Original Article

Mindful Machine Learning

Using Machine Learning Algorithms to Predict the Practice of Mindfulness

Sebastian Sauer,[1,2] Ricardo Buettner,[1] Thomas Heidenreich,[3] Jana Lemke,[4] Christoph Berg,[1] and Christoph Kurz[5]

[1] Institute of Business Psychology/ Institute of Management & Information Systems, FOM University of Applied Sciences, Munich, Germany
[2] Brain, Mind, and Healing Program, Samueli Institute, Alexandria, VA, USA
[3] Faculty of Social Work, Health, and Nursing, Esslingen University of Applied Sciences, Germany
[4] Institute of Transcultural Health Sciences, Viadrina University, Frankfurt/Oder, Germany
[5] Institute of Health Economics and Health Care Managemant, Helmholtz Zentrum München, Germany

Abstract: Mindfulness refers to a stance of nonjudgmental awareness of present-moment experiences. A growing body of research suggests that mindfulness may increase cognitive resources, thereby buffering stress. However, existing models have not achieved a consensus on how mindfulness should be operationalized. As the sound measurement of mindfulness is the foundation needed before substantial hypotheses can be supported, we propose a novel way of gauging the psychometric quality of a mindfulness measurement instrument (the *Freiburg Mindfulness Inventory*; FMI). Specifically, we employed 10 predictive algorithms to scrutinize the measurement quality of the FMI. Our criterion of measurement quality was the degree to which an algorithm separated mindfulness practitioner from nonpractitioners in a sample of N = 276. A high predictive accuracy of class membership can be taken as an indicator of the psychometric quality of the instrument. In sum, two findings are of interest. First, over and above some items of the FMI were able to reliably predict class membership. However, some items appeared to be uninformative. Second, from an applied methodological point of view, it appears that machine learning algorithms can outperform traditional predictive methods such as logistic regression. This finding may generalize to other branches of research.

Keywords: machine learning, predictive modeling, mindfulness, measurement, questionnaire

The idea that mindfulness may be able to alleviate a variety of psychosomatic symptoms has gained widespread currency in recent years (Chiesa, Calati, & Serretti, 2011; Hofmann, Sawyer, Witt, & Oh, 2010). Mindfulness as a state of consciousness may be characterized by an attentional stance that focuses on the present moment ("presence") with a nonjudgmental attitude toward all perceptions and sensations ("acceptance") that may be part of the actual experience (Sauer, Walach, et al., 2013). It implies a sense of being in the present, underlining the vivid awareness of sensory and mental experience, and calls for a voluntarily disruption of automatic mental processes of cognitive evaluation such as "good," "bad," "wanted," or "unwanted," as well as the suspension of labeling experience with emotional tones.

Several meta-analytical reviews have investigated the relations between mindfulness, health, and well-being and have corroborated the health-relevant effects of mindfulness (Eberth & Sedlmeier, 2012; Sedlmeier et al., 2012; Zainal, Booth, & Huppert, 2013). Symptoms for which health-relevant effects of mindfulness have been found beneficial include depression, anxiety, stress, pain, substance abuse, personality disorders, and elevated blood pressure, among others. Results from neurological research have begun to reveal the biological basis of mindfulness and have linked mindfulness back to several brain regions known to be involved in attention and emotion-regulation processes. These brain areas include the cingulate cortex or the insula (Farb, Anderson, & Segal, 2012; Tang & Posner, 2013; Teper & Inzlicht, 2013). A recent review suggests that mindfulness may buffer age-related neural decay, thus preventing the brain from aging prematurely (Gard, Hölzel, & Lazar, 2014).

In the light of mindfulness being a promising area of psychological research, rigorous examination is in place for scrutinizing the claims that mindfulness-based health interventions have the potential to improve health. One condition to test the health-related effects of mindfulness is that it can be validly measured. To date, more than a dozen instruments have been published, apparently confirming the notion that mindfulness can be measured (Baer, 2011; Sauer, Walach, et al., 2013).

Yet, no consensus on how mindfulness should be operationalized has emerged out of this variety of instruments. On the contrary, severe criticism of the common approaches has been put forth (Grossman, 2008, 2011). One criticism hinges on the fact that the instruments build on diverging, even contradicting ideas of mindfulness.

In the light of this, it is the aim of the current research to further the measurement of mindfulness. To this end, we suggest that the methodology that is currently employed, as sophisticated as it may be, is not sufficient and should be complemented by novel methods. Machine learning may provide such a method that is novel to psychometrics. Machine learning is an umbrella term for a dynamic array of algorithms for predictive modeling that can supplement more "traditional" statistical approaches. Albeit not recognized much in psychology to date, in other fields such as gene expression, machine learning algorithms are now being considered a standard instrument (Dudoit, Fridlyand, & Speed, 2002; Shipp et al., 2002; Ye et al., 2003).

It is still a subject of debate as to which of the many algorithms can be considered superior to others, or, strictly speaking, which algorithms fit with which problems and data. A practical recommendation (Hastie, Tibshirani, & Friedman, 2009) is, when in doubt, to employ several algorithms and to aggregate the results. This approach was taken in the present paper.

More specifically, we used 10 different predictive algorithms to test whether the items from a well-known mindfulness inventory would be able to predict a behavioral criterion, that is, whether the participants were practicing mindfulness on a regular basis or not. This correspondence hinges on the rationale that individuals undergoing regular mindfulness training should show a higher mindfulness trait level compared to individuals who do not. Thus, we strive to improve the measurement of mindfulness by testing the predictive validity of mindfulness items. Testing the psychometric validity using external behavioral criteria is by no means new. However, we used a variety of machine learning algorithms, an approach that is novel to the area of mindfulness and to many other areas of psychology, as well.

In sum, our approach consists of testing the psychometric quality of one mindfulness instrument using a method novel to psychometrics, that is, machine learning techniques. Our first research question is whether the items of the instruments predict a behavioral criterion (i.e., regular mindfulness practice). The degree to which the items correctly identify class membership (practice yes/no) is taken as an indicator of the psychometric quality of the item. Secondly, we were interested in whether the machine learning algorithms would outperform more "classical" methods such as logistic regression. Thus, we contrasted the machine learning algorithms with regression as a classical method in psychometrics in order to evaluate the purported predictive advantage of machine learning.

Current Mindfulness Measurement Practices

Utilizing psychometric scales is the standard approach for measuring mindfulness. Consequently, a substantial and growing number of scales have been published (Baer, 2011; Sauer, Walach, et al., 2013). The advantages of the psychometric approach are straightforward: convenient and quick application, well-known methodology, and empirical support (Mars & Abbey, 2010).

At the same time, there are shortcomings that may place considerable limits on the validity of the psychometric method (Grossman, 2008). One argument is that a wide array of conflicting definitions exist. For example, the concept of mindfulness put forward by Langer and Moldoveanu (2000) differs substantially from Brown and Ryan (2003) or from Kabat-Zinn (Bishop et al., 2004), to name some of the most influential researchers. As a result of different conceptions and theoretical backgrounds, the questionnaires differ greatly. The same is true for the number of supposed factors and the aspects deemed central to the construct (Grossman, 2011). The most widely used psychometric scales include the *Five Factor Mindfulness Inventory* (FFMQ; Bohlmeijer, Ten Klooster, Fledderus, Veehof, & Baer, 2011), the *Mindfulness Attention and Awareness Scale* (MAAS; Brown & Ryan, 2003), and the *Freiburg Mindfulness Inventory* (FMI; Kohls, Sauer, & Walach, 2009). We chose the FMI because it is a well-known instrument that combines two features. First, it incorporates both attentional and emotional components in the construct of mindfulness as opposed to the MAAS, for example. Second, it is shorter than some other instruments; for example, the FFMQ consists of 39 items.

Method

Sample

For increased sample size, we included two samples in this study; total sample size was $N = 276$. As Sample 2 was comparatively small, we refrained from analyzing each sample individually and report the combined results only. Instead, we used a statistical procedure (see below) as replication procedure rather than replicating the analysis on the second sample. Data can be accessed in Electronic Supplementary Material (ESM) 5.

Sample 1

Sample 1 ($N = 201$) was collected as part of an unpublished online study investigating the relations between mindfulness, health, and emotion. About two thirds of the sample (n = 129) did not indicate that they engaged in any mindfulness training, whereas n = 72 individuals (36%) reported practicing mindfulness on a regular basis. Examples of types of mindfulness training included Buddhist meditation or Yoga. Their mean ages were 35 years for nonpractitioners (SD = 13) and 39 years for practitioners (SD = 11); n = 141 (70%) persons were female and n = 59 (29%) persons were male (one missing value). The inclusion criterion for the mindfulness group was "regular mindfulness meditation practice."

Sample 2

Sample 2 (N = 75) was comprised of one half (n = 38) expert mindfulness practitioners (21 female, 17 male) and an age- and sex-matched half of (n = 37) nonmindfulness practitioners (28 female, 10 male). Practitioners were trained in different Buddhist mindfulness traditions. The mean age was 51 years in both groups (SD = 10 years). Care was taken to obtain distinct groups (i.e., mindfulness practitioners vs. nonpractitioners). The inclusion criteria were at least 5 years of ongoing mindfulness practice. See Sauer et al. (2012) for details.

Analytical Procedure

The analytical procedure consisted of predicting class membership (mindfulness practitioners vs. nonpractitioners) using 10 different algorithms. The predictive accuracy (see below) was taken as an indicator of the psychometric quality of the mindfulness measurement instrument.

One of the dangers of data modeling is overfitting. Overfitting refers to a situation in which a statistical model is excessively complex and thereby describes random noise as a systematic pattern. A possible solution to this problem is to split the data into a training set and a test set (Hastie et al., 2009). The model is then built on the training set only; for model results, the test set is used. We employed a more stringent variant known as the repeated k-fold cross-validation procedure (Kuhn, 2008). Here, the data are split into k distinct blocks of similar size. In our case, we split the data into k = 10 blocks and repeated this procedure for n = 10 repetitions. Then, in n runs, each of the k blocks was left out, and the model was fit to the remaining 9/10 of the sample. The model results, built on the "left out" 1/10, were then averaged.

All algorithms were taken from the R package *caret* (Kuhn, 2008). The statistical software R 3.0.1 on a Windows 7 computer was used as the statistical environment. Item scores were standardized to fall within the interval of 0 to 1. We used the default values of the statistical procedures (see R syntax for details, ESM 4).

To gauge the predictive quality of the mindfulness instrument, we computed the accuracy of predicting the class membership. More specifically, we split the predictive accuracy into the two aspects of sensitivity and specificity (Altman & Bland, 1994). Specificity is defined as the rate of mindfulness practitioners correctly identified as such ("true positives"). On the other hand, specificity refers to the rate of correctly identified nonpractitioners ("true negatives"). In addition, we computed Cohen's Kappa (Fleiss & Cohen, 1973) as this statistic corrects for chance concordance. For this statistic, 0 indicates no concordance between predicted and actual values. According to Landis and Koch (1977), $.41 < \kappa < .60$ can be seen as moderate concordance, with smaller values indicating only slight or poor concordance and higher values indicating substantial concordance.

Instruments

The psychometric instrument under investigation was the *Freiburg Mindfulness Inventory* short version (FMI; Kohls et al., 2009). The instrument was designed to measure mindfulness as a stable trait. A trait refers to a personality aspect consistent in various situations and stable for at least some weeks (Larsen & Buss, 2008). Whereas the scale was first developed to be a unidimensional scale, a recent study indicated a two-factor solution (Kohls et al., 2009). It has been validated not only by classical psychometric methods such as exploratory and confirmatory factor analyses but also using item response theory (Sauer, Walach, Offenbächer, Lynch, & Kohls, 2011). The instrument in its current form consists of 14 items with four answer options. The instrument has been employed in a number of studies (Eisendrath et al., 2008; Leigh, Bowen, & Marlatt, 2005; Sauer, Walach, & Kohls, 2010).

Predictive Models

The following 10 models were employed: Generalized linear model (*glm*; i.e., logistic regression), Neural networks (*nn*), Boosted classification trees (*ada*), Extreme learning machines (*elm*), Stochastic Gradient Boosting machine (*gbm*), k nearest neighbors (*knn*), Quadratic discriminant analysis (*qda*), Random Forests (*rf*), Support vector machine with linear kernel (*SvmLinear*), and Support vector machine with polynomial kernel function (*svmPoly*). We chose models based on their predictive track record (see ESM 1 for details).

Results

As expected, the overall mindfulness level of the mindfulness practitioner group was substantially higher than that

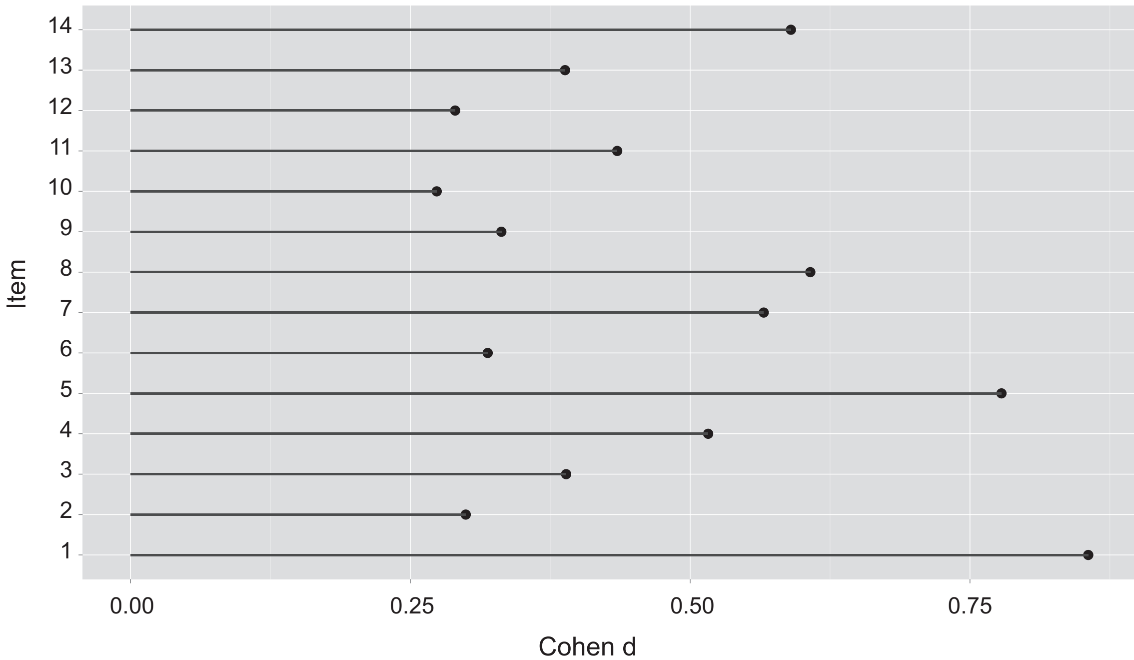

Figure 1. Cohen's *d* (mindfulness trait level differences) for each item. This figure shows the effect size difference (Cohen's *d*) in individual mindfulness levels between mindfulness practitioners and nonpractitioners.

of the nonpractitioners (the average across all items; see Figure C1 in ESM 3). Mindfulness trait levels differed statistically significantly between the practitioners and nonpractitioners, respectively, $t(241.61) = -7.37$, $p < .001$, $d = -.85$ (large effect; $-1.13 < d < -.56$). The picture was similar at the item level (cf. Figure C2). Figure 1 shows the effect size of the differences in mindfulness levels between the two groups for each item. The items that separated the mindfulness practitioners most strongly from the nonpractitioners in terms of Cohen's *d* were items 1, 5, and 14 (medium to strong effect sizes). Items 2, 10, and 12 showed a small effect size. Descriptive statistics for all items (Table B1), as well as a visualization of the correlation matrix (Figure C3), can be found in the ESM 2 and 3, respectively.

For the predictive models, the mean sensitivity across all 10 models was good ($M = .76$; $Mdn = .81$; $SD = 0.24$). The mean specificity was acceptable ($M = .67$; $Mdn = .70$; $SD = 0.23$. These values must be judged against the baseline rate of nonpractitioners of .60. Thus, the ratios were 1.27 (average sensitivity to baseline) and 1.12 (average specificity to baseline). All models except *ada* had sensitivity values above the baseline. All models except *ada* and *elm* had specificity values above the baseline (see Figure 2). The best algorithm in terms of accuracy was, by far, the *rf* algorithm (sensitivity = .99; specificity = 1). For Cohen's kappa, the picture was similar. The mean kappa was moderate in size ($M = .44$; $Mdn = .50$). Interestingly, the range of predictive accuracy [max – min] of the 10 algorithms differed substantially (range for sensitivity: .88; range for specificity: .76). It seems clear that relying on a single algorithm could lead to poor predictive performance. However, the variability (range) for accuracy was primarily due to one algorithm, which showed poor performance (*ada*). In sum, the data provided support for our hypothesis that using an aggregated accuracy statistic would be more adequate than relying on one algorithm alone. The machine learning algorithms outperformed the "classical" models (i.e., logistic regression, Figure 3), but only if *ada* was excluded.

As the *rf* algorithm allows to gauge the relative predictive importance of the items and because this model was the most accurate, it is interesting to know the importance of the items as computed by this algorithm (see details in ESM 1). The five items with the highest importance were 3, 2, 10, 12, and 8. Interestingly, the items with the lowest values (1, 6, 7, 9, and 11) did not include Item 13, which had been flagged as problematic in previous research (Sauer, Ziegler, Danay, Ives, & Kohls, 2013).

Discussion

In this paper we tested the predictive quality of the FMI in an attempt to gauge the validity of the scale using machine learning algorithms. This novel approach was applied to complement existing approaches of the measurement of mindfulness. In sum, our results indicate that the FMI is able to distinguish mindfulness practitioners from

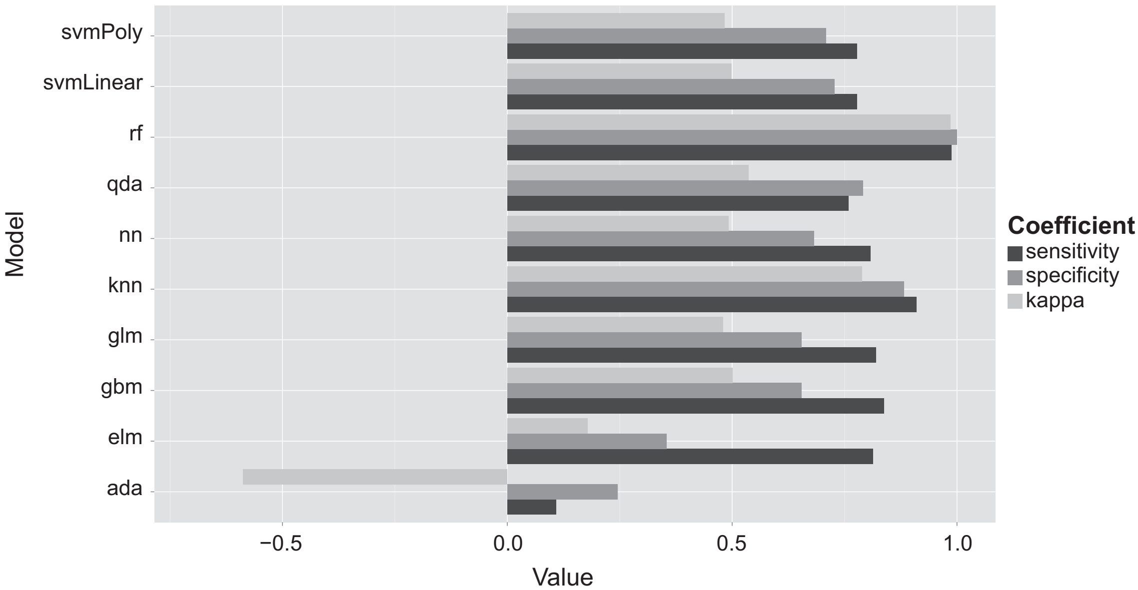

Figure 2. Results of predictive models. This figure shows sensitivity, specificity, and the kappa coefficient for each of the 10 predictive models.

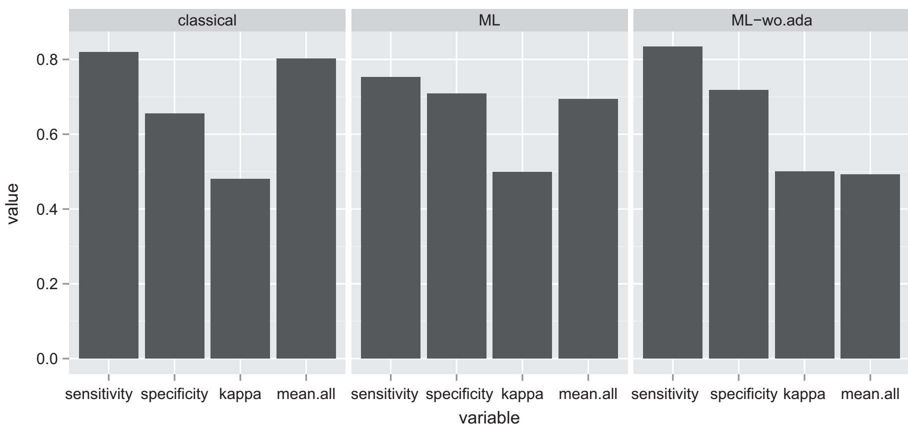

Figure 3. Comparison of predictive strengths of different types of models. The figure shows the predictive strength of (logistic) regression (*glm*) as a "classical" model (a), machine learning (ML) models (b), and ML models without *ada* (ML-wo.ada), a model which performed exceptionally weak (c). The values (i.e., sensitivity, specificity, kappa) of the ML models are based on the average performance of the rest of the models presented in this research.

nonpractitioners. This result can be taken as support for the FMI's criterion validity. However, substantial room for improvement exists. The contribution of the items was not homogeneous, and the importance of many items was weak.

The analytical method employed in this research relied on three aspects. First, we used well-acknowledged predictive algorithms novel to psychometrics. Second, we used a cross-validation approach (repeated *k*-fold cross-validation). Third, our results did not rely on only one algorithm but rather on the average of several (i.e., 10) algorithms, and thus can probably be regarded as stable.

With regard to our first research question, one can infer that the FMI items succeeded in predicting class membership. Our results support the predictive quality of the items; however, substantial room for improvement exists. Of particular interest is that only a few items substantially contributed toward separating the two groups (mindfulness practitioners vs. nonpractitioners). This result substantiates a similar notion regarding items from a different mindfulness questionnaire (Hoefling, Moosbrugger, Schermelleh-Engel, Heidenreich, & Höfling, 2011; Van Dam, Earleywine, & Borders, 2010). Item 13, which was flagged as problematic in previous research (Sauer, Ziegler, et al., 2013), was not among the most problematic items in the current study. However, given the argument that negatively coded items do not appear to simply convey the "same but reversed" effect, we propose that Item 13 still be treated as a revision candidate (Hoefling et al., 2011).

With regard to our second research question, we conclude that machine learning methods can outperform more classical approaches. In more detail, predictive algorithms differ in the accuracy with which they are able to separate mindfulness practitioners from nonpractitioners. Due to the particularities of the analytical problem and its data, some algorithms may obviously be severely misguided. In our case, *ada*, generally known as a high-performing algorithm, performed poorly. Whereas we can only speculate about the statistical reasons behind the performance problems, this result supports the hypothesis that aggregating algorithms is generally superior to relying on one standard algorithm alone, at least if the algorithm has no prior experience with the specific situation. On the other hand, *rf* performed extremely well, and the "classic psychological" algorithm, logistic regression (*glm*), yielded a somewhat intermediate performance. This is particularly interesting as psychological research makes wide use of regression analysis (Cohen, Cohen, West, & Aiken, 2013). Contrary to the customary practice, our results suggest that algorithmic pluralism is superior to algorithmic primacy. Note that the presented results rely on the replication results of the test sample.

To date, we can only speculate as to why *rf* performed so well (nearly all persons were correctly classified), whereas *ada* performed poorly (worse classification as expected by chance). A reason for this finding may be that *ada* penalizes wrong classification and puts emphasis on wrong classified cases in following rounds. So one may speculate that some participants were "odd" in the sense that they were not classifiable. Those persons have may scored highly on relevant items without pursuing mindfulness training. A reason for the success of *rf* may be that *rf* is robust as it is built on a repeated sampling rationale (both for variables and cases). As a baseline, this finding demonstrates that several models should be tested, and, if in doubt, averaged. More research is clearly needed. We recommend that future research similarly compute several models and compare the results.

Some words of caution need to be presented. First, one limitation of this study is that the quality of the online sample was limited. Limited sample quality may be likely to attenuate existing associations and the power of statistical tests. In addition, the allocation to groups in the online sample relied on self-report and was not controlled. Third, the predictive criterion (mindfulness practice or not) is only one of many plausible criteria. We would like to emphasize that this criterion is probably unable to produce a pure splitting of the groups by their mindfulness levels. The reason is that mindfulness level is less than perfectly correlated with (amount of mindfulness) practice. More specifically, we have not distinguished quantity (how much training) nor quality (what type of training) of mindfulness training. We would expect that differentiating quantity and quality of training will make the connection to mindfulness trait level even clearer. Furthermore, our operationalization of mindfulness was narrow (FMI only). Different instruments should be scrutinized to gauge the value of machine learning for mindfulness measurement.

The results and the limitations of the present research provide some ideas for new avenues of research, four of which we would like to spell out. (i) We urge researchers to make use of the predictive models that are widely used in other high-quality research branches such as gene expression or molecular biology in general. These results are not only of importance for the research on mindfulness but may help researchers in different psychological areas as well. It should be mentioned that the *R* package *caret* provides a unified (and simple) syntax for all of its approximately 180 models, thus providing comfortable and low-barrier access to researchers who do not have a background in machine learning (cf. Kuhn, 2008). We recommend selecting a bunch of methods as we did and to average the result. Random forests should definitely be among the selected algorithms. (ii) In a similar vein, we propose that different criteria be used as the target to be predicted by actual machine learning algorithms. Although mindfulness practice is perhaps one of the most straightforward criteria to choose from, other criteria such as drug intake or health incidents may provide empirically founded criteria (Leigh et al., 2005). (iii) In addition, it seems strongly warranted that future research strives to improve the instrument's items. Only some items contributed to group differentiation. The rest of the items should be considered as revision candidates. (iv) Many instruments for measuring mindfulness exist; their psychometric property should also be investigated using machine learning procedures given the successful application of the method in this study.

For the time being, we conclude that a multimodel predictive approach is advisable. Machine learning methods are useful for testing certain aspects of the validity of measurement instruments in psychology and in mindfulness research in particular. Given the widespread interest in and the promising preventative and curative health benefits of mindfulness, the measurement of mindfulness should be further investigated by means of machine learning techniques.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1015-5759/a000312

ESM 1. List of predictive models (text).
Description of the statistical rationale of the predictive models employed in the present study.
ESM 2. Table (text).
Table reporting detailed descriptive statistics.
ESM 3. 3 Figures (PDF).
Additional figures.

ESM 4. R file.
R Syntax
ESM 5. Data (csv format).
Data of sample and replication.

References

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal, 308*, 1552.

Baer, R. A. (2011). Measuring mindfulness. *Contemporary Buddhism, 12*, 241–261.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., ... Speca, M. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology Science and Practice, 11*, 230–241.

Bohlmeijer, E., Ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*, 308–320. doi: 10.1177/1073191111408231

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848. doi: 10.1037/0022-3514.84.4.822

Chiesa, A., Calati, R., & Serretti, A. (2011). Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. *Clinical Psychology Review, 31*, 449–464. doi: 10.1016/j.cpr.2010.11.003

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. London, UK: Routledge.

Dudoit, S., Fridlyand, J., & Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association, 97*, 77–87. doi: 10.1198/016214502753479248

Eberth, J., & Sedlmeier, P. (2012). The effects of mindfulness meditation: A meta-analysis. *Mindfulness, 3*, 174–189. doi: 10.1007/s12671-012-0101-x

Eisendrath, S. J., Delucchi, K., Bitner, R., Fenimore, P., Smit, M., & McLane, M. (2008). Mindfulness-based cognitive therapy for treatment-resistant depression: A pilot study. *Psychotherapy and Psychosomatics, 77*, 319–320.

Farb, N. A. S., Anderson, A. K., & Segal, Z. V. (2012). The mindful brain and emotion regulation in mood disorders. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie, 57*, 70–77.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*, 613–619.

Gard, T., Hölzel, B. K., & Lazar, S. W. (2014). The potential effects of meditation on age-related cognitive decline: A systematic review. *Annals of the New York Academy of Sciences, 1307*, 89–103. doi: 10.1111/nyas.12348

Grossman, P. (2008). On measuring mindfulness in psychosomatic and psychological research. *Journal of Psychosomatic Research, 64*, 405–408.

Grossman, P. (2011). Defining mindfulness by How poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: Comment on Brown et al. (2011). *Psychological Assessment, 23*, 1034–1040.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Elements* (Vol. 1), New York City, NY: Springer.

Hoefling, V., Moosbrugger, H., Schermelleh-Engel, K., Heidenreich, T., & Höfling, V. (2011). Mindfulness or Mindlessness? *European Journal of Psychological Assessment, 27*, 59–64. doi: 10.1027/1015-5759/a000045

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 169–183.

Kohls, N., Sauer, S., & Walach, H. (2009). Facets of mindfulness – results of an online study investigating the Freiburg mindfulness inventory. *Personality and Individual Differences, 46*, 224–230.

Kuhn, M. (2008). Building predictive models in R using the Caret package. *Journal of Statistical Software, 28*, 1–26.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33*, 363–374. doi: 10.2307/2529786

Langer, E. J., & Moldoveanu, M. (2000). The construct of mindfulness. *Journal of Social Issues, 56*, 1–9.

Larsen, R. J., & Buss, D. M. (2008). *Personality psychology*. New York City, NY: Mac Graw-Hill.

Leigh, J., Bowen, S., & Marlatt, G. A. (2005). Spirituality, mindfulness and substance abuse. *Addictive Behaviors, 30*, 1335–1341. doi: 10.1016/j.addbeh.2005.01.010

Mars, T. S., & Abbey, H. (2010). Mindfulness meditation practise as a healthcare intervention: A systematic review. *International Journal of Osteopathic Medicine, 13*, 56–66. doi: 10.1016/j.ijosm.2009.07.005

Sauer, S., Lemke, J., Wittmann, M., Kohls, N., Mochty, U., & Walach, H. (2012). How long is now for mindfulness meditators? *Personality and Individual Differences, 52*, 750–754.

Sauer, S., Walach, H., & Kohls, N. (2010). Gray's Behavioural Inhibition System as a mediator of mindfulness towards well-being. *Personality and Individual Differences, 50*, 506–551.

Sauer, S., Walach, H., Offenbächer, M., Lynch, S., & Kohls, N. (2011). Measuring mindfulness: A Rasch analysis of the Freiburg mindfulness inventory. *Religions, 2*, 693–706.

Sauer, S., Walach, H., Schmidt, S., Hinterberger, T., Lynch, S., Büssing, A., & Kohls, N. (2013). Assessment of mindfulness: review on state of the art. *Mindfulness, 4*, 3–17.

Sauer, S., Ziegler, M., Danay, E., Ives, J., & Kohls, N. (2013). Specific objectivity of mindfulness – a Rasch analysis of the Freiburg Mindfulness Inventory. *Mindfulness, 4*, 45–54. doi: 10.1007/s12671-012-0145-y

Sedlmeier, P., Eberth, J., Schwarz, M., Zimmermann, D., Haarig, F., Jaeger, S., & Kunze, S. (2012). The psychological effects of meditation: A meta-analysis. *Psychological Bulletin, 138*, 1139–1171. doi: 10.1037/a0028168

Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C. T., ... Golub, T. R. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine, 8*, 68–74. doi: 10.1038/nm0102-68

Tang, Y.-Y., & Posner, M. I. (2013). Tools of the trade: Theory and method in mindfulness neuroscience. *Social Cognitive and Affective Neuroscience, 8*, 118–120. doi: 10.1093/scan/nss112

Teper, R., & Inzlicht, M. (2013). Meditation, mindfulness and executive control: The importance of emotional acceptance and brain-based performance monitoring. *Social Cognitive and Affective Neuroscience, 8*, 85–92. doi: 10.1093/scan/nss045

Van Dam, N. T., Earleywine, M., & Borders, A. (2010). Measuring mindfulness? An Item Response Theory analysis of the Mindful Attention Awareness Scale. *Personality and Individual Differences, 49*, 805–810. doi: 10.1016/j.paid.2010.07.020

Ye, Q.-H., Qin, L.-X., Forgues, M., He, P., Kim, J. W., Peng, A. C., . . . Wang, X. W. (2003). Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature Medicine, 9*, 416–423. doi: 10.1038/nm843

Zainal, N. Z., Booth, S., & Huppert, F. A. (2013). The efficacy of mindfulness-based stress reduction on mental health of breast cancer patients: A meta-analysis. *Psycho-Oncology, 22*, 1457–1465. doi: 10.1002/pon.3171

Received November 19, 2014
Revised received May 31, 2015
Accepted July 6, 2015
Published online April 22, 2016

Sebastian Sauer
FOM University of Applied Sciences
Arnulfstr. 30
80335 Munich
Germany
Tel. +49 89 202 452-0
Fax +49 89 202 452-28
E-mail Sebastian.sauer@fom.de