A Data-Driven Exploration of Hypotheses on Disease Dynamics

Marcos L. P. Bueno[1,2(✉)], Arjen Hommersom[1,3], Peter J. F. Lucas[1,4], and Joost Janzing[5]

[1] iCIS, Radboud University Nijmegen, Nijmegen, The Netherlands
{mbueno,arjenh,peterl}@cs.ru.nl
[2] Department of Computer Science, Federal University of Uberlândia, Uberlândia, Brazil
[3] Faculty of Management, Science and Technology, Open University, Heerlen, The Netherlands
[4] LIACS, Leiden University, Leiden, The Netherlands
[5] Department of Psychiatry, Radboud UMC, Nijmegen, The Netherlands
Joost.Janzing@radboudumc.nl

Abstract. Unsupervised learning is often used to obtain insight into the underlying structure of medical data. In this paper, we show that unsupervised methods, in particular hidden Markov models, can go beyond this by guiding the generation of clinical outcome measures and hypotheses, which play a crucial role in medical research. The usage of the data-driven approach facilitates selecting which hypotheses to further investigate. We demonstrate this by using clinical trial data for psychotic depression treatment as a case study. The discovered latent structure and proposed outcome are shown to provide new insight into the heterogeneity of psychotic depression in terms of predictive symptoms.

Keywords: Machine learning · Psychiatry · Depression · Latent variable · Hidden Markov model · Unsupervised learning · Outcome measure

1 Introduction

Much about disease processes is unknown, as often the only available information about a disease are the patient's symptoms and signs. This results in an incomplete understanding of a medical disorder, which can be overcome by latent variable modeling. Latent variables can enhance our understanding of the problem domain by capturing unmeasured quantities (e.g. related to the underlying physiology) and their relationship to observed quantities [14], and might provide better fitted models [15]. Hence, by using latent variables, one can try to reconstruct the underlying structure of the process at hand by using observed data.

Unsupervised learning is the machine learning task that aims to generate representations of the underlying structure of the data. Applications of unsupervised

D. Riaño et al. (Eds.): AIME 2019, LNAI 11526, pp. 170–179, 2019.
https://doi.org/10.1007/978-3-030-21642-9_23

learning to medical data include, e.g., the discovery of underlying patient groups using clustering methods [7,8], which might help improve diagnosis and provide new insight into more effective treatment selection [1]. Yet, when applied to medical data, unsupervised techniques generate output that often makes experts confront themselves with questions like *what else can we do with this structure?*. We show in this paper that unsupervised learning methods, in particular hidden Markov models (HMMs) [9], can be used not only to describe the underlying structure but also to support the formulation of meaningful medical outcomes. Previous research suggested that the formulation of clinical outcomes might be guided by latent-variable models [5], with the advantage of reducing the hypothesis space to be explored by inspecting model properties. By using HMMs, we claim that one can explore hypotheses on disease dynamics by inspecting model characteristics such as transition dynamics, latent states, etc.

In order to illustrate the usage of HMMs on disease dynamics, we make use of data from a clinical trial originally designed to compare pharmacological treatments to psychotic depression (PD) [13]. PD is a severe medical condition that is associated with a high burden of disease and relatively low remission rates following pharmacological treatment [10]. Although recent research has considered PD as a homogeneous subtype of major depressive disorder [12], the possibility that this subtype itself is heterogeneous should also be considered, which would stimulate the development of subgroup adjusted prognostics and treatment modifications. In this work, we apply HMMs to one of the largest pharmacological trials of patients with PD conducted so far [13], aiming to explore potential differences in course characteristics.

The contributions of this paper are as follows. We present a procedure to guide the exploration of hypotheses on disease dynamics by means of HMMs. We then apply this methodology to yield insight into the dynamics of PD treatments by exploring clinically meaningful outcomes. The results are then assessed based on standard clinical response and remission in PD. To the best of our knowledge, this is the first effort into a more systematic approach for exploring hypotheses on disease dynamics based on probabilistic graphical models.

The remainder of this paper is organized as follows. In Sect. 2, the relevant work related to this paper is discussed. In Sect. 3, a method for exploring hypotheses on latent disease dynamics is proposed. In Sect. 4, the PD data used as case study is described. The experimental results are presented in Sect. 5. Section 6 summarizes the paper and suggests future work.

2 Related Work

Hidden Markov models have been extensively used in medicine in general, as well as in psychiatry. It is often the case that the number of latent states in HMMs is determined in advance, as researchers might be interested in a specific subset among all possible models. Previous research [4] used a two-state HMM to investigate the hypothesis that patients switch between two stable states (symptom-free versus depressed) in major depressive disorder. In the context of Alzheimer's disease, a four-state continuous-time HMM was developed

to investigate the relationship between cognition and psychotic symptoms [11]. However, one might argue that by not imposing an *a priori* number of or already known latent states, a larger set of possible models is considered, which can lead to more insight into disease dynamics, at a potential cost of having an increased difficulty interpreting the models.

The typical usage of HMMs is in prediction or as a model to describe the underlying structure of the data. While prediction is self-explanatory, the underlying structure is usually be seen as a set of clusters, thus it is a more abstract and more difficult to be used representation. A much more specialized usage of latent variables lies in the development of data-driven outcome measures [5], which was based on models other than HMMs (namely, the item response theory). Such data-driven approach to generating outcomes has the advantage that latent states might provide a more natural, compact and empirically-oriented way to measure multiple relationships between symptoms and other observables.

More recently, HMMs have been applied to electronic health records (see e.g. [6]). Such datasets are often large and heterogeneous, requiring models such as HMMs for gaining relevant insights.

3 Capturing Latent Disease Dynamics

In this section we discuss models suitable for capturing latent disease dynamics and propose a data-driven method for exploring medical outcomes.

3.1 Bayesian Networks and Hidden Markov Models

Hidden Markov models are models based on latent variables that are able to cope with uncertainty and sequential phenomena, which makes HMMs suitable for many biomedical problems [4,6,11]. In HMMs, the observable variables typically interact only via the latent (or state) variable [9], which is known as the naive-structure HMM. In this work we opt for modeling the observation space as a Bayesian network (BN), thus allowing for more general representations of symptom interaction. By such modeling, more insight into the problem can be obtained by a more concise latent-state representation [2].

3.2 State Trajectories

Before we describe how to use HMMs to obtain insight into disease processes, we introduce the notation. Let us denote by S the random variable representing the latent states to be modeled, where S takes values on the set $\text{dom}(S) = \{s_1, \ldots, s_k\}$. The remaining variables $\{X_1, \ldots, X_m\}$ are observable variables with associated domains $\text{dom}(X_i)$. In medical domains, each X_i will often refer to measured data such as symptoms, lab exams, medication, etc., while the latent variable S will refer to some state of the underlying disease (e.g. a disease remitting situation). The disease process of interest is assumed discrete over the time points $\{0, \ldots, T\}$, where the value of the latent variable and the

observables that hold at time t will be denoted by $S^{(t)}$ and $X_i^{(t)}$ respectively. For a discrete time interval $[t_1, t_2]$, the notation $S^{(t_1:t_2)}$ will be used.

HMMs can be used to predict the hidden states that better explain the observations [9]. Prediction is achieved by first computing the distribution of latent states at each week t conditional on the complete patient's symptom data (i.e. his or her data over all the weeks):

$$\gamma_t(s) = P(S^{(t)} = s \mid X_1^{(0:T)}, \ldots, X_m^{(0:T)}) \tag{1}$$

where $\gamma_t(s)$ is the notation used in standard forward-backward algorithms for HMMs [9]. After this has been done, the sequence of states for a given patient is obtained by selecting the most likely state at each time t:

$$\gamma_t^* = \arg\max_{s \in \mathrm{dom}(S)} \gamma_t(s) \tag{2}$$

for all $t \in \{0, \ldots, T\}$. This can be interpreted as "placing" patients in states. Note that the predicted states are the individually most likely states, obtained by maximizing Eq. 2 for each time point independently, in contrast to the so-called Viterbi path, where the maximization is applied over the probability of state trajectories over $\{0, \ldots, T\}$.

3.3 Exploring Medical Outcomes

Understanding disease dynamics in a multi-variate setting is challenging because of potential complex interactions between diseases, symptoms, and findings. Therefore, we propose to investigate the transition dynamics between latent states. This is convenient because each latent state can take into account multiple symptom dimensions at once, which makes reasoning over patient trajectory very feasible. Once the states are discovered, a detailed outcome measure that provide insight into treatment dynamics can be formulated.

We propose a procedure to build outcome measures in Fig. 1. The procedure selects a set of *baseline states* S_b based on a selection criterion. From the remaining states, a set of *target states* S_e are to be selected based on its own criterion. Once S_b and S_e are obtained, *state reachabilities* from S_b states to S_e states are calculated. By varying the time interval between two given states of S_b and S_e, the resulting probabilities $reach(i, j, t_1, t_2)$ indicate the temporal influence of a baseline state over a target state. Such state reachabilities can then be used to compose a rich outcome measure, e.g., by making $t_1 = 0$ and $t_2 \in \{1, \ldots, T\}$, which will result in a reachability trend as indicated in Fig. 1.

In this paper, a state is classified as a baseline state if one or more patients are predicted to be in this state at the process start. These states can be computed by first determining the state trajectories $\gamma_t(s)$ for each patient (see Eq. 1). Then we define $s \in S_b$ if and only if $\gamma_0^* = s$ holds for at least one patient. To determine the target states, we look at the model parameters, such that $P(s \rightarrow s) \geq \rho$, where $0 \leq \rho \leq 1$. In particular, we chose $\rho = 0.95$, resulting in states that are not in S_b and have a high self-transition probability.

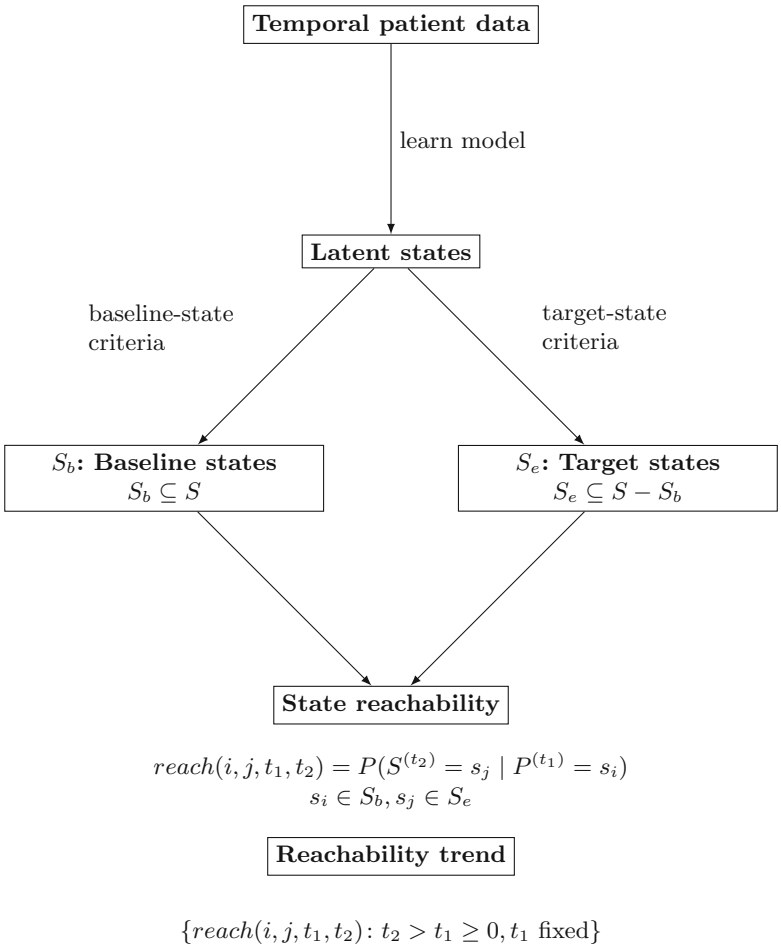


Fig. 1. Procedure to guide the generation of outcome measures based on latent states.

4 Data

4.1 Patients and Variables

All patients had participated in the DUDG (Dutch University Depression Group) study [13], a 7 week double-blind randomized clinical trial originally designed for comparing the effectiveness of venlafaxine, imipramine and venlafaxine plus quetiapine (V+Q, for brevity) in psychotic depression. The dataset originally included 122 participants aged 18–65 who met DSM-IV-TR criteria for a unipolar major depressive episode with psychotic symptoms and a 17-item Hamilton Depression Rating Scale (HAM-D [3]) score of at least 18 (both at the screening visit and at baseline). The 17-item HAM-D indicates severity of depression as follows:

normal (0–7), mild depression (8–13), moderate depression(14–18), severe depression (19–22), and very severe depression (greater than or equal to 23).

Because of insufficient information about the specific nature of psychotic symptoms, three patients were not included in the current study resulting in a dataset with 119 patients. From the total group, 59 (49,6%) were females; the mean age was 51.1 (SD 10.9) years. Forty patients were randomized to treatment with imipramine, 38 to venlafaxine and 41 to V+Q.

Severity of depression (HAM-D, continuous) and the presence of psychotic symptoms (dichotomized) were measured at baseline (i.e. before treatment starts) and weekly thereafter. A total of 98 patients completed the trial (34 in imipramine, 30 in venlafaxine, and 34 in V+Q). Data on patients who dropped out was imputed by the last-observation-carried-forward approach [13].

4.2 Depression Assessment

At the end of medical treatment, patients were assessed according to conventional criteria for response and remission of depression [13]. Response was defined as a reduction of at least 50% on the HAM-D score compared to baseline and a score of 14 or below, and remission as a score of 7 or below.

5 Experimental Results

5.1 Model of Observations

The observable variables in the HMM used in this work are modeled according to the BN shown in Fig. 2, which allows for a more expressive representation than the naive-Bayes structure by connecting Hal and Del via HAM-D. By doing so, we impose less independence assumptions than the naive solution, thus the model becomes more flexible in that more dependences can be induced from data. Hence, once in a state the observables are parameterized as follows: the psychotic symptoms are encoded as binary random variables, while the depressive symptom (the HAM-D score) is a conditional Gaussian distribution (conditioned on the state and on both psychotic symptoms, as shown in Fig. 2).

At any time point, the parameterization of each symptom is given by the factorization entailed by the BN structure of Fig. 2. This modeling dictates that HAM-D will be given by a mixture of four Gaussians, one for each configuration of Del and Hal (assuming the state is fixed). For a given state $s \in S$, the distribution of HAM-D can be obtained by marginalizing out Del and Hal and by applying the Bayesian network factorization as follows (we omit the time index as it is equal to t):

$$p(\text{HAM-D} \mid s) = \sum_{\text{Del,Hal}} p(\text{HAM-D}, \text{Hal}, \text{Del} \mid s) \tag{3}$$

$$= \sum_{\text{Del,Hal}} P(\text{Del} \mid s) P(\text{Hal} \mid s) p(\text{HAM-D} \mid \text{Del}, \text{Hal}, s) \tag{4}$$

Thus, the distribution of HAM-D conditional on state s is Gaussian as it is a linear combination of the Gaussians associated to the possible parent values.

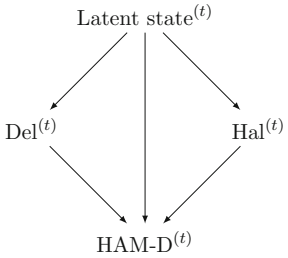


Fig. 2. BN structure of the space of observations. The domain of the state variable depends on the experiments.

5.2 Model Dimension

The number of latent states was obtained by balancing model fit and interpretability. Log-likelihoods were obtained from a 10-fold cross validation procedure, where models can have from two states up to the number of states obtained prior to model overfitting. Suppose we denote by $\mathcal{L}(k)$ [95% CI] the mean log-likelihood obtained by the model with k states. The obtained means are as follows: $\mathcal{L}(2) = -431[-441; -421]$, $\mathcal{L}(3) = -410[-422; -398]$, $\mathcal{L}(4) = -416[-436; -396]$, $\mathcal{L}(5) = -410[-431; -389]$, $\mathcal{L}(6) = -405[-423; -387]$, $\mathcal{L}(7) = -399[-418; -381]$, $\mathcal{L}(8) = -400[-417; -383]$. We do not show $\mathcal{L}$ values for $k > 8$ states as $\mathcal{L}$ approximately saturates at that point. As the values of $\mathcal{L}$ are to be maximized and the 95% CIs highly overlap for $k \geq 3$, then adding more than 3 states is not likely to lead to a significant improvement to $\mathcal{L}$. Hence, we choose $k = 3$ as the number of states. The selected number of states also takes into account that simpler models are preferred for the formulation of outcomes.

5.3 Identified States

The learned model has 3 latent states, as shown in Fig. 3 (top row), where in each latent state there is one distribution for each symptom measurement (i.e., Del, Hal and HAM-D). The states can be interpreted as follows:

- The **state Hallucinations (abbreviated as state H)** is associated with patients with high prevalence of hallucinations and moderate prevalence of delusions, with the highest mean HAM-D score and low variance.
- The **state Delusions (abbreviated as state D)** is associated with patients with high prevalence of delusions and low prevalence of hallucinations. Its mean HAM-D score is moderate and has wide tails.
- The **state No Psychosis (abbreviated as state NP)** is associated with patients with low prevalence of psychotic symptoms and moderate HAM-D score (though with high variance).

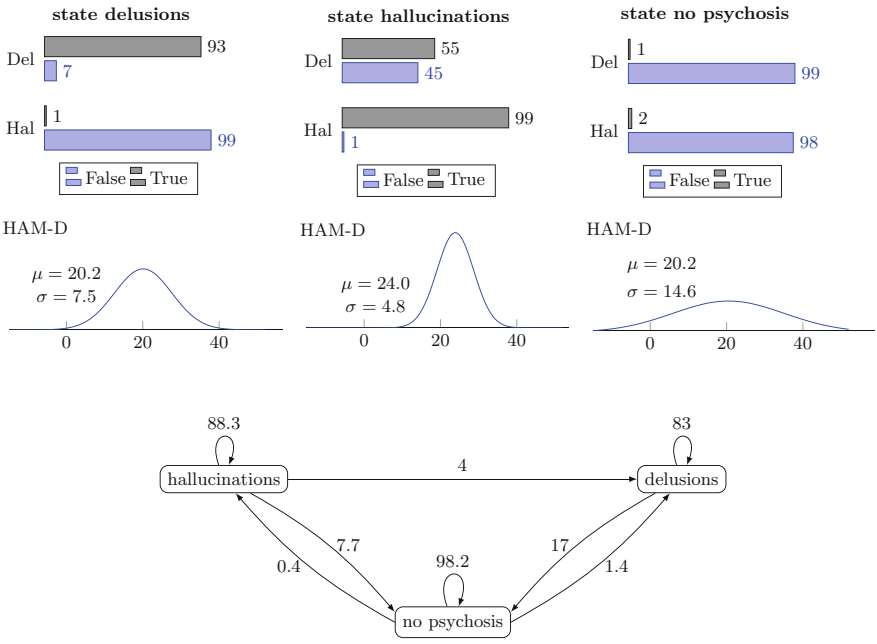
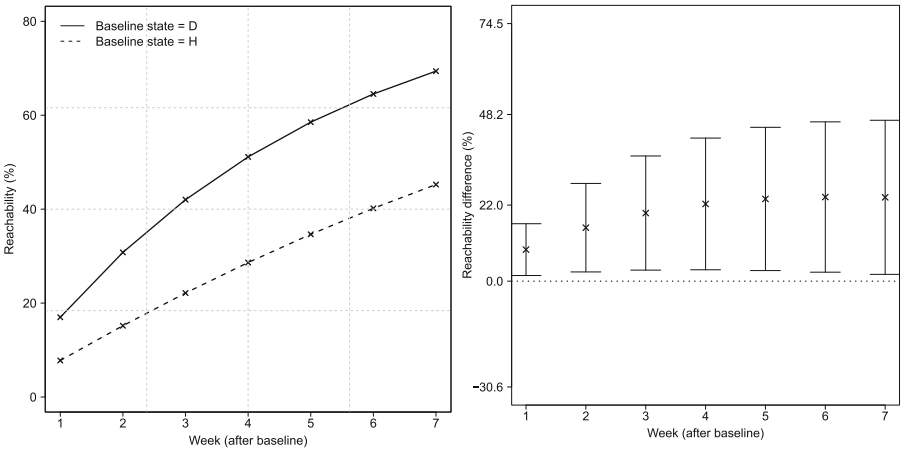


Fig. 3. Top: marginal distributions of symptoms in the latent states of the learned model (Del and Hal stand for symptom measurements). Bottom: state transitions.

5.4 Dynamics

Figure 3 (bottom row) shows the transition behavior of the learned HMM. The arcs indicate transition probabilities between latent states, e.g. the looping probability of 88.3% in state H represents the chance for remaining in such state over two adjacent weeks. Based on Fig. 3 (top row) and on the previous characterization of the states, D and H can be seen as starting states that are primarily distinguished based on the prevalence of hallucinations in patient. Later on, depending on their response to treatment, the patient will potentially move to state NP. The state NP can be seen as a healthier state due to the absence of psychotic symptoms, but the state does not imply depression remission or response due to its moderate mean HAM-D. In fact, the state NP characterizes a wide range of no-psychosis patients in terms of HAM-D score.

Figure 4-a shows the reachability trends given the baseline states, while Fig. 4-b shows the 95% bootstrap confidence intervals (BCIs) for the difference between the trends. In these cases, positive values indicate a stronger trend in favor of state D. The difference between the area under the curve of each trend was also computed, resulting in a 95% BCI equal to [0.17; 2.29]. The 95% BCI for the slope difference was [0.02; 0.17]. These results suggest that the initial state of the patient is relevant, i.e. starting in state D allows for a significantly stronger reachability to state NP than the reachability when starting in state H.



(a) Reachability trends (b) 95% BCIs of reachability differences

Fig. 4. Reachability trends based on different baseline states.

5.5 Validation of the Outcome

We now assess the claim that the state at baseline leads to significantly different state reachability. To this end, two distinct groups of patients were considered: patients with hallucinations at baseline (29 patients), and patients with no hallucinations at baseline (90 patients). The HAM-D scores of these groups at treatment endpoint were compared using a Mann-Whitney test for independent samples, which resulted in a p-value = 0.0007, suggesting that these two groups differ significantly (under a 95% confidence level). As a result, the psychotic symptom at baseline is predictive to depression recovery of patients in general.

6 Conclusions

This paper demonstrated that probabilistic graphical models can reveal insight into disease dynamics by considering not only the underlying structure, but also using meaningful outcome measures built from such structure. We illustrated the proposed methodology by applying hidden Markov models to psychotic depression treatment data, which were learned in a fully data-driven way.

The identified underlying symptom structure revealed two clinically significant results. First, the remission of psychotic symptoms preceded the decrease of depressive symptoms in PD treatment, which is in accordance with clinical observation. Second, it was shown that patients differed in their prognosis depending on the type of psychotic symptoms they exhibited at baseline (hallucinations versus delusions). Hence, our methodology allowed to shed light on the heterogeneity of psychotic depression. As future work, we will further investigate the clinical significance of the results, and will investigate the sensitivity of different treatment groups to treatment. The combination of graphical models and a data-driven approach can be easily integrated into the investigation of other disorders as well.

Acknowledgments. This work was partially funded by project "NORTE-01-0145-FEDER-000016" (NanoSTIMA). NanoSTIMA is financed by the North Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

1. Ahlqvist, E., et al.: Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diab. Endocrinol. **6**(5), 361–369 (2018)
2. Bueno, M.L.P., Hommersom, A., Lucas, P.J.F., Linard, A.: Asymmetric hidden Markov models. Intl. J. Approximate Reasoning **88**, 169–191 (2017)
3. Hamilton, M.: A rating scale for depression. J. Neurol. Neurosurg. Psychiatry **23**(1), 56–62 (1960)
4. Hosenfeld, B., et al.: Major depressive disorder as a nonlinear dynamic system: bimodality in the frequency distribution of depressive symptoms over time. BMC Psychiatry **15**(1), 222 (2015)
5. IsHak, W.W., et al.: The recovery index: a novel approach to measuring recovery and predicting remission in major depressive disorder. J. Affect Disord. **208**, 369–374 (2017)
6. Meier, J., Dietz, A., Boehm, A., Neumuth, T.: Predicting treatment process steps from events. J Biomed. Inform. **53**, 308–319 (2015)
7. Paoletti, M., et al.: Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of chronic obstructive pulmonary disease (COPD) phenotypes. J. Biomed. Inform. **42**(6), 1013–1021 (2009)
8. Paulus, M.P., et al.: Latent variable analysis of positive and negative valence processing focused on symptom and behavioral units of analysis in mood and anxiety disorders. J. Affect. Disord. **216**, 17–29 (2017)
9. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
10. Rothschild, A.J.: Challenges in the treatment of major depressive disorder with psychotic features. Schizophrenia Bull. **39**(4), 787–796 (2013)
11. Seltman, H.J., Mitchell, S., Sweet, R.A.: A Bayesian model of psychosis symptom trajectory in Alzheimer's disease. Int. J. Geriatr. Psychiatry **31**(2), 204–210 (2016)
12. Wijkstra, J., Lijmer, J., Burger, H., Cipriani, A., Geddes, J., Nolen, W.: Pharmacological treatment for psychotic depression. Cochrane Database Syst. Rev. **7** (2015)
13. Wijkstra, J., et al.: Treatment of unipolar psychotic depression: a randomized, double-blind study comparing imipramine, venlafaxine, and venlafaxine plus quetiapine. Acta Psychiatr. Scand. **121**(3), 190–200 (2010)
14. Yet, B., Perkins, Z., Fenton, N., Tai, N., Marsh, W.: Not just data: a method for improving prediction with knowledge. J. Biomed. Inform. **48**, 28–37 (2014)
15. Zhang, N.L., Nielsen, T.D., Jensen, F.V.: Latent variable discovery in classification models. Artif. Intell. Med. **30**(3), 283–299 (2004)