



Application of Dirichlet Bayesian Network Model in Classification of Gene Expression Data of Bipolar Disorder Patients Treated with Lithium

Sahar Souri Pilangorgi¹, Nader Salari^{2*}, Sodeh Shahsavari², Afshin Almasi²

¹Student Research Committee, Department of Biostatistics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

²Department of Biostatistics, Faculty of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

OPEN ACCESS

*Corresponding Author:

Dept. of Biostatistics, Faculty of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

Citation:

Souri Pilangorgi A, Salari N, Shahsavari S, Almasi A. Application of Dirichlet Bayesian Network Model in Classification of Gene Expression Data of Bipolar Disorder Patients Treated with Lithium. *Iranian biomedical journal* 2024; 28(7): 104.

ABSTRACT

Introduction: The Bayesian Dirichlet model is a method with score-based structural learning, leading to more accurate learning of the Bayesian network structure. In this study, the Bayesian Dirichlet Bayesian network was used to classify gene expression data in bipolar disorder with lithium treatment.

Methods and Materials: Gene expression data of 120 patients with bipolar disorder, including 47323 genes, were used, of which 120 received standard treatment and 120 received lithium treatment. To analyze and classify the data, the correlation and Bayesian networks were combined to create a model showing the interactions between thousands of genes in a network. The critical variables were then selected using partial least squares regression, and the correlation network was made using biclustering and principal component analysis. The Dirichlet Bayesian network model was created to classify the correlation network. Finally, the accuracy of the prediction model was evaluated using ROC curve analysis. R3.6.2 software was used to analyze the data.

Results: Our analysis discovered 10788 essential and significant genes using partial least squares regression. Nine homogeneous clusters were created using the quadratic algorithm, and the representative of each cluster was first selected with at least 75% of the changes in the data (correlation network). Then, the classification was performed using the Bayesian network method, the accuracy of which was 0.86, and the model's accuracy was 0.91.

Conclusion and Discussion: This study demonstrates the potential of a hybrid approach, which can be developed for network analysis of thousands of genes. This method can help study other diseases using existing datasets. The results of such experiments will also help determine the exact cause and molecular mechanisms of diseases.

Keywords: Bipolar disorder, Classification, Lithium