Predictive modeling in e-mental health: A common language framework

Dennis Beckera,⁎, Ward van Breda b, Burkhardt Funka, Mark Hoogendoorn a, Jeroen Ruwaardc,d, Heleen Riperc,d

a Institute of Information Systems, Leuphana University Lüneburg, Germany
b Faculty of Science, Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
c Department of Research & Innovation, GGZ inGeest, Amsterdam, P.O. Box 7057, Amsterdam MB 1007, The Netherlands
d Faculty of Behavioural and Movement Sciences, Department of Clinical, Neuro- and Developmental Psychology, Clinical Psychology Section, Vrije Universiteit Amsterdam, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands

A B S T R A C T

Recent developments in mobile technology, sensor devices, and artificial intelligence have created new opportunities for mental health care research. Enabled by large datasets collected in e-mental health research and practice, clinical researchers and members of the data mining community increasingly join forces to build predictive models for health monitoring, treatment selection, and treatment personalization. This paper aims to bridge the historical and conceptual gaps between the distant research domains involved in this new collaborative research by providing a conceptual model of common research goals. We first provide a brief overview of the data mining field and methods used for predictive modeling. Next, we propose to characterize predictive modeling research in mental health care on three dimensions: 1) time, relative to treatment (i.e., from screening to post-treatment relapse monitoring), 2) types of available data (e.g., questionnaire data, ecological momentary assessments, smartphone sensor data), and 3) type of clinical decision (i.e., whether data are used for screening purposes, treatment selection or treatment personalization). Building on these three dimensions, we introduce a framework that identifies four model types that can be used to classify existing and future research and applications. To illustrate this, we use the framework to classify and discuss published predictive modeling mental health research. Finally, in the discussion, we reflect on the next steps that are required to drive forward this promising new interdisciplinary field.

1. Introduction

Mental health problems have huge impacts on those affected, their social network, and society at large (Centre for Mental Health, 2010). By 2030, global mental health costs are expected to rise to about 6 trillion dollars per year, which by then will be more than the predicted health costs related to cancer, diabetes and respiratory diseases combined (Bloom et al., 2011). Ample research has been devoted to understanding the underlying causes of (specific) mental health problems, the factors that play a role in recovery, as well as the effectiveness of various therapies. This evidence-based movement has led to substantial improvements in mental health care. At the same time, however, there is a consensus that improvements should be made to more effectively address the global burden of mental health conditions.

E-mental health, the application of information technology in mental health care for the prevention and treatment of psychological disorders, may provide an answer to the global burden of mental health problems. The number of online treatment applications is increasing and can provide help to people who otherwise would remain untreated (Robinson et al., 2010; Titov et al., 2015). Recent advancements in computer and communication science have resulted in a rapid development of the field, which has led to new research opportunities and treatment models. Based on data that clients generate when using an online treatment, predictive models can be developed which support clients as well as therapists.

An example of a development fueled by the technological advancements includes so-called Ecological Momentary Assessments (EMA). Using EMA, a wealth of coarse- and fine-grained data is collected, from heart-rate sensors, physical activity sensors, and other mobile applications, to assess the dynamics of symptoms, affect, behavior and cognition over time, in the natural habitat of the patient (Trull and Ebner-Priemer, 2013). Initially designed as a pen and paper measurement, currently, EMA measures are dominantly collected using electronic devices such as the users’ smartphone. EMA enables
researchers to measure the users’ current state and behavior while engaged in their daily routine (Shifman et al., 2008; Wichers et al., 2011). Such type of data, however, is also partially collected when users engage into an online based treatment. This information allows to provide users with situational and personalized interventions (Burns et al., 2011) that are tailored to specific user needs. To analyze EMA data and other fine-grained types of data (e.g. log-level data), it is not straightforward to employ traditional statistical approaches such as t-tests, Analysis of Variance, or Ordinary Least Squares (OLS) regression. Although these methods allow to understand the relationship among measures and their influence on the treatment outcome, they neither account for the specific temporal (or sequential) nature of the data nor do they consider complex interaction effects among various measures. This is where new analytical methods come into play.

Promising methods include predictive modeling techniques such as Decision Trees, Bayesian approaches (Langley et al., 1992), Support Vector Machines (Cortes and Vapnik, 1995), and Artificial Neural Networks (Haykin, 2009). In order to apply these methods to the type of sequential data collected in e-mental health applications, the data has to be pre-processed (Hoogendoorn and Funk, 2017). The main objective of this pre-processing step is to derive meaningful variables that can then be used in predictive modeling. To succeed in leveraging the potential of predictive models for the purpose of developing effective interventions, each of the steps requires goal setting, understanding the data, construction of such models, as well as interpretation of the models, understanding user perception of these models, and finally deployment of the models developed. To fulfill all these requirements, an intense interdisciplinary collaboration between clinical researchers and computer scientists is mandatory.

The data science community has to understand therapists’ needs, their work process with clients, and decision points. On the other hand, therapists have to understand the technical capabilities, what data might be beneficial and what type of predictions can be derived from it. Only when both sides can communicate effectively, the potential of predictive modeling in improving treatment outcomes can be realized. The success of this collaboration will stand or fall with the development of a common language that we think is necessary to successfully plan predictive modeling in improving treatment outcomes can be realized. Only when both sides can communicate effectively, the potential of predictive modeling in improving treatment outcomes can be realized. The success of this collaboration will stand or fall with the development of a common language that we think is necessary to successfully plan

### 2. Methods in predictive modeling

In this section, we provide a short introduction to the predictive modeling domain. We define basic terminology, briefly describe supervised learning and evaluation methods to access the performance of such models.

#### 2.1. Terminology

Predictive modeling can be positioned under the broader umbrella term “statistical modeling” (Shmueli, 2010). While traditional statistical approaches focus on explaining data in terms of causality or identify relations, predictive modeling strives to find the model that provides the most accurate predictions. Predictions are derived from so-called attributes or features that are derived from observations. These features are not unlike the variables utilized in traditional statistical approaches. However, traditional statistical approaches, focusing on high explanatory power, do not necessarily lead to a high predictive power (Dawes, 1979; Forster and Sober, 1994).

#### 2.2. Supervised learning approaches

Building predictive models is often done through so-called supervised learning. Fig. 1 provides a general overview of the procedure of this approach. Supervised learning requires “historical” data providing predictive attributes and a target value (e.g. the occurrence of a depressive episode), that we want to predict from new data, where the attributes, but not the target values, are known. Targets can either be continuous values (referred to as regression) or categorical (classification). Data is commonly pre-processed to derive appropriate features from the raw attributes and to compose a training and test set. In addition, an assumption is made on the type of function that may describe the relationship between the features and the target. The set of all possible functions that can explain the observed data is called the hypothesis space. Next, a learning algorithm is applied that uses the training dataset in combination with an error metric to select a final hypothesis from the set of all hypotheses. To estimate the generalizability of this final hypothesis, it is evaluated against a test dataset that was not used in the learning process.

Various learning techniques and types of target functions exist. We follow the categorization of Witten and Frank (2005). They distinguish between (1) probabilistic modeling methods such as Naïve Bayes classifiers (see, e.g., Langley et al., 1992), (2) divide-and-conquer modeling methods such as decision trees (Safavian and Landgrebe, 1991), (3) extended linear modeling methods such as generalized linear models (e.g. logistic regression), support vector machines (Cortes and Vapnik, 1995), or artificial neural networks (see Haykin, 2009), and (4) instance-based learning methods such as k-nearest neighbors (Altman, 1992). A discussion of these methods is beyond the scope of this paper, but the references provide a good starting point for detailed information.

Supervised learning is a so-called bottom-up approach as it is driven by the data. It is complemented by the top-down approach, which is knowledge-driven. In the latter approach, existing theories are formalized into computational models that connect theoretical concepts from a domain to executable code. Therefore, the top-down approach starts from a theory and typically utilizes model structures that are based on theoretical and empirical insights into the problem domain. There are a multitude of techniques ranging from more mathematical types of modeling such as systems of differential equations (Zeigler et al., 2000) to rule-based or agent-based systems (Van der Hoek and
Wooldridge, 2008).

2.3. Evaluating predictive models

The purpose of predictive modeling is to find a model that generalizes well beyond the training data. Evaluation methods are different for regression and classification models. When evaluating regression models, two objective functions that are often used are the mean absolute error (MAE) and the mean squared error (MSE). The right type of error measure is dependent on what type of result is most desirable. The MAE is more suitable if you do not want your performance measure to be highly sensitive to outliers, and the MSE is more suitable if large errors are particularly undesirable.

For evaluating classification models, a great variety of measures such as accuracy (percentage of correctly classified cases), precision (fraction of correct classifications out of all cases attributed to a certain class by the predictive model), recall (fraction of cases correctly found to be in that class out of all elements in the data that were in the class) are used. When the classification performance of a model depends on a threshold value, as is often the case, the receiver operating characteristic (ROC) curve and the Area Under the Curve (AUC) can be used as performance measures (see e.g. Hanley and McNeil, 1982).

As mentioned, the generalizability of models is often tested on a test dataset, which is kept apart from the training set used to fit the model. When data is limited, a validation method called cross-validation can be used. The method divides the original sample set into k subsets, where k – 1 subsets are used as training samples and one subset is used as a validation and/or test sample. The process is repeated k times, where each of the subsets acts as the test set once, and the performance results are averaged (Olson and Delen, 2008).

3. Framework

To describe and categorize existing applications of predictive modeling within the mental health domain, we propose a framework. The framework (Fig. 2) provides a common language for researchers and forms the basis for informing treatment decisions and designing new interventions and experiments. The framework has four major elements, which are described below: time (phases), data, decisions, and model types.

3.1. Time

Predictive modeling in mental health naturally tries to uncover meaningful predictive sequential relationships. Hence, the framework is arranged along core phases of the timeline of the psychological intervention. We refer to these phases using the generic, yet recognizable, terms pre-intervention, intervention, and post-intervention to make them applicable across a wide range of mental health settings.

3.2. Data

Along the different phases, the framework identifies different types of data that are collected from participating clients, therapists and the technical systems involved. To predict states related to an individual, not only the data related to that individual could be relevant, but possibly the data from other individuals who have gone through this (or a similar) intervention can also hold predictive power.

Different data are collected in the different phases. In the pre-intervention phase questionnaires or interviews are used to determine the severity of past and current problems. Additionally, socio-demographic data, personality traits, motivation and attitude towards the treatment can be assessed as part of this assessment process.

The largest amount of data is usually collected in the intervention phase. This encompasses, for example, the client’s text responses to exercises in the interventions, their logs to the health systems, and other system interactions including technical loggings. Daily diaries are an option for recording the client’s development during the intervention (e.g., Bolger et al., 1989; Burton et al., 2009; Jacelon and Imperio, 2005). Smartphone-based EMA enables clients to regularly report relevant variables a few times a day (Wichers et al., 2011). Smartphones can also serve as a rich source of data when usage data, such as call logs and app usage, and sensor data, such as GPS and accelerometer data, are collected (e.g. Lane et al., 2010; Eagle et al., 2009). The high sampling frequency can quickly lead to a large amount of data per client. Whether and when more data leads to better predictive models is the subject of debate (Stange and Funk, 2016). The intervention often ends with a final screening, which can consist of a variety of questionnaires to estimate the remaining symptom severity and to evaluate the improvement of the client.

The post-intervention phase can consist of follow-up screenings and additional or repeated interventions to reduce the risk of relapse (Kessler et al., 2005; Vittengl et al., 2009).

3.3. Decisions

The data described in the previous subsection fuel recommendations and therapeutic decisions. These decisions may influence the course of the treatment and the client’s adherence and should aim to improve the therapeutic outcome.

In the pre-intervention phase, one has to decide what intervention and what level of guidance would best suit the client’s needs. With moderate symptoms, for example, the client might be best of with a preventive self-help intervention without any personal guidance. When symptoms are severe or the when problems are complex, a guided or blended intervention would perhaps be more appropriate.

During the intervention phase, earlier decisions may need to be corrected on the basis of new information. For instance, the level of guidance might be increased if a client is at risk of dropping out. Keeping the intervention relevant for the client by personalizing it based on his or her behavior could prevent this. The right timing of EMA (Armey et al., 2015; Moskowitz and Young, 2006; Smyth and Stone, 2003) and Ecological Momentary Interventions (EMI) (Heron and Smyth, 2010; Runyan et al., 2013) also represent important decisions during the intervention phase. An optimal timing can minimize the intrusiveness and enhance the client’s perception of the intervention.

In the post-treatment phase, additional interventions can be scheduled to maintain the outcome and prevent relapse (Kessler and Chiu, 2005; Vittengl et al., 2009). Usually, Internet-based treatment during this phase also consists of symptom screening and interventions to guide clients after their active treatment (Kok et al., 2014; Lord et al., 2016). Decisions that are seen in all phases are adherence measures that include motivational messages and feedback as well as reminders or contact to the therapist.

3.4. Model types

Predictive models can serve a variety of purposes. Traditionally, predictive models have been limited to prediction based on the data of a certain stage (e.g. at the pre-intervention) to predict what is going to happen in the next stage (e.g. success of the intervention). However, given the higher granularity and quality of the current assessments, more detailed models have become available that may enable predictions over shorter time periods. These models can drive personalized interventions that are fully tailored to observations collected from a specific client. The framework recognizes this by distinguishing four model types. It arranges the model types on the basis of the phase of treatment, the type of data that is available to the models, and the clinical decisions that are made on the basis of the model output. Table 1 presents an overview of the four model types.

Type 1 models predict risk of mental illness and can be used to identify best treatment options. There is extensive research on the predictive
power of variables with respect to risk assessment and diagnosis, focusing on a wide variety of variables related to socio-demographic characteristics (see e.g. Tovar et al., 2014; Mittendorfer-Rutz et al., 2014), personality traits (see e.g. Magidson et al., 2014), and illness characteristics (see e.g. Otto et al., 2001; Wade et al., 1998). In the past, this type of research has often employed standard statistical methods like OLS regression and logistic regression. Data available to Type 1 models can be collected through questionnaires, EMA assessments, or shared electronic health records. The time-span of the data collection varies from 1 h (one-time screening instrument administrations), to

Table 1
Proposed model types of the framework.

| Model type | Predictors | Purpose | Usage |
|------------|------------|---------|-------|
| 1          | Pre-intervention | Risk assessment and diagnosis | Identify clients at risk for mental health problems | Support diagnostic process | Support selection of intervention | Estimate level of guidance, EMA, EMI | Support selection of therapy by therapist | Facilitate personalized therapy | Support selection of individual decisions, such as next intervention, screening, motivation and level of guidance | Identify risk of drop-out | Adapt intervention to maximize short-term outcomes |
| 2          | Pre-intervention and intervention | Short-term trends | Track therapeutic progress | Support selection of therapy by therapist | Facilitate personalized therapy | Support selection of individual decisions, such as next intervention, screening, motivation and level of guidance | Identify clients with high relapse risk | Facilitate personalized after-care | Adapt after-care to maximize long-term outcomes |
| 3          | Pre-intervention and intervention | Predict therapy outcome | Stabilize results, prevent relapse | Support selection of therapy by therapist | Facilitate personalized therapy | Support selection of individual decisions, such as next intervention, screening, motivation and level of guidance | Adapt intervention to maximize treatment outcomes | Identifying clients with high relapse risk | Facilitate personalized after-care | Adapt after-care to maximize long-term outcomes |
days (e.g., a one-week diary), or weeks (e.g., EMA assessments).

Type 2 models predict short-term changes of the health status and treatment adherence to optimize treatment and treatment progress. As clients progress through the intervention, more data becomes available, as illustrated in Fig. 2 by the little-dotted boxes. Whenever new data arrive, the client-specific model can be updated, which, on average, increases the power to predict the short-term evolution of key concepts such as the valence of mood or ruminations. Based on predicted short-term changes, the next steps of the intervention can be planned. Furthermore, such models might contribute to the client’s adherence to the intervention. When the current engagement level is understood and the predicted probability for drop-out is high, motivational interventions can be applied, such as direct contact with the therapist, or a reminding message.

Type 3 models predict the outcome of the intervention phase, based on available data such as socio-demographic data and observations made during the intervention phase. Here, models that estimate and minimize dropout risk are also relevant, but in this case, the focus should be on drop-out over a longer term (which may require qualitatively different interventions than those deployed to reduce short-term drop-out).

Type 4 models aim to predict relapse. This class of models uses data from the pre-intervention phase and the intervention phase to provide a prediction of relapse risk (Kessing, 1999). Type 4 models can help to determine whether patients need after-care, or to decide on an optimal assessment schedule to regularly assess the stability of short-term treatment results.

4. Applying the framework to published research

In this section, we use the introduced framework to categorize published e-mental health research in which predictive models were applied, to illustrate that 1) the framework works in an insightful way and is useful, and 2) the framework covers work done within the field. Discussed papers, their key characteristics, and the relationship to the framework are summarized in Table 2.

4.1. Model Type 1: risk assessment

Type 1 models aim for the prediction of mental health problems. One way of achieving this goal is to apply data mining techniques to (re-)analyze existing epidemiological data, to identify illness risk factors that might not be easily detected with more traditional statistical techniques. For example, Ankarali et al. (2007) compared the performance of classification trees, a supervised modeling technique, to standard logistic regression in determining social-demographic risk factors for postpartum depression in women. The classification tree identified six risk factors, while the logistic regression model found only three. Another example is the study by Van der Werf et al. (2006) who used a large Dutch epidemiological database to create a mathematical model of the recovery from depression in the general population.

Relevant data for identifying people with health risks can also be collected through popular social media, such as Facebook and Twitter. In a study conducted by Gittelman et al. (2015), for example, Facebook “likes” were collected to predict mortality, using Principal Component Analysis (to reduce the high-dimensional predictor space), followed by Bootstrap Regression (a regression technique that is less vulnerable to violations of statistical assumptions). The model with Facebook likes and basic demographic features was better than either one alone. In this context, advances in automated text analysis are also promising. For example, Pestian and Nasrallah (2010) explored whether text mining techniques could be used to identify genuine suicide notes. In their study, the predictive model proved to be better in this task than mental health professionals (78% vs. 63%, respectively).

Technological progress allows to continuously acquire data on an individual level. For example, several studies suggest that depressive symptoms could possibly be monitored unobtrusively, without explicit user input, using predictive models built from smartphone log-file data. Building on a pilot study of Burns et al. (2011), Saeb et al. (2015) analyzed the connection between users’ daily movement patterns, estimated from their phones’ GPS records, and the presence of depressive symptoms. Various features of the users’ movements were derived from these data, such as the variance in location visits. To determine client-specific locations of importance, Saeb et al. (2015) used a variant of the K-means clustering algorithm (Arthur and Vassilvitskii, 2007), an unsupervised learning technique that finds optimal partitions of multivariate data. Patterns in location visit clusters were found to be correlated with depressive symptoms as assessed by a self-report questionnaire. As a result, the presence of clinical levels of depression could be detected with a high degree of accuracy. The relevance of mobility and activity monitoring was further demonstrated in studies of Kim et al. (2015) and Osmani et al. (2013).

Several other studies explored the usefulness of smartphone logfiles to detect mental health problems in healthy participants. Doryab et al. (2014) used phone call logs and phone light sensor data as proxies of users’ social and sleeping behavior. Findings suggested that clear changes in outgoing calls were associated with changes in depressive symptoms. Similar findings were reported by Mesty et al. (2015), who concluded that predicting a possible range of depression, stress and anxiety is more feasible than predicting absolute values. Demirci et al. (2015) investigated the relationship between smartphone usage and sleep quality, depression, and anxiety in a student sample. Increased smartphone usage was found to be related to more symptoms of depression, anxiety, and lower sleep quality. Similar results were reported by Ma et al. (2012) and Likamwa et al. (2013), who were able to predict mood scores with an accuracy of 50% and 93%, respectively, based on variables that were collected on the smartphone.

The predictive accuracy of the models may be further increased by capturing contextual information, either through unobtrusive assessment or through prompted self-report questionnaires. Panagiotakopoulos et al. (2010) collected such data from 27 anxiety clients over 30 days, five times a day (allowing participants to make additional ratings freely at any time). Using Bayesian networks, they were able to infer stress levels from EMA data with an average accuracy of 84%.

Voice recordings, which can be unobtrusively sampled through the microphones of smartphones, might provide another source for health screening applications. Lu et al. (2012) analyzed voice recordings to estimate stress levels and found changes in pitch to be correlated with stress levels. Chang et al. (2011) showed that voice analysis programs can estimate current affection or stress levels, using smartphone-based voice analysis software. Van der Sluis et al. (2012) found that, in addition to pitch, other speech features such as amplitude, zero crossings, power, and high-frequency power are also useful predictors of stress levels of patients with post-traumatic stress disorder (PTSD).

As suggested by the studies discussed, Type 1 models most probably will find their way into future e-mental health in the form of smartphone applications. For this, it is important that the promising preliminary findings, which are often based on small-sample pilot studies, are corroborated by adequately powered independent replication studies. The findings of Likamwa et al. (2013), for example, could not be replicated in a follow-up study (Asselbergs et al., 2016).

4.2. Model Type 2: short-term predictions during treatment

Type 2 models target the prediction of patients’ states as they evolve during the intervention. Studies have shown that tracking patients’ states can improve treatment adherence and treatment outcomes by providing a feedback cycle (Lambert, 2010; Miller et al., 2006). For this, traditional self-report questionnaires can be administered regularly, for instance, every two weeks. If necessary, these coarse-grained assessments can be complemented by more fine-grained daily
Table 2: Predictive models used in e-mental health.

| Model Type | Study | Method | Data | Prediction | Comment |
|------------|-------|--------|------|-----------|---------|
| Postpartum depression | Ankarali et al. (2007) | Classification tree, logistic regression | Social-demographic data | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Time-to-event data | Van der Werf et al. (2006) | Sequential-phase model | Estimation | The relationship among mobile phone-GPS sensor features, EMA ratings, and the PHQ-9 scores is analyzed. |
| Facebook likes | Gittelman et al. (2015) | Principal Component Analysis, linear regression | Facebook likes | Estimation | Facebook likes indicate habits and activities, which are used to predict life expectancy. |
| Suicide notes | Pestana and Nazareth (2010) | Support vector machine, Bernoulli model | Estimation | Certain traits are extracted from the user's data to estimate the probability of suicide. |
| EMA | Both et al. (2010), Both and Hoogendoorn (2011), Both, Hoogendoorn and Klein (2012) | Dynamic model | Correlation between EMA and contextual data | Estimation | The relationship among mobile phone-GPS sensor features, EMA ratings, and the PHQ-9 scores is analyzed. |
| GPS movement data | Deer et al. (2014) | K-means for location clustering, elastic net regularization for prediction of depressive symptoms | Estimation | The relationship among mobile phone-GPS sensor features, EMA ratings, and the PHQ-9 scores is analyzed. |
| Activity measures | Kim et al. (2015) | Hierarchical model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Sleep quality, depression and anxiety score | Pestana and Nazareth (2010) | Linear model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Doryab et al. (2014) | Tertius algorithm (Flach and Lachiche, 2001) | Estimation | The relationship among mobile phone-GPS sensor features, EMA ratings, and the PHQ-9 scores is analyzed. |
| Smartphone data | Mestry et al. (2015) | Correlation | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Sleep quality, depression and anxiety score | Demirci et al. (2015) | Correlation | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Ma et al. (2012) | Factor graph | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Likamwa et al. (2013) | Regression model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Panagiotakopoulos et al. (2010) | Bayesian network | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Depressive symptoms | Lu et al. (2012) | Mixture of hidden Markov models | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Emotion recognition | Chang et al. (2011) | Support vector machine | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Stress level of post-traumatic stress disorder patients | Van der Sluis et al. (2012) | Dynamic model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Asselbergs et al. (2016) | Dynamic model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Noble (2014) | Dynamic model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |
| Mood | Patten (2005) | Dynamic model | Estimation | The probability of postpartum depression in a woman, along with factors that are associated with the severity of depression. |

(continued on next page)
| Model type | Study | Method | Data | Prediction | Comment |
|------------|-------|--------|------|------------|---------|
| 2          | Bremer et al. (2017) | Diary data | Data mining and Bayesian regression | Data | To infer the current mood. |
| 2          | Osmani et al. (2013) | Smartphone measures | Hierarchical Poisson regression modeling | Current mood | To correlate smartphone-measured activity with depressive state. |
| 3          | Karyotaki et al. (2015) | Demographics | Logistic regression | Estimation of drop-out risk factors | To identify drop-out risk factors. |
| 3          | Kegel and Flückiger (2014) | Self-esteem, mastery, clarification, global Alliance | Logistic regression | Treatment dropout | To predict treatment dropout. |
| 3          | Meulenbeek et al. (2015) | Socio-demographic, personal, and illness-related variables | Logistic regression | Treatment outcome | To predict treatment outcome. |
| 3          | Proudfoot et al. (2013) | Perceived self-efficacy questionnaire | Logistic regression | Symptom improvement | To predict symptom improvement. |
| 3          | Van et al. (2008) | Hamilton Rating Scale for Depression | Logistic regression | Treatment failure | To predict early improvement for therapy outcome. |
| 3          | Priebe et al. (2011) | Therapeutic relationship | Logistic regression | Treatment outcome | To predict treatment outcome. |
| 3          | Van Gemert-Pijnen et al. (2014) | Login frequency | Logistic regression | Prediction of outcome after therapy | To predict improvement after therapy. |
| 3          | Whitton et al. (2015) | Collected data about used program features | Logistic regression | Outcome prediction | To predict treatment outcome. |
| 3          | Bennett et al. (2011) | Session based outcome questionnaire | Various methods | Treatment resistance | To predict treatment resistance. |
| 4          | Kessing (1999) | ICD-10 Depression rating | Cox-regression | Relapse risk | To predict relapse risk. |
| 4          | Busch et al. (2012) | Demographics, medication, clinical data | Hierarchical logistic regression | Alcohol relapse risk | To predict alcohol relapse risk. |
| 4          | Farren and McElroy (2010) | Demographics, previous drinking characteristics, comorbidity | Logistic regression | Alcohol relapse risk | To predict alcohol relapse risk. |
| 4          | Pederson and Hesse (2009) | Demographics, previous drinking characteristics | Logistic regression | Alcohol relapse risk | To predict alcohol relapse risk. |
| 4          | van Voorhees et al. (2008) | Mood, social and cognitive characteristics | Logistic regression | Alcohol relapse risk | To predict alcohol relapse risk. |
| 4          | Gustafson et al. (2014) | GPS position | Regression trees | Alcohol relapse risk | To predict alcohol relapse risk. |
| 4          | Chih et al. (2014) | Weekly assessed EMA ratings | Bayesian network model | Relapse risk | To predict relapse risk in coming week. |
| 4          | Aziz et al. (2009) | Ambient measures | Bayesian network model | Relapse risk | To predict relapse risk in coming week. |

*Studies are listed in the order in which they were presented in the main text.*
smartphone-based assessments (Torous et al., 2015). The distinction between Type 1 and Type 2 models is not always clear-cut. When Type 1 models and applications are used during treatment to make short-term predictions of health states to inform decisions on the course of treatment, the models should be classified as Type 2. Type 2 models can also be more advanced, since more detailed data from the ongoing treatment is available to inform the modeling process (i.e., log-data of web-based treatment platforms, therapist input, and extensive diagnostic data).

Type 2 models explicitly consider the patient in the context of treatment. Daugherty et al. (2009) used oscillating differential equations to predict hypomanic, stable, and depressive episodes in patients diagnosed with bipolar disorder. Their model incorporates the effects of medication treatment and behavior coupling between patients. Medication treatment slows the rapid mood changes and dampens the amplitude of mood oscillations. Patients with a similar cycle synchronize over time, whereas patients with an opposite cycle remain in an opposite cycle.

Becker et al. (2016) exploit phone usage data and previous EMA measurements to predict the reported value for mood for an experiment with a group of students. For phone usage data they include app usage, activity levels, number of phone calls, and number of text messages sent. They apply a variety of data mining techniques including Support Vector Machines, Lasso and Linear regression, and Bayesian Hierarchical Regression. They create a general model as well as user level models and are able to achieve a root mean squared error of 0.83 using the Lasso Regression approach with the user level approach.

Similarly, van Breda et al. (2016) try to predict mood for the next day using self-reported EMA data of depressed patients during previous days. They optimize the number of days taken into account in their features to predict the mood value for the next day. They build individual patient models and use a combination of linear regression models (using a so-called bagging approach).

Another possibility to understand clients’ mood trajectories are online diaries that are part of online interventions. Bremer et al. (2017) demonstrate that the clients’ mood on a specific day can be inferred based on their diary data. In a 2-phase modeling approach, they employ techniques from text mining to first extract activities that were likely described in the diary entries by the user. In the subsequent step, the set of relevant activities is used to successfully predict the clients’ mood using an ordered logit model.

Another clear Type 2 model is the ‘virtual patient’ model of Both et al. (2010). This model describes the relationships between different client states such as mood, thoughts, and coping skills as a system of differential equations, which allows simulation of the development of patient states over time. As discussed in our overview of predictive modeling approaches, such a model is developed using a top-down approach, where domain knowledge is formalized in a computational model. With the model, the development of internal states can be simulated to predict mood during treatment. The model incorporates hypothesized working mechanisms of different therapeutic interventions, such as activity scheduling or cognitive behavior therapy (CBT), to enable simulations of the potential effects of these interventions, which in turn could be used to identify the intervention that could benefit the patient the most (Both and Hoogendoorn, 2011). To do this effectively, additional model states are necessary to account for external variables that are not controllable by the client (Both, Hoogendoorn and Klein, 2012). Similar approaches, using different computational techniques, have been proposed by Touboul et al. (2010), Demic and Cheng (2014), Noble (2014), and Patten (2005).

4.3. Model Type 3: predicting treatment outcome

Type 3 models predict treatment outcome (including drop-out) during active treatment. Rather than focusing on the relationship between short-term symptom dynamics and treatment decisions, these models aim to predict the outcome of the full therapeutic process (i.e., the post-test results) from data that are collected before and during treatment. This outcome has two aspects: 1) the change in health symptoms (which is typically known as the vector of continuous mental health questionnaire scores), and 2) whether or not treatment was provided as intended (which is typically a binary variable, flagging the treatment as drop-out for patients who did not complete a pre-determined minimum percentage of treatment).

Using regression techniques, clinical researchers have been exploring Type 3 models, using predictors that were collected at baseline (pre-test) assessments (e.g., DeRubeis et al., 2014; Huibers et al., 2015; Karyotaki et al., 2015; Kegel and Flückiger, 2014; Meulenbeek et al., 2015), mental health questionnaire responses that were collected during treatment (e.g., Proudfoot et al., 2013; Van et al., 2008), or repeated measures of the therapeutic relationship between the patient and the therapists (e.g., Priebe et al., 2011). In these studies, researchers tend to focus more on the importance of individual predictors on a population level (i.e., risk factors), rather than on the predictive power of the model as a whole.

Recent studies suggest that logfiles of electronic treatment delivery systems might also provide useful detailed predictors of treatment outcome. The number of logins, actions per login, completed modules, and time spent are metrics that can identify differences between adherers and non-adherers (Donkin et al., 2013). Van Gemert-Pijnen et al. (2014), for example, found user login frequency to be correlated with depressive symptoms after treatment. Whitton et al. (2015) examined the correlation between the usage of program features and outcomes in a study of a web-based self-help intervention for symptoms of depression and anxiety. The usage of diary functions and SMS reminding was not correlated with outcome, nor did the number of interventions (both started and completed) have an impact on symptom reduction. The symptom tracking functionality, which enabled users to track their improvement, also had no influence on the final treatment result. However, the use of the reminder function of the symptom tracking tool had a positive influence on the outcome.

Advanced data mining techniques have been used to build Type 3 models as well. For example, Bennett et al. (2011) used (naïve) Bayesian classifiers and Random Forest decision trees (see Breiman, 1984) to predict treatment outcomes from health questionnaires that were collected at each treatment session. Early reductions in symptom levels were found to be a significant predictor of treatment outcome. Using 10-fold cross-validation, the predictive accuracy of the various classification models varied between 60% and 76%.

Early detection of clients at high risk for treatment resistance could also be helpful, for example, to decide on the optimal level of therapist guidance in a web-based treatment. Perlis (2013) used self-reported socio-demographic and clinical variables to predict treatment resistance. With multivariate models such as logistic regression, naïve Bayes and Support Vector Machines, the authors achieved an AUC of 0.71, indicating fair predictive performance.

An approach that utilizes free text written by patients is proposed in Hoogendoorn et al. (2017). They extracted features from the text messages sent by patients suffering from an anxiety disorder to their therapist as part of an anxiety treatment. Features included word usage, sentiment, response rate, length of email, phrasing of the emails, and topics that patients wrote about in their emails. They were aimed at predicting a reliable improvement in the so-called Social Phobia Measure and were able to achieve AUCs of around 0.83 halfway through the therapy and similar scores at the end, which is significantly better than using baseline data only.

4.4. Model Type 4: models for relapse prediction

Type 4 models focus on predicting the risk of relapse, that is, the re-occurrence of symptoms in the long term. Typically, the relapse risk determines the amount of aftercare required. The design of interventions and treatment platforms for relapse prevention training is similar.
to that of treatment platforms and interventions used in regular treatment (Barnes et al., 2007; Holländere et al., 2013; Lobban et al., 2015). To stabilize the condition of patients, aftercare can be provided in the form of screenings and interventions.

As with Type 3 models, clinical researchers have used traditional regression techniques such as logistic regression to predict relapse in patients diagnosed with a variety of disorders, including depression (e.g., Kessing, 1999), bipolar depression (Busch et al., 2012), and alcohol misuse (Farren and McElroy, 2010; Pedersen and Hesse, 2009). More advanced statistical techniques have also been applied. For instance, Patten (2005) analyzed data from several clinical depression trials to estimate the parameters of a Markov model, and Van Voorhees et al. (2008) identified risk factors through regression tree modeling.

Type 4 models can support mental health specialists to assess the risk of relapse in their patients, to scale-up after-care when needed. More likely, however, is that these models will then be able to take current contextual information of the patient into account, resulting in more options to personalize interventions. According to Juarascio et al. (2015), context-aware interventions are expected to appear for treatment and relapse prevention for a variety of health problems, due to the rapid developments in sensor technology and mobile applications. Gustafson et al. (2014) and Chih et al. (2014) implemented Bayesian network modeling in an EMA/EMI smartphone application aimed at reducing relapse for recovering alcohol-dependent individuals. In this app, relapse-prevention interventions are triggered when the risk of relapse occurrence is estimated to be high based on the EMA ratings (including unobtrusive contextual variables). In a validation study, the model achieved an AUC of 0.91 in predicting relapse in the following week. A similar EMA/EMI approach was proposed by Aziz et al. (2009), who developed a rule-based support agent that monitors client conditions to trigger EMI when the risk of relapse is predicted to be high, based on the client’s mental health state, social interactions, estimated substance use, and physiological conditions.

5. Discussion

In this paper, we provided a brief introduction to predictive modeling methods, introduced a common framework for understanding applications of predictive modeling in mental health, and applied the framework to categorize published mental health research in which predictive modeling techniques were applied. Doing so, we aimed to contribute to the development of a common language between clinical researchers and members of the data mining community.

With this framework, opportunities can be identified to extend and improve an emerging field. For instance, our preliminary literature review suggests that e-mental health researchers should perhaps focus more on the validity of model predictions rather than the more traditional goal of identifying specific predictors. Demonstrating the relevance of a specific predictor has theoretical relevance. However, single predictors rarely provide a basis for predictions of variables that are relevant to clinical practice. In practice, often only a fraction of the variance of the target variable of interest is explained. In comparison with clinical researchers, data miners are more used (and more tolerant) to focusing on the accurate prediction of high-value target variables while focusing less on the specific predictors that play a role in this prediction. It would be interesting to learn more about the relative merits of this approach in e-mental health applications.

We also identified room for improvement in the proper application of predictive modeling methodology. We found that several studies do not use independent test sets to evaluate experimental modeling techniques. Proper testing (validation) of models is of utmost importance to investigate the generalizability of proposed models. Patterns in available data can often be modeled well, but more importantly testing a model on to unseen data is critical to test generalizability, as explained in the first section of this paper. The current wealth of data collected in e-mental health applications provides many opportunities to use test sets for model validation, and we urge researchers to apply this technique more. Furthermore, utilization of predictive modeling can have pitfalls such as estimation of a spurious correlation due to an inadvertently introduced bias or a systematic error during study design (Pearl, 2009; Sjölander, 2009). Especially, with the larger amount of data collected during online intentions this effect can magnify (Khoury and Ioannidis, 2014). This emphasis the need for a common language and understanding of the subject.

Some identified studies, relevant to mental health care, could not be readily classified with the proposed framework. These studies focused on the modeling of mental processes, without explicit links to psychotherapeutic intervention. By focusing on the therapeutic purpose of the modeling, the framework stresses the need for empirical validation of model performance in clinical settings, which is more stringent (i.e., prediction errors are more serious when they provide the basis for clinical treatment decisions). Nonetheless, these studies apply predictive modeling to processes that are relevant to mental health care. The framework may need to be extended to incorporate this type of research more easily.

We would like to stress that our review of available predictive modeling mental health research should not be considered exhaustive. Although we feel that the papers identified are representative for the field, we cannot rule out that important publications were missed, especially since the number of modeling studies seems to be rising each month. In our view, this illustrates the need for conceptual frameworks such as the one proposed in this paper, so that the growing body of research initiatives can be more easily understood, categorized, and evaluated.

In this paper, we could only provide a brief overview of data mining methods. While our framework should promote a shared understanding of high-level research goals, we also acknowledge that productive collaboration in this field requires researchers to gain a deeper and better understanding of data mining methods. To promote this, we suggest that authors make a deliberate effort to thoroughly explain core aspects of data mining methods applied when they publish predictive modeling mental health care research (i.e., to contribute to the development of a common language by providing basic descriptions of core methods to which clinical researchers may be less familiar).

E-mental health is a rich interdisciplinary research field that enables many new research approaches, of which predictive modeling appears to be most promising. The framework proposed in this paper might serve to bridge the conceptual gap between psychologists and predictive modelers. The framework provides a common language for classifying predictive modeling mental health research, which may help to promote constructive and productive interdisciplinary research and to identify new research opportunities.

Conflict of interest

There are no conflicts of interest.

References

Abu-Mostafa, Y., 2013. Learning From Data. Caltech. AMLBook Retrieved from. http://work.caltech.edu/lectures.html.
Allman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. 46 (3), 175–185.
Ankarali, H., Canan, A., Akkus, Z., 2007. Comparison of logistic regression model and classification tree: an application to postpartum depression data. Expert Syst. Appl. 32 (4), 987–994, 2007. http://dx.doi.org/10.1016/j.eswa.2006.02.022.
Armey, M.F., Schatten, H.T., Haradhvala, N., Miller, I.W., 2015. Ecological momentary assessment (EMA) of depression-related phenomena. Curr. Opin. Psychiatry 4, 21–25. http://dx.doi.org/10.1097/JCO.2015.01.002.
Arthur, D., Vassilvitskii, S., 2007. K-means++. the advantages of careful seeding. In: Presented at: Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1027–1035 (Philadelphia, PA, USA).
Asselbergs, J., Ruwaard, J., Efjys, M., Schrader, N., Sijbrandij, M., Ripper, H., 2016.
Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an exploratory study. J. Med. Internet Res. 18 (3), e72. http://dx.doi.org/10.2196/jmir.5505.

Ariz, A.A., Klein, M.C.A., Treur, J., 2009. Modeling an ambient agent to support depression relapse prevention. In: Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT Workshops 2009. vol. 3. pp. 335–340. http://dx.doi.org/10.1109/WI-IAT.2009.5224765.

Barnes, C., Harvey, R., Mitchell, P., Smith, M., Wilhelm, K., 2007. Evaluation of an online relapse prevention program for bipolar disorder. Dis. Manag. Health Out. 15 (4), 215–224. http://dx.doi.org/10.1016/j.dmk.2007.05.003.

Becker, D., Bremer, V., Funk, B., Ansellberg, J., Ripper, H., Ruwarz, J., 2016. How to predict mood? Delving into features of smartphone-based data. In: American Conference on Information Systems! AMCIS 2016: San Diego, pp. 2016.

Bennett, C., Douh, T., Bragg, A., Luellen, J., Van Regenmorter, C., Lockman, J., Reiserer, R., 2011. Data mining session-based patient reported outcomes (PROs) in a mental health setting: toward data-driven clinical decision support and personalized treat- ment. In: Proceedings - 2011 1st IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology, HISB 2011, pp. 229–236. http://dx.doi.org/10.1109/HISB.2011.15.

Bloom, D.E., Caferro, E.T., Jané-Etloi, E., Abrahams-Gessel, S., Bloom, L.R., Fatima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., Weinstein, C., 2011. The Global Economic Burden of Noncommunicable Diseases. World Economic Forum, Geneva.

Bolger, N., DeLongis, A., Kessler, R.C., Schilling, E.A., 1989. E

Barnes, C., Harvey, R., Mitchell, P., Smith, M., Wilhelm, K., 2007. Evaluation of an online relapse prevention program for bipolar disorder. Dis. Manag. Health Out. 15 (4), 215–224. http://dx.doi.org/10.1016/j.dmk.2007.05.003.
