Mining Mental States using Music Associations

Rajat Agarwal[1*], *Ravinder Singh*[1*], *Suvi Saarikallio*[2], *Katrina McFerran*[3], *Vinoo Alluri*[1]

[1]International Institute of Information Technology, Hyderabad
[2]University of Jyväskylä
[3]University of Melbourne

rajat.agarwal@research.iiit.ac.in, ravinder.singh@research.iiit.ac.in, suvi.saarikallio@jyu.fi, k.mcferran@unimelb.edu.au, vinoo.alluri@iiit.ac.in

Abstract

Owing to the stigmatization of mental illnesses such as depression in India [1], there is a need for indirect unsuspecting ways to identify risk for depression and provide timely intervention. Healthy-Unhealthy Music scale (HUMS) [2] is one such assessment tool developed on Australian population that uses music engagement as an indicator of anxiety levels and potential high-risk for depression as assessed by Kessler's Psychological Distress Scale (K10). The current study aims to ascertain its validity in an Indian setting followed by applying machine learning approaches to predict mental well-being from music associations. A diverse group comprising Indian adult population was assessed using HUMS and mental well-being and proneness to depression measures. HUMS structure investigated via Exploratory factor analyses, and concurrent validity tested with correlations to depression risk and wellbeing revealed high external validity and applicability of HUMS in Indian adult population. Furthermore, very low in-sample error for models like Support Vector Machines (SVM) with nonlinear kernels suggests an underlying pattern between HUMS responses and K10 score. Finally, a two-class model resulted in out of sample accuracy of 81%. To conclude, HUMS demonstrates high generalizability and hence applicability in Indian adult population and potential for employing ML models to capture the underlying pattern.

Introduction

Mental disorders are on the rise all over the world owing to unprecedented social and economic changes. In India, mental illness in general accounted for 31 million Disability-adjusted life-year (6% of overall disease burden) of which 37% is due to depressive disorders [1] As per the latest estimates, 57 million people were affected by depression in India and is projected to be the leading cause of disease burden in by the end of the next decade by WHO.

Despite the unprecedented prevalence of depression in India, the stigma associated with depression poses to be one of the biggest challenges [1]. Consequently, depression goes unacknowledged, undetected and hence untreated engendering marked impact in crucial areas of life including family, work place, and social relations. Depression is to a great extent preventable and treatable provided it is detected at early stages. Therefore, there is a need to find indirect and unsuspecting ways to identify risk for depression and provide timely intervention.

Traditional assessments like K10 have some questions that can be perceived to be invasive to personal privacy and the responder may or may not be willing to answer them truthfully, especially in India, owing to the stigma. Music exists in all cultures and music listening strategies have been found to be representative of internal states and personal tendencies [3] including indicators of ill-health such as internalized symptomatology and depression [4] [5] [6] [7] as well as externalized symptoms and antisocial behaviors [8] particularly in youth. The Healthy-Unhealthy Music Scale (HUMS) was developed, albeit on Australian adolescents, as a non-intrusive tool which allows for indirectly measuring susceptibility to depression and levels of psychological distress based on music listening strategies [2]. The HUMS or Healthy-Unhealthy music scale is a 13-item (Table 1) long assessment that aims to assess music engagement strategies of the responder on a 5-point Likert scale with 1 representing Highly Disagree and 5 Highly Agree.

Using factor analysis, the 13-items are divided into two main dimensions depicting "*Healthy*" and "*Unhealthy*" music listening strategies. It was observed that individuals scoring high on the *Unhealthy* factor also scored high on K10 as evidenced by the significant positive correlation. Therefore, HUMS can be a viable alternative to detecting risk of depression and psychological distress in a non-invasive fashion. However, the validity of HUMS has not been investigated in an Indian context. Furthermore, with the advent of more sophisticated AI/ML models, responses to HUMS items have great potential to be an apt set of features for AI/ML based software that is capable of flagging people with depressive tendencies and psychological distress.

The objectives of the current study are to:

1. Examine the applicability of HUMS as an indirect screening tool to measure well-being and risk of depression in an adult Indian population using statistical methods used in the original study [2]
2. Employ machine learning models like SVM, Neural network, Linear and Logistic regression to establish predictive validity of HUMS.
3. Employ models like SVM and deep learning to assess feasibility of AI based psychological distress and risk of depression diagnosis using HUMS.

10.21437/SMM.2019-12

	Healthy Unhealthy Music Scale
U1	When I listen to music I can get stuck in bad memories
U2	I like to listen to songs over and over even though it makes me feel worse
U3	It can be hard to stop listening to music that connects me to bad memories
U4	I hide in my music because nobody understands me, and it blocks people out
U5	When I try to use music to feel better I can actually end up feeling worse
U6	Music gives me an excuse not to face up to the real world
U7	Music makes me feel bad about who I am
U8	Music leads me to do things I shouldn't do
H1	I feel happier after playing or listening to music
H2	Music gives me the energy to get going
H3	When I'm feeling tense or tired in my body music helps me to relax
H4	Music helps me to relax
H5	Music helps me connect with other people who are like me

Table 1: Items of HUMS. U: Unhealthy items, H: Healthy items as revealed by Factor Analysis in the original study [2]

Method

1.1. Participants

A total of 292 adults participated in the survey (mean age = 26.88, sd = 6.02). Data was collected through online surveys. The survey comprised 13 items from HUMS, 10 from standard K10 survey and 14 from standard Mental Health Continuum short form (MHCSF) assessment. Half of the sample (ID1 - 151 individuals) belonged to a working population and other half (ID2 - 141 individuals) comprised mainly students engaged in post-graduate and doctoral studies. The analyses were performed on the combined dataset (ID = ID1 + ID2). Data from the original study [2] comprised 211 Australian adolescents with a mean age of 13.75 years and a standard deviation of .72 years. (AD). We chose adults instead of adolescents as consent for minors to take part in the survey is an elaborate process which requires parents' permission. Moreover, previous research indicates that there are no significant differences between adolescents and adults in terms of the patterns that specifically link music listening patterns to depression [2,6].

1.2. Testing Validity

All the analyses were done on the data using standard statistical tools from R and Python.

First, exploratory factor analysis was performed on both the individual Indian datasets (ID1 and ID2) separately and combined (ID = ID1+ID2) to identify the latent HUMS structure and compared with the results of the original study. Subsequent concurrent validity analysis was performed by finding correlations between the Factor scores derived from the above step and standard assessments obtained using K10 and MHCSF.

1.3. Modelling psychological distress and depression from HUMS

In order to explore the feasibility of predicting K10 score using the individual HUMS items, we performed stepwise linear regression which allowed to assess the relative importance of each question in predicting K10 scores.

Following this, the participants were grouped into 4 categories [9] based on their K10 scores. Various machine learning approaches including SVM (linear), SVM (Non-linear), multi-layer feedforward neural networks were used to predict K10 from HUMS. These approaches were compared based on in sample error and out sample error (10-fold cross validation with 10% held out data). The analyses described in this section was performed on combined datasets (ID + AD).

Subsequently, due to inherent imbalance in collected data of fewer samples in the higher ranges of K10 distress category, various techniques like under-sampling and oversampling of data were attempted. Few ensemble method based classifiers were also evaluated. In addition, to overcome the inherent imbalance, we further repeated the analyses on 2 categories (by combining the first two and last two categories of the 4 class model [9]) rather than 4 in an attempt to make the model simpler and capable of learning with less amount of data. A 2-category classification in fact allows for identifying higher-risk individuals who fall into the second category of moderate to high K10 scores and would most benefit from further clinical evaluation and possible early intervention of the then identified condition.

Results

In order to evaluate the intrinsic dimensionality of our data sets we employed one of the most commonly used estimation methods, which is based on the Eigenvalues satisfying the Kaiser criterion (1960). As a result, two intrinsic dimensions were observed as in the original study. Factor analyses revealed highly similar loading patterns in the individual and combined Indian datasets and were close to being identical to the underlying structure of AD. Hence, the combined Indian dataset ID was used for subsequent analyses in order to increase statistical power of hence obtained results. Figure 1 displays the factor loadings for both ID and AD for the "Healthy" and "Unhealthy" factors. The two factors will be referred to as *Unhealthy* and *Healthy* for the reminder of the study.

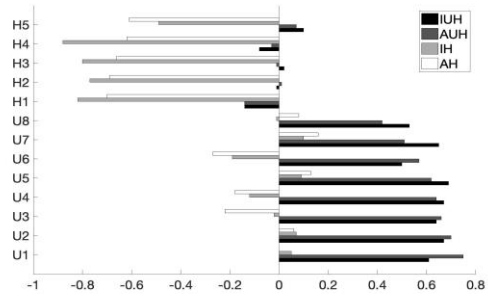

Figure 1: Factor loadings of the Australian and Indian sample. The loadings for the healthy factor have been reversed in sign for representational purposes. IUH: Indian Unhealthy Factor, AUH: Australian Unhealthy Factor, IH: Indian Healthy Factor, AH: Australian Unhealthy Factor

Table 2 displays the correlation pattern between the factors for both ID and AD. Pearson correlations revealed significant positive correlation between *Unhealthy* factor scores and K10 scores. However, *Healthy* factor scores were found to correlate significantly with well-being as measured by MHCSF only in ID.

	ID		AD	
	H	U	H	U
H	1	.17**	1	.16*
U	.17**	1	.16*	1
K10	.06	.52***	.19*	.68***
MHCSF	.30**	-.15*	.08	-.49***

*p < .05; **p < .01; ***p < .001

Table 2: Pearson correlations between HUMS *Healthy, Unhealthy*, K10 and MHCSF. U: *Unhealthy* factor score, H: *Healthy* factor score

Multivariate linear regression performed using the unhealthy items (U1-U8) revealed that a relatively high amount of variance could be explained in AD (adjusted R^2 = 47%) while a moderate amount could be explained in ID (adjusted R^2 = 31%). Adding the Healthy items (H1-H5) in the model resulted in an increase in variance explained, albeit marginal and non-significant. These results hint at the existence of potential non-linear combinations of HUMS items in predicting K10.

Due to high concurrent validity of HUMS with K10 in an Indian setting and highly similar factor structure between both the Indian and Australian datasets, subsequent analyses were performed on combined data (AD + ID). The combined data was divided into 4 classes based on K10 score as described in [9]. The Unhealthy items were used as predictors for K10 scores. The classification accuracy using SVM and other linear and non-linear modeling approaches can be seen in Table 3.

Model	In sample accuracy (training)	Out sample accuracy (testing)
SVM - Linear	58.3	56.2
SVM - RBF (SVM - RBF with all HUMS items)	93 (99.2)	54.5 (55.3)
Neural Net (13,13,3)*	70.4	56.2
Logistic regression	59.2	56

Table 3: Classification accuracy of various classification models using Unhealthy HUMS items for K10 4-class K10 predictions using the combined dataset. *size of hidden layers in the network

Since SVM – RBF (radial basis function) demonstrated highest in-class accuracy, we repeated the analysis with all HUMS items as predictors and achieved a training accuracy of 99% (row 2 of Table 3).

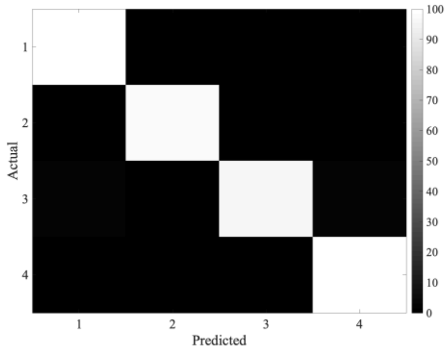

Figure 2 – Confusion matrix for in-sample error in case of SVM with RBF kernel. Class Labels 1: No risk 2: Mild risk 3: Moderate Risk 4: High Risk for depression and/or mental disorders.

The confusion matrix (Figure 2) for the in-sample accuracy of SVM-RBF with HUMS as predictors illustrates near perfect accuracy for every class. This presents some additional evidence of existence of a non-liner mapping between HUMS and K10. However, the accuracy for test data is limited to 55% (Table 3). The accuracy did not improve much even after extensive hyper parameter tuning of SVM (non linear). Hyper parameter tuning for neural network was not attempted due to limited training data being available. One potential explanation for the low generalizability is the imbalance of class-wise training data. Since the data was not collected from a clinically diagnosed sample, the probability of obtaining sufficient data in classes 3 & 4 is relatively low in comparison to class 1 & 2 (n=274 for class 1, n=107 for class 2, n =57 for class 3 and n=65 for class 4).

Hence, in order to minimize the under representation of classes 3 & 4, techniques such as oversampling data of the minority class (class 3 & 4) data or undersampling that of the majority category data were employed [10]. Furthermore, since classes 3 & 4 represent high-risk population who require further clinical evaluation, the classification problem was converted to a 2-category problem by combining classes 1 and 2, and classes 3 and 4 respectively. Table 4 displays the results of the undersampling and oversampling techniques for both 4-class and 2-class classification. Specifically, in order to minimize overfitting and increase generalizability of the models, training was performed on ID and tested on AD and vice-versa.

Table 4: Comparison of various undersampling and oversampling techniques and classifiers with unbalanced data. The results are out of sample accuracy percentages by training on ID data and testing on AD

Model	4-class	2-class
SVM (rbf) with Random oversampling[3]	45.97	78.67
SVM (rbf) with SMOTE[3]	47.87	78.20
SVM (rbf) with ADASYN[3]	43.60	75.83
SVM (rbf) with random undersampling[3]	48.34	76.78
SVM with undersampling using NearMiss[3]	25.59	70.62
BalancedRandomForestClassifier[3]	47.39	81.04

As can be seen in Table 4, the results of a 2-class classification range are found to be significantly higher than chance level reaching a maximum accuracy of 81%. Similarly, the model trained on AD and tested on ID revealed a 2-class accuracy of 77%. These results imply that music listening strategies indeed have universals, which can be learnt using machine learning models.

Conclusions

HUMS demonstrates high validity in Indian adult population to assess psychological distress and potential risk for depression and other mental disorders. Very low training error and high accuracy points towards the existence of a non-linear function that maps the HUMS responses to K10 score. Existence of such a function adds to the aptness of HUMS as a mental state detector.

To add to this, the high predictive power of the machine learning models, especially in the 2-class problem (as evidenced by significantly high classification accuracy for the moderate to high-risk groups), further increases the generalizability of HUMS in a global context.

Thus, HUMS can be used as a comprehensive and valid instrument to employ in the Indian context as a non-invasive tool to assess mental well-being thereby circumventing the stigma associated with direct assessment and discussion related especially to depression. This has potential implications in a corporate setting wherein detecting internal states and potential risky behavior in employees is vital for constructive intervention.

Acknowledgements

We acknowledge Sarath Subramaniam in helping in data collection.

References

[1] World Health Organization. 2017b. Depression in India. Let's Talk. World Health Organization, http://www.searo.who.int/india/depression_in_india.pdf.

[2] Saarikallio, Suvi, Gold, Christian & McFerran, Katrina (2015). Development and validation of the Healthy- Unhealthy Music Scale, *Child and Adolescent Mental Health* 20, No. 4, 2015, pp. 210–217.

[3] Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes. *Psychological Science*, 29(7), 1145–1158.

[4] Miranda, Dave & Gaudreau, Patrick & Debrosse, Régine & Morizot, Julien & Kirmayer, Laurence. (2012). Music Listening and Mental Health: Variations on Internalizing Psychopathology. *Music, Health, and Wellbeing*. 10.1093/acprof:oso/9780199586974.003.0034.

[5] Katrina S. McFerran & Helen Shoemark (2013). How musical engagement promotes well-being in education contexts: The case of a young man with profound and multiple disabilities. *International Journal of Qualitative Studies on Health and Well-being*, 8:1, DOI: 10.3402/qhw.v8i0.20570

[6] McFerran, K. S., & Saarikallio, S. (2013) Depending on music to feel better: Being conscious of responsibility when appropriating the power of music. *The Arts in Psychotherapy*, http://dx.doi.org/10.1016/j.aip.2013.11.007

[7] Garrido, Sandra & Eerola, Tuomas & Mcferran, Katrina. (2017). "Group Rumination: Social Interactions Around Music in People with Depression." *Frontiers in Psychology*. 8. 490. 10.3389/fpsyg.2017.00490.

[8] Mulder, J., Bogt, T., Raaijmakers, Q. et al., (2007) Music Taste Groups and Problem Behavior. *Journal of Youth and Adolescence* 36: 313. https://doi.org/10.1007/s10964-006-9090-1

[9] Victorian Population Health Survey. Melbourne: Department of Human Services, Victoria; 2001.

[10] Guillaume Lemaitre, Fernando Nogueira, Christos K. Aridas." Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*", Journal of Machine Learning Research 7* (2016) 1-5, https://arxiv.org/pdf/1609.06570.pdf

[11] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

[12] McFerran, K.S., & Saarikallio, S. (2013). Depending on music to make me feel better: Who is responsible for the ways young people appropriate music for health benefits. The Arts in Psychotherapy, 41, 89–97.