

# Automatic Assessment of Mental Health Using Phone Metadata

**Vivek K. Singh**

Rutgers University, USA. [V.singh@rutgers.edu](mailto:V.singh@rutgers.edu)

**Teng Long**

Rutgers University, USA. [tl505@scarletmail.rutgers.edu](mailto:tl505@scarletmail.rutgers.edu)

## ABSTRACT

An accurate and timely assessment of mental health of individuals is an important building block toward healthy well-functioning societies. Traditional methods for assessing mental health and wellbeing are surveys administered by health-care professionals. Such surveys are labor intensive, costly, and typically intermittent. With the growth in mobile sensing, health informatics, and data science, this work argues a case for using mobile phone meta-data and data analytics to automatically infer an individual's mental health. Specifically, based on a clinically utilized health measure as ground truth (MHI-5) and ten-week phone usage metadata corpus for 45 participants we report that automated machine learning algorithms utilizing phone based features can achieve reasonable accuracy (around 80%) at automatically classifying the level of mental health of an individual. Results could pave way for cheaper, automated, health assessments with timely escalations to mental health professionals and interventions when required.

## KEYWORDS

Mental Health, Ubiquitous Computing, Mobile Computing, Health Informatics, Data Science

## INTRODUCTION

Mental health is a fundamental aspect of human lives and an accurate and timely assessment of mental health has been identified as an important enabler of individual and societal success (Wang et al., 2007). Keeping track of individual wellbeing can allow for understanding the dynamics of the process and support timely interventions (if required) to counter any issues before they become serious health risks. The state of the art in assessing mental health, however, still remains manual, self-reported, surveys administered periodically by health care professionals. Besides being costs for the individual and the healthcare system, such a process also involves sporadic sampling (typically once every few months to years), and most such tests are initiated *after* some serious events or episodes have taken place in an individual's life.

Inspired by the recent trends in Information Science involving advancement of mobile sensing for logging individual information, data science, health informatics, and social informatics, e.g. (Nicholas et al., 2013; Robinson, 2015; Shilton, 2012), this work explores the use of passive phone use meta-data to automatically infer an individual's mental health. The underlying hypothesis in this work is that social behavior is intrinsically linked with one's mental wellbeing (Taylor & Brown, 1988) and mobile phone's provide a rich window into the social behavior of an individual (Shmueli et al., 2014). Hence, appropriately characterizing an individual's social behavior (via phone) and building predictive machine learning models using it could yield important insights into creating *passive* (i.e. without manual attention or effort) predictors for mental health.

This builds upon a recent line of work in ubiquitous computing and social informatics that utilizes phone-based data to automatically infer an individual's traits and characteristics (e.g. Abdullah et al., 2015; Singh & Agarwal, 2016) and contributes to the line of work in health informatics where mobile phones have been used to automatically detect specific mental health related disorders (e.g., stress (Saeb et al., 2015)). Instead of focusing on specific disorders, though, this work aims to utilize mobile phones to infer general purpose mental health using clinically robust instruments.

Based on a ten-week field and lab study involving 45 participants, we report that (1) multiple phone-based behavioral features are intrinsically associated with mental health and general wellbeing; and (2) automated machine learning algorithms utilizing phone based features can achieve reasonable accuracy (around 80%) at automatically classifying the level of mental health and general wellbeing of an individual. The results pave way for better understanding of the underlying phenomena as well as the creation of automatic tools, which with refinement, could be used to trigger appropriate interventions.

The rest of the paper is organized as follows. First, we present background and related work in this area and identify the research questions. Second, we describe the details of the study conducted. Third, we present the data analysis and machine learning based classification approach. Then, we discuss the results, including the implications and the limitations. Lastly, we summarize the findings and identify future work.

## **RELATED WORK AND RESEARCH QUESTIONS**

Over the last decade, multiple efforts have tried to understand and predict mental and physical health using personal data. While some efforts have focused on online social network data (De Choudhury et al., 2013), others have argued the value in using both online and offline user behavior to model mental and physical health (Bogomolov et al., 2014).

Here we discuss the related work in terms of the conceptual arguments connecting social processes with health, identify different trends in using mobile phones for healthcare, and identify the gaps to situate the contributions of this work.

### ***Social Processes and their connections with health***

There is extensive evidence that social capital and social support can help reduce feelings of stress, thus also minimizing its negative effects on physical health (Almedom, 2005; Uchino, 2004; Nabi et al., 2013). The effects of social support on health can be explained by at least two different models. The main effects model proposes that social support has a direct effect on health such that strong social support networks enhance health and well-being by minimizing stress, whereas weak social networks and a lack of social support act as stressors that have negative health effects (Underwood, 2000). Alternatively, the buffering hypothesis proposes that those with greater levels of social support enjoy greater health and wellbeing than those with poor social support when facing stressors in their lives because of the “buffer” space or cushioning provided by the social support (Cobb, 1976).

### ***Using phones to survey user state***

A number of recent efforts have tried to utilize mobile phones to administer fine-grained surveys that understand user state. For instance, (Servia-Rodriguez et al., 2017) discusses a large-scale study aimed at obtaining user mood state, multiple times a day, over time. Similarly, multiple efforts have used PANAS (Positive Negative Affect Sampling) and Ecological Momentary Assessments (EMAs) to understand user situation in a fine-grained manner (Turner et al., 2008; Silk et al., 2011). The primary goal of such studies, typically, is to collect more fine-grained data about user, rather than prediction, and most of such apps focus on aspects that change quite frequently over time (e.g. mood or affect) rather than longer-term clinical attributes like mental or general health.

### ***Using phones to administer interventions***

Many mental health apps focus on administering clinical self-report questionnaires to detect change of moods, stress level and depression level, followed by real-time interventions like text messaging, video, interactive voice, email prompt and virtual reality treatment are provided (e.g., Donker, 2013). Over the last decade, many such studies have suggested that the use of mobile phone and mobile phone apps for interventions in mental health settings (e.g., Harrison, 2011).

We consider building robust health status detectors as a pre-requisite for creating useful interventions. However, the current state of the art in detection is not robust enough for clinical purposes. Further, given the wide variety in mental and physical health and conditions and the associated nuances, there is a lack of general purpose approaches to automatically predict mental and physical health in a robust manner. Hence, we leave the intervention strategies as outside the scope of our current discussion and focus on detection aspects in this work.

### ***Using phones to predict user stress, depression, and moods***

Some recent efforts have proposed the use of GPS data and phone usage to predict depressive symptoms (Saeb et al., 2015; Canzian & Musolesi, 2015). In (Ferdous et al., 2015), the authors have shown that app usages can predict stress levels within the workplace context. In (Sano et al., 2015), the authors have reported the use mobile phone based features to automatically infer the Perceived Stress Score (PSS) for the participants. Similarly (Servia-Rodriguez et al., 2017) make use of mobile phone data to predict daily mood and affect levels of the participants.

### ***Research Questions***

Mental health includes a wide gamut of processes, states, and disorders. These include temporal states like mood and affect, symptoms like stress, and disorders like anxiety (agoraphobia; generalized anxiety disorder; panic disorder; post-traumatic stress disorder; social phobia), mood-based disorders (bipolar disorder including bipolar I and II; dysthymia; major depressive disorder), and substance disorders (alcohol and drug abuse and dependence) (Wang et al., 2007). Each of these marks a specialized set of conditions - as a means of analogy in the physical health, they vary from liver cirrhosis, to flu, to high blood pressure - and hence specialized instruments targeting each disorder is quite different from each other. Past literature has largely focused on detecting specific mental health disorders (Saeb et al., 2015; Servia-Rodriguez et al., 2017) or used ad-hoc (i.e. non-clinical) interpretation of overall health and well-being (Lane et al., 2011). For instance, (Lane et al., 2011) defines “wellbeing” as a weighted average of sleep, social activity, and physical activity. While Lane et al.'s effort at quantifying overall wellbeing is in the right spirit, there as yet few attempts that quantify and try to predict overall mental health and wellbeing using clinically-robust scales.

The SF-36 Health Survey asks 36 questions to measure functional health and well-being from the patient's point of view. It is a practical, reliable and valid measure of physical and mental health that can be completed in five to ten minutes (SF-36V2 Health Survey; Ware & Sherbourne, 1992). It is considered a generic health survey because it can be used across age (18 and older), disease, and treatment group, as opposed to a disease-specific health survey, which focuses on a particular condition or disease (SF-36V2 Health Survey). The survey is meaningful to patients, clinicians, researchers and administrators across the health care spectrum and has various applications including measuring health improvement or decline and assessing treatment effectiveness (Ware & Sherbourne, 1992). It is also reported that patients are more accepting of the test because it does not single out mental health concerns but rather integrates them with questions about the overall health (SF-36V2 Health Survey).

In this work, mental health is measured by the mental health sub-scale included in the SF-36 survey. The questions associated with the sub-scale were originally taken from the full-length (38-question) mental health inventory (MHI) survey. The subscale correlates significantly with the MHI-38 score and has been used as a standalone scale in its own right, known as the MHI-5. We refer to the score as the MHI-5 score in this work.

While there have been multiple recent attempts in the area of mobile phone based mental health measurement (e.g. (Saeb et al., 2015)), this work contributes to the literature by focusing on the following research questions:

***(RQ1) Do certain mobile phone-based behavioral features have intrinsic associations with a person's mental health levels as measured via clinically-robust (MHI-5) survey instrument?***

***(RQ2) Can such mobile phone data and analytics techniques be used to create automated predictive models for an individual's mental health as measured via a clinically-robust (MHI-5) survey instrument?***

## **STUDY**

This study explores the associations between call log data and mental health based on a ten-week field + lab study involving 45 participants. The mental health levels of each participant is measured based on a "ground truth" in-lab survey. The survey is considered "ground truth" in the current study due to its validity tested by other efforts in clinical settings as well as manual validation of the data (no missing values, improper selection of same choices etc.) captured in the current study by the authors.

Motivated by the existing literature in mental health as well as phone based behavioral assessment, a set of call-based features are defined. These features are computed in the dataset by preprocessing raw phone usage metadata. Machine learning based classification approaches have been applied to examine the interconnections between the defined call-based features and two aforementioned health levels.

### **Participants**

The participant's for the study were recruited using flyers, email announcements, and social media posts in the area surrounding a major North American university. A total of 59 participants completed the study. Some of the participants did not complete all the surveys, and some did not enter their unique identifying code consistently across surveys, resulting in a set of 45 participants for whom we have phone usage data as well as the mental health scores from the survey of interest.

Out of the 45 participants in the resulting dataset, 30 were male, and 15 were female. The most common age group was 16-29 years, most common marital status was single, the most common education level was "some college", and the median annual family income was in the range US \$50,000-\$74,999.

### **Method**

The data obtained in this study was collected as part of the <Anonymized> Well-being study. In this study, the participants were invited to attend three in-person sessions which involved filling out several surveys pertaining to health, well-being, cooperation and some demographic data. The survey of relevant to our study are the ones on Mental Health Inventory (MHI-5) subscale from the Short Form Health Survey (SF-36), and demography (details follow).

In the initial session, the participants were asked to read and sign a consent form to join the study and install a study client app on to their cell phones. The study started with an orientation program where participants were trained on how to use the app. Participants could opt out of the study at any point. The survey order was randomized for different participants. The participants were also instructed on the process of uninstalling the app and stopping data collection in the exit session.

All personnel involved in the study underwent human object training and Institutional Review Board certification. In order to protect participants' personal information, all participants were identified using their phone's unique IMEI number which was kept constant throughout the course of the study. This unique identifier was anonymized via one-way cryptographic hashing before any analysis was undertaken.

The participants were compensated up to a sum of \$100 (\$20, \$30, and \$50 respectively for the three in-person sessions) on successful completion of the study.

**Data Collection**

Two types of data were collected from the participants in this study. First, the phone-based data were collected through the app installed on each participants' phone. Second, different survey information were collected via in-person survey sessions.

The phone-based data included (call/SMS) log data and physical location data (GPS). However, in this work we consciously choose to focus on call-based features to be feature-phone compatible. Besides keeping the models and their interpretations simpler, the focus on feature phones allows for wider applicability across in low- and middle- income countries, for many of which mental health issues are among most disabling conditions (Wang et al., 2007). The call log data considered in this study includes date and time of communication event (call), incoming or outgoing direction, duration, and *anonymized* identifiers for both sender and receiver. Actual call audio were *not* collected in this study.

**Health Measures**

Mental health is measured by the mental health sub-scale included in the SF-36 survey. The questions associated with the sub-scale were originally taken from the full-length (38-question) MHI survey. The subscale correlates significantly with the MHI-38 score and has been used as a standalone scale in its own right, known as the MHI-5. The questions in MHI-5 are summarized in Table 1. The questions were assigned scores as prescribed by the SF-36 scoring guide (including score reversals, where required) (SF-36V2 Health Survey, Ware & Sherbourne, 1992).

**Table 1: Questions in the MHI-5 Survey**

S.No.	Question
1	During the past month, how much of the time were you a happy person?
2	How much of the time, during the past month, have you felt calm and peaceful?
3	How much of the time, during the past month, have you been a very nervous person?
4	How much of the time, during the past month, have you felt downhearted and blue?
5	How much of the time, during the past month, have you felt so down in the dumps that nothing could cheer you up?

**Demographic Descriptors**

The participants were surveyed about their demography. We collected the following information: age, gender, marital status, level of education, and level of family's income.

**Phone-based Features**

There has been a growing interest in using social mobile behaviors to infer people's wellbeing. From a conceptual perspective, these efforts are inspired by the literature connecting social capital, social support and mental health (Almedom, 2005; Uchino, 2004). Specifically, we focus on social capital (Putnam et al., 1994), its variants on “bridging” and “bonding” capital as suggested by Williams (2006), and we connect the concepts of “bonding” and “bridging” to those of “strong” and “weak” ties as proposed by Granovetter (1973).

From a practical perspective, prior work on characterizing user behavior via phones (e.g. (Wang et al., 2014; Saeb et al., 2015; Abdullah et al., 2014; Canzian & Musolesi, 2015)) has found that conversation, activity, mobility and sleep have significant correlations with mental health. For instance, the Student Life Project (Wang et al., 2014) reported that the students who engage in more frequent and longer conversations in the day epoch are less likely to feel stressed. Also, they suggest that students with less interaction in the day might have less collaboration with others thus leading to long hours of studying alone and getting less sleep.

Based on a survey of above-mentioned literature we have identified five categories of features (total 19 features) to characterize social behavior. We have defined these features based on the working assumption that certain individual traits that manifest themselves in the long-term social and mobile behavior patterns of the users might be predictive of one's mental and physical health status.

**(A) Call Volume**

Prior literature has connected social connections with mental health in both conceptual and empirical terms (Almedom, 2005; Saeb et al., 2015). Hence, we consider call count, total call duration, distinct users called and distinct users called (outgoing). A summary of the identified features, their explanation, as well as descriptive statistics are presented in Table 2.

**(B) Interaction Dynamics**

An individual's interaction behavior might provide prescriptive clues to her mental and physical processes. Recent literature on phone-based latent attribute identification has used a number of such behavioral features (De Choudhury et al., 2013; Ghosh

& Singh, 2016). Based on this literature we consider the following features: In Out Ratio, Missed Calls Percentage, Call Response Rate, Call Response Latency, and Longest Call Percentage.

**Table 2: Features considered in the study, including their literature support, explanation, and median and mean values in the dataset.**

Category (with supporting literature)	Feature	Explanation	Median	Mean
Call volume (Almedom, 2005; Saeb et al., 2015)	Call Count	Total calls made	274.00	452.31
	Total Call Duration	Time spent on calls (in seconds)	35842.00	79651.89
	Dinstinct Users Called	Unique users called	47.00	50.78
	Dinstinct Users Called (Outgoing)	Unique users called in outgoing calls	30.00	50.78
Interaction Dynamics (de Montjoye et al., 2013; Ghosh & Singh, 2016)	In Out Ratio	Incoming to Outgoing calls ratio	54.08	65.12
	Missed Call Percentage	Percentage of incoming calls not picked.	15.84	17.35
	Call Response Rate	Percentage of missed calls responded back within an hour	37.93	37.19
	Call Response Latency	Average response time (seconds) for missed calls responded	2801.12	2838.63
	Longest Call Percentage	Percentage of overall time spent on single longest call	7.60	10.13
Diversity and Novelty in Contacts (Sih et al., 2004; Singh and Ghosh, 2017)	Call Diversity	Shannon Entropy based measure of spreading calls across different contacts	72.00	71.98
	Call Loyalty	Percentage of calls made to top-three contacts	53.00	51.18
	Number of New Contacts	Number of contacts not present in first four weeks of data	17.00	19.64
	Number of New Contacts (Outgoing)	Number of outgoing call contacts not present in first four weeks of data	12.50	14.23
	Percentage Time Spent with New Contacts	Percentage of time spent talking with New Contacts	8.49	13.51
Tie Strength (Uchino, 2004; Underwood, 2000)	Call Strong Ties	Percentage of calls made to 33% most frequent contacts	84.00	81.82
	Call Weak Ties	Percentage of calls made to 33% least frequent contacts	5.00	6.04
Temporal Rhythms (R. Wang et al., 2014; Abdullah et al., 2014)	DayNight 12am Call Ratio	Ratio of calls made in PM phase to AM phase	3.58	4.47
	DayNight 12am Call Ratio Reverse	Ratio of calls made in AM phase to PM phase	0.28	0.28
	WeekdayWeekend Call Ratio	Ratio of calls made on weekdays to the weekend	2.26	2.44

### (C) Diversity and Novelty in Contacts

Prior literature has stressed the value of diversity as well as change in maintaining one's health and well-being (Sih et al., 2004; Singh et al., 2013). For instance, a combination of an animal's behavioral traits (e.g., exploration, loyalty) have been demonstrated to have significant predictive power on an animal's social stature, survivability, metabolism rates, reproductive success, and other life outcomes (Sih et al., 2004; Wilson et al., 1994). Similarly, recent literature on mobile computing has connected diversity of contacts and novelty over time with individual outcomes (Ghosh & Singh, 2016; Singh et al., 2013). Hence, we have identified the following features to characterize diversity and novelty in the participant's contacts.

First is Call Diversity, which is based on Shannon Entropy and is a measure of how evenly a user's social engagements are distributed between different contacts (Shmueli et al., 2014). A user with low diversity distributes her communication unevenly across contacts, whereas a user with high diversity spends time evenly across many contacts.

$D_i = \sum_j p_{ij} \log_b p_{ij}$ , where  $p_{ij}$  is the percentage of communication (call) between individual 'i' and 'j', and 'b' is the total number of contacts.

Call Loyalty quantifies the level of concentration (or preference) that an individual exhibits towards her strongest ties. While a highly "loyal" individual might have almost her entire social activity focusing on top three contacts, this ratio might be miniscule for another individual with a wider social graph (Singh et al., 2013). In a way, this aspect is tied to the "bonding" social capital derived from strong ties as defined by Granovetter and Putnam, and operationalized by Williams (Granovetter,

1973; Putnam et al., 1994; Williams, 2006). Further we capture the novelty in one’s social patterns via the following features: Number of New Contacts, Number of New Contacts (Outgoing), and Percentage of Time Spent with New Contacts.

**(D) Tie Strength**

Significant prior literature connects strong and weak ties with mental and general health (Thoits, 2011; Uchino, 2004). The main effects model proposes that social support has a direct effect on health such that strong social support networks and satisfactory social support enhance health and well-being, whereas weak social networks and a lack of social support act as stressors that have deleterious health effects (Underwood, 2000). To quantify the relative role of strong and weak ties in the one’s social network and its connections with mental and physical health, we identify the following features: Call Strong Ties and Call Weak Ties. These features quantify the percentage of communication effort that a user devotes to his/her preferred contacts. Following (Singh & Ghosh, 2017), we consider top one-third (33%) most frequent contacts as “strong ties” and bottom third contacts as “weak ties”. It is expected that a user will spend at least 33% of her time with her top third of contacts. However, a higher score (e.g. 80%) could be an indicator of a person’s preference toward engaging significantly more with strong ties rather than spreading the interaction effort more equitably. Similarly, the Call Weak Ties feature quantifies the percentage of communication effort that a user devotes to his/her less preferred (bottom third) contacts. A high weak tie engagement ratio would suggest that the user spends significant communication effort to interact with the less preferred contacts.

**(E) Temporal Rhythms**

Prior literature has connected circadian (biological rhythms) with stress as well as other cognitive functions (Wang et al., 2014; Abdullah et al., 2014). Hence we characterize daily and weekly patterns and the underlying variations in social activities as follows. DN12am Call Ratio, DN12am Call Reverse Ratio, and Weekday Weekend Call Ratio.

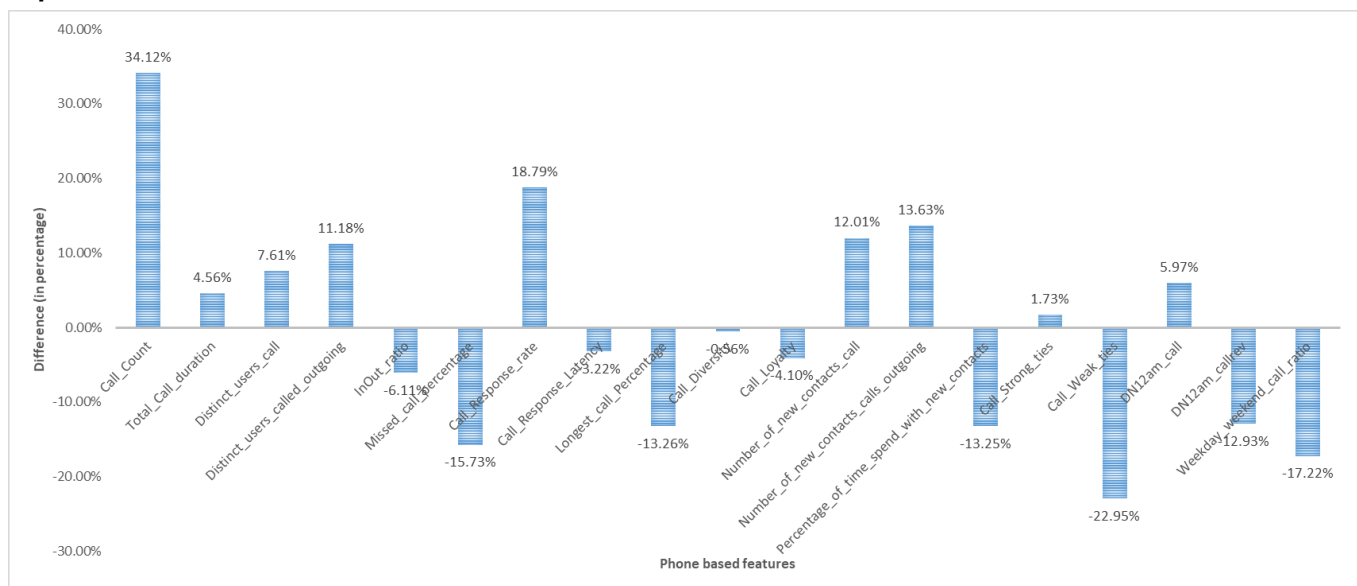
**ANALYSIS AND RESULTS**

**Preprocessing**

The data collected from the Anonymized Wellbeing study came with missing data and in different scales. Before moving on with any data analysis, missing data was imputed from the whole dataset and replaced with the median value of the corresponding feature. After imputation, all the real-valued features were standardized by removing the mean and scaling them to unit variance. Categorical variables, for example demographic features, were encoded by dummy variables for each category.

To build a classification model for high/low mental health score, instances were labeled into two categories via median split. The median value of the sample is chosen because there is no standardized median scores available for the specific survey instrument used in our work. Median split resulted in 2 classes (i.e. labels). There are 22 positive instances (higher mental health score) and 23 negative instances for the mental health classification

**Impact of Phone Use based Features on Mental Health**



**Figure 1: Difference in the feature values for the High and Low mental health groups.**

In Figure 1, we show a comparison of the average values obtained for different features in the two categories as follows:  $difference = \frac{mean(HighMentalHealth) - mean(LowMentalHealth)}{Mean(HighMentalHealth)}$ . As can be seen, the values for 11 out of 19 phone-based features showed a difference of at least 10% between the two groups. Focusing the attention on the top three features in both positive and negative directions, we find that a higher call count, call response rate, and the number

of new contacts (outgoing) are associated with higher mental health. Conversely, a higher percentage of calls made to weak ties, a higher percentage of calls made on weekdays compared to weekends, and a high missed call percentage are associated with lower values of mental health. In general, more social activity, and reciprocity are connected with better mental health. (We discuss some of the implications in the Discussion section.)

**Classification: Feature Selection**

Feature selection is the process of automatically selecting a subset of the defined features to create optimal classification models. Such feature selection is crucial especially for high-dimensional and small sample data to achieve interpretability, enhance generalization and reducing overfitting (Guyon & Elisseeff, 2003). In this study, feature selection is based on optimal subset selection according to mutual information between each feature and the target variable (Peng et al., 2005). Subset of features are selected to be highly correlated with the target variable. We used the implementation in SciKit Learn. The selected features are shown in Table 3.

**Table 3: Features selected for different machine learning models.**

Model	Features
Only Demography Features	Income, Gender
Only Phone Log Features	Call count, Total call duration, Call response rate, Number of new contacts in calls, Number of new contacts in outgoing calls
Demography + Phone Log features	Call count, Total call duration, Call response rate, Number of new contacts in calls, Number of new contacts in outgoing calls

**Classification: Training and Evaluation**

The selected features were then passed to different machine learning algorithms for classification. Specifically, we used three well-established methods implemented in Scikit Learn, namely Logistic Regression, (Gaussian) Naive Bayes, and Random Forest. We used a 10-fold cross validation procedure to evaluate the models. We also used Zero-R (Constant or Majority) without any cross validation as a baseline to facilitate interpreting the results. The following table offers a comparison of the results. It worth noting that AUCROC stands for the Area Under the Receiver Operating Characteristic Curve) (Bewick et al., 2004).

**Table 4: Performance of different machine learning models.**

Method	Demography Only		Phone Log features Only		Demography + Phone Log features	
	<i>AUCROC</i>	<i>Accuracy</i>	<i>AUCROC</i>	<i>Accuracy</i>	<i>AUCROC</i>	<i>Accuracy</i>
Logistic Regression	0.52	0.53	0.75	0.76	0.75	0.76
Naïve Bayes	<b>0.63</b>	<b>0.62</b>	0.64	0.64	0.64	0.64
Random Forest	0.57	0.58	<b>0.80</b>	<b>0.80</b>	<b>0.80</b>	<b>0.80</b>
Zero-R	0.50	0.51	0.50	0.51	0.50	0.51

As shown in Table 4, the “phone log features only” and “demographic + phone log features” based models yielded the best case accuracy of 80% in predicting mental health level. Also as seen in Table 3, the features selected in both models are the same. It is hence unsurprising that the performance of the model remains the same. The classification performance (accuracy of 80%), marks a 56% relative improvement over a baseline majority classifier (Zero-R) and a 29% relative improvement over the “demography only” model.

**DISCUSSION**

The first research question in this work was: *(RQ1) Do certain mobile phone-based behavioral features have intrinsic associations with a person’s mental health levels as measured via clinically-robust (MHI-5) survey instrument?*

As shown in Figure 1, we find multiple features to vary significantly (greater than 10%) between the high and low mental health groups. Focusing the attention on the top three features in both positive and negative directions, we find that a higher *call count*, *call response rate*, and the *number of new contacts (outgoing)* are associated with higher mental health.

This is largely in line with the existing literature on mental health as well as the theoretical construct of social capital (Putnam et al., 1994). The link between social isolation and reduced psychological well-being is well established in sociology, dating

back to Durkheim (Durkheim, 1951(originally published in 1897)). Smaller social networks, fewer close relationships, and lower perceived adequacy of social support have all been linked to depressive symptoms (Barnett and Gottlieb, 1988). From a conceptual perspective a higher social capital is generally associated with higher mental health (Kawachi and Berkman, 2001). However, there have also been reported studies connecting the “dark side” of social capital for mental health (Brown and Harris, 1978).

The results of this study connect a higher call count with better mental health. A higher call count indicates that the person is more social and has access to more number of friends for emotional support when needed (Putnam et al., 1994). A higher call response rate is in line with existing literature and people who are easy to contact can be said to engage with a wider social network than those who remain unavailable (Singh & Ghosh, 2017). Lastly, a higher number of new contacts (outgoing) also points to a constantly growing network of contacts and again a higher social capital.

The top three predictors of lower mental health scores were a *higher percentage of calls made to weak ties*, a *higher percentage of calls made on weekdays compared to weekends*, and a *high missed call percentage*. A higher percentage of calls made to *weak ties* needs to be contrasted with the choice to make more calls to *strong ties*. While strong ties are expected to provide emotional support to the individual, weak ties often provide information and access to resources (Granovetter, 1973). The results from this study suggest that focusing the communication on strong ties (as opposed to weak ties) is important for better mental health. Similarly, the choice to make more calls during the weekdays needs to be contrasted with the calls made over the weekend. For most people, weekday activities are more likely to involve work and weekend activities to be focused on more leisure and social events. Hence, it is not surprising that a higher ratio of weekend calls is found to be connected with better mental health. Lastly, a higher missed call percentage might reflect a person’s lower tendency to reciprocate in social networks. Reciprocity is a key component of social capital (Putnam et al., 1994) and a lower social capital is often connected with lower mental health (Nabi et al., 2013; Uchino, 2004).

The second research question in this work was: *(RQ2) Can such mobile phone data and analytics techniques be used to create automated predictive models for an individual’s mental health as measured via a clinically-robust (MHI-5) survey instrument?*

As shown in Table 4, the phone log features based models yielded the best case accuracy of 80% in predicting mental health level which was significantly higher than a demography data based model (62%). One way to interpret these results is that having mobile sensing data (phone-based features for 10 weeks) may allow for the creation of a detailed model for personal behavior which goes beyond static demographic descriptors (age, gender, income level) for predicting mental health scores. The classification performance (accuracy of 80%), marks a 56% improvement over a baseline majority classifier. Thus, providing clear evidence of interconnections between the mobile features and mental and physical well-being and also motivating further work in this direction.

Note though, that we do not consider the current prediction levels to be sufficient for completely automated monitoring and decision making. However, they could act as a “triaging” step and medical experts could be called in to look more closely based on the initial results from such a system. We further identify certain limitations in the current study. This study focuses on a modest sample (N = 45) of participants from a relatively healthy university population. We also note that none of the results obtained here are causal, and hence the directionalities of the explanation could indeed be inverse. For instance, less healthy individuals could make less friends, or individuals with less friends may succumb to health issues more readily. Taking into account these limitations, we will be cautious in generalizing the results obtained until they are verified at scale over samples of more representative populations. Despite these limitations, this study is the first of its kind. To our knowledge, there have been no previous studies undertaken that analyze the link between stable clinically robust mental health instruments (namely MHI-5 subscale of SF-36) and phone-based behavioral features.

We also note the privacy and ethical considerations in giving a person a health score based on passive data collection (Tufekci, 2015). We suggest an explicit opt-in policy (as also adopted in this work) and use of such tools as merely a triage step before trained professionals can step in and take a closer look at the individual circumstances. Given the current health-care setup, we posit that automated tools that can reduce the costs and burdens on individuals and health care professionals could be a useful way forward. While a more nuanced policy debate is required, rather than abstaining from presenting such outcomes, we choose to raise awareness about these new prospects and informing the policy debate around them.

The results open the doors to a methodology that, with refinements and validation, could be used at scale. Mobile phones are now actively used by more than 3 billion users, hence the proposed method could potentially be applied to estimate the mental health levels for billions of individuals. In future, this work could also have multiple implications for health care professionals, health informatics professionals, mobile phone service providers, social scientists and individuals themselves. For example, the suggested methodology could help health care professionals study mental health at scale in the society. Similarly, as mobile phones are increasingly used, these methods could be used to trigger preventive interventions before serious lapses or incidents take place.



## CONCLUSIONS AND FUTURE WORK

This work establishes a link between stable clinically robust mental health instrument (namely MHI-5 subscale from SF-36) and phone-based behavioral features. Specifically, it demonstrates how a selection of phone based behavioral features can be used to build prediction models whose AUCROC and accuracy are promising and encouraging (around 80%) at recognizing different levels of mental health. While we consider the results to be preliminary, they could be enhanced by future work by including a larger number of participants, more detailed phone-based features, and considering larger time durations. Furthermore, the suggested phone-based model could be expanded to study and predict other personal and societal behaviors and outcomes such as trust, cooperation, and happiness. Taken together such methods open ways to employing mobile sensing, data analytics, and social informatics for better human health and wellbeing.

## REFERENCES

- Abdullah, S., Matthews, M., Murnane, E. L., Gay, G., & Choudhury, T. (2014). Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing(pp. 673–684).
- Almedom, A. M. (2005). Social capital and mental health: An interdisciplinary review of primary evidence. *Social science & medicine*, 61(5), 943–964.
- Barnett PA, Gotlib IH. Psychosocial functioning and depression: distinguishing among antecedents, concomitants, and consequences. *Psychol Bull.* 1988;104:97–126.
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13: receiver operating characteristic curves. *Critical care*, 8(6), 508.
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. S. (2014). Daily stress recognition from mobile phone data, weather conditions and individual traits. In Proceedings of the 22nd acm international conference on multimedia (pp. 477–486).
- Brown GW, Harris TO. *Social Origins of Depression: a Study of Psychiatric Disorder in Women*. London: Tavistock; 1978.
- Canzian, L., & Musolesi, M. (2015). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In Proc. 2015 acm international joint conference on pervasive and ubiquitous computing (pp. 1293–1304).
- Cobb, S. (1976). Social support as a moderator of life stress. *Psychosomatic medicine*, 38(5), 300–314.
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Icwsn* (p. 2).
- de Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. In *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 48–55).
- Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of medical Internet research*, 15 (11), e247.
- Durkheim E. *Suicide*. New York: Free Press; 1951. (Originally published 1897.)
- Ferdous, R., Osmani, V., & Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. In *Pervasive computing technologies for healthcare (pervasivehealth), 2015 9th international conference on* (pp. 225–228).
- Ghosh, I., & Singh, V. K. (2016). Predicting privacy attitudes using phone metadata. In *International conference on social computing, behavioral-cultural modeling, and prediction, sbp-brims*. Springer.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360–1380.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3 (Mar), 1157–1182.
- Harrison, V., Proudfoot, J., Wee, P. P., Parker, G., Pavlovic, D. H., & Manicavasagar, V.(2011). Mobile mental health: review of the emerging field and proof of concept study. *Journal of mental health*, 20 (6), 509–524.
- Herbert, T. B., & Cohen, S. (1993). Stress and immunity in humans: a meta-analytic review. *Psychosomatic medicine*, 55 (4), 364–379.
- Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban health*, 78(3), 458–467.
- Lane, N. D., Mohammad, M., Lin, M., Yang, X., Lu, H., Ali, S., . . . Campbell, A. (2011). Bewell: A smartphone application to monitor, model and promote wellbeing. In *5th international icst conference on pervasive computing technologies for healthcare* (pp. 23–26).
- Nabi, R. L., Prestin, A., & So, J. (2013). Facebook friends with (health) benefits? exploring social network site use and perceptions of social support, stress, and well-being. *Cyberpsychology, Behavior, and Social Networking*, 16 (10), 721– 727.
- Nicholas, D., Clark, D., Rowlands, I., & Jamali, H. R. (2013). Information on the go: A case study of europeana mobile users. *Journal of the American Society for Information Science and Technology*, 64 (7), 1311–1322.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27 (8), 1226–1238.
- Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern italy*. Princeton university press.

- Robinson, L. (2015). Multisensory, pervasive, immersive: Towards a new generation of documents. *Journal of the Association for Information Science and Technology*, 66(8), 1734–1737.
- Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet Research*, 17(7), e175.17
- Sano, A., Phillips, A. J., Amy, Z. Y., McHill, A. W., Taylor, S., Jaques, N., . . . Picard, R. W. (2015). Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In *Wearable and implantable body sensor networks (bsn)*, 2015 IEEE 12th international conference on (pp. 1–6).
- Servia-Rodriguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017). Mobile sensing at the service of mental well-being: A large-scale longitudinal study. In *Proc. International conf. on world wide web* (pp. 103–112).
- SF-36V2 HEALTH SURVEY Health Survey. (2016). (Available online. <http://campaign.optum.com/optum-outcomes/what-we-do/health-surveys/SF-36V2-health-survey.html>)
- Shilton, K. (2012). Participatory personal data: An emerging research challenge for the information sciences. *Journal of the American Society for Information Science and Technology*, 63(10), 1905–1915.
- Shmueli, E., Singh, V. K., Lepri, B., & Pentland, A. (2014). Sensing, understanding, and shaping social behavior. *IEEE Transactions on Computational Social Systems*, 1(1), 22–34.
- Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: an integrative overview. *The quarterly review of biology*, 79(3), 241–277.
- Silk, J. S., Forbes, E. E., Whalen, D. J., Jakubcak, J. L., Thompson, W. K., Ryan, N. D., . . . Dahl, R. E. (2011). Daily emotional dynamics in depressed youth: A cell phone ecological momentary assessment study. *Journal of experimental child psychology*, 110(2), 241–257.
- Singh, V. K., & Agarwal, R. R. (2016). Cooperative phoneotypes: exploring phone-based behavioral markers of cooperation. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 646–657).
- Singh, V.K., & Ghosh, I. (2017). Inferring Individual Social Capital Automatically via Phone Logs. In *Proceedings of the ACM Human Computer Interaction*, 1 (CSCW), Article 95.
- Singh, V. K., Freeman, L., Lepri, B., & Pentland, A. S. (2013). Predicting spending behavior using socio-mobile features. In *Social computing (socialcom)*, 2013 international conference on (pp. 174–179).
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, 103(2), 193.
- Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of health and social behavior*, 52 (2), 145–161.
- Tufekci, Z. (2015). Algorithmic harms beyond facebook and google: Emergent challenges of computational agency. *J. on Telecomm. & High Tech. L.*, 13, 203.
- Turner, M., Love, S., & Howell, M. (2008). Understanding emotions experienced when using a mobile phone in public: The social usability of mobile (cellular) telephones. *Telematics and Informatics*, 25 (3), 201–215.
- Uchino, B. N. (2004). *Social support and physical health: Understanding the health consequences of relationships*. Yale University Press.
- Underwood, L. G. (2000). *Social relationships and health. Social support measurement and intervention: A guide for health and social scientists*, 1.18
- Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Borges, G., Bromet, E. J., . . . others (2007). Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO World Mental Health Surveys. *The Lancet*, 370 (9590), 841–850.
- Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., . . . Campbell, A. T. (2014). Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp.3–14).
- Ware Jr, J. E. (2000). SF-36 health survey update. *Spine*, 25 (24), 3130–3139.
- Ware Jr, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Medical care*, 473–483.
- Williams, D. (2006). On and off thenet: Scales for social capital in an online era. *Journal of Computer-Mediated Communication*, 11 (2), 593–628.
- Wilson, D. S., Clark, A. B., Coleman, K., & Dearstyne, T. (1994). Shyness and boldness in humans and other animals. *Trends in Ecology & Evolution*, 9(11), 442–446