TAG2RISK: HARNESSING SOCIAL MUSIC TAGS FOR CHARACTERIZING DEPRESSION RISK

Aayush Surana[1]   **Yash Goyal**[1]   **Manish Shrivastava**[1]
Suvi Saarikallio[2]   **Vinoo Alluri**[1]

[1] International Institute of Information Technology, Hyderabad, India
[2] Department of Music, Art and Culture Studies, University of Jyväskylä, Finland

{aayush.surana, yash.goyal}@research.iiit.ac.in, suvi.saarikallio@jyu.fi
{m.shrivastava, vinoo.alluri}@iiit.ac.in

ABSTRACT

Musical preferences have been considered a mirror of the self. In this age of Big Data, online music streaming services allow us to capture ecologically valid music listening behavior and provide a rich source of information to identify several user-specific aspects. Studies have shown musical engagement to be an indirect representation of internal states including internalized symptomatology and depression. The current study aims at unearthing patterns and trends in the individuals at risk for depression as it manifests in naturally occurring music listening behavior. Mental well-being scores, musical engagement measures, and listening histories of Last.fm users (N=541) were acquired. Social tags associated with each listener's most popular tracks were analyzed to unearth the mood/emotions and genres associated with the users. Results revealed that social tags prevalent in the users at risk for depression were predominantly related to emotions depicting *Sadness* associated with genre tags representing *neo-psychedelic-, avant garde-, dream-pop*. This study will open up avenues for an MIR-based approach to characterizing and predicting risk for depression which can be helpful in early detection and additionally provide bases for designing music recommendations accordingly.

1. INTRODUCTION

According to reports from the World Health Organization, an estimated 322 million people worldwide were affected from depression, the leading cause of disability [1]. Recent times have witnessed a surge in studies on using social multimedia content, such as those from Facebook, Twitter, Instagram, to detect mental disorders including depression [2–6]. Music plays a vital role in mental well-being by impacting moods, emotions and other affective states [7]. Musical preferences and habits have been associated with the individual's need to satisfy and reinforce their psychological needs [8, 9]. Empirical evidence exists linking musical engagement strategies to measures of ill-health including internalized symptomatology and depression [10, 11]. Also, increased emotional dependency on music during periods of depression has been reported [12]. Specifically, the Healthy-Unhealthy Music Scale (HUMS), a 13-item questionnaire was developed for assessing musical engagement strategies that identified maladaptive ways of using music. Such strategies are characterized by using music to avoid other people, resort to ruminative thinking and ending up feeling worse after music engagement. Such unhealthy musical engagement was found to correlate with higher psychological distress and was indicative of depressive tendencies [13]. Furthermore, the high predictive power observed from the machine learning models in predicting risk for depression from HUMS further bolsters its efficacy as an indirect tool for assessing mental states [14]. Research suggests that such musical engagement does not always lead to alleviating depressive symptoms [15]. This indeed calls for developing intervention strategies that allow for altering music listening behavior to suit the individual's state, traits, and general musical preferences which may lead to a positive outcome. Thus, it is of vital importance not only to identify individuals with depressive tendencies but also to unearth music listening habits of such individuals that will provide bases for designing music recommendations accordingly.

Past research studying the link between music listening habits and depression has been done using self-reported data and controlled listening experiments wherein participants may have wished to conform to social expectations, or their responses might be influenced by how they want other people to perceive them thereby resulting in demand characteristics [16]. This has also been identified as a limitation by Nave et al. [8], who have proposed collecting data in more ecologically valid settings, such as user listening histories from music streaming platforms which are a better reflection of the users' true preferences and behaviours. To date no studies have looked at the link between active music listening and depression using the music listening histories of users which motivates us for this study.

In this age of big data, online music streaming platforms such as Last.fm, Spotify, and Apple Music provide access to millions of songs of varying genres and this has allowed
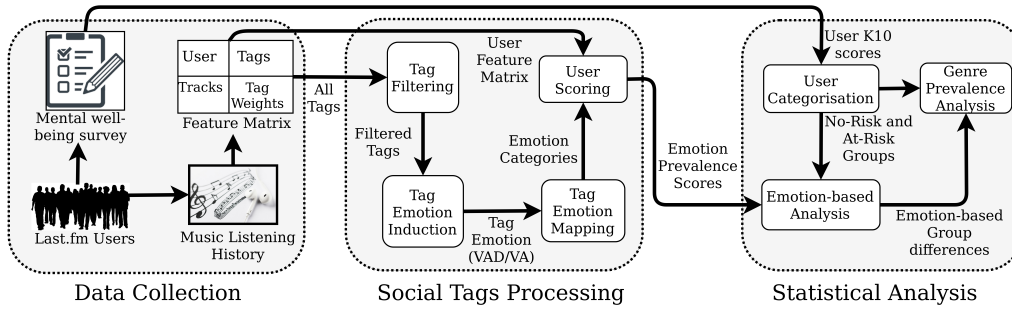
Figure 1: Methodology

for assessing users' features from their digital traces on music streaming platforms. To the best of our knowledge, Last.fm is the only platform that makes it possible to extract the listening history of users and other metadata describing their listening behavior using its own public API. Last.fm has been used extensively by researchers for various purposes such as music emotion classification, user behavior analysis, and social tag categorization [17, 18]. Last.fm has an abundance of social tags that provide a wide range of information about the musical tracks including audio low- and high-level feature description, emotions and experiences evoked, genre, etc. These tags have been found to predict short-term user music preferences [19] and in successfully predicting next played songs in the design of a recommendation system [20]. Our aim is to identify the tags and their respective occurrences in the listening behavior of individuals at risk for depression, which makes Last.fm an apt choice for this study. The data was collected using an online survey comprising of Last.fm music listening histories, in addition to music engagement strategies (HUMS), and mental well-being scores of the participants. Specifically, each track in the data was semantically represented by the tags assigned to it. We leverage these representations of tags as social descriptors of music to uncover emotional experiences and concepts that are associated with users with risk for depression.

1.1 Research Objectives and Hypotheses

In this study we investigated whether people's music listening history, in terms of social tags, could be used to predict a risk for depression. Our research questions were:

- What are the social tags associated with music chosen by At-Risk users?

- What emotions do these tags signify in the context of musically evoked emotions?

- What genres are mostly associated with At-Risk users?

- How well can we classify users as At-Risk given user-specific social tags?

We expected the social tags linked with At-Risk listeners to contain emotions with low arousal and low valence, being typical of depressive mood. Owing to the lack of research

associating music genres and risk for depression [15], this part of the study was exploratory.

2. METHODOLOGY

The methodological approach and procedure of our study is illustrated in Figure 1. The steps of data collection, processing, and analysis are described below.

2.1 Data Collection

An online survey was designed wherein participants were asked to fill their Last.fm usernames and demographics followed by standard scales for assessing their mental well-being, musical engagement strategies and personality. Participants were solicited on the Last.fm groups of social media platforms like Reddit and Facebook. The inclusion criterion required being an active listener on Last.fm for at least a year prior to filling the survey. The survey form required the users' consent to access their Last.fm music history.

2.1.1 Participants

A total of 541 individuals (Mean Age = 25.4, SD = 7.3) were recorded to be eligible and willing to participate in the study consisting of 444 males, 82 females and 15 others. Most of them belonged to the United States and the United Kingdom accounting for about 30% and 10% of the participants respectively. Every other country contributed to less than 5% of the total participants.

2.1.2 Measure of Well-Being, Musical Engagement, and Personality

The Kessler's Psychological Distress Scale (K10) questionnaire [21] was used to assess mental well-being. It is a measure of psychological distress, particularly assessing anxiety and depression symptoms. Individuals scoring 29 and above on K-10 are likely to be at severe risk for depression and hence, constitute the "At-Risk" group. Those scoring below 20 are labeled as the "No-Risk" group [22] as they are likely to be well. There were 193 participants in the No-Risk group and 142 in the At-Risk group. The HUMS survey was administered to assess musical engagement strategies which resulted in two scores per participant, *Healthy* and *Unhealthy*. Personality information was obtained using the Mini-IPIP questionnaire [23]

which results in scores for the Big Five traits of Personality namely *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness* and *Neuroticism*. HUMS and personality data were collected in order to identify if specific personality traits engage more in *Unhealthy* music listening and as additional measures to assess internal validity.

2.1.3 Music Listening History

Each participant's music listening history was extracted using a publicly available API. The data included tracks, artists, and social tags associated with the tracks. For each participant, the top *n* (n=500,200,100) tracks based on play-counts were extracted centered around the time *t* (t = ± 3 months,2 months) they filled in the questionnaire. The reason for varying *n* and *t* was to find converging evidence in music listening behavior in order to make our results more robust. For each track, the top 50 social tags based on tag weight (number of times the tag has been assigned to the track) were chosen for subsequent analysis.

2.2 Social Tags Processing

2.2.1 Tag Filtering

Music-related social tags are known to be descriptors of genre, perceived emotion, artist and album amongst others. It is therefore important to filter them to organize them according to some structure and interpretable dimensions for the task at hand. The purpose of this preprocessing step was to retrieve tags that could be mapped onto a semantic space representing music-evoked emotions. To this end, we used four filtering stages: first, include lower-casing, removal of punctuation and stop-words, spell-checking and checking for the existence of tag words in the English corpus; second, retain tags that are most frequently assigned adverbs or adjectives via POS (Part Of Speech) tagging since POS tags representing nouns and pronouns do not have emotion relevance in this context; third, remove tags containing 2 or more words to avoid valence shifters [24] and sentence-like descriptions from our Last.fm corpus; fourth, manually filter them by discarding tags without any mood/emotion associations.

2.2.2 Tag Emotion Induction

To project the tags onto an emotion space, we used dimensional models that represent the emotions. Multiple research studies have shown the usefulness of both two-dimensional and three-dimensional models to represent emotions [25–27]. We therefore used both these models for further analysis in order to check for trends and the effect of the third dimension when dealing with emotions.

The first model is one of the most popular dimensional models, the Russell's Circumplex Model of Affect [28], where an emotion is a point in a two-dimensional continuous space representing *Valence* and *Arousal* (VA). *Valence* reflects pleasantness and *Arousal* describes the energy content of the emotion. The second model is an extension of the Russell's model with an added *Dominance*

dimension (VAD), which represents control of the emotional state. The VAD model has been a popular framework used to construct emotion lexicons in the field of Natural Language Processing. The projection in the VAD space is based on semantic similarity and has been largely used to obtain affective ratings for large corpora of English words [29] [30]. Another common emotion model is the VAT model wherein the third dimension represents *Tension* (VAT) and has been used in retrieving mood information from Last.fm tags [17]. However, Saari et. al.'s [17] approach was based on tag co-occurrence rather than semantic similarity. Moreover, a subsequent study by the same authors reported a positive correlation (r=0.85) between *tension* and *dominance* [31]. Also, multiple studies have supported the use of the VAD space for analysing emotions in the context of music [32, 33]. We therefore have chosen the VAD framework for the purpose of our study. Since VA dimensions alone were found to sufficiently capture musical emotions [26], we also repeat our analysis based on the VA model to observe the effect of the omitted *Dominance* dimension.

The tags were projected onto the VAD space using a word-emotion induction model introduced by Buechel and Hahn [29]. We used the FastText embeddings of the tags as input to a 3-layer multi-layer perceptron that produced VAD values ranging from 1 to 9 on either of the dimensions. FastText has shown better accuracy for word-emotion induction [29] when compared to other commonly used models like Word2vec and GloVe. Moreover, Fast-Text embeddings incorporate sub-word character n-grams that enable the handling of out-of-vocabulary words. This results in a large advantage over the other models [34]. In addition, FastText works well with rarely occurring words because their character n-grams are still shared with other words. This made it a suitable choice since some of the user-assigned tags may be infrequent or absent in the training corpus of the embedding model. We used the same approach to project the tags onto the VA space by changing the number of nodes in the output layer from 3 to 2.

Both the models were trained using the EN+ dataset which contains *valence*, *arousal* and *dominance* ratings (on a 9-point scale) for a majority of well-known English words [35]. This module resulted in an n-dimensional vector (n=3 for VAD, n=2 for VA) for each tag. The remainder of the pipeline describes the 3-dimensional VAD vector processing. The same procedure is repeated for the VA scores.

2.2.3 Tag Emotion Mapping

The social tags were grouped into broader emotion categories. These categories consisted of 9 first-order factors of Geneva Emotional Music Scale (GEMS) [36]. These were *Wonder, Transcendence, Nostalgia, Tenderness, Peacefulness, Power, Joyful activation, Tension and sadness*. Table 1 in the supplementary material displays the factor loadings for these first-order factors of GEMS. GEMS contains 40 emotion terms that were consistently chosen to describe musically evoked emotive states across

a wide range of music genres. These were subsequently grouped to provide a taxonomy for music-evoked emotions. This scale has outperformed other discrete and dimensional emotion models in accounting for music-evoked emotions [37]. In order to project these 9 emotion categories onto the VAD space, we first obtained the VAD values for the 40 emotion terms. Next, the VAD values were weighted and summed according to the weights provided in the original GEMS study to finally obtain VAD values for each of the emotion categories. Figures 1 & 2 in supplementary material display the projections of these emotion categories onto the VAD and VA spaces. Each of the tags are then assigned the emotion category based on the proximity in the VAD space as evaluated by the euclidean distance.

2.3 User-Specific Emotion Prevalence Score

After every user's tags had been mapped onto the 9 emotion categories, we calculated an *Emotion Prevalence Score* $S_{u,c}$ for every user. This represents the presence of tags belonging to that particular emotion category in the user's listening history.

$$S_{u,c} = \frac{\sum_{j \epsilon V_{\text{tr}}} (N_{j,c} \times tr_{\text{u,j}})}{\sum_{i \epsilon T_{\text{u}}} tr_{\text{u,i}}} \quad (1)$$

where

$$N_{j,c} = \sum_{k \epsilon Tags_{\text{c}}} \frac{tw_{\text{j,k}}}{\sum_{l \epsilon V_{\text{tg}}} tw_{\text{j,l}}} \quad (2)$$

c : emotion category
$N_{j,c}$: the association of track j with c
T_u : all tracks for user u
V_{tg} : all tags obtained after tag filtering
V_{tr} : all tracks having at least one tag from V_{tg}
$tr_{u,i}$: playcount of track i for user u
$tw_{j,k}$: tag weight of tag k for track j
$Tags_c$: all tags in V_{tg} which belong to c

Since the objective of this work was to identify which of the 9 categories are most characteristic of At-Risk individuals when compared to No-Risk individuals, we performed group-level statistical tests of difference as described in the following section.

2.4 Emotion-based Analysis : Group Differences and Bootstrapping

For each emotion category, we performed a two-tailed Mann-Whitney U (MWU) Test on the *Emotion Prevalence Scores* between the No-Risk and At-Risk groups. For a category, the group having higher mean rank from MWU Test indicates a stronger association of the category with that group. For the emotion categories that exhibited significant differences ($p < .05$), we further performed bootstrapping to account for Type I error and ensure that the observed differences are not due to chance. Bootstrapping (random sampling) with replacement was performed with 10,000 iterations. Each iteration randomly assigned participants to the At-Risk or No-Risk group. The U-statistic

for each iteration was calculated. As a result, we obtain a bootstrap distribution for the U-statistic from which we estimate the significance of the observed statistic.

2.5 Genre-Prevalence Analysis

To further analyse the types of music associated with the tags of emotion categories, we explored genre-related social tags. In order to select the genre-related tags from our data, we collected the results of the multi-stage model proposed by Ferrer et al. [38] which assigned tags of Last.fm to different semantic layers namely genre, artist, affect, instrument, etc. In order to understand the underlying genre tag structure and obtain broader genre categories, we employed the approach described by Ferrer et al. [39] to cluster genre tags (details in Equation 1 in the Supplementary material). In this, music tags were hierarchically organized revealing taxonomy of music tags by means of latent semantic analysis. The clusters thus obtained were labelled based on the genre-tags constituting the core points of the cluster [40].

For the emotion categories that exhibited significant group differences, the genre tags co-occurring with its tags were used to calculate a user-specific *Genre Prevalence Score* for each genre-tag cluster. The formula used was similar to *Emotion Prevalence Scores* with the change in definition of the following terms: c represents the genre cluster, T_u is the set of all tracks for user u which have a tag belonging to the particular emotion category and V_{tg} is the set of all genre tags. Finally, we performed a biserial correlation between *Genre Prevalence Scores* for each genre-tag cluster and the users' risk for depression (represented as a dichotomous variable with 0 = No-Risk; 1 = At-Risk).

3. RESULTS

3.1 Internal Consistency and Criterion Validity

The Cronbach's alphas for *Unhealthy* scores obtained from HUMS and K10 scores were found to be relatively high at 0.80 and 0.91 respectively. A significant correlation ($r=0.55$, $df=539$, $p<0.001$) between *Unhealthy* Score and K10 was found which is in concordance with past research studies in the field [13]. Also, in line with previous research [41, 42], a significant positive correlation was observed between K10 score and Neuroticism ($r=0.68$, $p<0.0001$) adding to the internal consistency of the data and confirming construct validity. As can be seen in Figure 2, the At-Risk group displayed higher mean and median *Unhealthy* score compared to No-Risk while *Healthy* scores were comparable. Partial correlations between *Unhealthy*, *Healthy*, and K10 are presented in Table 1. K10 scores exhibit significant positive correlation only with *Unhealthy* for both the groups. The moderate correlation between *Healthy* and *Unhealthy* scores for the No-Risk population indicates that both of these subscales capture a shared element, most likely active music listening.

Figure 2: Boxplot of HUMS scores for No-Risk and At-Risk Groups.

	No-Risk		At-Risk	
Scales	Healthy	Unhealthy	Healthy	Unhealthy
Healthy	1.0	0.36**	1.0	-0.14
Unhealthy	0.36**	1.0	-0.14	1.0
K10	0.07	0.26**	-0.11	0.22*

Table 1: Partial Correlation Values between HUMS & K10. (*p<0.01 & **p<0.001)

3.2 Emotion-based Results

The data ($t=\pm3$,$n=500$) consisted of 3,80,261 social tags. The tag filtering process resulted in a final set of 1254 unique tags (Mean=109, SD = 24 tags per user) which were then mapped onto the VA and VAD emotion spaces. Figure 3 in supplementary material displays the tags closest to each of the Emotion Categories based on VA and VAD models.

Figure 3: Boxplot of Emotion Prevalence Scores for No-Risk and At-Risk based on VA.

Figure 3 illustrates the *Emotion Prevalence Scores* for both groups for VA mapping (Supplementary Figure 4 displays the same for VAD, showing a similar distribution). The overall pattern appears similar between both the groups with minor observable differences for the emotion categories *wonder, transcendence, tenderness, tension, and sadness*. Table 2 displays the emotion categories that exhibited significant differences between the groups (MWU U-statistic and bootstrap p-values in Table 2 of Supplementary material). The At-Risk group consistently exhibits higher Prevalence Scores in *Sadness* while the No-Risk group vacillates between *Wonder* and *Transcendence*. The most significant difference was observed in *Sadness* (VA model,$t=\pm3$,$n=100$) with a significantly greater *Emotion Prevalence Score* for the At-Risk group (Median

= 0.0117) than the No-Risk group (Median = 0.0091), U=11414.5, p=0.009. Significant difference was also observed for *Tenderness* with greater *Emotion Prevalence Score* for the At-Risk group (Median = 0.1271) than the No-Risk group (Median = 0.1189), U=11905.0, p=0.04. On the other hand, the *Emotion Prevalence Score* in *Wonder* (VA model,$t=\pm2$,$n=100$) was significantly greater for the No-Risk group (Median = 0.0131) than the At-Risk group (Median = 0.0086), U=16270.0, p=0.003. The word-clouds of tags comprising *Sadness* and *Tenderness* are displayed in Figure 4a and Figure 4b. A score per tag is computed for each group (Equation 2 in Supplementary material). A rank was assigned to the tag based on the absolute difference of the tag scores between No-Risk & At-Risk groups. The size of the tag in the word-cloud is directly proportional to its rank in the category. Supplementary figures 5 and 6 depict word-clouds for *Transcendence* and *Wonder*.

(a) Sadness (VA,±3,100) (b) Tenderness (VAD,±3,500)

Figure 4: Wordclouds for emotion categories associated with At-Risk group.

We also assessed the predictive power of social tags for risk of depression by classifying participants into At-Risk or No-Risk groups using their tag information (feature details in Equation 4 in Supplementary material). The SVM model with 'rbf' kernel (C=2301, gamma=101) gave the best results with a 5-fold cross-validation accuracy of 66.4%.

3.3 Genre-Prevalence Results

Out of the 5062 tags assigned to the genre layer in [38], 94% (4766) of the tags were present in our data. The clustering of the genre tags resulted in 17 clusters and is displayed in Table 3 of Supplementary material. Figure 7 in Supplementary material displays mean genre prevalence scores between both groups for these 17 clusters. Overall, genre-cluster representing *indie-,alternative-pop/rock* represented by Cluster 4 is predominant in both groups. Genre prevalence scores were then evaluated specific to the tracks associated with the emotion categories that exhibited most significant group differences, that is, *Wonder* and *Sadness* (VA model, $t=\pm3$, $n=100$). For *Sadness*-specific tracks, the highest correlation (r=0.2, p<0.01) was observed between the Genre-Prevalence scores in the cluster representing *neo psychedelic-, avant garde-, dream-pop* and K-10 scores. Also, genre clusters representing *electronic rock* (r=0.17, p<0.01), *indie-, alternative-pop/rock* (r=0.12, p<0.05), and

Group	Top Tracks	VAD		VA	
		$t=\pm3$	$t=\pm2$	$t=\pm3$	$t=\pm2$
At-Risk	n=100	Sadness*		Sadness**	
	n=200	Sadness*	Sadness*, Tenderness*	Sadness*	Sadness*, Transcendence*
	n=500	Sadness*, Tenderness*	Tenderness*	Sadness*	Sadness*, Transcendence*
No-Risk	n=100		Transcendence*		Wonder**
	n=200	Transcendence*	Transcendence*	Wonder*	Wonder**
	n=500	Transcendence*	Transcendence*	Wonder*	Wonder**

Table 2: Emotion Categories with Significant Differences between At-Risk and No-Risk groups.*p<0.05; **p<0.01

world music (r=-0.11, p<0.05) demonstrated significant correlations for *Tenderness*. For *Wonder* (VA model, $t=\pm2$, n=100), the K-10 scores exhibited significant negative correlation with *Genre Prevalence Scores* of clusters representing *black metal* (r=-0.11, p<0.05) and *neo-progressive rock* (r=-0.13, p<0.05).

4. DISCUSSION

This study is the first of its kind to examine the association between risk for depression and social tags related to music listening habits as they occur naturally as opposed to self-reported or lab-based studies. A clear difference in the music listening profiles was observed between the At-Risk group and the No-Risk group, particularly in terms of the emotional content of the tags. *Sadness* was significantly more prevalent in the At-Risk group and the word-cloud of sadness was highly illustrative of other low-arousal, low-valence emotions such as *dead*, *low*, *depressed, miserable, broken*, and *lonely*. The stronger association of the At-Risk group with sadness is in concordance with the past research studies in the field [43] and confirms our hypothesis. The At-Risk group is attracted to music that reflects and resonates with their internal state. Whether this provides emotional consolation as an adaptive resource or whether it only worsens repetitive negative feelings and fuels rumination, remains an open question. Nonetheless, statistically, such listening style can be seen as a highly predictive factor of psychological distress.

In addition, *Tenderness*, which represents low-arousal and high-valence, was also more prevalent in the At-Risk group, especially for shorter-term ($\pm$ 2 months) music listening habits. *Tenderness* appears to be more significant in the shorter time period in addition to *Sadness*, possibly indicating that At-Risk people tend to oscillate between positive and negative states within a general state of low arousal. These findings appear to be very much in line with the results found by Houben et al. [44] who found high levels of emotional inertia and emotional variability to be linked with depression and ill-being. The consistent results related to *Sadness* in our study reflect the overall inert states in which the At-Risk tend to be. On the other hand, the *Tenderness* results reflect their tendency to jump to positive affective states while retaining low arousal, thereby demonstrating emotional variability. Furthermore, the omission of the *Dominance* dimension causes most of the tags to shift from *Tenderness* to *Transcendence* and *Transcendence* to *Wonder*, which explains

the results in a reversal of the group association as evidenced in the results. Nevertheless, *Sadness* appears to be the predominant state as it is largely consistent for ±3 months as well as for ±2 months of music listening histories.

The At-Risk group also exhibited a tendency to gravitate towards music with genre tags such as *neo-psychedelic-*, *avant garde-*, *dream-pop* co-occurring with *Sadness*. Such genres are characterized by ethereal-sounding mixtures that often result in a wall of sound comprising electronic textures with obscured vocals. Similarly, the genres co-occurring with *Tenderness* (VAD model) or *Transcendence* (VA model) comprise similar mixtures with heavy synthesizer-based sounds (such as mellotron) which result in sounds that seem otherworldly. Such out-of-this world soundscapes have been also associated with transcendent druggy and mystical imagery and immersive experiences [45]. These results strengthen the claim that depression may foster musical immersion as an escape from a reality that is perceived to be adverse. This is somewhat in line with prior research that has linked depression with the use of music for avoidant coping [46]. On the other hand, music listening history of the No-Risk group was characterized by an inclination to listen to music tagged by positive valence and higher arousal as characterized by *Wonder* with a predilection for *Heavy metal* and *Progressive Rock* genres.

The use of only single word tags in the third stage of tag filtering is one limitation of this study which is due to lack of compatibility of the word emotion induction model with multi-word tags. Our results could potentially be extended to find significant differences in emotional concepts after considering multi-word social tags. We achieve a decent classification accuracy of 66.4% which is significantly above the chance level which indicate that social tags indeed may be indicative of At-Risk behavior. This may further be improved by considering additional descriptors of music such as acoustic features and lyrical content of the tracks. Another future direction is to incorporate the temporal evolution of these emotion categories in the listening histories to characterize depression, since past research suggests depression to be a result of gradual development of daily emotional experiences [47]. This study is intended to be one of many to come that will be helpful in early detection of depression and other potential mental disorders in individuals using their digital music footprints.

5. AUTHOR CONTRIBUTIONS

Aayush Surana and Yash Goyal are joint first authors with equal contribution.

6. REFERENCES

[1] WHO, "Depression and other common mental disorders: global health estimates," World Health Organization, Tech. Rep., 2017.

[2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[3] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 2014, pp. 51–60.

[4] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 626–638.

[5] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 2098–2110.

[6] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," *EPJ Data Science*, vol. 6, no. 1, pp. 1–12, 2017.

[7] M. Baltazar and S. Saarikallio, "Toward a better understanding and conceptualization of affect self-regulation through music: A critical, integrative literature review," *Psychology of Music*, vol. 44, no. 6, pp. 1500–1521, 2016.

[8] G. Nave, J. Minxha, D. M. Greenberg, M. Kosinski, D. Stillwell, and J. Rentfrow, "Musical preferences predict personality: evidence from active listening and facebook likes," *Psychological Science*, vol. 29, no. 7, pp. 1145–1158, 2018.

[9] L. Qiu, J. Chen, J. Ramsay, and J. Lu, "Personality predicts words in favorite songs," *Journal of Research in Personality*, vol. 78, pp. 25–35, 2019.

[10] K. S. McFerran, S. Garrido, and S. Saarikallio, "A critical interpretive synthesis of the literature linking music and adolescent mental health," *Youth & Society*, vol. 48, no. 4, pp. 521–538, 2016.

[11] S. Garrido, T. Eerola, and K. McFerran, "Group rumination: Social interactions around music in people with depression," *Frontiers in psychology*, vol. 8, p. 490, 2017.

[12] K. S. McFerran, "Contextualising the relationship between music, emotions and the well-being of young people: A critical interpretive synthesis," *Musicae Scientiae*, vol. 20, no. 1, pp. 103–121, 2016.

[13] S. Saarikallio, C. Gold, and K. McFerran, "Development and validation of the healthy-unhealthy music scale," *Child and adolescent mental health*, vol. 20, no. 4, pp. 210–217, 2015.

[14] R. Agarwal, R. Singh, S. Saarikallio, K. McFerran, and V. Alluri, "Mining mental states using music associations," *depression*, vol. 2, p. 6, 2019.

[15] J. Stewart, S. Garrido, C. Hense, and K. McFerran, "Music use for mood regulation: self-awareness and conscious listening choices in young people with tendencies to depression," *Frontiers in psychology*, vol. 10, p. 1199, 2019.

[16] D. M. Greenberg and P. J. Rentfrow, "Music and big data: a new frontier," *Current opinion in behavioral sciences*, vol. 18, pp. 50–56, 2017.

[17] P. Saari and T. Eerola, "Semantic computing of moods based on tags in social media of music," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2548–2560, 2013.

[18] C. Laurier, M. Sordo, J. Serra, and P. Herrera, "Music mood representations from social tags." in *International Society for Music Information Retrieval (ISMIR) Conference*, 2009, pp. 381–386.

[19] K. Gupta, N. Sachdeva, and V. Pudi, "Explicit modelling of the implicit short term user preferences for music recommendation," in *European Conference on Information Retrieval*. Springer, 2018, pp. 333–344.

[20] M. Polignano, P. Basile, M. de Gemmis, and G. Semeraro, "Social tags and emotions as main features for the next song to play in automatic playlist continuation," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 2019, pp. 235–239.

[21] R. C. Kessler, G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky, "Short screening scales to monitor population prevalences and trends in non-specific psychological distress," *Psychological medicine*, vol. 32, no. 6, pp. 959–976, 2002.

[22] L. S. Sakka and P. N. Juslin, "Emotion regulation with music in depressed and non-depressed individuals: Goals, strategies, and mechanisms," *Music & Science*, vol. 1, p. 2059204318755023, 2018.

[23] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality." *Psychological assessment*, vol. 18, no. 2, p. 192, 2006.

[24] L. Polanyi and A. Zaenen, "Contextual valence shifters," in *Computing attitude and affect in text: Theory and applications.* Springer, 2006, pp. 1–10.

[25] R. Trnka, A. Lačev, K. Balcar, M. Kuška, and P. Tavel, "Modeling semantic emotion space using a 3d hypercube-projection: an innovative analytical approach for the psychology of emotions," *Frontiers in psychology*, vol. 7, p. 522, 2016.

[26] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011.

[27] Z. Zhu, J. Li, X. Deng, Y. Hu *et al.*, "An improved three-dimensional model for emotion based on fuzzy theory," *Journal of Computer and Communications*, vol. 6, no. 08, p. 101, 2018.

[28] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[29] S. Buechel and U. Hahn, "Word emotion induction for multiple languages as a deep multi-task learning problem," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 1907–1918.

[30] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[31] P. Saari, M. Barthet, G. Fazekas, T. Eerola, and M. Sandler, "Semantic models of musical mood: Comparison between crowd-sourced and curated editorial tags," in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW).* IEEE, 2013, pp. 1–6.

[32] M. Buccoli, M. Zanoni, G. Fazekas, A. Sarti, and M. B. Sandler, "A higher-dimensional expansion of affective norms for english terms for music tagging." in *ISMIR*, 2016, pp. 316–322.

[33] F. H. Rachman, R. Sarno, and C. Fatichah, "Music emotion classification based on lyrics-audio using corpus based emotion." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 8, no. 3, 2018.

[34] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[35] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 english lemmas," *Behavior research methods*, vol. 45, no. 4, pp. 1191–1207, 2013.

[36] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: characterization, classification, and measurement." *Emotion*, vol. 8, no. 4, p. 494, 2008.

[37] J. K. Vuoskoski and T. Eerola, "Domain-specific or not? the applicability of different emotion models in the assessment of music-induced emotions," in *Proceedings of the 10th international conference on music perception and cognition*, 2010, pp. 196–199.

[38] R. Ferrer and T. Eerola, "Looking beyond genres: Identifying meaningful semantic layers from tags in online music collections," in *2011 10th International Conference on Machine Learning and Applications and Workshops*, vol. 2. IEEE, 2011, pp. 112–117.

[39] ——, "Semantic structures of timbre emerging from social and acoustic descriptions of music," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2011, no. 1, p. 11, 2011.

[40] P. Langfelder, B. Zhang, and S. Horvath, "Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r," *Bioinformatics*, vol. 24, no. 5, pp. 719–720, 2008.

[41] D. N. Klein, R. Kotov, and S. J. Bufferd, "Personality and depression: explanatory models and review of the evidence," *Annual review of clinical psychology*, vol. 7, p. 269, 2011.

[42] S. H. McKenzie, U. W. Jayasinghe, M. Fanaian, M. Passey, D. Lyle, G. P. Davies, and M. F. Harris, "Socio-demographic factors, behaviour and personality: associations with psychological distress," *European journal of preventive cardiology*, vol. 19, no. 2, pp. 250–257, 2012.

[43] S. Garrido and E. Schubert, "Music and people with tendencies to depression," *Music Perception: An Interdisciplinary Journal*, vol. 32, no. 4, pp. 313–321, 2015.

[44] M. Houben, W. Van Den Noortgate, and P. Kuppens, "The relation between short-term emotion dynamics and psychological well-being: A meta-analysis." *Psychological bulletin*, vol. 141, no. 4, p. 901, 2015.

[45] M. Goddard, B. Halligan, and N. Spelman, *Resonances: noise and contemporary music.* A&C Black, 2013.

[46] D. Miranda and M. Claes, "Music listening, coping, peer affiliation and depression in adolescence," *Psychology of music*, vol. 37, no. 2, pp. 215–233, 2009.

[47] D. Miranda, P. Gaudreau, R. Debrosse, J. Morizot, and L. J. Kirmayer, "Music listening and mental health: Variations on internalizing psychopathology," *Music, health, and wellbeing*, pp. 513–529, 2012.