

# A First Step towards a Clinical Decision Support System for Post-traumatic Stress Disorders

Sisi Ma, PhD, Isaac R Galatzer-Levy, PhD, Xuya Wang, MS,  
David Fenyö, PhD, Arieh Y Shalev, PhD  
New York University School of Medicine, NY, NY, USA

## Abstract

*PTSD is distressful and debilitating, following a non-remitting course in about 10% to 20% of trauma survivors. Numerous risk indicators of PTSD have been identified, but individual level prediction remains elusive. As an effort to bridge the gap between scientific discovery and practical application, we designed and implemented a clinical decision support pipeline to provide clinically relevant recommendation for trauma survivors. To meet the specific challenge of early prediction, this work uses data obtained within ten days of a traumatic event. The pipeline creates personalized predictive model for each individual, and computes quality metrics for each predictive model. Clinical recommendations are made based on both the prediction of the model and its quality, thus avoiding making potentially detrimental recommendations based on insufficient information or suboptimal model. The current pipeline outperforms the acute stress disorder, a commonly used clinical risk factor for PTSD development, both in terms of sensitivity and specificity.*

## Introduction

Chronic PTSD is distressful and debilitating<sup>1</sup>, following a non-remitting course in about 10% to 20% of trauma survivors<sup>2,3</sup>. The early response may provide sufficient information to identify individuals at high risk<sup>4,5</sup> as multiple studies to date have identified numerous risk indicators of chronic PTSD, many of which are retrievable shortly after trauma exposure including early symptoms of PTSD, depression or dissociation, physiological arousal (e.g., heart rate), early neuroendocrine responses, gender, lower socio-economic status, the early use of opiate analgesics, the occurrence of traumatic brain injury, and a progressively growing number of genetic and transcriptional factors<sup>6-10</sup>. Despite these discoveries, the individual prediction of PTSD remains elusive, thereby leaving a major gap between scientific discovery and practical application.

One reason for such a gap is the current use of computational models that do not match the disorder's inherent complexity in etiology. As attested by its numerous risk factors, the etiology of PTSD is multi-causal, multi-modal and complex. As such, the longitudinal course of PTSD reflects a converging interaction of numerous, multimodal risk factors. Moreover, specific risk markers and their relative weight can vary between individuals and traumatic circumstances. For example, head injury increases the likelihood of developing PTSD<sup>11</sup> but does not occur to many survivors. Similarly, the contribution of female gender to the risk of developing PTSD varies between traumatic events<sup>12</sup> and with specific genetic risk alleles<sup>13</sup>. One other challenge faced by practitioners is that the risk of developing PTSD often has to be determined based on incomplete data, due to the constraint of clinical resources. Thus, to accurately predict PTSD in individuals, one must account for complex and variable interactions between putative markers and develop a method that is robust to missing data.

Machine learning techniques are well suited for knowledge discovery and outcome prediction for diseases that have complex etiology and multifaceted manifestation like PTSD. Machine learning methods, especially supervised learning methods, can discover structures and interactions that are informative for prediction from high dimensional, multi-modular data<sup>14-16</sup>. However, the implementation of machine learning methods to clinical observations for individual level prognostics and diagnostics is limited in psychiatry.

In the current study, as an effort to bridge the gap between the existing academic knowledge about PTSD and individual level diagnostics and prognostics, we designed and implemented a clinical decision support pipeline to provide clinically relevant recommendation for trauma survivors. To meet the specific challenge of early prediction, this work uses data obtained within ten days of a traumatic event. The current pipeline outperforms the acute stress disorder, a commonly used clinical diagnostic criterion, both in terms of sensitivity and specificity. The decision support pipeline is interactive in nature. When the quality of the predictive model is unsatisfactory, the pipeline recommends more information to be collected for re-evaluation rather than making a potentially detrimental decision based on insufficient data.

## Methods

### Data

This study used data collected for the Jerusalem Trauma Outreach and Prevention Study (J-TOPS<sup>2, 17</sup>; ClinicalTrial.Gov identifier: NCT0014690). Participants were adults admitted to the emergency department (ED) following potentially traumatic events. Eligible participants were screened by short telephone interviews, and those with confirmed PTEs as per DSM-IV PTSD criteria A1 and A2 received structured, telephone-based interviews approximately ten days ( $9.6 \pm 3.9$  days) after trauma exposure. Participants with acute PTSD symptoms in the first assessment were additionally invited for clinical interviews,  $29.5 \pm 4.6$  days after trauma exposure. Participants of the first clinical assessment were re-evaluated five, nine, and fifteen months after the traumatic event. For detailed procedures, see Shalev et al<sup>17</sup>. Participants provided oral and written consent for all phases of the study. The study's procedures were approved and monitored by the Hadassah University Hospital's Institutional Review Board.

For the purpose of this study, we included individuals who had initial data available at ten days and at least two additional time points ( $n = 957$ ). The initial traumatic event exposure included motor vehicle accidents (84.1%), terrorist attacks (9.4%), work accidents (4.4%) and other accidents (2.0%). Participants included in this study did not differ from those who were not included in gender distribution, age, general distress, initial PTSD symptoms and the frequency of exposure to new traumatic events during the study<sup>2</sup>.

To characterize the stress response after the exposure to the traumatic events, PTSD symptom trajectory was constructed with latent growth mixture modeling (LGMM) on PTSD symptom severity measured at different times after the trauma (see figure 1 in reference 5). The LGMM identified three distinct stress response types/trajectories: non-remitting (17% of the participants), slow remitting (27% of the participants), and rapid remitting (56% of the participants). The goal of the study is to predict whether an individual will display a non-remitting stress response or a remitting stress response and recommend appropriate course of action using data collected in the ED and 10 days following the trauma.

To predict whether an individual will display a non-remitting stress response, 20 features measured in the ED and at 10 days after the trauma were selected as predictors. These variables were selected because they were shown to contain predictive information (measured by the area under the ROC curve) regarding PTSD symptom trajectory in a previous study<sup>5</sup>. All the selected features can be measured at low cost during emergency room visits and through follow-up interview by phone, therefore, ideal for a clinical decision support system. The features used in the current study are listed in table 1.

Table1. Features used to construct predictive models.

Measurement time	Feature Name	Mean(SD)/%
Emergency Department	Amount of Time in the ER (hrs)	3.51(5.19)
	Age (years)	36.23(11.92)
	Loss of Consciousness	9.62%
	Tranquilizers	1.96%
	Head Injury	29.61%
	Motor Vehicle Accident	84.10%
10 days post trauma	Pain	6.49(2.59)
	Patient: Clinical Global Impression	3.96(1.73)
	Clinician: Clinical Global Impression	3.74(1.4)
	Total Depression Score	16.61(5.63)
	Total PTSD Score	9.41(3.63)
	Nightmares	35.44%
	Avoid thinking about reminders	72.62%
	Social Support	71.97%
	Derealization	33.75%
	Scared	74.32%
	Trouble Concentrating	59.05%
	Reexperience of the event	39.59%
	Want Clinical Support	50.51%
	Worthlessness	2.03(1.59)

### *The Decision Support Pipeline*

The goal of the proposed decision support pipeline is to assist practitioners in assigning proper treatment options for individuals who are recently exposed to a traumatic event. More specifically, upon receiving relevant information collected from an individual, the proposed decision support system would provide one of the following three recommendations: (r1) High clinical risk, recommend clinical monitoring or treatment. (r2) low clinical risk, no clinical monitoring or treatment recommended (r3) insufficient information to determine clinical risk, recommend more information to be collected.

The decision support pipeline consists of three components, illustrated in figure 1. The first component is a database containing data collected from individuals who were exposed to a traumatic event and followed for a period of 14 month such that their long-term stress response trajectory could be determined. In our implementation, we used data collected for the Jerusalem Trauma Outreach and Prevention Study (J-TOPS) as the database. Since the goal is the early detection of high risk individuals for PTSD, we included data collected in the emergency department and within 10 days after the trauma. The second component of the decision support pipeline is a prediction algorithm that models the relationship between the stress response trajectory and other information (demographic, clinical, psychiatric, etc.) in the database. The prediction algorithm produces a predictive model as well as quality metrics associated with the predictive model. In our implementation, upon receiving the information (features) about an individual, the database was queried such that the data for all individuals that have the same features measured were returned. The predictive model was built on this subset of the database. We used the support vector machine with radial basis function kernel<sup>18-20</sup> as the predictive algorithm, and sensitivity and specificity computed from leave-one-out cross validation on the validation set as the quality metrics of the model. We consider predicting the individual would display remitting stress response when he/she would actually display non-remitting stress response ten times as costly as predicting the individual would display non-remitting stress response when he/she would actually display remitting stress response. This cost information was taken into account by the prediction algorithm when building the predictive model. The third component of the decision support pipeline is a recommendation generator. The recommendation generator generates recommendations based on the prediction and quality metric(s) computed by the predictive model. In our implementation, we used a simple decision rule: if the leave-one-out cross validation sensitivity (the first quality metric) is greater than or equal to  $s_1$  and the leave-one-out cross validation specificity (the second quality metric) is greater than or equal to  $s_2$ , make the recommendation according to the prediction made by the predictive model, otherwise make the recommendation that more information is needed to determine the stress response. . The quality metrics and the decision rule are critical in our pipeline. The quality metrics estimate the performance of the predictive model built for individual patients (consists of different features). And, the decision rule produces clinical recommendations only when the predictive model is determined to be high quality. This allows the algorithm to recommend potentially detrimental clinical decisions based on predictions that have low quality. The decision support pipeline was implemented in Python.

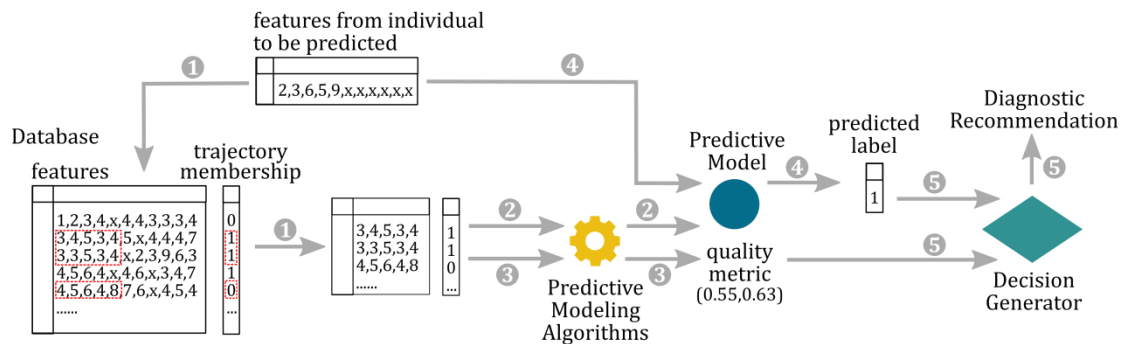


Figure 1: The workflow of the decision support pipeline. 1) A subset of data from the database is retrieved according to what features are available for the individual to be predicted. 2) A predictive model is built on the subset of the data by the predictive modeling algorithm. 3) Quality metrics is computed for the predictive model. 4) PTSD trajectory membership is predicted by the predictive model. 5) The decision generator generates recommendation based on the quality metric and the predicted trajectory.

### Evaluation of the Decision Support Pipeline

Leave-one-out cross validation was implemented to evaluate the performance of the decision support pipeline, i.e. each individual's PTSD trajectory membership was predicted using the model built with data collected from the rest of the individuals as the database. Three metrics were computed to evaluate the performance of the pipeline: (1) Percent of decisive decision: the percent of individuals that the pipeline is able to make a decisive decision on, i.e. recommending (r1) or (r2); (2) Sensitivity; and (3) Specificity. We evaluated the performance of the pipeline with 143 different sets of thresholds for the quality metrics (leave-one-out cross validation sensitivity and specificity computed with the data in the database).

We also calculated the predictive performance of acute stress disorder (ASD, defined according to the DSM-IV<sup>21</sup>) measured around 10 days after trauma exposure. ASD is commonly used in the clinical setting to identify people who are at high risk for PTSD development in the initial period of time following trauma exposure.<sup>22</sup> ASD is a binary variable computed from ASD symptom measurements. The predictive performance of ASD was compared to the predictive performance of our pipeline. Sensitivity and specificity obtained from leave-one-out cross validation computed from the entire cohort was reported.

## Results

### Predictive Performance of ASD

In the current patient cohort, 17% of the individuals developed non-remitting stress response. Predicting the post-traumatic stress response trajectory membership using 10-days' ASD results in a sensitivity of 0.527 and specificity of 0.695.

### Predictive Performance of the Decision Support Pipeline

The predictive performance of the decision support pipeline was evaluated with 143 sets of different threshold values (leave-one-out cross validation sensitivity and specificity computed on the validation set) for the quality metrics. The sensitivity, specificity and percent of decisive decisions of the pipeline corresponding to the 143 different threshold values were shown in figure 2. Not surprisingly, the percent of decisive decision decreases as the thresholds for the quality metric increases. On the other hand, sensitivity and specificity do not change linearly as the thresholds increase. In figure 2, the performance corresponds to sets of thresholds that result in more than 60% of decisive decisions were highlighted with dashed line. Within that same threshold region, the range of sensitivity on the testing set is 0.616-0.667, and the range of specificity on the testing set is 0.697-0.726 (see figure 3 for sensitivity and specificity for individual threshold values, with the Pareto front highlighted). The performance of the decision support pipeline outperforms that of the ASD both in terms of sensitivity and specificity within this threshold region.

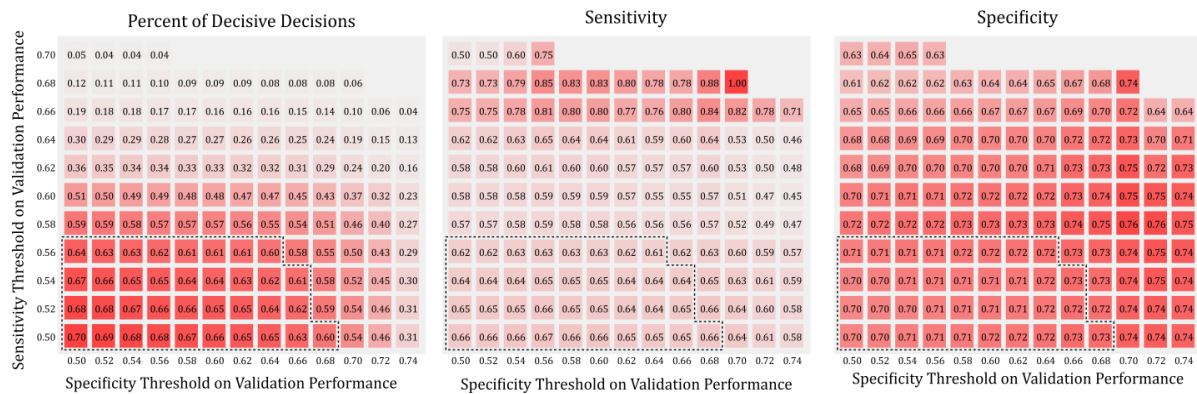


Figure 2: Model performance evaluated at different sets of quality metric (validation set sensitivity and specificity) thresholds. Three panels show percent decisive decisions, sensitivity and specificity on the testing set. Each square shows the performance of the pipeline under a particular threshold setting. The color of the square (different shades of red) represents the performance, with red mapped to the best performance. Black dashed line encloses threshold region where the model makes decisive decision for more than 60% of the individuals. The most stringent threshold settings (top right corner of individual panel) did not result in any decisive decisions. These results were left blank in the figure.

## Discussion

The main contribution of this study is the design and implementation of a decision support pipeline that automatically generates clinical recommendations for individuals who have recently been exposed to trauma. To the best of our knowledge, this pipeline is the first to provide individual level and personalized prognostics for trauma exposure with special consideration for the quality of the predictive model. The current pipeline outperforms the ASD, a commonly used clinical diagnostic criterion, both in terms of sensitivity and specificity. One distinctive characteristic of the decision support pipeline is the inclusion of quality metrics for the predictive models. The main purpose of the quality metric is to evaluate the quality of models constructed with different predictors, since in the clinical setting, not all predictor are available for all patients. If the quality of a predictive model (constructed with a subset of features) is inadequate, the pipeline recommends more information to be collected for re-evaluation, rather than making a decision based on the prediction generated by a suboptimal model. Some of the features used for prediction are routinely obtained in the emergency department, whereas other features can be obtained through phone interviews around 10 days after trauma exposure at low financial cost.

Therefore, this decision support pipeline can be readily incorporated into a clinical setting with minimal modification of the existing clinical workflow. We are currently working on implementing software based on the proposed pipeline. The software would automatically integrate clinical observations from sources including the EHR, diagnostic questionnaires and clinical interviews, to generate appropriate recommendations. When the information regarding a patient is not sufficient for prognosis, the software would generate a list of additional information to be collected and re-evaluate when new information is available. We believe the current pipeline and the aforementioned software could be a valuable addition to a clinician's toolkit, especially when insufficient information is available.

One direction to expand the current work is to test the performance of the pipeline in a different population. In the current study, the predictive model was constructed with the J-TOPS population and the evaluation of pipeline is conducted on the same population. This may result in optimistic performance estimation of the pipeline. Furthermore, a certain percentage of the J-TOP patients only have data collected in less than 2 time points and were exclude in the study, which might result in bias of the performance estimation. To address these issues, we are currently collecting data in the Bellevue hospital center in New York as an independent test set.

In the proposed decision support pipeline, all three components can be modified to further improve the predictive performance. The first component of the pipeline, the database for building the predictive model, can be constructed to suit specific diagnostic purposes. For example, if the goal is to predict stress response trajectory in the general population, including individuals from a diverse demographic and socio-economic background would likely increase the predictive performance of the algorithm. On the other hand, if the goal is to predict stress response trajectory in a specific population, for example, war veterans, it is more appropriate to restrict the database to the target population. Moreover, other variables that are potentially predictive of PTSD trajectory could be included in the database as potential predictive features of PTSD to improve the predictive performance and robustness of the prediction. Similarly, for the second part of the pipeline, any predictive algorithms can be implemented to generate the predictive model. If the database consists of predictive features for distinct domains (e.g. demographic, psychiatric measurements, clinical measurements and physiological measurements) utilizing ensemble methods<sup>23, 24</sup> might result in better predictive performance. In addition, the current implementation builds the predictive model based on subset of the database where the features present in the testing set are available. Exploring different data imputation methods<sup>25-27</sup> may result in better utilization of information contained in the database, thus resulting in improved performance. Also, different methods for model performance estimation can be implemented. Instead of leave-one-out cross validation, cross validation or cross indexing<sup>28, 29</sup> can be utilized as performance estimation with limited computational resource. Lastly, for decision generator, the last component of the pipeline, one can expand the possible recommendations and design customized functions to convert the prediction of the predictive algorithms to actionable recommendations<sup>15</sup>.

## Conclusion

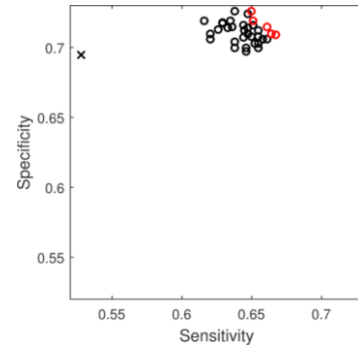


Figure 3: Sensitivity and Specificity for predictive models. Circles show models produced by the decision support pipeline that makes decisive decisions in more than 60% of the individual. Red open circles shows the Pareto front. Black x shows the predictive performance of ASD.

The current study demonstrated the feasibility of early individual level prediction for chronic PTSD. As a first step towards a decision support system for chronic PTSD, the proposed pipeline provides clinically relevant recommendations based on personalized predictive models. To avoid making potentially detrimental clinical recommendations based on insufficient information, the quality of the predictive model was computed and integrated into the decision support pipeline. The pipeline outperforms ASD, a criterion often used in the clinical setting.

## Acknowledgement

Isaac Galatzer-Levy is supported by K01MH102415 from the National Institute of Mental Health. Arieh Shalev is supported by MH071651 from the National Institute of Mental Health. The authors are grateful to the high performance computing facility at New York University Langone Medical Center for providing computational resources for conducting the analytical experiments.

## Reference

1. Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. *The Journal of clinical psychiatry*. 2000;61 Suppl 5:4-12; discussion 3-4. PubMed PMID: 10761674. eng.
2. Galatzer-Levy IR, Ankri Y, Freedman S, et al. Early PTSD Symptom Trajectories: Persistence, Recovery, and Response to Treatment: Results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *PLoS one*. 2013;8(8):e70084.
3. Peleg T, Shalev AY. Longitudinal studies of PTSD: overview of findings and methods. *CNS spectrums*. 2006;11(8):589.
4. Karstoft K-I, Galatzer-Levy IR, Statnikov A, et al. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC psychiatry*. 2015;15(1):30.
5. Galatzer-Levy IR, Karstoft K-I, Statnikov A, et al. Quantitative forecasting of PTSD from early trauma responses: A Machine Learning application. *Journal of psychiatric research*. 2014;59:68-76.
6. Ozer EJ, Best SR, Lipsey TL, et al. Predictors of Posttraumatic Stress Disorder and Symptoms in Adults: A Meta-Analysis. *Psychological Bulletin*. 2003;129(1):52-73.
7. Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol*. 2000 Oct;68(5):748-66. PubMed PMID: 11068961.
8. Karl A, Schaefer M, Malta LS, et al. A meta-analysis of structural brain abnormalities in PTSD. *Neuroscience and biobehavioral reviews*. 2006;30(7):1004-31. PubMed PMID: 16730374. eng.
9. Etkin A, Wager TD. Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American journal of psychiatry*. 2007;164(10):1476-88.
10. Boscarino JA, Erlich PM, Hoffman SN, et al. Higher FKBP5, COMT, CHRNA5, and CRHR1 allele burdens are associated with PTSD and interact with trauma exposure: implications for neuropsychiatric research and treatment. *Neuropsychiatric disease and treatment*. 2012;8:131-9. PubMed PMID: 22536069. Pubmed Central PMCID: 3333786. Epub 2012/04/27. eng.
11. Bryan CJ, Clemans TA. Repetitive traumatic brain injury, psychological symptoms, and suicide risk in a clinical sample of deployed military personnel. *JAMA psychiatry (Chicago, Ill)*. 2013 Jul;70(7):686-91. PubMed PMID: 23676987. Epub 2013/05/17. eng.
12. Kessler RC, Sonnega A, Bromet E, et al. Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*. 1995;52:1048-60.
13. Ressler KJ, Mercer KB, Bradley B, et al. Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. *Nature*. 2011;470(7335):492-7.
14. Bishop CM. *Pattern Recognition. Machine Learning*. 2006.
15. Duda RO, Hart PE, Stork DG. *Pattern classification: John Wiley & Sons*; 2012.
16. Murphy KP. *Machine learning: a probabilistic perspective: MIT press*; 2012.
17. Shalev A, Ankri Y, Peleg T, et al. SA (2012) Early Treatment for PTSD: Results from the Jerusalem Trauma Outreach and Prevention Study (J-TOPS). *Archives of General Psychiatry*.69:166-76.

18. Boser BE, Guyon IM, Vapnik VN, editors. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory; 1992: ACM.
19. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2011;2(3):27.
20. Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. 2003.
21. Association AP, Association AP. Diagnostic and statistical manual of mental disorders (DSM). Washington, DC: American psychiatric association. 1994:143-7.
22. Bryant RA, Friedman MJ, Spiegel D, et al. A review of acute stress disorder in DSM-5. *Depression and anxiety*. 2011;28(9):802-17.
23. Kapoor A, Picard RW, Ivanov Y, editors. Probabilistic combination of multiple modalities to detect interest. *Pattern Recognition, 2004 ICPR 2004 Proceedings of the 17th International Conference on*; 2004: IEEE.
24. Oza NC, Tumer K. Classifier ensembles: Select real-world applications. *Information Fusion*. 2008;9(1):4-20.
25. Engels JM, Diehr P. Imputation of missing longitudinal data: a comparison of methods. *Journal of clinical epidemiology*. 2003;56(10):968-76.
26. Myrtveit I, Stensrud E, Olsson UH. Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods. *Software Engineering, IEEE Transactions on*. 2001;27(11):999-1013.
27. Shrive FM, Stuart H, Quan H, et al. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC medical research methodology*. 2006;6(1):1.
28. Reunanen J, editor *Model selection and assessment using cross-indexing*. *Neural Networks, 2007 IJCNN 2007 International Joint Conference on*; 2007: IEEE.
29. Statnikov A, Aliferis CF, Hardin DP. *A Gentle Introduction to Support Vector Machines in Biomedicine: Theory and Methods*: world scientific; 2011.