treatments: A Symptom Clustering Approach. JAMA Psychiatry. 2017;74(4):370. doi:10.1001/jamapsychiatry. 2017.0025
9. Romera I, Delgado-Cohen H, Perez T, Caballero L, Gilaberte I. Factor analysis of the Zung self-rating depression scale in a large sample of patients with major depressive disorder in primary care. BMC Psychiatry. 2008;8:4. doi:10.1186/1471-244X-8-4
10. Li Y, Aggen S, Shi S, et al. The structure of the symptoms of major depression: exploratory and confirmatory factor analysis in depressed Han Chinese women. Psychol Med. 2014;44(07):1391-1401. doi:10.1017/S003329171300192X
11. van Loo, Hanna M et al. "Data-driven subtypes of major depressive disorder: a systematic review" BMC medicine vol. 10 156. 4 Dec. 2012, doi:10.1186/1741-7015-10-156
12. Miller, Diane B and James P O'Callaghan. "Personalized medicine in major depressive disorder -- opportunities and pitfalls" Metabolism: clinical and experimental vol. 62 Suppl 1,0 1 (2012): S34-9.
13. Barros, Rodrigo C., Basgalupp, M. P., Carvalho, A. C. P. L. F., Freitas, Alex A., A Survey of Evolutionary Algorithms for Decision-Tree Induction. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 42, n. 3, p. 291-312, May 2012.
14. M. Basgalupp, A. de Carvalho, R. C. Barros, D. Ruiz and A. Freitas "Lexicographic multi-objective evolutionary induction of decision trees", Int. J. Bio-Inspired Comput., vol. 1, no. 1–2, pp.105 -117 2009
15. Z. Fu, B. L. Golden, S. Lele, S. Raghavan and E. Wasil "Diversification for better classification trees", Comput. Oper. Res., vol. 33, no. 11, pp.3185 -3202 2006
16. Podgorelec, V., Kokol, P., Stiglic, B. et al. Journal of Medical Systems (2002) 26: 445. https://doi.org/10.1023/A:1016409317640
17. Vikas Chaurasia, Saurabh Pal," Early Prediction of Heart Diseases Using Data Mining Techniques, Carib.j.SciTech, Vol.1, 208-217, 2013
18. C. Bratu, C. Savin and R. Potolea "A hybrid algorithm for medical diagnosis", Proc. Int. Conf. Comput. Tool (EUROCON), pp.668 -673 2007
19. Babic, S.H., Kokol, P., Stiglic, M.M., Fuzzy decision trees in the support of breastfeeding, Proceedings of the 13th IEEE Symposium on Computer-Based Medical Systems CBMS'2000, pp. 7-11, 2000.
20. Caballero L, Aragonès E, García-Campayo J, Rodriguez-Artalejo F, Ayuso-Mateos JL, olavieja MJ. Cross-Sectional Study of the Prevalence, Characteristics, and Attribution of Somatic Symptoms in Patients with Major Depressive Disorder Seeking Primary Health Care in Spain. Psychosomatics. 2007
21. Flint J, Chen Y, Shi S, Kendler KS (2012). Epilogue: lessons from the CONVERGE study of major depressive disorder in China. Journal of Affective Disorders 140, 1–5
22. Zung W, Durham NC. A Self- Rating Depression Scale. Archives of General Psychiatry. 1965;12:63–70.
23. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006;163(1):28-40.
24. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR*D Project results: a comprehensive review of findings. Curr Psychiatry Rep. 2007;9(6):449-459.

25. clinicaltrials.gov. Sequenced Treatment Alternatives to Relieve Depression (STAR*D). NCT00021528. https://clinicaltrials.gov/ct2/show/NCT00021528. Accessed January 13, 2017.
26. Rush AJ, Trivedi MH, Stewart JW, et al. Combining medications to enhance depression outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. Am J Psychiatry. 2011;168(7):689-701.
27. clinicalTrials.gov. Combining Medications to Enhance Depression Outcomes (CO-MED). NCT00590863. https://clinicaltrials.gov/ct2/show/NCT00590863. Accessed January 13, 2017.
28. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56-62.
29. Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. K., ... Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological Psychiatry, 54, 573-583.
30. Kroenke K, Spitzer R, Williams W. The PHQ-9: Validity of a brief depression severity measure. JGIM, 2001, 16:606-616
31. Vaingankar JA, Subramaniam M, Chong SA, et al. The positive mental health instrument: development and validation of a culturally relevant scale in a multi-ethnic Asian population. Health Qual Life Outcomes. 2011;9:92. Published 2011 Oct 31. doi:10.1186/1477-7525-9-92
32. Beck, A.T., Epstein, N., Brown, G., & Steer, R.A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology, 56, 893-897.
33. Guy, W. (2000). "Clinical Global Impressions (CGI) Scale, Modified". In Rush, John A.; Task Force for the Handbook of Psychiatric Measures. Handbook of Psychiatric Measures (1st ed.). Washington, DC: American Psychiatric Association. ISBN 978-0-89042-415-5. OCLC 43483679
34. B.S. Linn, M.W. Linn, L. Gurel. Cumulative illness rating scale. J Am Geriatr Soc, 16 (1968), pp. 622-626