Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Marisa H. Olsen Date

BlūBot: Integration of an mHealth application and EMR system to Increase Frequency of Depression Screenings in Non-Clinical Settings

By

Marisa Hall Olsen

Degree to be awarded: Master of Public Health

Applied Public Health Informatics

K.C. Decker, M.S.

Committee Chair

Chris Karr, M.A.

Committee Member

BlūBot: Integration of an mHealth application and EMR system to Increase Frequency of Depression Screenings in Non-Clinical Settings

By

Marisa Hall Olsen

B.S.
Georgia Southern University,
2001

Thesis Committee Chair: K.C. Decker, M.S.

Adjunct Faculty
Rollins School of Public Health, Emory University

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in Applied Public Health Informatics
2017

Abstract

BlūBot: Integration of an mHealth Application and EMR System to Increase Frequency of Depression Screenings in Non-Clinical Settings

By Marisa Hall Olsen

Introduction: Detecting and treating depressive disorders is a public health priority. About one in ten patients seen by primary care physicians has a depressive disorder, yet these disorders largely remain underdiagnosed. Post diagnosis, a study found that 67% of respondents with mental illness were interested in monitoring their symptoms through applications on their phones. With the wearable market expected to increase from 275 million devices in 2016 to 477 million devices in 2020, providers and researchers need a system to capture and use health data from these devices. The purpose of this thesis is to develop a prototype for integrating data from an mHealth application, collected near continuously, that screens for depression and translates information into an open electronic medical record system (EMR) for clinical and research analysis.

Methods: The BlūBot app collects raw mobile phone sensor data and transforms it into location and mobile device use features significantly correlated with PHQ-9 scores. A predictive model is trained on user PHQ-9 responses to generate predictions. One user carried a mobile device that collected data points near continuously for 10 weeks. To compare accuracy of the BlūBot model, 66 days of observations (collected from December 28, 2016 through March 3, 2017) were used to calculate the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

Results: Infrastructure was created to inject app data into an OpenEMR system and to alert provider of depressive episodes. When compared to a running average of PHQ-9 responses (RMSE=1.99, MAE=1.52), the BlūBot model did not perform as well (RSME= 2.48, MAE=1.97). This may be due to a lack of overall variation in the user scores and a series of spikes in the user scores during week 5 of data collection.

Conclusion: The BlūBot prototype can be used for depression screenings in between doctor visits, providing intervention opportunities when depressive episodes occur. Steps should be taken to further validate the model and improve upon the application through user studies. This prototype, including its application, infrastructure, and repository, can all be repurposed and expanded upon to incorporate data from various wearables and mHealth applications into EMR/EHR systems for public health.

BlūBot: Integration of an mHealth application and EMR system to Increase Frequency of Depression Screenings in Non-Clinical Settings

By

Marisa Hall Olsen

B.S.
Georgia Southern University,
2001

Thesis Committee Chair: K.C. Decker, M.S.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in Applied Public Health Informatics
2017

Acknowledgements

I would like to thank my thesis committee for their expertise and guidance. Thank you to my thesis chair, KC Decker, who consistently pushed me to go further than I ever thought possible and who always knew how to talk me down when I got stuck. Also, a big thank you goes out to my field advisor, Chris Karr, for making all of this possible. I truly appreciate all of the non-billable hours he dedicated to making my idea a reality. I would not have been able to implement my project without his technical skills and resources.

I would also like to thank Flo Wagner, Andy Coppolino, Toby Goldsmith, MD, and Dan Hoke, for sharing your subject knowledge and for connecting me to additional resources.

Thanks to everyone in the Rollins School of Public Health EMPH program, both colleagues and faculty, who offered guidance and support during this process, especially Mark Conde and Laurie Gaydos.

Thanks to my family and friends, for all of their love and support. I would like to thank my daughter, Bryn, who tucked herself into bed many nights while I held Skype calls with my field advisor; I could not ask for a more patient and loving daughter. I owe a special thanks to my husband, Scott, who never let me give up and who took care of the household during the countless hours I spent working on my thesis. He has been my rock through all of this.

Contents

List of Figures

List of Tables

1 Chapter 1: Introduction

1.1 Background and Rationale

Detecting and treating depressive disorders is a priority for public health in terms of disease prevention and health promotion because they are associated with an increase in prevalence of chronic diseases and disability.[1] It is projected that by 2020, depression will be second only to heart disease in its global burden of disease, based on disability-adjusted life years.[2] The estimated financial burden of major depressive disorder in the United States is $210 billion a year, mainly due to indirect costs of the illness.[3] Although depression can be successfully treated with psychotherapies and medication, a majority of adults do not seek care.[1] About one in ten patients seen by primary care physicians has a depressive disorder, yet these disorders largely remain underdiagnosed.[4] The estimates of those undiagnosed are even higher in populations with chronic diseases, like diabetes.[5] In accordance with organizational recommendations, the rate of screening for depression has increased in the primary care setting over the past three years, shifting away from diagnoses limited to the mental health care setting.[6] Measurement-based screening tools used in primary care settings can be beneficial to a patient diagnosed with depression by detecting suboptimal response to treatments and guiding alternative treatment recommendations.[7] However, some patients in a primary care setting are uncomfortable discussing depression with primary care physicians (PCPs), as they feel a PCP is not the appropriate person in which to disclose this information.[8] This attitude may impede progress to reaching recommended goals of increased depression screenings by primary care physicians, so it is imperative to find ways to educate patients and screen for depression in ways that are preferable to the patient.[9] The advent of mobile health (mHealth) applications has

introduced novel approaches for depression screening and treatment beyond the doctor's office. Mobile health applications have presented the opportunity to collect continuous, longitudinal data from patients outside of the clinical setting. These types of applications have demonstrated feasibility for a range of psychiatric illness including schizophrenia, bipolar disorders, substance abuse disorders, anxiety disorders, and depression.[10, 11] Sensors imbedded into mobile devices offer a wealth of data related to psychiatric illnesses that normally would not be available to practitioners. Despite the advantages of real-time data collection in "real-life," this information is not used in clinical interactions or decision-making because data from these mental health mobile applications are not easily shared with practitioners or researchers.[12] Because the variety of data from mobile device sensors is only expected to increase in the near future, now is the time to devise a way to harness the full potential of these devices and applications.[11] By connecting data generated from mobile mental health applications to an electronic health record (EHR) or electronic medical record (EMR), practitioners can get a better idea of patient's mental health between visits. There are many barriers to integrating patient generated data, like those created by mHealth applications, into an EHR/EMR. This project will create an mHealth application and supporting infrastructure to collect data via a mobile device to predict depression and alert health providers to occurrences of depression through an EMR system.

1.1.1 Review of Literature

> "Grief and fear, when lingering, provoke melancholia."
>
> Hippocrates, 460–377 B.C.

In a review of depression, Wong and Licinio created a timeline reflecting the cornerstones of the history of major depression: Between 460 and 377 BC, the Hippocratic classification of diseases believed that melancholia was caused by the "humour" of black bile and was treated by removing or purging blood. In the 1600's, the Anatomy of Melancholy was written by Richard Burton, describing the Renaissance views on melancholy.[13] According to Wong and Licinio, while the existence of melancholia or depression has been documented since ancient times, our current understanding of the disorder did not surface until the latter half of the 19th century. During this time period, Emil Kraepelin and Sigmund Freud had two different views on depression: Kraepelin classified depression as a disease based on clinical and anatomical concepts, while Freud saw depression as internalized anger or loss.[13] In 1900, the International Classification of Diseases was introduced and by 1910, the origins of psychotherapy emerged.[13] The 1950's ushered in the birth of psychopharmacological treatments as well as the first edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM). The first self-rating scale for depression, the Beck Depression Index (BDI 21), was used in 1961.[13] By the late 1970's and early 1980's, it was shown that psychotherapy sessions, alone or in combination with new antidepressant drugs, called selective serotonin re-uptake inhibitors (SSRIs), could be effective in the treatment of depression. SSRIs remain the treatment of choice, but the 2000's ushered in the experimental use of an anesthesia called ketamine for improving symptoms of depression [13, 14] The proliferation of the World Wide Web created a new environment in which to treat depression. Web-based therapies proved to be successful alternatives to traditional face-to-face sessions, allowing the expansion of depression screening and intervention into the realm of mHealth applications.[11, 15]

1.1.2 Diagnosis of Depression

Depression is not easily diagnosed; there are no blood tests or scans that can give a definitive diagnosis. The assessment "should include a physical examination, patient history, psychological testing, and interviews."[16] Classification of depressive disorders is based on subjective descriptions of symptomology as outlined in the American Psychiatric Association's DSM.[13] By the end of the 20th century, depression was typically diagnosed through structured interviews in a doctor's office. These diagnostic interviews were designed to elicit verbal descriptions of symptoms that were congruent with DSM criteria.[17] There were several issues with this methodology. Recall was affected by both lapses in time between the experience and the assessment, as well as cognitive distortions, observed with depression, which cause some individuals to attend to negative experiences and thoughts.[18, 19] Those who sought assessment sometimes inflated or over exaggerated their symptomology in order to receive a diagnosis and treatment. It is likely that by the time a person comes in for assessment, the individual has been impaired by the severity of symptoms, which makes the illness more difficult to treat. Furthermore, traditional methods of assessment, administered by mental health professionals in a clinical setting, were not reflective of the individual's typical environment, which called ecological validity into question.[19] The very nature of these structured interviews conducted by clinicians have been shown to influence the responses of individuals, rather than reflect the patient's every day experiences. In 1993, Mokros conducted a study to compare the presence of three symptoms essential to a diagnosis of depression: sadness, irritability and an inability to feel pleasure, as determined by a clinical evaluation, versus the Experience Sampling Method (ESM). Agreement between these two methods ranged from 29%- 57%, with disagreements being bidirectional.[17]

In recent years, the Department of Health and Human Services (HHS) and the United States Preventive Services Task Force (USPSTF) have recommended screening for depression in primary care visits by using a screening tool like the Patient Health Questionnaire (PHQ-9).[20] The PHQ-9 is a self-report consisting of nine questions related to depressive symptoms, which can be used for screening, diagnosing, monitoring, and determining the severity of depression. The questionnaire is useful to clinical practice because it is brief, easy to score, and can be used repeatedly to assess changes in depression in response to treatment.[21] Although there has been some skepticism as to the usefulness of these screenings, a recent systematic review found evidence that depression screenings can be useful to postpartum women and the general population.[20]

1.1.4 The Role of Technology in Depression Screening

As mobile technology has become highly integrated in daily life, mobile devices have the capacity for ecological momentary assessment (EMA), specifically the experience sampling method (ESM), where behavior is captured in real-time or closer in time to experiences in one's natural environment.[22] In the early 2000's, research on mood disorders shifted toward ESM and EMA, which resulted in some important findings. One study found that other methods of screening and diagnosis may be less accurate because depressed people had a tendency to recall past events and moods with a negative bias and overestimated their stress levels and depressive symptoms.[18] The repeated assessments built in to EMA/ESM allowed users to correlate symptoms and improvements in mood to various aspects of their lives.[19, 23] Research has shown that daily self-ratings entered into a mobile mental health app provided comparable results comparable to traditional depression screening tools.[9, 10] In fact, it has been suggested that these types of devices may hold an advantage over traditional screening tools because there

is a shorter recall period which may compensate for measurement errors.[24] In 2015, Torous et al. found that when a subset of the PHQ-9 was administered via an mHealth app, three times a day, respondents were more likely to report suicidal ideation than on a paper based test.[25] That said, one limitation that EMA assessments have in common with traditional assessments is that self-reports may be confounded with mood disorders, which can skew results.[23]

1.1.5 Automating Assessments

The development and validation of depression lexicons have allowed researchers to use analysis of text to screen for depression. There are several types of text that can be evaluated to indicate depression, such as questions posted to mental health sites, blog posts, social media posts, and text messages.[22, 26] In fact, a study by Neuman et al. found that a text analysis methodology, known as Pedisis, was correct 84% of the time in classifying signs of depression in blog posts analyzed.[26] These depression lexicons serve as a foundation for text mining and mental health analytics incorporated into mHealth applications.

A majority of adults in the United States own smartphones and those with mental illnesses own smartphones at a rate similar to the national average.[11, 12] Furthermore, one study found that 67% of respondents with mental illness were interested in monitoring their symptoms through applications on their phones.[12] Mobile devices come equipped with many sensors, including cameras, accelerometers, gyroscopes, magnetometers, proximity sensors, light sensors, barometers, thermometers, and pedometers. Newer mobile devices may even have biometric sensors such as heart rate monitors and fingerprint sensors.[27] Wearable devices containing sensors have been used to collect physiological data and transmit this information to an mHealth application via Bluetooth technology in order to measure stress levels between psychotherapy sessions.[28] Mobile device sensors can be used to provide details about an

individual's mood and behavior. GPS sensors can detect location, accelerometers measure movement and gross motor activity, while call and text-messaging logs can monitor social interaction. Voice and tone can also be recorded to estimate mood.[11, 19] Similar to the way clinicians assess depressive symptoms, mHealth applications can use behavioral observations to offer a new way to screen for depression without the constraints of traditional methods.[11, 19]

According to a 2015 paper by Torous, Staples, and Onnela, while the volume, velocity, and variety of data generated by a mobile device offers one of the biggest opportunities in psychiatry, it also presents the biggest challenge. Clinical decision support tools, like algorithmically generated decision trees, have been successfully applied to large volumes of complex data to determine if people were walking, standing, or driving. Decision trees divide data into classes based on a set of rules, which can predict outcomes based on complex information, like that of mental health data.[11] A clinically relevant screening tool can be created by pairing mobile device data with appropriate analytic methodologies, so that behavioral observations can be used to detect signs of depression in a manner similar to that of a psychiatric assessment.[11] The potential opportunity here is that with the use of mobile device data, subtleties that may indicate impending depression can be detected and headed off before the symptoms become more serious.[19]

Several studies have been conducted to investigate the feasibility and validity of screening for depression using passive data from mobile devices in combination with various statistical models.[29, 30] In 2011, Morturu et al. conducted research to determine the link between sleep, sociability, and mood, based on mobile device sensor data. The researchers found that individuals who were less sociable were more likely to exhibit poor mood and the study discovered a bidirectional relationship between sleep and mood.[31] This study served as a

foundation for the development of the Ginger.io app, that detects patterns of depression and offers coaching to its users for a fee.[32]

There were two relevant studies published in 2015 based on a mobile device sensor data-collection app called Purple Robot. Saeb et al. introduced the GPS sensor features known as location entropy and circadian movement to a mobile device-based app for depression and discovered a stronger relationship between the location features and PHQ-9 scores than daily EMA ratings and PHQ-9 scores.[33] In a related study, the results showed a strong correlation between PHQ-9 scores and "normalized entropy, location variance, home stay, circadian movement, and phone usage duration and frequency."[30]

Also in 2015, Ben-Zeev et al. conducted a study to investigate whether mobile device sensor data could be repurposed to screen for depression.[19] In this study, the embedded mobile device sensors were configured to passively track behavior as users carried their phones as they normally would, without asking users to input any data. Over a period of ten weeks, the mobile devices passively collected data and the subjects completed daily stress self-reports as well as mental health assessments related to stress, depression, and loneliness, before and after the data collection. Data collected by the mobile device sensors indicated that geospatial activity, sleep duration, and variability in geospatial activity were associated with daily stress levels reported by participants. Speech duration, sleep duration, and geospatial activity were associated with depression levels over the duration of the study. The first model fitted to the data indicated that geospatial activity and sleep duration were inversely correlated with daily stress levels.[19]

1.1.5 Integration of Mobile Device Data into EMR

Although published literature on the interface between a passive depression app and EMR are lacking, there are some studies that share similar concepts and demonstrate the lack of

interoperability between mobile device applications and electronic medical or health records. A study by Park, et al. developed applications to use mobile device sensors for patient monitoring in a hospital setting. In this study, the researchers developed a methodology to integrate data from mobile devices into the hospital EMR. Figure 1 shows the technical architecture design for the interface used in this study:[27]

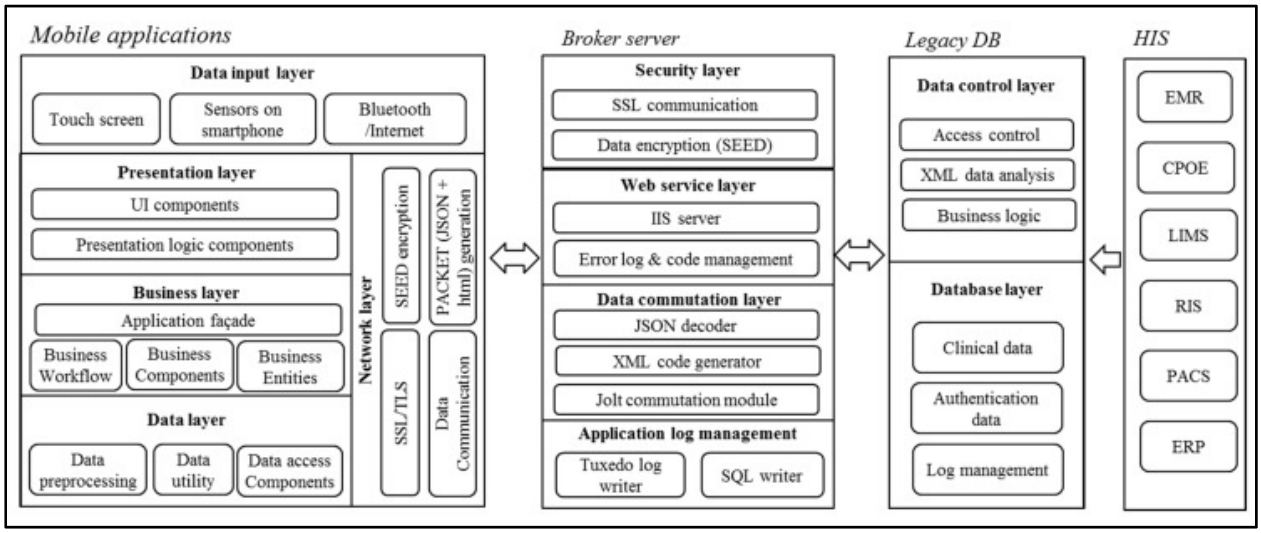


Figure 1. Technical Architecture Design for Park et al. Study

The research revealed technical issues that inhibit the use of this kind of data. The type of data generated from mobile device sensors had not been used previously in a hospital setting, so there was no standardization for this kind of data. Also, security vulnerabilities in the wireless networks and applications in which protected health information was transmitted were discovered.[27]

Some mHealth applications use a web portal or patient portal to allow the patient to enter and send information to providers, while others allow patients to email or text information.[34, 35] The developments of wireless technologies such as 3G and Wi-Fi have been explored for use in telemedicine in fields such as Parkinson's and Alzheimer's. These wireless networks have also

been used for monitoring patient health through sensors worn on the body or placed in the environment.[36] In 2015, Epic made changes to their MyChart EHR app, so that it could sync with Apple's Healthkit and FitBit via a patient portal.[37]

The challenge of interoperability between EMRs and mHealth technology persists, but a possible solution to extract data from information siloes without standardization of systems has been introduced. Health Level 7 (HL7) is a group of widely accepted patient health interoperability standards, but HL7 V2 and V3 have both faced challenges and limitations in implementation.[38] Complaints about HL7 V2 include a lack of: a uniform application data model, official approach for modeling data elements and messages, specifications for application and user roles, and a precise, scalable design.[38] According to Kasthurirathne et al., V3 made improvements, but still faced criticism surrounding the difficulty of implementing its underlying Reference Information Model (RIM). Additionally, V3 still faced slow adoption, problems with usability in specialist domains, and complex documentation.[38] A new HL7 health data exchange standard, called Fast Health Interoperable Resources (FHIR), has been introduced in an effort to overcome these criticisms and achieve plug and play interoperability between many types of systems.

FHIR is designed to address one of the biggest shortcomings of the Consolidated Clinical Document Architecture (C-CDA) created for HL7 V3: lack of granularity. Instead of sharing an entire summary document in order to obtain one data point on a patient, FHIR can communicate with a system to request and retrieve one piece of a patient's health data, like medication history, for instance, but it also can be used to create documents that can be used like C-CDAs.[38, 39] FHIR uses a Representational State Transfer (REST) architecture application programming interface (API) and web data format standards (Extensible Markup Language (XML), JavaScript

Object Notation (JSON)) to obtain data from resources.[38] Efforts are underway to encourage the adoption of FHIR to support interoperability.[40] A 2016 Interoperability Proving Ground project is currently exploring ways to screen for depression using PHQ-9 and interface that data with EMR using HL7 standards such as FHIR.[41] The FHIR specification is currently in “Draft Standard for Trial Use" (DTSU) and is expected to be published in 2017.[42]

1.2 Problem Statement

Depression is one of the leading causes of disability in people 15 years and older.[43] Between 2005-2010, the incremental economic cost of major depressive disorder rose 21.5%, from $172.3 billion to $210.5 billion dollars.[3] According Cassano and Fava, depressed individuals experience a notable amount of subjective suffering and their ability to function socially or at work becomes impaired. Depressive disorders are often comorbid with prevalent chronic diseases such as high blood pressure and diabetes; they can also affect the outcome of comorbid diseases like cancer and cardiovascular disease.[4] Cassano and Fava state that depression rivals the other major illnesses of this century in terms of burden of morbidity. Furthermore, people who experience depressive disorders have an increased risk of mortality, particularly due to suicide.[4] Depression is the most prominent risk for suicide; two-thirds of people who commit suicide suffer from this illness.[4] To put the risk of suicide from depression into a broader context, suicide accounts for only 0.9% of deaths in the general population, while 21% of patients with recurrent depression attempt suicide.[4]

The severity of depression covers a broad range, from temporary bouts of sadness, to major depressive disorder. One must meet the symptom criteria published in the DSM for major depressive disorder in order to be diagnosed with clinical depression.[44] According to the

DSM-V, to meet the criteria for diagnosis of major depressive disorder, one must experience five or more of the recognized symptoms (in addition to other criteria), a majority of time, over a two-week period, with at least one symptom being either (1) depressed mood or (2) loss of interest or pleasure (excluding symptoms clearly attributable to other medical conditions).[45] See list of recognized symptoms in Appendix A.

According to the NIMH, depression is a public health priority due to its detrimental impact on society and because effective treatment options for the disease exist. An increase in depression screenings is imperative to finding cases that have gone undetected.[4] The U.S. Preventive Service Task Force (USPSTF) recommends an increase in the screenings of adults for depression when the appropriate support services are in place.[43]

Furthermore, because depression can be chronic, but not necessarily static in its symptomology, there is a need to evaluate symptoms outside of the doctor's office. Bouts of depression are likely to be missed between appointments with physicians.[19] Even individuals being treated for depression may go 3 months or more between appointments.[11] Although there is a wealth of data relative to mental health being collected by mobile device devices, it lacks clinical applicability because the information is trapped within the device. This knowledge gap goes beyond simply integrating data from a mobile device; there are challenges with security, HIPAA, and validity of the screening apps themselves, which must be considered. There is a need for more continuous assessment outside of periodic doctor visits and the data collected on patients' mobile devices may hold the key to accomplish more frequent depression screenings, but this data must be translated into validated screening tools in order to be useful to the patient and provider.

1.3 Purpose

The purpose of this thesis is to develop a prototype for integrating data from an mHealth application, collected near continuously, that screens for depression and translates information into an open electronic medical record system (EMR) for clinical and research analysis.

Definitions:

APPLICATION PROGRAMMING INTERFACE (API): A collection of functions, commands, objects, and protocols which programmers use to develop software or interact with an external system. By using an API, developers do not have to create code from scratch.[46]

MOBILE HEALTH (mHealth): "Medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices."[47]

MOBILE DEVICE: A mobile electronic device with extensible computational capabilities, like a smartphone or tablet. Its abilities can be enhanced an expanded by third parties beyond what is included by the manufacturer of the device.

ELECTRONIC MEDICAL RECORD (EMR): "An EMR contains the standard medical and clinical data gathered in one provider's office."

ELECTRONIC HEALTH RECORD (EHR): "EHRs are designed to contain and share information from all providers involved in a patient's care. EHR data can be created, managed, and consulted by authorized providers and staff from across more than one health care organization."[48]

1.4 Significance

While there is no substitute for the judgment of a mental health professional, the advancement of technology has offered alternative methods for timely preliminary depression screening that are applicable to public health. Data from mobile devices, when connected to a PCP's EMR, can be used as a screening tool for primary care practices without the usual constraints of administering assessments during a PCP visit, thus achieving the recommendations of Healthy People 2020 and the USPSTF.[26] By incorporating depression data from a mobile device application into an EMR, there is a potential to capture the cases of depression missed in individuals who are part of an EMR network. These patients can then be identified and treated for the disorder, thereby yielding a significant financial savings and reduction of the diseases comorbid with depression. Additionally, the data integrated into the EMR, produced by the app, can be used for evaluating outcomes of treatments and improving treatment methods for depression. This information can be used to estimate the prevalence of depression and to target interventions for those specific demographic groups and geographic areas with higher prevalence.

2 Chapter 2: Methodology

2.1 Introduction

Clinicians and researchers are interested in data generated from wearable devices and mHealth applications, but that data is hard to collect and use practically in a health care setting.[41, 49] The overall goal of this project is to utilize, modify, and extend existing technologies, so that data generated by an mHealth app can be sent to and stored within an EMR

system. The application uses data to detect patterns of depression by generating a model for predicting depression. These predictions will be compared to the user's self-reported scores in order to determine how closely the generated predictions align the user's self-reports. When depression is detected, an alert will be sent to the EMR to notify a user's health provider. The prototype will be tested to determine if automated messages are triggered by PHQ-9 composite scores of 10 or greater.

2.2 Data Types and Description

A literature review was performed in PubMed for terms related to mental health applications and EMR. Terms included "mHealth," "mental health app," "depression screening," "PHQ-9," and "EMR integration." Given that passive depression screening is a relatively new field and experts are still working on methods to interface mHealth apps with EMR, subject matter experts (SMEs) were interviewed for this project, in addition to performing a literature review. Interviews with SMEs from Emory University/Emory Brain Health Center included a Psychiatrist, an IT Director, and informaticians who work with their proprietary EMR system. One of these informaticians had recently completed a pilot study involving the administration of the PHQ-9 via an mHealth application and integrating those scores into the EMR. Interviews were unstructured, but centered on workflow, visualization design, HIPAA, message standardization, business associate agreements, costs of proprietary EMR systems, adoption, and security. In addition, Chris Karr, an mHealth application developer and researcher, was consulted numerous times via Skype throughout the project. He served as an expert on these types of mHealth applications and as a technical resource to help implement the project design. Conversations with these SMEs resulted in a number of important issues that need to be

addressed when designing a cloud-based EMR integration. Feedback from the SMEs was taken into consideration during project design.

Passive data can be collected continuously and autonomously without user intervention; this includes sensor data, system data, communication data, and public and personal online data (Karr, PDK Conceptual Overview, 2017). The general approach to creating a model for this study was influenced by the Saeb, et al. Purple Robot study because it demonstrated a good example of how passive sensor data can be used to predict depression.[30] Saeb et al. used PHQ-9 scores as the dependent variable in their study, finding that changes in location and phone usage are correlated with changes in PHQ-9 scores. Passive data was preprocessed to extract features related to GPS location and phone usage. The first procedure for the GPS location data is to determine if the data was sampled from a stationary state (e.g., sitting at home, watching TV) or a transitional state (e.g., riding in a car). The movement speed at each location sample is estimated by calculating its time derivative and by establishing a threshold to delineate between stationary and transitional states.[30] Movement speeds above 1 km/h are considered to be transitional.

Clustering is the second procedure applied to stationary state location data samples to determine the locations in which users spend a majority of their time, like home, work, etc.[30] While the Saeb et al. model used repeated k-Nearest Neighbors for clustering, BlūBot uses density-based spatial clustering of applications with noise (DBSCAN) in order to make the clustering process more tractable on the mobile device. Points that are closely packed together are grouped together and those that lie alone, in low-density regions, are marked as outliers.[50]

Location features known as home stay, location variance, circadian movement, and normalized entropy were found to be significantly correlated with PHQ-9 scores.[30, 33] Home stay refers to the amount of time a person spends at home, with the location where time is spent between the hours of midnight and 6AM designated as home.[30] Location variance is the variability in a person's GPS location, as calculated by a logarithm of the sum of statistical variance of the longitude and latitude of stationary state location data.[30] The logarithm to calculate location variance is:

$$Location\ Variance = \log\left(\sigma_{lat^2} + \sigma_{long^2}\right) (1)$$

Circadian movement measures the degree in which a person's order of locations follows a 24-hour, or circadian, rhythm and is measured with a Lomb-Scargle Periodogram.[30, 51] Entropy is based on the concept of entropy from information theory and is defined as the variability of time spent by users at the location clusters.[30, 52]:

$$Entropy = -\sum ipilog\ pi(2)$$

"Where each *i*=1, 2, …, *N* represented a location cluster, *N* denoted the total number of location clusters, and *pi* was the percentage of time the participant spent at the location cluster *i*."[30] When time spent is spread more evenly across various location clusters, it is known as high entropy, while time spent less equally across location clusters is deemed low entropy. Entropy may capture information reflective of a person's occupation or lifestyle that makes it difficult to compare between subjects, so this feature was normalized in the Saeb et al. to compare across groups.[30] Normalized entropy is calculated by dividing the entropy by its maximum value:

$$\text{Normalized Entropy} = \text{Entropy}/\log \text{N}(3)$$

Entropy may capture information indicative of a person's occupation or lifestyle that makes it difficult to compare between subjects, so this feature was normalized in the Saeb et al. to compare across groups. "Normalized entropy" refers to the frequency with which a person visits different locations and how the frequency is distributed over various locations.[30] People with depressive symptoms are more likely to favor some locations over others and are likely to visit fewer locations. Less variability in location and more time spent at home are associated with symptoms of depression, such as decreased motivation, reduction in activity, and social withdrawal.[30]

Mobile device usage frequency and duration is also correlated with PHQ-9 scores. Mobile device use frequency is determined by the number of times in a day a user interacts with the device, while usage duration is the total number of seconds that a user spends interacting with the mobile device. This usage data was measured by the number times the device screen was on, omitting events that lasted less than 30 seconds.[30] Features found to be significantly correlated with PHQ-9 scores were retained for the prediction model in this study. A linear regression model was used to estimate the user's PHQ-9 scores based on the features extracted using device sensor data.[30]

Saeb et al. collected the sensor data from the study and then analyzed it offline, once the data collection was completed. However, an mHealth application producing a near continuous stream of PHQ-9 predictions is desirable when monitoring depression from an EMA standpoint. The application created for this project, known as BlūBot, was built using Passive Data Kit (PDK), an API for gathering and transmitting sensor data on mobile devices

(https://passivedatakit.org/#what). The BlūBot component of the overall technical infrastructure serves 3 primary purposes:

1. Continually collects sensor data from the mobile device
2. Prompts the user to complete the PHQ-9 daily
3. Generates continuous predictions informed by both the passive data and user-provided PHQ-9 scores

In an effort to provide feedback to the user, visualizations of the data used to inform model predictions were developed within the app. A display containing a graph of the history of daily PHQ-9 scores was added to the home page of the app, including a record of both the user composite scores and model-predicted scores. Data stream displays were also added, which include GPS location and screen on/off events.

Mirroring the Saeb study, the PHQ-9 was used as the gold standard for depression screening. The PHQ-9 consists of eight main questions and a follow-up question related to self-harm. Items are rated on a Likert scale, according to severity, from 0-3. The questionnaire and its scoring can be found in its entirety in Appendix A. The data from mobile device sensors were used to approximate the results of a PHQ-9 screening. The information was used to calculate a composite score, which was categorized to reflect the severity of depressive symptoms: not depressed (0 to 4), mildly depressed (5 to 9), moderately depressed (10-14) and severely depressed (15-30). The BlūBot application includes displays based on the actual PHQ-9, so that the user can provide feedback to inform the model's performance.

2.3 Project Design

This project is designed to produce a working prototype to collect passive data and PHQ-9 user responses, translate that information into a model to predict depression, and send that PHQ-9 history to an EMR to inform health care providers of depressive episodes. This project was comprised of several design goals:

- Use open-source resources whenever possible
- Create an mHealth app to:
    - Collect passive data related to location and phone usage every five minutes
    - Process the passive data on the mobile device to generate features significantly correlated to PHQ-9 scores
    - Remind user to provide daily PHQ-9 responses
    - Use those responses and feature information to generate PHQ-9 score predictions
    - Send the data to a repository that will connect to an EMR
    - Create a user interface that includes PHQ-9 history and data stream information
    - Allow user to customize model settings
- Develop the supporting cloud infrastructure:
    - The outcome of this project is a prototype consisting of an mHealth app and supporting cloud infrastructure that uses data from the user's mobile device to detect indicators of depression, and then sends an alert to the user's EMR in order to notify the user's provider
- Inject data from the mHealth app repository into an EMR
    - Create a method to send PHQ-9 user responses and prediction history to an EMR
    - Link patient identifier in EMR to identifier on mHealth app

- Trigger an alert within the EMR to notify health care provider when a user's PHQ-9 score is 10 or greater (indicating moderate depression)
- Modify EMR dashboard to include embedded visualizations of the patient's PHQ-9 history

2.4 Procedures

2.4.1 Ethics

The Emory Institutional Review Board (IRB) determined that this project was a non-human subject study, therefore, was exempt from review.

2.4.2 Database and Server Procurement

Chris Karr, a member of Saeb's team who created the Purple Robot mobile app and server infrastructure, provided technical advice and the server and database resources, in an attempt to replicate Saeb's findings on a per-participant basis. An Amazon Web Service (AWS) Elastic Compute Cloud (EC2) server stored and processed data from the BlūBot mobile application. EC2 "is a scalable, user-configurable compute service that supports multiple methods for encrypting data at rest."[53] A PostgreSQL object-relational database was configured for data storage on the EC2 server.[54] In addition to the PostgreSQL database, a Django web application server was hosted on EC2 to perform the data analysis and provide embedded visualizations for the OpenEMR system. A HIPAA-compliant cloud configuration was not implemented for this project, due to study constraints.

2.4.3 App Development

After exploring several mHealth applications designed to screen for depression, including Purple Robot, a decision was made to create a new Android application to suit the goals of this project. This new application applied the techniques used by Purple Robot in the Saeb et al. study, but re-implemented them in a self-contained application to support model generation and data visualization that Purple Robot was incapable of doing on its own. The new application uses Passive Data Kit to collect data about device on/off screen state and GPS location to inform model predictions, but it also captures user PHQ-9 responses for training the model. Material, a visual design language (https://material.io/), guided the design of the application to display visualizations of data collected from the data stream. Although the BlūBot application is similar in architecture to the platform used for Purple Robot, several improvements have been made in terms of its interface simplicity, ease of deployment, and visualizations of collected sensor data. BlūBot was deployed to testers using HockeyApp, a testing and feedback service for debugging and deploying pre-release mobile software.[55] Software testers enroll in HockeyApp by providing an email address to the developer, who then signs them up for testing the product.

2.4.4 Electronic Medical Record System

After exploring EMR system options through a Google search, OpenEMR was chosen for this project because it is a free and open-source electronic health records and medical management application. It is free to download, use, and modify and documentation and support forums are included at no charge. OpenEMR has been used as an alternative to expensive, proprietary systems in low resource settings. While the product lacks the functionality of more expensive systems, it still supports HL7 messaging and is ONC ambulatory certified. OpenEMR is currently in the process of obtaining certification for ONC Meaningful Use Stage 2.[56]

The OpenEMR software was downloaded from the website using instructions for standard installation (open-emr.org). The OpenEMR system was modified to accept new types of data generated by the BlūBot application. Using administrative features, a patient alert was created to ensure that PHQ-9 scores would be reviewed by the provider. Because a composite score of 10 or greater indicates moderate depression, a threshold was set to alert the provider, via the clinical messaging system, if a user's PHQ-9 composite score or the model predicted score met or exceeded this number.[21]

2.4.5 Data Standards

Because HL7 is the accepted standard for electronic health messaging, a Google search was performed with the search terms "HL7 and PHQ-9." The search results led to a post within HL7's FHIR website containing examples for encoding a PHQ-9 questionnaire into a FHIR compatible document using Logical Observation Identifiers Names and Codes (LOINC®) standards (https://hl7.org/fhir/cqif/questionnaire-cqif-example.json.html). LOINC® is a no-cost international data standard which codifies observations, measurements, and laboratory tests, in an effort to create data exchangeability for the purpose of clinical care, outcomes management, and research.[57] In 2010, Vreeman, McDonald, and Huff described the effort to encode patient assessments and surveys, like the PHQ-9, into LOINC® specifications. The questions, answers, and composite scores pertaining to the PHQ-9 are now available in the LOINC® database.[58]

Unfortunately, after searching through OpenEMR discussion forums, it became apparent that OpenEMR does not support FHIR at this point in time, due to the lack of an API.[59] According to the discussion boards, collaborators within the OpenEMR community have been working on creating an operational API, but have stalled on its development.[60] That said, with FHIR emerging as the next big health care data standard, the PHQ-9 document was drafted in

compliance with FHIR. Ideally, there will be an API for OpenEMR available in the near future or an integration engine can be built between the app and the system to support communication between the two. Despite the technical limitations of OpenEMR, workarounds were created to allow the achievement of project goals, as described below.

2.5 Instruments

While BlūBot was developed and tested on a range of Android mobile devices, from an earlier Nexus model to the Pixel, the Nexus 5x was chosen for the project because the user owned and actively used this model. The Nexus 5x was carried with a "dummy" user for a period of ten weeks, ensuring that the device was powered at all times. Throughout the ten-week period, the BlūBot application collected data from sensors embedded in the mobile device related to phone usage and location. The application took a sample of the GPS location every 5 minutes and measured phone usage based on the screen on and off events. The GPS location is stored on the device to measure entropy and circadian movement, based on the original features used in the Saeb et al. Purple Robot study. These features, which were found to be significantly correlated with depression, were incorporated into the BlūBot prediction model. Unlike the Purple Robot study model, built retrospectively on previously collected data, BlūBot collects data on a near continuous basis, and uses that passively-gathered data to generate depression prediction models. PHQ-9 responses are also collected through a self-report feature and used to inform the predictive model. In contrast to the Saeb et al. model, which generates a group model based on previously collected data and additional analytical tools (MATLAB), BlūBot generates an individualized model on the mobile device itself, without the need for additional off-line processing or analytical tools.

The BlūBot application employs a store-and-forward architecture, so that the data was collected and stored locally on the mobile device and then uploaded to the BlūBot Data Repository when network connectivity was available. BlūBot transmitted sensor data, user PHQ-9 responses, and generated PHQ-9 model predictions to the BlūBot Data Repository using commercial-grade web encryption. From there, The BlūBot Data Repository processed the incoming data and stored the raw information in a PostgreSQL database. Although MD5 hashing and encryption were excluded to speed up prototype development, it is recommended that future versions incorporate these protections to anonymize personally identifiable information prior to and transmission.

2.6 Data Analysis

Raw data was collected from the Nexus 5x mobile device and processed within the BlūBot application on the mobile device using an algorithm to extract features for the model predictions (see Figure 2). Data points were generated by segmenting the user's PHQ-9 scores and mobile device sensor data history into discrete time units. Those data points were then used to calculate the Ordinary Least Squares regression. To create a new prediction for a particular date, the system worked backwards from that particular prediction date and extracted all the relevant features for each step. The algorithm would obtain the data history and then use the data points to ascertain the linear equation for the model. Finally, the current date's features were extracted to generate the PHQ-9 prediction for that day.

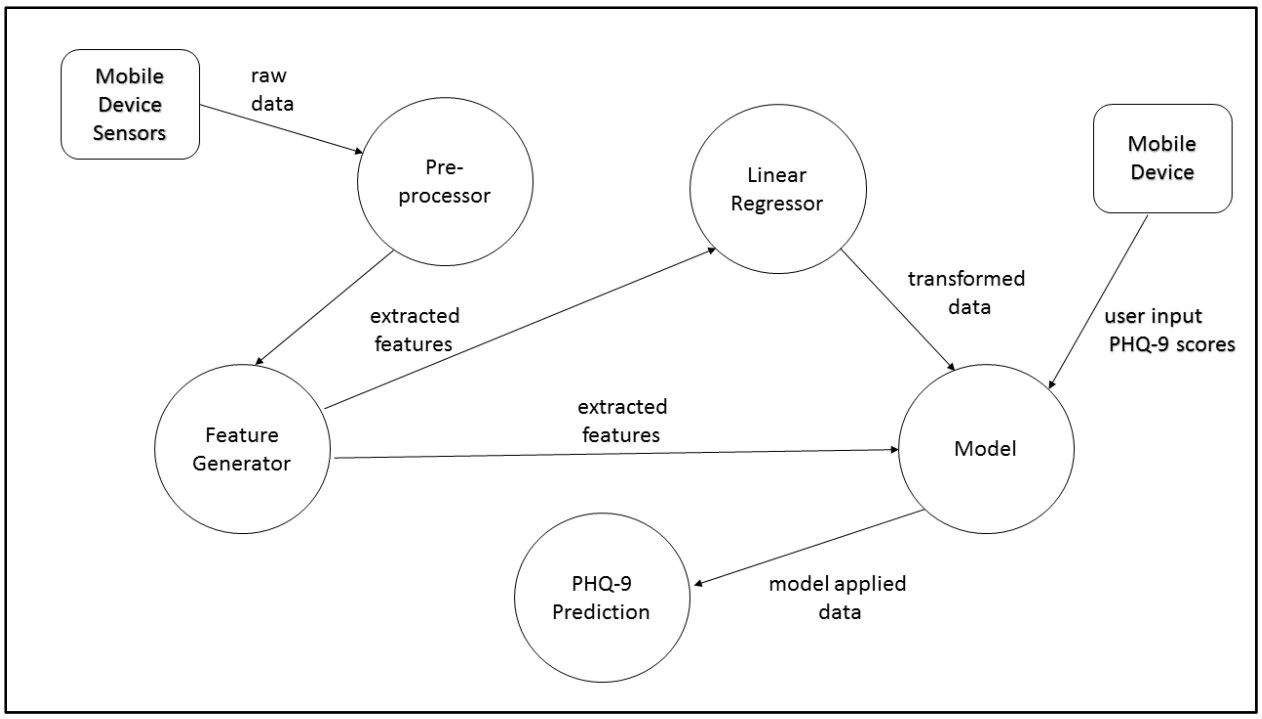


Figure 2. Data Flow: Transformation of raw data into PHQ-9 predictions

The model can be configured to train on one day (86400000 ms), one week (604800000 ms), two weeks (1,209,600,000 ms), or one month (2,629,746,000 ms) of evaluation data, but there were only enough data observations collected in the study to run the model on a one-day or one-week training evaluation window with all 10 features intact.

Data related to the user PHQ-9 scores and various model predictions were exported from the BlūBot server into a text file and then imported to a spreadsheet in Microsoft Excel 2016. The total file size downladed was 407 KB for the entire 10 weeks. Once in Excel, the data was cleaned, removing any duplicates and any model prediction values that did not have a corresponding user score for the same date. Because multiple predictions can be created for the same day, the latest generated prediction for each day was selected for analysis, in order to maintain consistency. Upon inspecting the PHQ-9 graphs generated in BlūBot, it appeared as if the one-day training and evaluation window model was a better fit to the PHQ-9 user responses

than the one-week training and evaluation window, which was later verified in analysis. Between December 15, 2016 and March 11, 2017, a total of 82 user responses and corresponding one-day BlūBot training model predictions were collected. Of note, some of the dates were missing from the dataset for the one-week model, most likely due to crashes within the app while generating predictions, resulting in a smaller number of observations for the same time period of data collection.

Only consecutive days that contained both a user and model predicted score were retained for the initial analysis. Additonally, the model composite score predicted on the first of these consecutive dates was removed because there was no user score for the previous day, resulting in a maximum composite score prediction of 27. There is an argument to be made for removing outliers in small data sets because it can skew the results of model fit tests.[61] Therefore, data was filtered so that only the one-day training and evaluation-based model and the corresponding PHQ-9 user responses for each date remained. This resulted in 66 observations between December 28, 2016 and March 3, 2017.

In order to analyze the data, a statistical procedure was chosen that would determine how well the model predictions fit the actual user responses, known as root mean square error (RMSE):[61]

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}$$

This statistical analysis is commonly used in machine learning and forecasting to measure the error between observed and predicted values, including studies using artificial intelligence to

predict depression.[62] Actual PHQ-9 user scores were compared to composite scores estimated by the model by calculating the RMSE. The variance between the user scores and model predictions was totaled, squared, then divided by the total number of observations. The square root of the result was taken. Variance, Standard Deviation, Correlation Coefficient, R^2, and Mean Absolute Error (MAE) were also calculated to provide additional measures of comparison. These calculations were repeated for comparison of the average running user score, which had been used as a placeholder until the BlūBot model was applied to the data.

3 Chapter 3: Outcome

3.1 Introduction

Several important issues emerged from discussions with SMEs. According to a psychiatrist interviewed for this project, one of her patients uses a mood tracking app, but in order for the psychiatrist to view the data, the patient has to generate print-outs from the application and bring them to appointments. At best, the print-outs could be scanned as PDF and uploaded to the EMR (Dr. Toby Goldsmith, personal communication, May 4, 2016). According to an informatician working on a mHealth application/EMR integration, the EMR is the weakest link, not the application. There are no fields built into the EMR for patient generated data because each entry must be associated with a provider (Andy Coppolino, personal communication, June 14, 2016). The SMEs interviewed for this study had been involved in a pilot study that administered the PHQ-9 to patients on an iPad in the clinic's waiting room, upon check-in for their appointments. The results were integrated into the patients' electronic medical records for the provider to review upon encounter. Based on this pilot study experience, SMEs

identified several important issues that must be considered when integrating data from an mHealth app into an EMR:

1. Workflow Analysis: A workflow study should be performed to determine how each user interacts with the data.
2. Design: The design of the application must be conceptualized and tested prior to implementation. The design of the EMR dashboard and how the information is to be stored and displayed within the system will be based on workflow studies and stakeholder input. (Flo Wagner, personal communication, June, 29, 2016).
3. Engineering: Engineers build to agreed specifications, while informatician must translate user needs to engineers and create scenarios for them.
4. Security: Security reviews with 3rd party must be completed to ensure security requirements are met (Dan Hoke, personal communication, May 31, 2016).
5. Business Associate Agreement (BAA): A legal BAA is necessary so that the institution shares liability and responsibility for protecting PHI with a service provider.
6. Pilot Agreement: A Pilot Agreement is made with proprietary systems to address such issues as making firewall adjustments and VPN configurations.
7. Institutional Approval and Support: To expand the beyond pilot study, the project must go through institutional system, committee, etc. to get priority and funding (Flo Wagner, personal communication, June, 29, 2016).

Based on the information provided by SMEs and the literature review, a solution was created and implemented to inject data from a mental health app into an EMR. An overview of the project architecture can be seen in Figure 3.

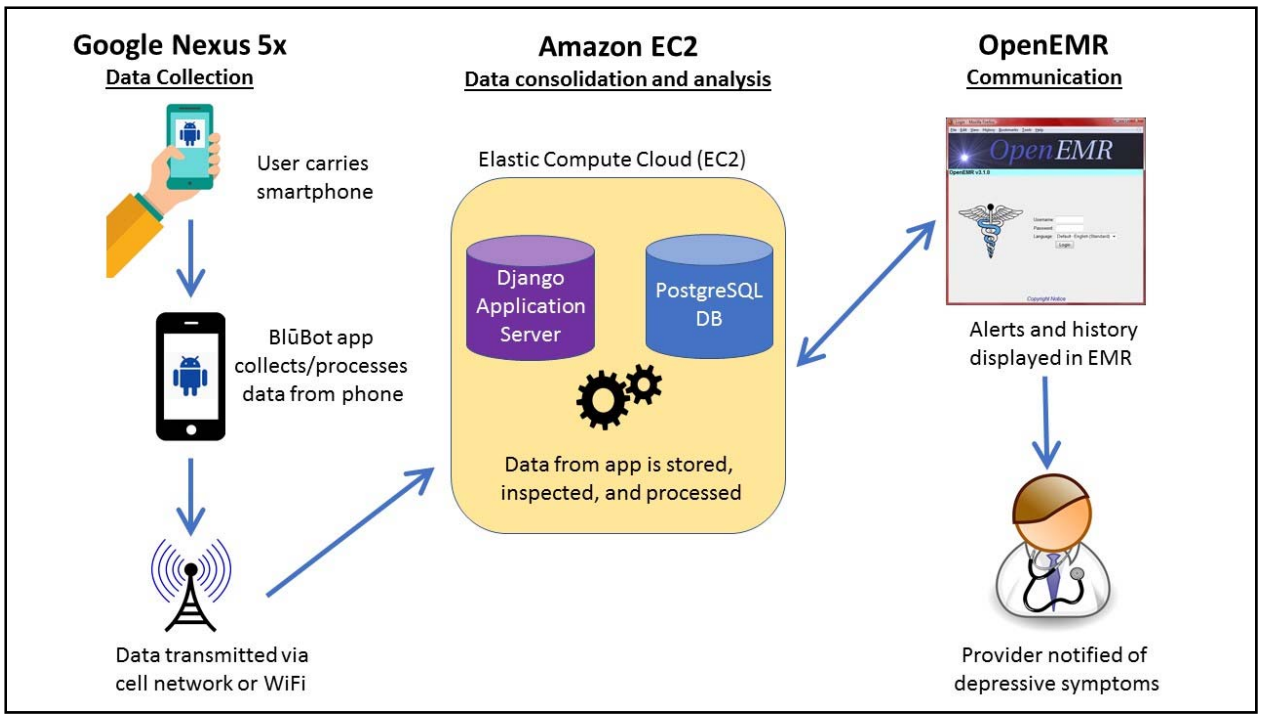


Figure 3. Overview of the BlūBot and OpenEMR integration architecture

3.2 Project Outcome/Deliverables

3.2.1 BlūBot App

There are four data streams within the BlūBot application: GPS sensor data, screen on/off status, and PHQ-9 scores (separated into user responses and the auto-generated PHQ-9 score predictions, as calculated by the model). Each data stream has its own database within the application that resides on the mobile device. The local storage of data allows the application to build a history for each data stream in order to generate a prediction model trained on historical observations. A data flow diagram of the BlūBot application can be seen in Figure 4.

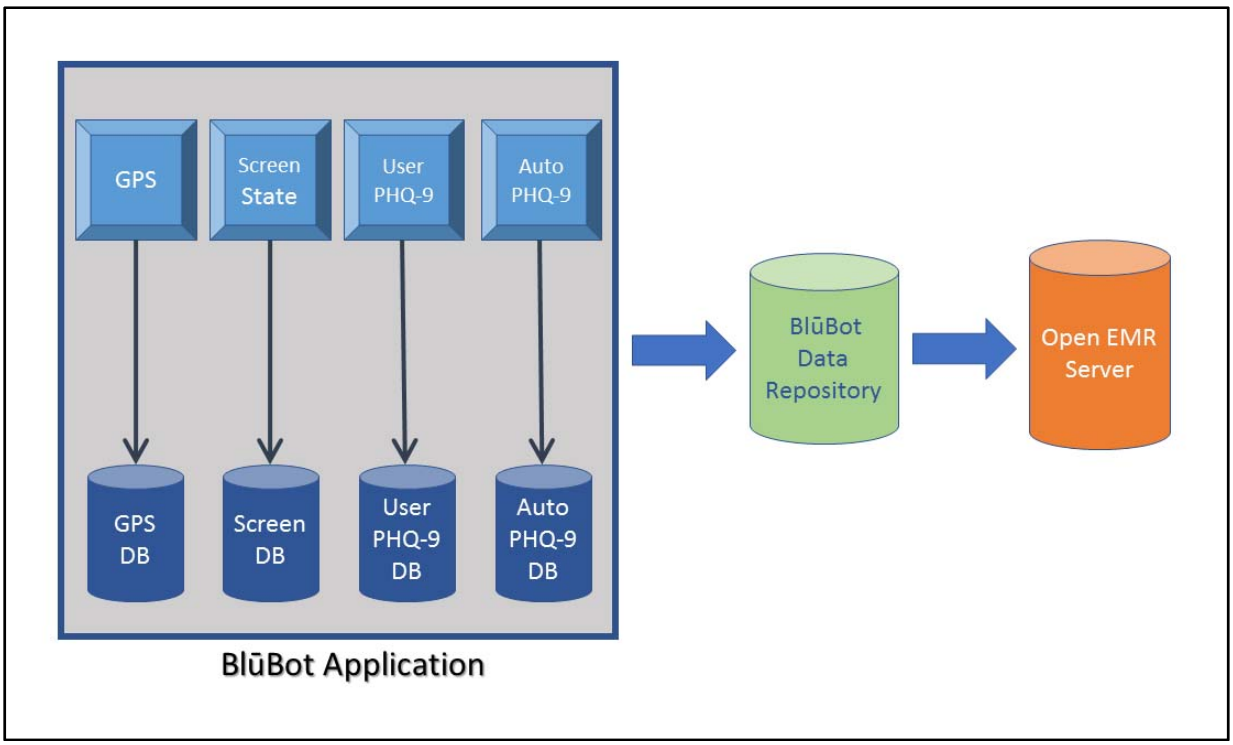


Figure 4. Data flow diagram for BlūBot application

On the BlūBot user's side, displays were created to visualize the data sources in the app. The home screen of the app shows a graph of the user's history of PHQ-9 scores, with the date listed on the X axis and the total composite score on the Y axis. The brown line on the graph represents the user composite scores and the dashed blue line represents the composite scores, as predicted by the application model (Figure 5).

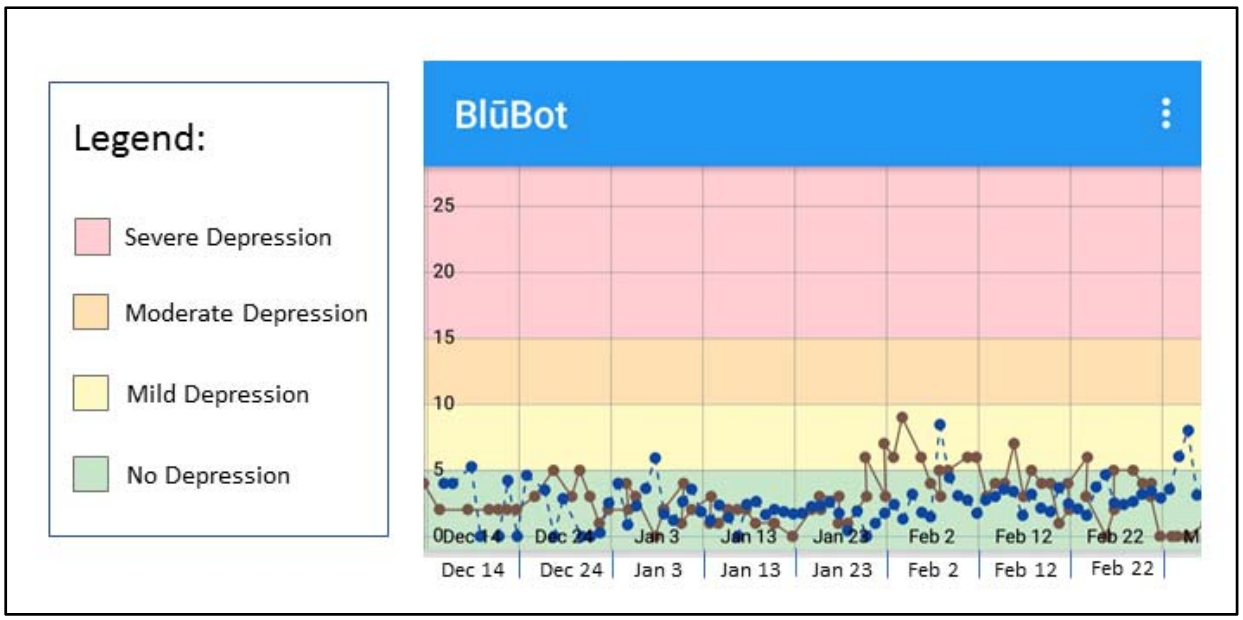


Figure 5. BlūBot Graph: User and Model-Predicted PHQ-9 Scores
The brown line represents the user PHQ-9 composite scores and the dashed blue lines represent the model predicted composite score.

Located below the PHQ-9 graph is a breakdown of the scores for the 9 individual questions, with X representing the composite score. Scores with a brown person icon on the left are the ones generated by the user. The scores with a blue robot icon to the left are the model-predicted scores (Figure 6). Questions 1-9 are ordered from left to right, with the composite score at the end of the row.

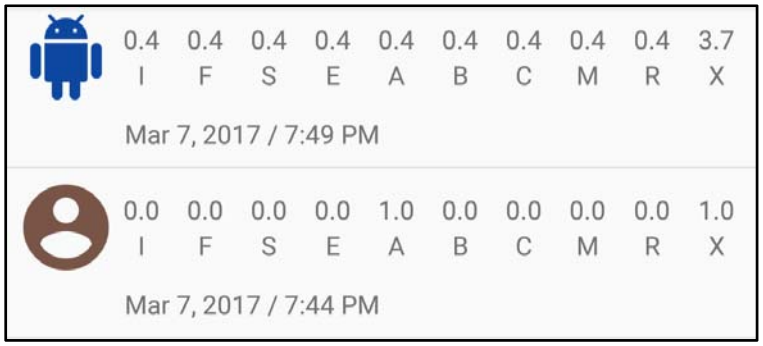


Figure 6. Itemized PHQ-9 scores: predicted and actual
The blue robot icon represents model predictions and the brown person icon represents the user input scores.

From the home screen, the user can choose “Data Stream” from the menu in the upper right hand corner. Choosing this option brings up a series of displays that visualize the data being generated by the app that is used to inform the model predictions.

Device Location Display

Within the device location display lies a heatmap showing the locations a user has visited. This map is informed by the GPS location clustering within the application. The larger and darker the circle is, the more time that has been spent at that location (Figure 7).

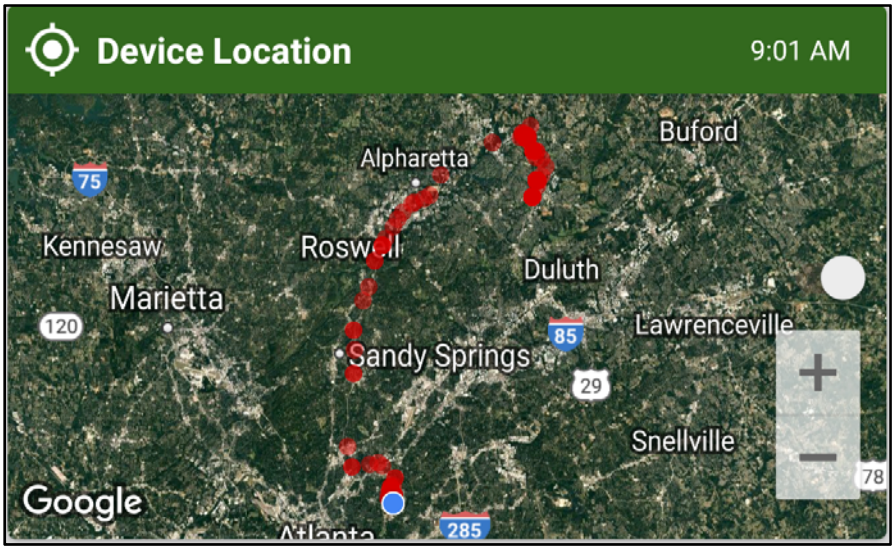


Figure 7. Example of Device Location in BlūBot user display

Screen State Display

This displays the amount of time the mobile device screen has been on or off (Figure 8). There are three conditions for screen state: on, off, or doze. When the screen is activated for a duration of 30 seconds or longer, it is considered a screen “on” session and is recorded by the application. “Doze” refers to a power-management state where the screen’s functionality may be limited to save power on the device. “Unknown” refers to instances in which information is not available.

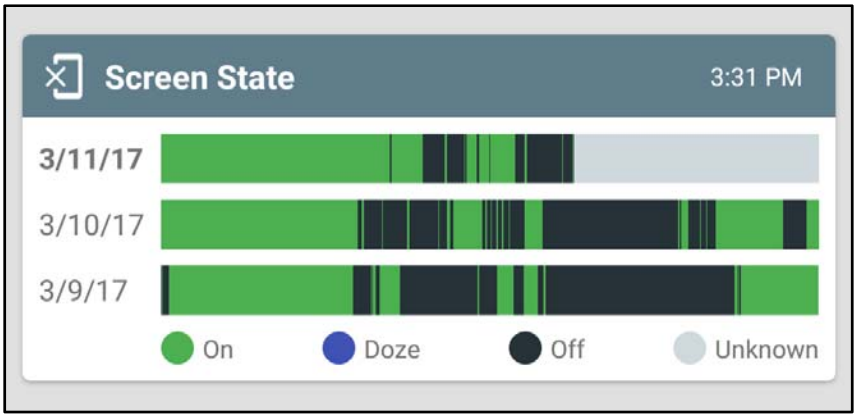


Figure 8. Screen State Display

Model Settings

The BlūBot application settings allow the user to customize the features, the model type, and the training and evaluation duration to be applied to the collected data. The user can choose which type of predictive model to apply to the data: an average of the PHQ-9 scores or the "Saeb Sensor Model" (referred to as the "BlūBot Model" in this study). Also, the user can change the predictive model settings, so that the data training and evaluation window can be set to one day, one week, two weeks, or one month. Within the predictive model settings, model features that were found to be significantly correlated with PHQ-9 scores can be activated or inactivated (Figure 9). These include location variance, location cluster count, location entropy, normalized location entropy, home stay ratio, in-transit ratio, total distance covered, screen activation count, and screen active duration.

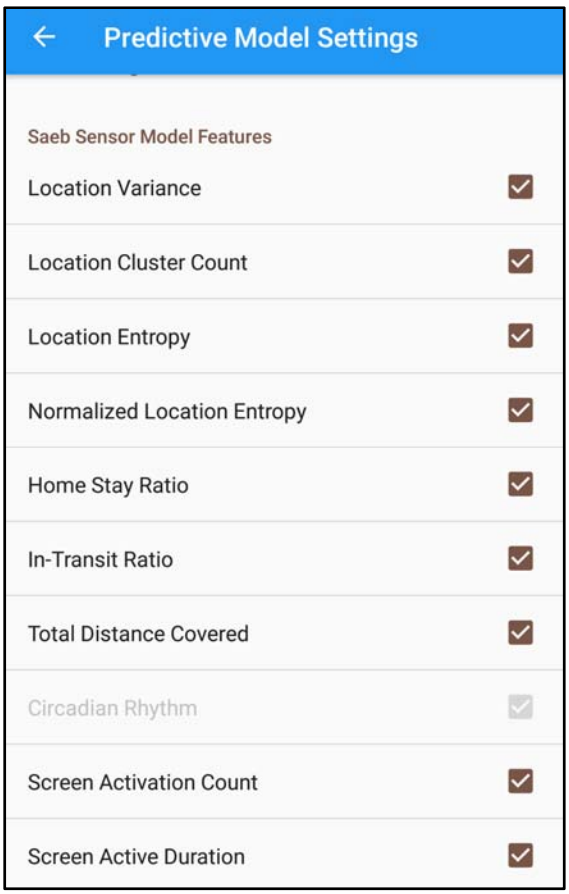
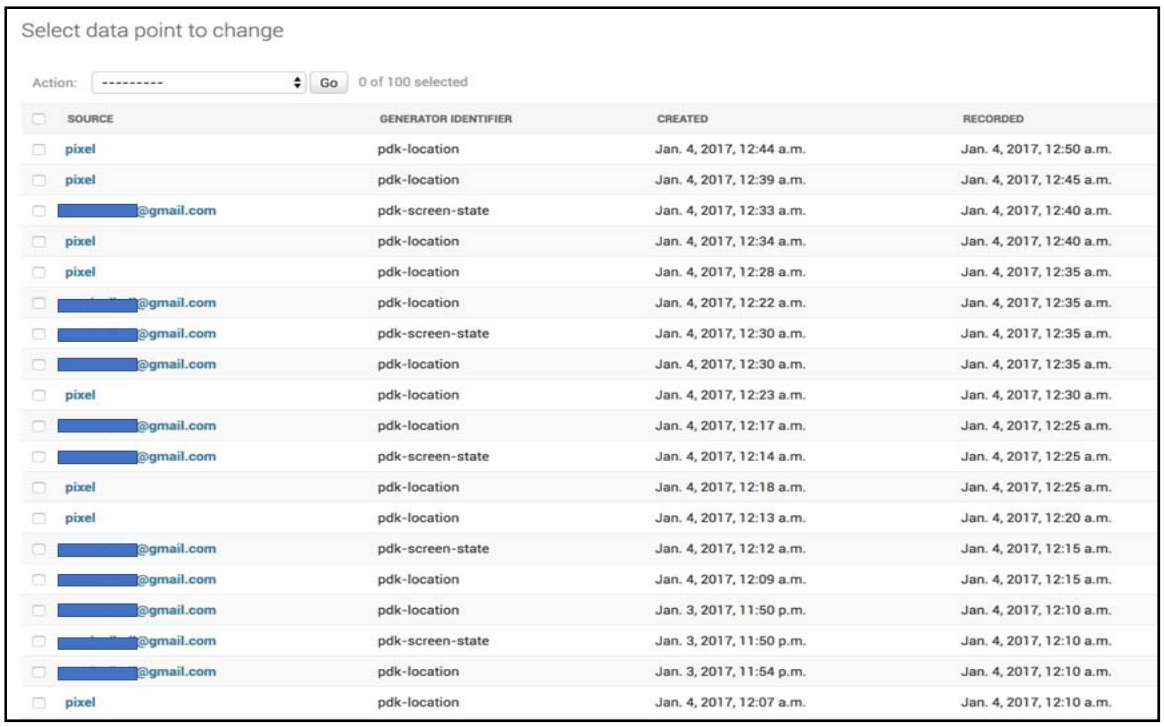


Figure 9. BlūBot predictive model settings

3.2.2 Application Server

Data collected from the mobile device is sent to the BlūBot (PDK) repository. Each data point contains information about the source (ID associated with user), the generator identifier (GPS or screen state), the timestamp in which the data point was created by the user and the timestamp in which the data point was recorded in the database (Figure 10).



Select data point to change

Action: --------- Go 0 of 100 selected

SOURCE	GENERATOR IDENTIFIER	CREATED	RECORDED
pixel	pdk-location	Jan. 4, 2017, 12:44 a.m.	Jan. 4, 2017, 12:50 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:39 a.m.	Jan. 4, 2017, 12:45 a.m.
[redacted]@gmail.com	pdk-screen-state	Jan. 4, 2017, 12:33 a.m.	Jan. 4, 2017, 12:40 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:34 a.m.	Jan. 4, 2017, 12:40 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:28 a.m.	Jan. 4, 2017, 12:35 a.m.
[redacted]@gmail.com	pdk-location	Jan. 4, 2017, 12:22 a.m.	Jan. 4, 2017, 12:35 a.m.
[redacted]@gmail.com	pdk-screen-state	Jan. 4, 2017, 12:30 a.m.	Jan. 4, 2017, 12:35 a.m.
[redacted]@gmail.com	pdk-location	Jan. 4, 2017, 12:30 a.m.	Jan. 4, 2017, 12:35 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:23 a.m.	Jan. 4, 2017, 12:30 a.m.
[redacted]@gmail.com	pdk-location	Jan. 4, 2017, 12:17 a.m.	Jan. 4, 2017, 12:25 a.m.
[redacted]@gmail.com	pdk-screen-state	Jan. 4, 2017, 12:14 a.m.	Jan. 4, 2017, 12:25 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:18 a.m.	Jan. 4, 2017, 12:25 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:13 a.m.	Jan. 4, 2017, 12:20 a.m.
[redacted]@gmail.com	pdk-screen-state	Jan. 4, 2017, 12:12 a.m.	Jan. 4, 2017, 12:15 a.m.
[redacted]@gmail.com	pdk-location	Jan. 4, 2017, 12:09 a.m.	Jan. 4, 2017, 12:15 a.m.
[redacted]@gmail.com	pdk-location	Jan. 3, 2017, 11:50 p.m.	Jan. 4, 2017, 12:10 a.m.
[redacted]@gmail.com	pdk-screen-state	Jan. 3, 2017, 11:50 p.m.	Jan. 4, 2017, 12:10 a.m.
[redacted]@gmail.com	pdk-location	Jan. 3, 2017, 11:54 p.m.	Jan. 4, 2017, 12:10 a.m.
pixel	pdk-location	Jan. 4, 2017, 12:07 a.m.	Jan. 4, 2017, 12:10 a.m.

Figure 10. Data points within the BlūBot repository

These data points were used to inform the model predictions. The architecture of the underlying components and the journey the data takes through this system can be seen in Figure 11.

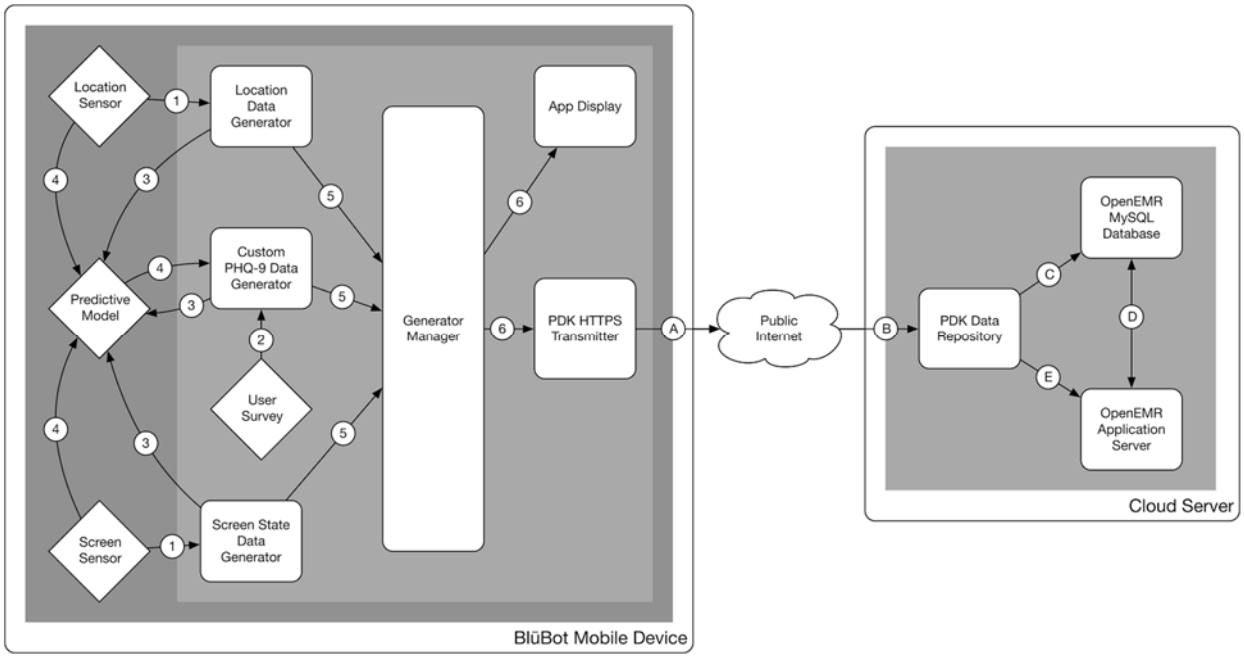


Figure 11. BlūBot predictive model data flow
(Chris Karr, personal communication, March, 2017). See Appendix C for Legend.

3.2.3 Injecting Data into EMR

Because there is no suitable API currently available for use with OpenEMR, an additional Django component, named emr_integration, was created to interface the BlūBot data repository and the OpenEMR server. This integration layer is used to accomplish four objectives:

1. Map BlūBot user identifier to the corresponding patient within OpenEMR
2. Insert mental health alert messages into OpenEMR system
3. Create PHQ-9 record in OpenEMR database
4. Visualize patient PHQ-9 data in OpenEMR

3.2.3.1 Mapping User Identifier to Patient Identifier

In order to pair BlūBot user data with a patient identifier, a Registration object type was added to the emr_integration app that mapped the BlūBot app user identifier to the identifiers

used in OpenEMR. This Registration model also includes fields to establish a BlūBot user's OpenEMR doctor and health group so that Mental Health Alert messages specified both the patient who generated the PHQ-9 predictions and the provider assigned to receive the warning messages.

3.2.3.2 Creating Alert Messages

To generate alert messages, a scheduled background task was created. This background task inspected new PHQ-9 data points within the repository every minute to identify scores exceeding a predefined threshold (composite score 10 or greater). When this threshold was exceeded, or self-harm was indicated in response to question #9, a message labeled as a "Mental Health Alert" was pushed to OpenEMR (Figure 12). The Django application server directly injects message content into the OpenEMR database, circumventing the OpenEMR application layer in the database layer.

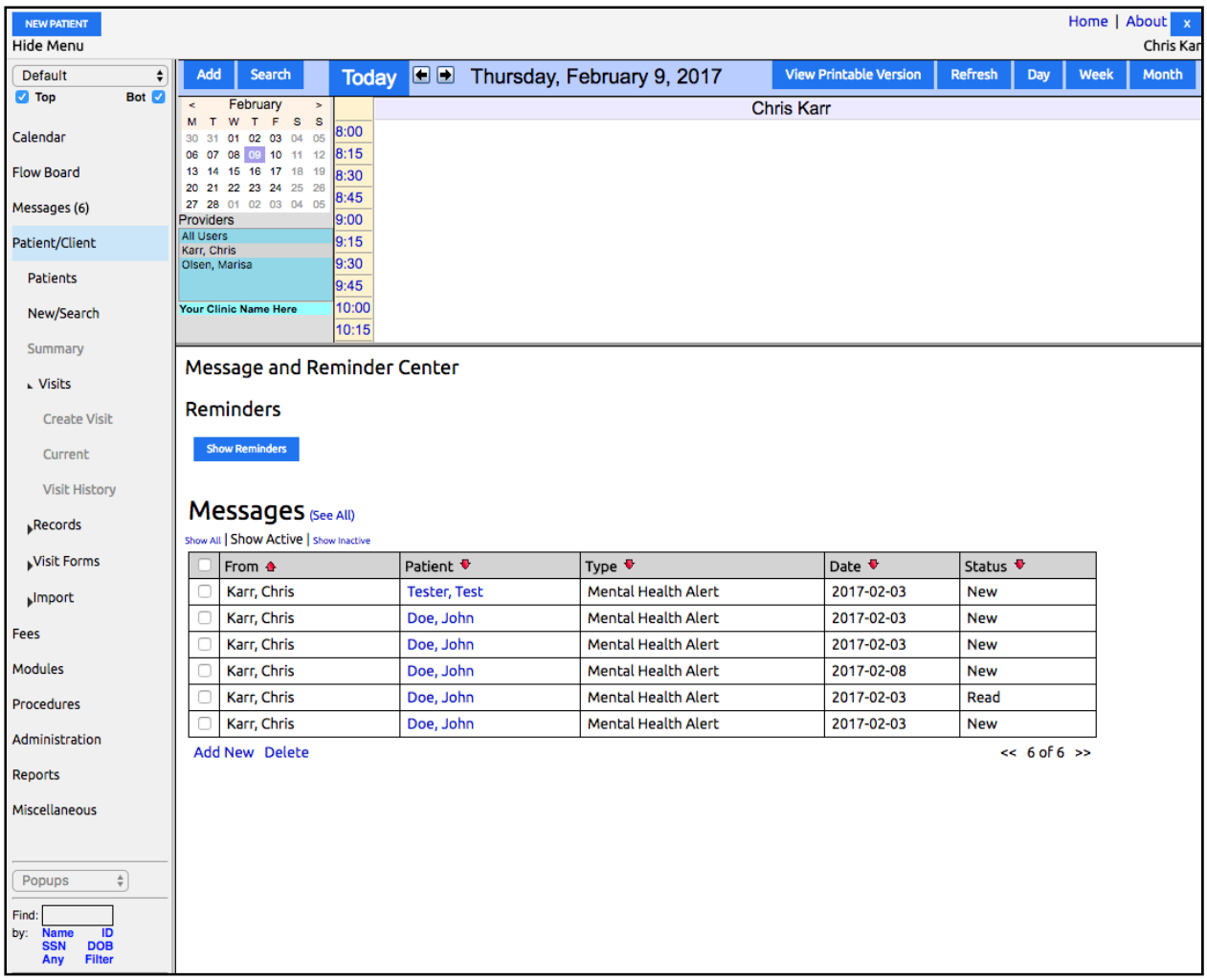


Figure 12. Health care provider view of Message and Reminder Center Dashboard

Information about messages is displayed in the Message and Reminder Center. The summary includes the provider's name in the "From" field, the patient's name, the type of message, the date received, and the status of the message, indicating whether or not it has been read. "Mental Health Alert" is the type of message created for this study and its presence alerts the provider that a patient has a PHQ-9 composite score of 10 or higher.

Upon opening the message, the provider will see the patient's PHQ-9 composite score (Figure 13).

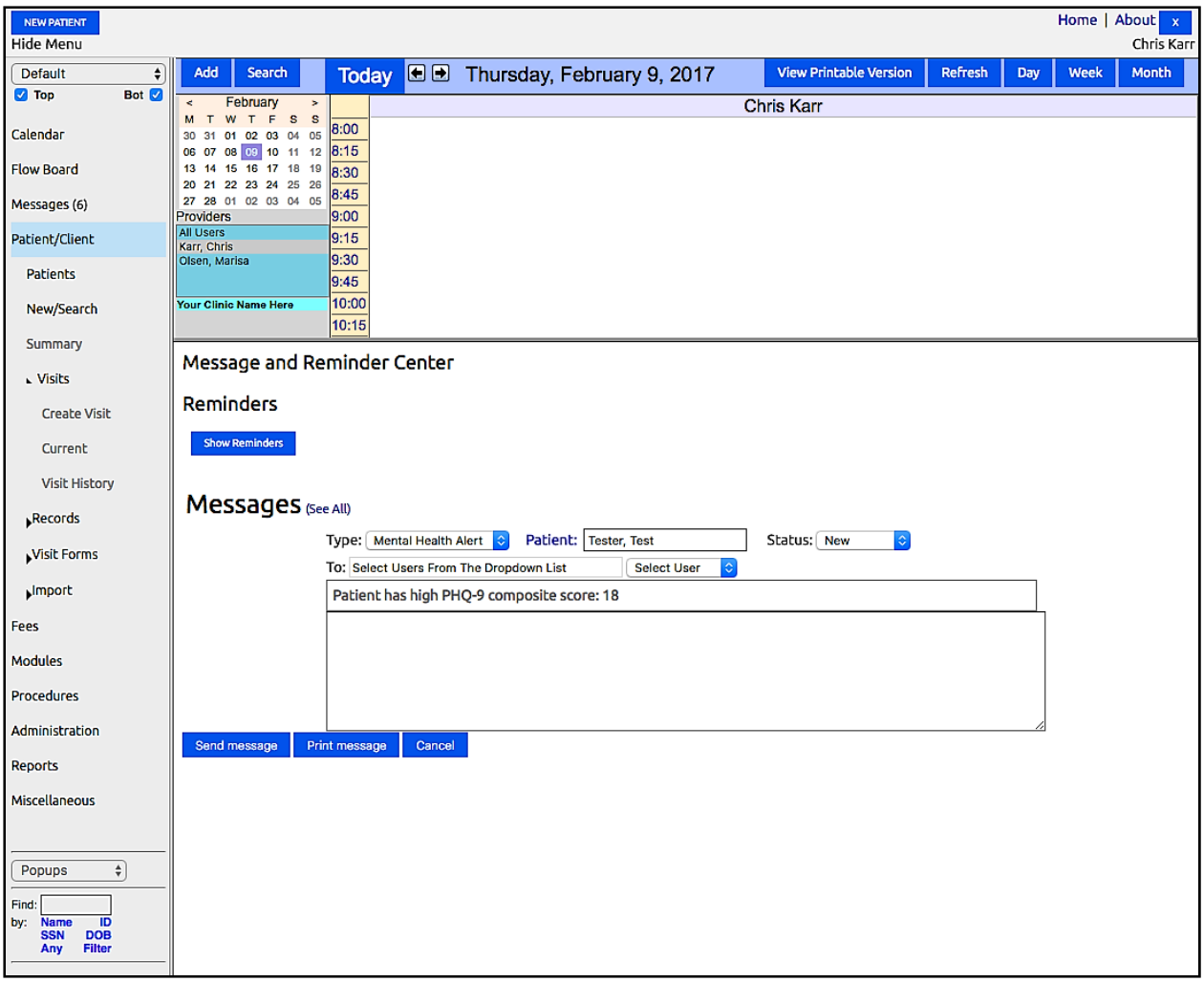


Figure 13. Mental Health Alert message contents

PHQ-9 composite scores of 5-9, 10-14, 15-19, and 20+ represent mild, moderate, moderately severe, and severe depression, respectively.[21] Providers are expected to follow up with patients who have a Mental Health Alert message.

3.2.3.3. Uploading PHQ-9 Report into OpenEMR

Although C-CDA uploads are supported within OpenEMR, FHIR specifications were used because of the advantages FHIR documents have, in terms of interoperability. Under the document link on the patient dashboard, a new file folder was created entitled “Mental Health

Screeners." The PHQ-9 document was uploaded into this folder, which did not require the creation of a new patient encounter for each upload. A new FHIR-enabled document was uploaded daily. This was accomplished by leveraging OpenEMR's file management option to include a link instead of a file. Links were uploaded to the OpenEMR database that referred back to the BlūBot (PDK) repository, which contains a transformer that converts PDK-encoded PHQ-9 data into a clinical document. This real-time data transformer was implemented as a view within the emr_integration app.

3.2.3.*4. Creating Provider View in OpenEMR*

In order to give the health care provider a live view of a patient's PHQ-9 score history within OpenEMR, a different method was needed because scheduling a background task would not work in this instance. Instead, space was made within the OpenEMR provider view to encode an IFRAME that would display information from BlūBot repository. This Mental Health History graph (Figure 14) was incorporated into OpenEMR by modifying the OpenEMR source code to implement a new widget containing the visualization. This widget embedded an HTML IFRAME in its body that referred to the graph view in the PDK repository for its content. This custom widget was the only direct modification made to the OpenEMR system in order to accomplish the project's integration goals.

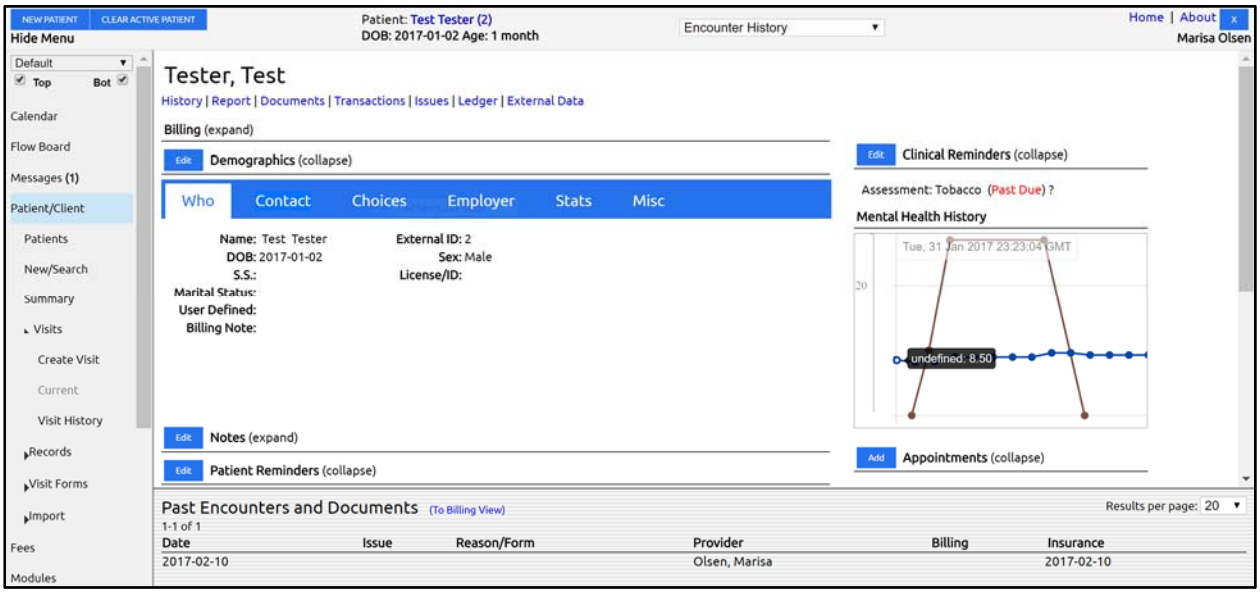


Figure 14. View of Patient Dashboard with Mental Health History embedded

Based on clinical workflow analysis findings from an informatician, the OpenEMR dashboard was modified to show a patient's PHQ-9 history over time as a line graph to illustrate any serious depressive episodes that were experienced over time. The patient ID is linked to the BlūBot identifier. In order to accomplish this, a visualization was created as a view within the emr_integration app; it used the OpenEMR patient identifier as an argument and queried the PDK repository for the PHQ-9 data points for the past two weeks. This information was then graphed on a standard line chart using the Rickshaw JavaScript library.

3.3 Model Fit

There were only enough observations collected in the study to run analyses for the one-day training and evaluation model and the one-week training and evaluation model, incorporating all of the significant Saeb sensor model features (omitting circadian movement).

Upon first glance, the one-day training and evaluation window (Figure 15) seemed to be a better fit than the one-week option (Figure 16), so data was initially analyzed using the one-day model.

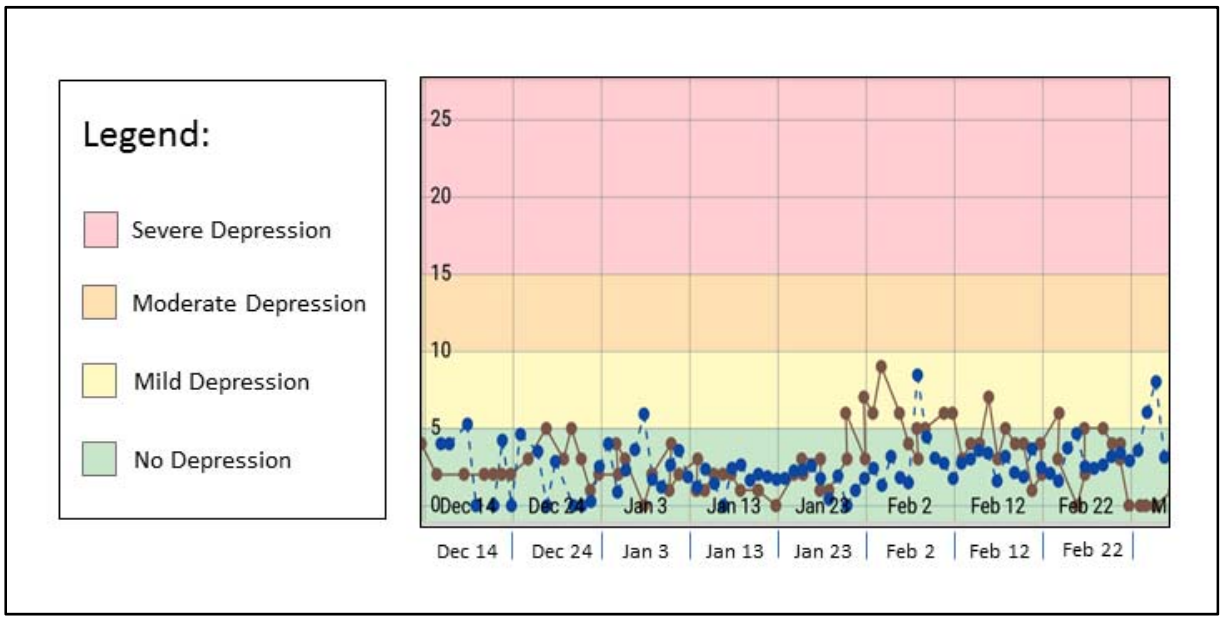


Figure 15. BlūBot graph showing one day training model performance
The brown line represents the user PHQ-9 composite scores and the blue dashed line represents the model predicted responses by date.

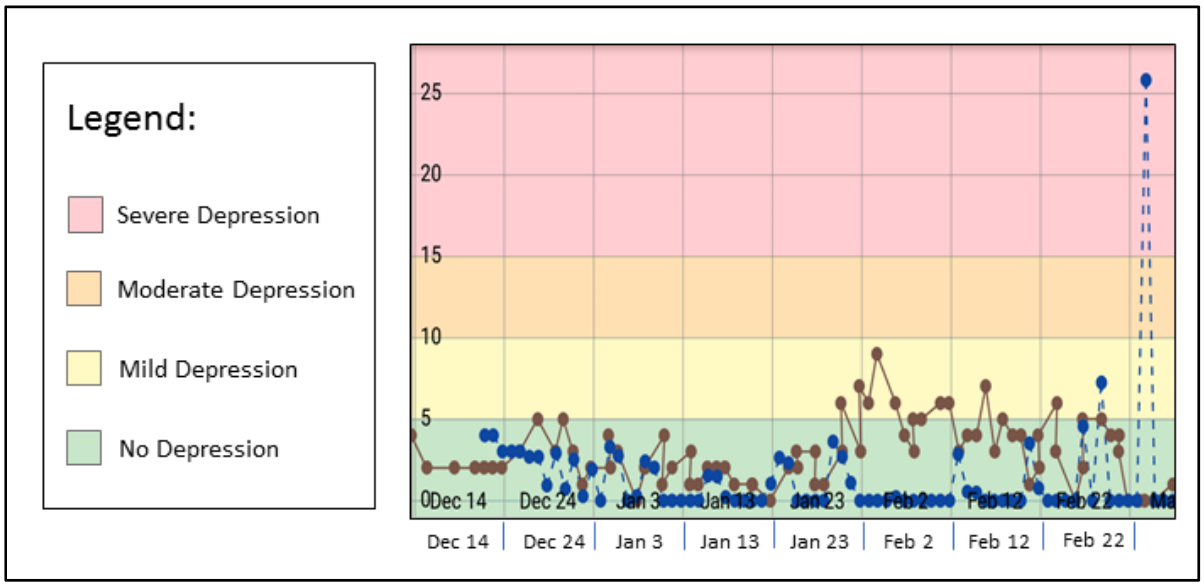


Figure 16. BlūBot: one-week training and evaluation window model performance

The brown line represents the user PHQ-9 composite scores and the blue dashed line represents the model predicted responses by date.

As mentioned previously in the methodology, only data points for consecutive days that contained both a user and model-predicted score were used for the initial analysis, in addition to removing the first prediction in the series because it had no user PHQ-9 response for the previous day. This omission gave the model the appearance of better fit (Figure 17).

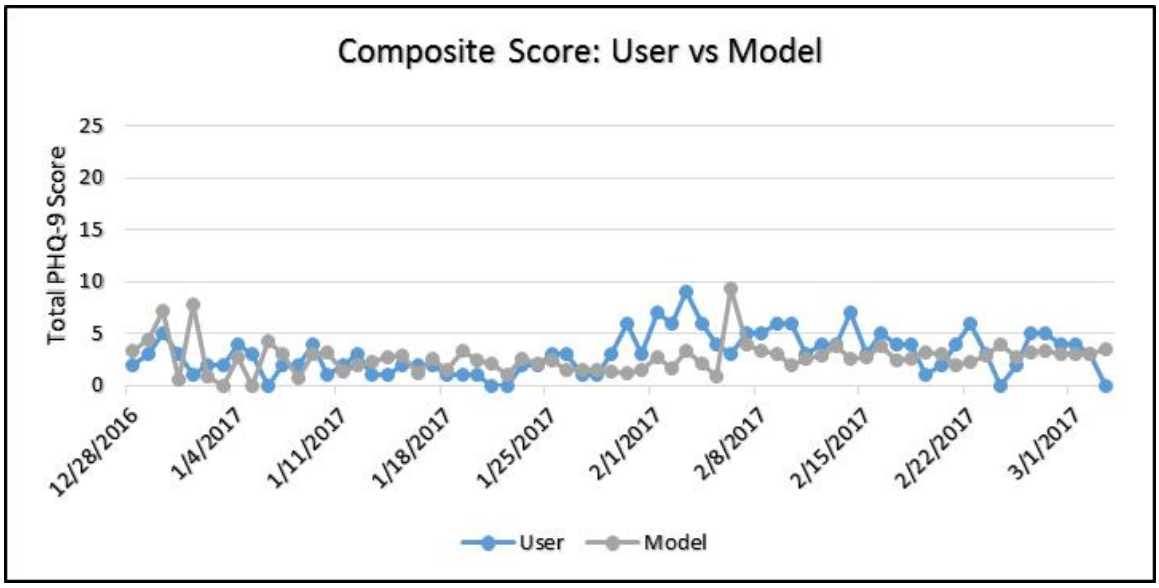


Figure 17. Comparison of the user PHQ-9 scores to model predictions
(Based on consecutive days, omitting initial day that did not have a user response for the previous day, Variance 3.84).

This modified set of data was analyzed for RMSE, variance, standard deviation, R^2 and mean absolute error (MAE). To compare the BlūBot model performance against the average running total of PHQ-9 scores, these same measures were repeated for the additional data set. The results of this analysis can be seen in Table 1.

Model Type	RMSE	Variance	Standard Deviation	R^2	MAE
BlūBot Model	2.48	3.84	1.96	0.0026	1.97
Running Average	1.99	0.25	0.50	0.0178	1.52

Table 1. Statistical measures: BlūBot Model vs running average of user PHQ-9 scores

A graph representing the absolute variance between the BlūBot model and the running average is depicted in Figure 18.

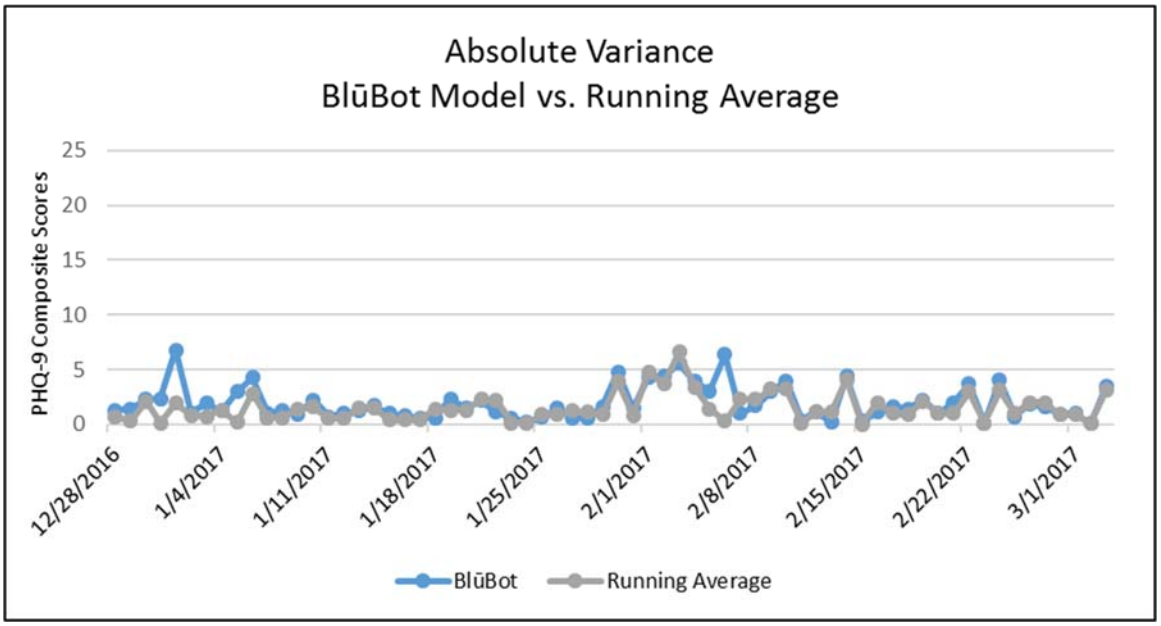


Figure 18. Absolute errors: BlūBot one-day model vs running average of PHQ-9 scores (BlūBot Model Variance = 2.534569, MAE = 1.93; Running Average: Variance =2.29, MAE = 0.749)

The same statistical analysis was also applied to the BlūBot one-week training and evaluation window model (604800000 ms) (RMSE = 6.16). There were less observations in this set (OBS= 37) for the same time period of December 28, 2016 through March 3, 2017.

3.4 Summary

Data was passively collected via sensors on a mobile device, using the BlūBot app, as a user carries the device throughout day-to-day activities. Each day, the user is reminded to complete the PHQ-9 on the app. The application stores data related to GPS location, screen on/off status, and PHQ-9 scores on the mobile device. The data is sent from the mobile device to a data repository for processing. PHQ-9 responses and features indicating the presence of depression are used to create both an actual and model-predicted PHQ-9 composite score for the user. PHQ-9 guidelines recommend that health care providers create a treatment plan for patients with a composite score of 10 or greater, so the system is set to trigger a mental health alert message to the provider's inbox within OpenEMR when a composite score reaches this threshold.[63] The data collected in the BlūBot data repository is also used to generate a patient's PHQ-9 history, which is embedded in the OpenEMR dashboard. This prototype has demonstrated how passive data collected and processed from sensors within a mobile device can be connected to an EMR to alert providers when a user exhibits indicators of depression and/or suicide.

4 Chapter 4: Discussion

4.1 Introduction

This project created a prototype to interface an mHealth application used to screen for depression with an EMR. This innovative way to screen for depression can be used to comply

with recommendations issued by the USPSTF and to meet Healthy People 2020 objectives pertaining to mental health.[43, 64]

4.2 Summary of Project

Depression is a serious public health concern that places a high burden on society. More frequent screening is recommended to discover undiagnosed cases of depression and to refer individuals for treatment. This project introduces a novel approach to help providers screen for depression by expanding upon existing mobile device and EMR technology.

An existing mHealth approach was reconfigured so that a model could be trained on an individual basis to generate predictions of depression on a mobile device. The model based on the one-day training and evaluation window produced the best fit for the user PHQ-9 data, based on RMSE. The study consisted of one test user whose PHQ-9 responses were used to train the model and generate predictions. The PHQ-9 results for this individual were fairly consistent, with only one spike of mild depression (composite score of 9) throughout the study. The model did not anticipate that spike beforehand because it had based its new predictions on the user scores collected up to that point, which had been stable. Following the spike in the user score, it took the model three days to predict a score of equal severity (9.4). It is worth noting that the user reported that the fifth week of the study, when the PHQ-9 score peaked at 9, was an extraordinarily stressful week. The anomaly in the scores for this week may have affected the performance of the model over time.

When the dates with gaps in user responses are included in analysis, the RMSE increases to 3.45664. A model-predicted score of 27 was observed on December 27, 2016; there was no user PHQ-9 response recorded on the previous day (Figure 19).

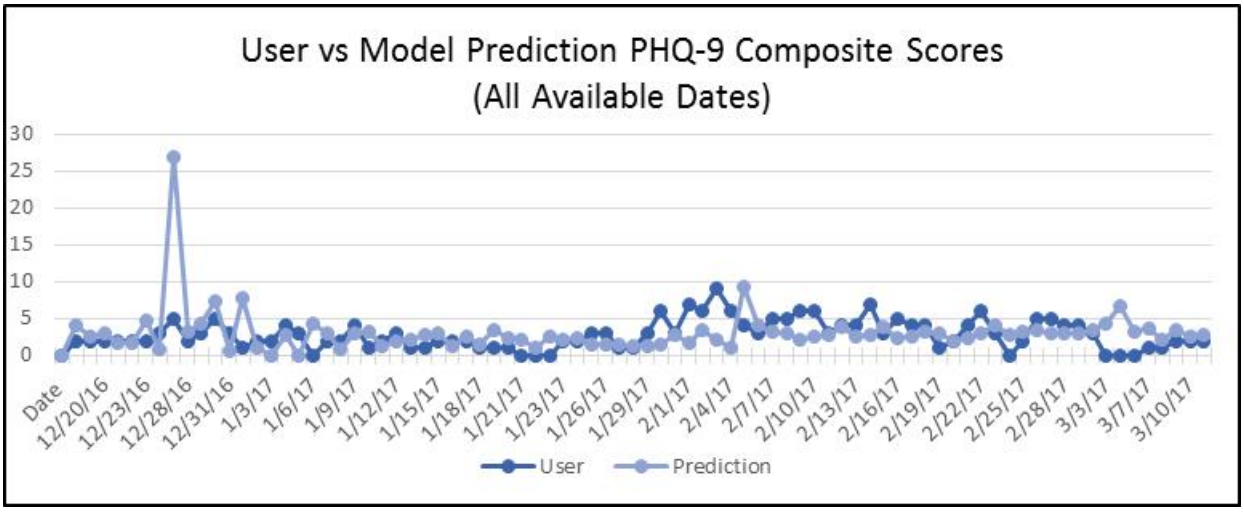


Figure 19. User PHQ-9 scores: all dates with corresponding model predicted scores

To determine if the model improved the accuracy of its predictions over time, the days in which data was collected was broken into one-week segments. The first week of data collection was excluded as there were no model predictions for that time period. RMSE was calculated for each week. It appears that the model's accuracy fluctuated over the course of the ten weeks, but ultimately resulted in a model that performed better in the tenth week than the first (Figure 20). This fluctuation may be partially attributed to a spike in the user scores during the fifth week of data collection, which will be discussed later.

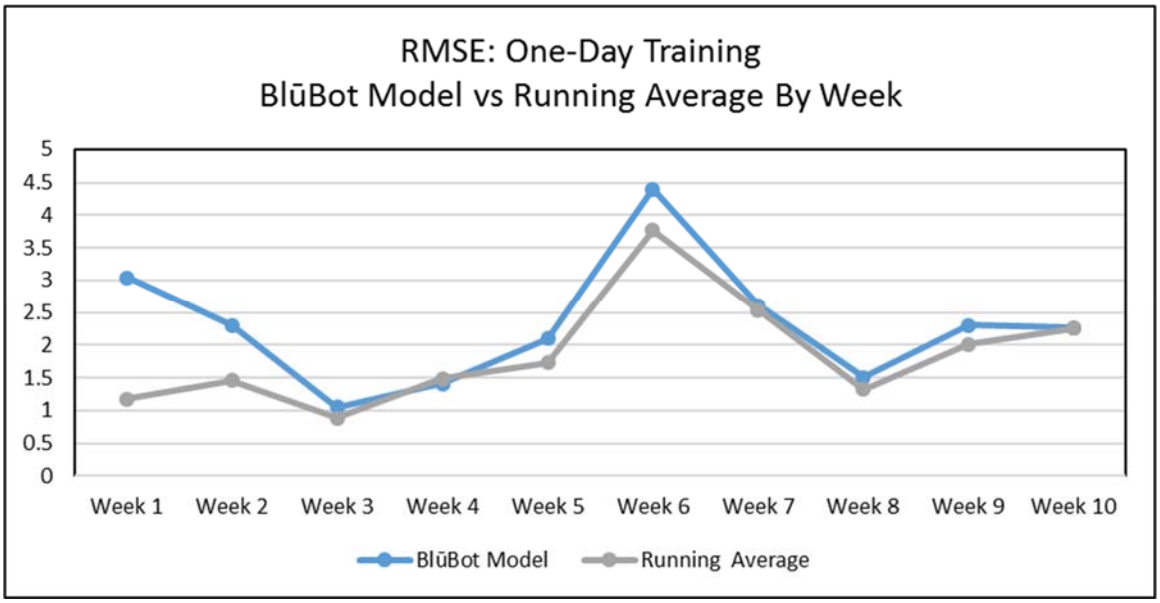


Figure 20. RMSE One-day and one-week training models, evaluated over consecutive weeks

An open-source EMR was modified to allow data from the BlūBot application to be injected into its system. A Django application server processed and analyzed data from the app to be sent to OpenEMR. LOINC® and FHIR standards were used for PHQ-9 score documentation in an effort to ensure interoperability with other systems. The BlūBot endeavor resulted in a process to collect, analyze, and communicate data related to depressive indicators, captured via mobile device, to an EMR/provider. With further study and modification, as described in the implications sections below, this approach could be used in a clinical or research setting.

4.3 Implications

4.3.1 Clinical Implications

The BlūBot prototype introduces a new depression screening tool based on features that have been shown to be significantly correlated with depression. While the depression screening model employed by BlūBot has not been validated in an adequate number of

subjects, it has the potential to be used as a clinical support tool, once validated. Even without validation, the tool may be used as a supplement for the paper-based PHQ-9 given in a clinical setting. This will give the provider a window into their patients' mental well-being during the gaps between face-to-face encounters, detecting missed cases of depression, or monitoring the progress of those previously diagnosed with depression. Health care providers or support staff can then connect with patients for follow-up screenings and/or provide support and guidance for the treatment of depression.

Additionally, the data collected from the BlūBot app can be used to inform healthcare quality and performance measures. Agencies such as the Agency for Healthcare Research and Quality (AHRQ), and the National Committee for Quality Assurance (NCQA), have a vested interest in creating, measuring, and improving standards of care.[65, 66] Statistics about diagnoses and course of treatment provide insight to treatment outcomes and inform evidence-based care.[66] The NCQA currently includes an initiative to measure depression care and treatment outcomes through the utilization of the PHQ-9 as a depression screener.[67]

Use of mobile devices to collect health information between doctor visits can help clinicians and health care staff to make better informed decisions, improve collaboration with patients for treatment plans, increase workflow efficiency, and reduce health costs.[68] This concept can be applied to monitor information related to other chronic health conditions, like blood glucose level for diabetics, heart rate and blood pressure for patients with a history of heart disease, and seizure activity in epileptics. Patients can also improve mental health outcomes by completing surveys related to other mental health disorders, such as bipolar disorder and generalized anxiety disorder, which could alert clinicians to potential problems when connected to an EMR.

4.3.2 Implications for Public Health

The innovation proposed in this project may be used, both directly and indirectly, to achieve Healthy People 2020 objectives for Mental Health and Mental Disorders: Mental Health Status Improvement and Treatment Expansion.[64] Depending on the targeted demographic, the following objectives listed in Table 2 could potentially be improved:

Category	Objective	Title
Mental Health Status Improvement	MHMD-1	Reduce the suicide rate
Mental Health Status Improvement	MHMD-2	Reduce suicide attempts by adolescents
Mental Health Status Improvement	MHMD-4	Reduce the proportion of persons who experience major depressive episodes (MDEs)
Treatment Expansion	MHMD-6	Increase the proportion of children with mental health problems who receive treatment
Treatment Expansion	MHMD-9.2	Increase the proportion of adults aged 18 years and older with major depressive episodes (MDEs) who receive treatment
Treatment Expansion	MHMD-11	Increase depression screening by primary care providers

Table 2. Healthy People 2020 Objectives potentially impacted

4.3.3 Other Implications

Outside of the clinical and public health setting, this design has practical uses for research. Research subjects could be recruited through their mobile device. Use of mHealth apps could improve efficiency and reduce the need for research site visits.[68] Because much of the

components in this project are open source, it would be relatively simple for a researcher with a limited budget to set up the BlūBot application for study participants and then link that data to a research server. Analysis of this data could be used to determine outcomes of treatment interventions for depression, for example.

The BlūBot screening tool presented in this project is not a substitute for a professional diagnosis of clinical depression and cannot be coded as such within an EMR. Use of the application should include a disclaimer stating as such. Even with a disclaimer, the ethical question of a provider's liability to a patient must be addressed. Further validation of the model could offer a supplement for patient screening outside of the provider's office. This can be used to screen for cases of depression that are missed in underserved populations. The PHQ-9 results, as reported by the BlūBot application, offer valuable information to researchers and clinicians about mood and attitudes toward self-harm. Through analysis of this information, trends can be discovered that may provide insight to a patient's mental state and patterns of behavior over time.

4.4 Limitations

There are several limitations for this project. Financial constraints necessitated the use of open-source products, which all have limitations. OpenEMR has no built-in fields to accept patient generated data, so significant configurations had to be made in order to work this information into the system. OpenEMR also lacks a suitable API, so the original idea of requesting and receiving information via a FHIR API had to be modified. Scripts had to be written and data had to be processed within the Django application server and injected into the OpenEMR system. Similar workarounds could be created for other non-proprietary EMR systems. It is customary for clinical practices to use proprietary EMR systems, so modifying

them to accept patient generated data can be time consuming and costly. However, because several proprietary EMR systems already distribute their own APIs, a BlūBot integration may actually be easier to accomplish than the solution implemented here for OpenEMR.

4.4.1 Incorporating PHQ-9 into OpenEMR

There have been other tools in the Open EMR community used to integrate PHQ-9 data into the OpenEMR system. Some users have incorporated the PHQ-9 into Open EMR under the "Visit Forms" section of the patient dashboard. The questionnaire can be completed in this way, with two caveats: One, a new patient encounter must be created in order to access the questionnaire. Two, the composite score is not automatically calculated within the EMR, so it must be hand entered. However, recently, an open-source plug-in became available for OpenEMR that will automatically calculate the PHQ-9 composite scores.[69] Like many EMRs, OpenEMR is encounter-centric and data entry requires the creation of a new encounter that is attached to a clinician for billing and accountability purposes. To avoid creating an encounter for each daily PHQ-9 response and prediction sent from the BlūBot repository to the OpenEMR system, an alternative approach was necessary. In the previously mentioned Brain Center pilot study, the PHQ-9 scores were uploaded into the clinical notes for a GE Centricity EMR system. Given the challenges of duplicating this within OpenEMR, the PHQ-9 scores have been incorporated into the system by embedding a graph on the patient-view dashboard and uploading a report into documents.

4.4.2 Model Performance and Validity

The training of the model to predict depression was limited to one individual for this study. That said, the data in this study was collected over a longer period of time than in previous, related studies. This study also had the added benefit of using continuously collected

data in real-time to make predictions and incorporating PHQ-9 feedback from the user to inform the model. Use of the BlūBot application as a screening tool assumes that the application can detect patterns of depression which mimic the results of the PHQ-9. Verifying the validity of this application for this purpose is beyond the scope of this study. While electronic versions of the PHQ-9 have been validated as a comparable test to the paper version, modifications from the original PHQ-9 were made to accommodate daily responses within the BlūBot app for this study.[9] The original PHQ-9 asks the responder to consider his or her mood over the course of the past two weeks. Due to time constraints, this project was not able to take measurements at two-week intervals. Responses were needed more frequently in order to hone the model for the application.

Additionally, a follow-up study has been published by Saeb et al., since the inception of the BlūBot project, which indicates the general, group-based model utilized in the Purple Robot study is not generalizable to all people.[70] Because the BlūBot system is designed to train models on an individual, rather than a group, the generalizability issues from Saeb's study have been mitigated. Even if the individual model cannot be validated, the BlūBot application could still be useful for improving public health. One option is to consider omitting the model predictions and just send the user's PHQ-9 responses to OpenEMR. An electronic questionnaire based on the last two weeks is a validated model for depression screening.[9] This may also be useful if compliance is a concern. The user could be expected to answer the PHQ-9 every two weeks, instead of daily, as the questionnaire's design originally intended. The BlūBot app reminder could be set to prompt the user to a bi-weekly submission, rather than daily. The goal of obtaining more frequent depression screenings would still be met in this way; cases missed between provider screenings could be detected.

Initially, the circadian movement feature used in the Saeb et al. model was not included in the BlūBot implementation because the code available to perform this measure was written in Python and not for use, as-is, in Android applications. However, after the initial data analysis was completed, a solution to incorporate the circadian movement feature into the BlūBot model was discovered. The app was updated and the model predictions were regenerated with the circadian movement feature enabled. Although additional data had been generated by this point in time, the statistics were run again using the same time frame as the previous analyses (December 28, 2016 through March 3, 2017). Incorporating the circadian movement feature did not improve the model. Whereas the original model had an RMSE of 2.48, the model including the circadian movement feature had an RMSE of 3.82. A comparison of the two model versions can be seen in Table 3.

Model Type	RMSE	Variance	Standard Deviation	R^2	MAE
BlūBot Model	2.48	3.84	1.96	0.0026	1.97
BlūBot Model with Circadian Movement	3.81	3.84	1.96	0.0008	2.26

Table 3. BlūBot model performance with and without the circadian movement feature enabled

It is also interesting to note that while the original model predicted a PHQ-9 composite score of 27 on December 27, 2016 (following a non-response from the user on December 26), the model including circadian movement predicted a score of 27, two days later, on December 29, 2016 (Figure 20).

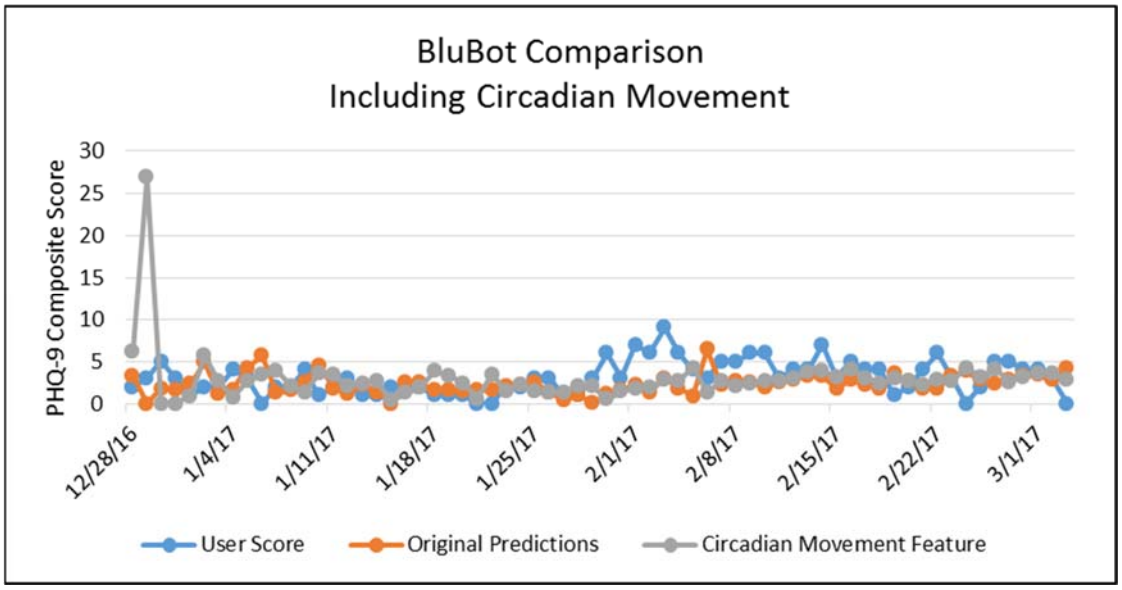


Figure 21. BlūBot performance with and without circadian movement feature enabled

When the 27th and 28th of December are removed from the one-day training and evaluation model analysis that includes the circadian movement feature, it results in a similar RMSE (2.40) to the model without. It is hard to say what effect adding the circadian movement feature to the model had. In machine learning, adding additional features does not ensure that the model will improve.

4.4.3 Interoperability

The integration layer between the BlūBot repository and OpenEMR is not generalizable to other EMR systems because each operates under its own programming language and data standards. That said, the infrastructure created for the BlūBot app and its repository can be used with other proprietary EMR systems by modifying the EMR itself. Furthermore, generalizability can be supported by EMRs with APIs that the BlūBot repository can be programmed to communicate with. Some of these interoperability limitations may be overcome with the use of FHIR implementation. Because documentation of PHQ-9 scores are written in

specifications used for FHIR, it is plausible that any EMR/EHR system with a FHIR-compatible API can share this information with other FHIR enabled systems.

4.4.4 HIPAA Compliance

The mobile device, server, and database were not configured to be compliant with the HIPAA security rule. If this methodology were to be implemented and contain protected health information (PHI) being sent or received by a covered entity, additional configuration and security settings would be required. Although data transmission encryption was supplied through Passive Data Kit's SSL-secured HTTPS API, encrypting data at rest was not addressed for the data stored on the device awaiting transmission or for the data on the server awaiting analysis or visualization. In order to configure EC2 for PHI, there are several issues that must be addressed.

Encrypting Data at Rest

On-Server Security

Generally speaking, this can be addressed in the following ways:

Perform application or field level encryption at the point it is processed within application or in the database hosted in EC2, with the following approaches:

- Utilize standard libraries in application framework (i.e. Java, .NET)
- Take advantage of Transparent Data Encryption features in Microsoft SQL or Oracle
- Integrate other third party and software as a service (SaaS)-based solutions into applications
- Third party software from AWS Marketplace Partners, or native file system encryption tools can be used to encrypt data at rest using file-level or full disk encryption
- Applications running in Amazon EC2 can be integrated with AWS KMS SDKs[53]

On-Device Security

The mobile device itself should be either password protected or protected using biometric identification methods (or a combination of multiple authentication factors) to prevent unauthorized persons from obtaining access to PHI data stored and displayed in the BlūBot application.

Encrypting Data in Transit:

- For sending data between EC2 and external sources, like internet environments: customers should use industry-standard transport encryption mechanisms such as TLS or IPsec virtual private networks (VPNs)
- Network traffic containing PHI within the Amazon EC2 internal network must also be encrypted
- Applications and protocols that do not support encryption, can send PHI through encrypted tunnels using IPsec or similar implementations between instances
- Amazon EC2 instances that customers use to process, store, or transmit PHI are run on Dedicated Instances, which uses Amazon VPC on hardware that is dedicated to an individual customer and are physically isolated from other AWS accounts[53]

It is also worth mentioning that OpenEMR has weak security practices applied to its system. In fact, this weakened level of security is what allowed the manipulations to the system that were used to achieve study objectives.

4.5 Recommendations and Next Steps

4.5.1 Recommendations

Future studies should evaluate how changing the model features affects the model's performance. Another feature which might provide more insight to training the model is text messaging data. The log of text messages sent and received by the user, captured by the BlūBot app, can be monitored and the content of those messages could be analyzed using text mining. It may also be possible to include analyses of social media posts and speech to look for indicators of depression. Incorporating some type of therapy or coaching could also be included in future versions of the BlūBot application.

To clarify questions surrounding the validity of the BlūBot model, further studies are recommended to train and test the model in a larger number of users, over a longer period of time, in order to verify statistical significance. Future studies should also implement additional model training and evaluation configurations. Modifying BlūBot to gather PHQ-9 responses on a biweekly, rather than daily, basis will improve the strength of validity because the questionnaire will be answered at the interval established by the original PHQ-9 research. The model's predictions can then be compared to a validated measure of depression. However, collecting user PHQ-9 data for a two-week model training and evaluation timeframe will take a considerably longer amount of time. It takes six two-week time windows to generate a model with ten predictive features enabled (not including circadian movement); this would have to be repeated several times.

It would also be useful to train and test the model on users with more variability in their PHQ-9 scores, perhaps those suffering from more severe forms of depression. Based on the outcome of the analysis, the model is a good fit for data with little variance. It does

not predict spikes in advance and it takes a couple of days to catch up. It is hard to say how accurate the model would be in people with a wider range of PHQ-9 scores. Over time, it may be possible to account for some of the variance by adding or subtracting to the overall score, if the variance is consistently off in the same direction.

Because the Nexus 5x is a mobile device with considerable computational abilities, it would be prudent to conduct a wider study, employing more types of Android devices, in the real world. This would help determine if the approach to store data and train the model on the mobile device is feasible across the Android device spectrum, especially on lower-end devices.

4.5.2 Next Steps

The innovation demonstrated within this thesis is on the forefront of what is next in health informatics. Because of the potential information that is missed between doctor's appointments, practitioners and researchers are expressing an interest in patient generated data through wearable devices and mHealth applications. In 2017, The Office of the National Coordinator for Health Information Technology (ONC) presented a draft paper on the framework of using patient generated health data (PGHD) in clinical and research applications, including an overview of the potential opportunities and challenges of using this type of data. A majority of Americans own a smartphone and carry it everywhere with them, resulting in the ability to collect health-related data passively, without notifying the user, then display visual feedback to the user and health care provider that can used to modify behaviors.[68]

The use of PGHD is being incentivized by health care legislation.[68] There are several incentives being offered from the Centers of Medicare and Medicaid Services (CMS) that encourage providers to capture and use PGHD collected outside of the clinical setting (Table 3).

Program	Incentive
Medicare and Medicaid Electronic Health Record Incentive Programs - Stage 3 and Modifications to Meaningful Use in 2015-2017 (MU3)	Covered provider can receive credit in the program when PGHD are incorporated into the certified EHR technology for >5% patients seen eligible providers[71]
Community-Based Care Transitions Program; Chronic Care Management Services	Reimbursement for non-face-to-face care coordination for beneficiaries with multiple chronic conditions and for transitional care programs; can be supported through the use of PGHD[72, 73]
Merit-Based Incentive Payment System (MIPS): Medicare Access and Children's Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA)	Optional measure for using PGHD to support the goal of coordinating care through patient engagement[74]

Table 4. *CMS incentives for capturing and using PHGD*

There are many challenges to incorporating PGHD into clinical and research practice which can be mitigated through development of policies. As of now, there are not many mHealth apps that are approved for FDA use. Apps that are not regulated through the FDA are held to less rigorous standards and are typically somewhat exempt from HIPAA security and privacy rules.[68] The Department of Health and Human Services (HHS) has addressed the privacy and security issues surrounding the use of social media and mHealth technologies to collect health data in their report entitled "Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA" also referred to as the "NCE" report.[75] Even non-HIPAA covered entities may be subject to the Federal Trade Commission's FTC's Section 5 authority that prohibits the use of deceptive or unfair practices by both HIPAA covered and non-HIPAA covered entities.[68] The federal government can continue to encourage use of PGHD by providing health IT incentives and making changes to delivery systems. Current incentive models are moving away from payment for individual services to an overall management of a patient's health, which may include monitoring patients outside of clinical settings.[68]

Furthermore, funding needs to be put toward research investigating the accuracy, validity, and authenticity of data generated through mHealth applications in an effort to reduce providers' hesitancy to use PGHD.[68]

4.6 Conclusion

The study resulted in the creation of a mobile app that trained a statistical model using device sensor data and user self-reports to detect patterns of depression (RMSE of 2.48). The composite score generated by the BlūBot application model is about +/- 1.25 points off from the actual user PHQ-9 scores. Given that the maximum PHQ-9 composite score is 27, the predictions generated by the one-day training and evaluation model could serve as a valuable tool to detect episodes of depression between doctor's visits.

The study was successful in developing an infrastructure and methodology to send information related to a user's mental health status to an EMR. This, in itself, could prove useful to a single clinical practice, but could be further modified in order to be used in health information exchanges. Due to the lack of interoperability between EMR systems, the exact process developed for this project cannot be duplicated for all EMR/EHR systems and the methodology proposed does not create a solution of true interoperability between all types of systems. However, this prototype is valuable because the BlūBot app and its repository can be repurposed for integration with other EMR systems; provided modifications are made on the EMR side, as demonstrated here.

A solution to work around barriers preventing integration of PGHD into an EMR system has implications that expand beyond the scope of depression. The increasing popularity of the use of mHealth applications and wearable devices creates a wealth of data about the health and behaviors of a population of individuals. Yet until recently, this resource has largely remained

untapped due to technological limitations and lack of standardization. There are still uncertainties surrounding the implementation of this type of methodology and questions remain about the usefulness of this type of patient generated data, but this project opens the door to possibilities that were previously unavailable. The solution proposed could influence the future of research, precision medicine, public health, and clinical medicine.

If the study were to be repeated in the future, there are a few things that may be done differently. The concepts behind creating this prototype are on the forefront of health technology; if there have been people working on these concepts, they are not likely to have published their work, or may not want to share their findings for proprietary reasons. Given the limited amount of peer-reviewed literature related to the topics presented in this study, more time should be invested in reviewing and engaging in online communities of practice instead of searching PubMed. Future research should focus more on stakeholder engagement to obtain insight in terms of app development, workflow, and feasibility. Stakeholders can also validate the need and desire for a prototype like the one created for this study. Further market research is recommended to discover upcoming EMR developments; representatives from the health care software industry and government agencies should be consulted regarding the oversight of EMR developments. In future studies, given more time and resources, data should be collected from more subjects on more than one type of mobile device, with a large enough sample size to provide significance. Furthermore, obtaining access to an EMR vendor that offers an API would simplify the process described in this current study.

5 Chapter 5: Executive Summary

5.1 Overview

The increasing popularity of the use of mHealth applications and wearable devices creates a wealth of data about the health and behaviors of a population of individuals. Yet, until recently, this resource has largely remained untapped due to technological limitations and a lack of standardization. About one in ten patients seen by primary care physicians has a depressive disorder, yet these disorders largely remain underdiagnosed.[4] Post diagnosis, a study found that 67% of respondents with mental illness were interested in monitoring their symptoms through applications on their phones.[12] The innovation proposed within this thesis is line with developments being made on the forefront of health informatics.

5.2 Innovation

Because individuals do not see their health providers on a regular basis and the symptomology of depression is not static, there is a need for more continuous assessment outside of periodic doctor visits. Mobile devices have given patients the ability to conveniently collect and use their own generated data, known as Patient Generated Health Data (PGHD), but this information is not easily shared beyond the device that captured it. The development of mHealth applications enables the translation of PGHD into valid health screenings tools that can be beneficial to both patient and provider. The BlūBot prototype creates a solution for automating depression screenings outside of the clinical setting. In the future, score predictions from the BlūBot app may take the place of paper based screenings and may be reimbursable.

Whether it be this type of simulated mental health screening, fitness tracker data, or surveys completed by patients or caregivers, the ONC is investing in the development of a framework to incorporate PGHD into EHR/EMR systems.[68] With the wearable market

expected to increase from 275 million devices in 2016, to 477 million devices in 2020, providers and researchers will be missing an enormous opportunity to capitalize on PGHD if they do not have a system to capture and use data from these devices.[68]

5.3 Purpose

The purpose of this thesis is to develop a prototype for integrating data collected near continuously, via an mHealth application that screens for depression, into an open electronic medical record system (EMR) for clinical and research analysis. One goal of this prototype study is to develop an mHealth application that can predict episodes of depression by training a model based on user PHQ-9 responses and passive mobile device sensor data. Once the model is trained, the application will be able to predict depression using passive data alone, without user intervention. The other goal is to create an infrastructure and process to inject data from said mHealth app into an EMR to alert providers when a patient is experiencing depression.

5.4 Findings

An analysis was performed on the data to compare the user responses to each corresponding date's prediction. In order to analyze the data, a statistical procedure was chosen that would determine how well the model predictions fit the actual responses from the user, known as root mean square error (RMSE).[61] The RMSE gives an idea of the amount of variance between the actual and model-predicted values. With an RMSE of 2.48, the model typically will be able to accurately predict the composite scores within +/- 1.25 points. Even with a model-predicted PHQ-9 composite score off by 1-2 points, the model performs well enough that it can be used to alert providers about depressive episodes in their patients.

5.5 Incentives for Use

There are several incentives being offered by CMS programs that encourage providers to capture and use PGHD collected outside of the clinical setting. A summary of these can be seen in the table below.

Program	Incentive
Medicare and Medicaid Electronic Health Record Incentive Programs - Stage 3 and Modifications to Meaningful Use in 2015-2017 (MU3)	Covered provider can receive credit in the program when PGHD are incorporated into the certified EHR technology for >5% patients seen eligible providers[71]
Community-Based Care Transitions Program; Chronic Care Management Services	Reimbursement for non-face-to-face care coordination for beneficiaries with multiple chronic conditions and for transitional care programs; can be supported through the use of PGHD[72, 73]
Merit-Based Incentive Payment System (MIPS): Medicare Access and Children's Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA)	Optional measure for using PGHD to support the goal of coordinating care through patient engagement[74]

Additionally, the BlūBot prototype can used to meet public health objectives for mental health established by Healthy People 2020.[64] This includes the following objectives:

- MHMD-1: Reduce the suicide rate
- MHMD-2: Reduce suicide attempts by adolescents
- MHMD-4: Reduce the proportion of persons who experience major depressive episodes (MDEs)
- MHMD-6: Increase the proportion of children with mental health problems who receive treatment
- MHMD-9.2: Increase the proportion of adults aged 18 years and older with major depressive episodes (MDEs) who receive treatment
- MHMD-11: Increase depression screening by primary care providers

5.5 Recommendation

Based on the outcome of the analysis, the BlūBot one-day training and evaluation model is a good predictor of depressive episodes; aligning closely with user-generated PHQ-9 scores. The BlūBot prototype can be a useful tool for screening for depression in between doctor visits and provide opportunities for intervention from health care providers when depressive episodes occur. Steps should be taken to further validate the model and improve upon the application through user studies. Once the model has been validated in a significant sample size, efforts should be made to integrate the BlūBot app into other EMR/EHR systems. This prototype, including its application, infrastructure, and repository, can all be repurposed and expanded upon to incorporate data from various wearables and mHealth applications into EMR/EHR systems for public health. The configuration setting of the app allows users to customize the features included in generating model predictions. The reporting document, created using FHIR specifications, supports interoperability with FHIR-enabled systems.

6. References

1. Strine TW, Mokdad AH, Balluz LS, et al. Depression and anxiety in the United States: findings from the 2006 Behavioral Risk Factor Surveillance System. Psychiatr Serv 2008; 59(12). http://www.ncbi.nlm.nih.gov/pubmed/19033164 (accessed Aug 18 2016).

2. Chapman DP, Perry GS. Depression as a major component of public health for older adults. Prev Chronic Dis. 2008; 5(1). (accessed July 2016).

3. Greenberg P, Fournier A, Sisitsky T, Pike C, Kessler R. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). Journal of Clinical Psychiatry 2015; 76(2). http://www.psychiatrist.com/jcp/article/Pages/2015/v76n02/v76n0204.aspx (accessed Aug 2016).

4. Cassano P, Fava M. Depression and public health: an overview. J Psychosom Res 2002; 53(4). https://www.ncbi.nlm.nih.gov/pubmed/12377293 (accessed Oct 2016).

5. Li C, Ford ES, Zhao G, Ahluwalia IB, Pearson WS, Mokdad AH. Prevalence and correlates of undiagnosed depression among U.S. adults with diabetes: the Behavioral Risk Factor Surveillance System, 2006. Diabetes Res Clin Pract 2009 Dec 25 2008; 83(2). https://www.ncbi.nlm.nih.gov/pubmed/19111364 (accessed Sep 2016).

6. Lewandowski RE, O'Connor B, Bertagnolli A, et al. Screening for and Diagnosis of Depression Among Adolescents in a Large Health Maintenance Organization. Psychiatr Serv 2016;67(6):636-41 doi: 10.1176/appi.ps.201400465published Online First: Epub Date]|.

7. Jackson WC. The benefits of measurement-based care for primary care patients with depression. J Clin Psychiatry 2016; 77(3). https://www.ncbi.nlm.nih.gov/pubmed/27046318 (accessed Aug 2016).

8. Keller AO, Valdez CR, Schwei RJ, Jacobs EA. Disclosure of Depression in Primary Care: A Qualitative Study of Women's Perceptions. Women's Health Issues 2016 Aug 13 2016; (1878-4321 (Electronic)). (accessed Aug 2016).

9. Bush NE, Skopp N, Smolenski D, Crumpton R, Fairall J. Behavioral screening measures delivered with a smartphone app: psychometric properties and user preference. J Nerv Ment Dis 2013; 201(11). https://www.ncbi.nlm.nih.gov/pubmed/24177488 (accessed Nov 2016).

10. Torous J, Baker JT. Why Psychiatry Needs Data Science and Data Science Needs Psychiatry: Connecting With Technology. JAMA psychiatry 2015; 73(1). https://www.ncbi.nlm.nih.gov/pubmed/26676879 (accessed Aug 19 2016).

11. Torous J, Staples P, Onnela JP. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015; 17(8). https://www.ncbi.nlm.nih.gov/pubmed/26073363 (accessed Aug 18 2016).

12. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014; 2(1). https://www.ncbi.nlm.nih.gov/pubmed/25098314 (accessed Jan 21).

13. Wong ML, Licinio J. Research and treatment approaches to depression. Nat Rev Neurosci 2001;2(5):343-51 doi: 10.1038/35072566published Online First: Epub Date]|.

14. Lee EE, Della Selva MP, Liu A, Himelhoch S. Ketamine as a novel treatment for major depressive disorder and bipolar depression: a systematic review and quantitative meta-analysis. Gen Hosp Psychiatry 2015 Jan 15 2015; 37(2). https://www.ncbi.nlm.nih.gov/pubmed/25698228 (accessed Feb 2017).

15. Renton T, Tang H, Ennis N, et al. Web-based intervention programs for depression: a scoping review and evaluation. J Med Internet Res 2014; 16(9). https://www.ncbi.nlm.nih.gov/pubmed/25249003 (accessed Feb 28 2017).

16. Moragne W. DIAGNOSIS AND TREATMENT. Depression: Lerner Publishing Group, 2001:44.

17. Mokros H. Communication and Psychiatric Diagnosis: Tales of Depressive Moods From Two Contexts. Health Communication 1993;5(2):113

18. Beck AT. Thinking and Depression. I. Idiosyncratic Content and Cognitive Distortions. Arch Gen Psychiatry 1963; 9(4). https://www.ncbi.nlm.nih.gov/pubmed/14045261 (accessed Oct 2016).

19. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015; 38(3). https://www.ncbi.nlm.nih.gov/pubmed/25844912 (accessed Sep 2016).

20. O'Connor E, Rossom RC, Henninger M, et al. Screening for Depression in Adults: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force. Evidence Syntheses. Rockville, MD: Agency for Healthcare Research and Quality (US) CTI - U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews, 2016.

21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sept 15 2001; 16(9). https://www.ncbi.nlm.nih.gov/pubmed/11556941 (accessed Sep 2016).

22. Cheng PG, Ramos RM, Bitsch JA, et al. Psychologist in a Pocket: Lexicon Development and Content Validation of a Mobile-Based App for Depression Screening. JMIR Mhealth Uhealth 2016; 4(3). https://www.ncbi.nlm.nih.gov/pubmed/27439444 (accessed Jul 20).

23. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. Clin Psychol Rev 2012; 32(6). https://www.ncbi.nlm.nih.gov/pubmed/22721999 (accessed Aug 2016).

24. Kim J, Lim S, Min YH, et al. Depression Screening Using Daily Mental-Health Ratings from a Smartphone Application for Breast Cancer Patients. J Med Internet Res 2016 Aug 4 2016; 18(8). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4990716/ (accessed Aug 2016).

25. Torous J, Staples P, Shanahan M, et al. Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. JMIR mental health 2015; 2(1). https://www.ncbi.nlm.nih.gov/pubmed/26543914 (accessed Sep 2 2016).

26. Neuman Y, Cohen Y, Assaf D, Kedma G. Proactive screening for depression through metaphorical and automatic text analysis. Artif Intell Med 2012;56(1):19-25 doi: 10.1016/j.artmed.2012.06.001published Online First: Epub Date]|.

27. Park YR, Lee Y, Lee G, Lee JH, Shin SY. Smartphone applications with sensors used in a tertiary hospital-current status and future challenges. Sensors (Basel) 2015 Apr 27 2015; 15(5). https://www.ncbi.nlm.nih.gov/pubmed/25923933 (accessed Sep 2016).

28. Winslow BD, Chadderdon GL, Dechmerowski SJ, et al. Development and Clinical Evaluation of an mHealth Application for Stress Management. Front Psychiatry 2016; 7. https://www.ncbi.nlm.nih.gov/pubmed/27507949 (accessed July 2016).

29. Weidt S, Wahle F, Rufer M, Horni A, Kowatsch T. [MOSS- Mobile Sensing and Support Detection of depressive moods with an app and help those affected]. Ther Umsch 2015 Sep 1 2015; 72(9). https://www.ncbi.nlm.nih.gov/pubmed/26323953 (accessed Aug 25 2016).

30. Saeb S, Zhang M, Karr CJ, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015; 17(7). https://www.ncbi.nlm.nih.gov/pubmed/26180009 (accessed Aug 2016).

31. Using Social Sensing to Understand the Links Between Sleep, Mood, and Sociability. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing; 2011; Cambridge, MA. IEEE.

32. Ginger.io. Good care starts with good evidence. [Web Page] https://ginger.io/evidence/ (accessed Sep 2, 2016).

33. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The Relationship between Clinical, Momentary, and Sensor-based Assessment of Depression. 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth). Istanbul, Turkey: ICST, the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2015:229.

34. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: A Novel mHealth System for Supporting Self-Care in Management of Complex and Chronic Conditions. JMIR Mhealth Uhealth 2013 Jul 2013; 1(2). https://www.ncbi.nlm.nih.gov/pubmed/25100682 (accessed Sep 16 2016).

35. Richmond SJ, Keding A, Hover M, et al. Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. BMC Psychiatry 2015; 15(1471-244X (Electronic)). https://www.ncbi.nlm.nih.gov/pubmed/25886377 (accessed Sep 6 2016).

36. Darwish A, Hassanien AE. Wearable and implantable wireless sensor network solutions for healthcare monitoring. Sensors (Basel) 2011; 11(6). https://www.ncbi.nlm.nih.gov/pubmed/22163914 (accessed Sep 2016).

37. Pennic F. Novant Health To Integrate Epic MyChart with FitBit & HealthKit. http://hitconsultant.net/2015/08/21/novant-health-to-integrate-epic-mychart-with-fitbit-healthkit/ (accessed February 13, 2017).

38. Kasthurirathne SN, Mamlin B, Kumara H, Grieve G, Biondich P. Enabling Better Interoperability for HealthCare: Lessons in Developing a Standards Based Application Programing Interface for Electronic Medical Record Systems. J Med Syst 2015; 39(11). https://www.ncbi.nlm.nih.gov/pubmed/26446013 (accessed Feb 2017).

39. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Feb 17 2016; 23(5). https://www.ncbi.nlm.nih.gov/pubmed/26911829 (accessed Feb 2017).

40. Wagholikar KB, Mandel JC, Klann JG, et al. SMART-on-FHIR implemented over i2b2. J Am Med Inform Assoc 2016 Jun 06 2016. https://www.ncbi.nlm.nih.gov/pubmed/27274012 (accessed Sep 16 2016).

41. HealthIT.gov. PHQ-9 Reporting Trial: Interoperability Proving Ground. [Web Page] https://www.healthit.gov/techlab/ipg/node/4/submission/196 (accessed Sept 1, 2016).

42. FHIR Infrastructure Work Group. FHIR Timelines. https://www.hl7.org/fhir/timelines.html (accessed February 7, 2017).

43. Siu AL, Force USPST. Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. JAMA 2016; 315(4). https://www.ncbi.nlm.nih.gov/pubmed/26813211 (accessed Aug 20 2016).

44. Hall-Flavin DK. What does the term "clinical depression" mean? [Web Page] http://www.mayoclinic.org/diseases-conditions/depression/expert-answers/clinical-depression/faq-20057770 (accessed Sept 1, 2016).

45. American Psychiatric Association. Depressive Disorders. Diagnostic and Statistical Manual of Mental Disorders. Fifth ed. Washington D.C.: American Psychiatric Publishing, 2013.

46. Christensson P. API Definition. http://techterms.com/definition/api (accessed January 22, 2017).

47. World Health Organization. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. Global Observatory for eHealth Series: World Health Organization, 2011.

48. HealthIT.gov. What is an Electronic Medical Record (EMR)? [Web Page] https://www.healthit.gov/providers-professionals/electronic-medical-records-emr (accessed Sept 16, 2016).

49. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016; 23(3). https://www.ncbi.nlm.nih.gov/pubmed/27018263 (accessed Jan 2017).

50. M. E, Kriegel H, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96): AAAI Press, 1996:226--31.

51. VanderPlas JT, Ivezic Z. Periodograms for Multiband Astronomical Time Series. Astrophys J 2015 Oct 10 2015; 812(1). <Go to ISI>://WOS:000364234700018 (accessed Mar 7 2017).

52. Shenkin PS, Erman B, Mastrandrea LD. Information-theoretical entropy as a measure of sequence variability. Proteins 1991; 11(4). https://www.ncbi.nlm.nih.gov/pubmed/1758884 (accessed Mar 7 2017).

53. Amazon Web Services. Architecting for HIPAA Security and Compliance on Amazon Web Services. 2017 March 2017. https://d0.awsstatic.com/whitepapers/compliance/AWS_HIPAA_Compliance_Whitepaper.pdf (accessed March, 2017).

54. The PostgreSQL Global Development Group. PostgreSQL: About. https://www.postgresql.org/about (accessed March 2, 2017).

55. HockeyApp. HockeyApp-The Platform for Your Apps. [Web Page] https://hockeyapp.net/ (accessed Mar 2 2017).

56. OpenEMR. OpenEMR. http://www.open-emr.org/ (accessed Oct 20, 2016).

57. Vreeman DJ, McDonald CJ, Huff SM. Representing Patient Assessments in LOINC®. AMIA Annual Symposium Proceedings 2010 Nov 13 2010; 2010. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041404/ (accessed Feb 18 2017).

58. Logical Observation Identifiers Names and Codes (LOINC®). 44249-1 PHQ-9 quick depression assessment panel [Reported.PHQ]. [Online Database] http://s.details.loinc.org/LOINC/44249-1.html?sections=Comprehensive (accessed December 5, 2016).

59. OpenEMR Discussion. https://sourceforge.net/p/openemr/discussion/ (accessed October 20, 2016).

60. OpenEMR. OpenEMR: Active Projects. http://www.open-emr.org/wiki/index.php/Active_Projects (accessed Dec 21, 2016).

61. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. Geoscientific Model Development 2014; 7(3). <Go to ISI>://WOS:000341600100030.

62. D'Mello SK. Chapter 5 - Automated Mental State Detection for Mental Health Care In: Luxton DD, ed. Artificial Intelligence in Behavioral and Mental Health Care. San Diego: Academic Press, 2016:117-36.

63. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. Psychiatric Annals 2002;32(9):509-15

64. Office of Disease Prevention and Health Promotion. Healthy People 2020. [Web Page] https://www.healthypeople.gov/2020/topics-objectives/topic/mental-health-and-mental-disorders/objectives (accessed February 15, 2017).

65. Agency for Healthcare Research and Quality. Agency for Healthcare Research and Quality: A Profile. https://www.ahrq.gov/cpi/about/profile/index.html (accessed February 9 2017).

66. The National Committee for Quality Assurance. About NCQA. http://www.ncqa.org/about-ncqa#sthash.R1lrN9cL.dpuf (accessed February 14, 2017).

67. The National Committee for Quality Assurance. HEDIS Depression Measures Specified for Electronic Clinical Data Systems. http://www.ncqa.org/hedis-quality-measurement/hedis-learning-collaborative/hedis-depression-measures (accessed February 9, 2017).

68. Service AF. Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024: Draft White Paper for a PGHD Policy Framework. https://www.healthit.gov/sites/default/files/Draft_White_Paper_PGHD_Policy_Framework.pdf (accessed Feb 19, 2017).

69. OpenEMR Forums: Mental Health Forms. https://sourceforge.net/p/openemr/discussion/202504/thread/0db2469f/?limit=25&page=1#7c1d (accessed February 13, 2017).

70. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016; 4(e2537). https://doi.org/10.7717/peerj.2537 (accessed Feb 15 2017).

71. Centers for Medicare & Medicaid Services. Federal Register | Medicare and Medicaid Programs; Electronic Health Record Incentive Program-Stage3 and Modifications to Meaningful Use in 2015 Through 2017. https://www.federalregister.gov/articles/2015/10/16/2015-25595/medicare-and-medicaid-programselectronic-health-record-incentive-program-stage-3-and-modifications- p-1678. (accessed March 16, 2017).

72. Centers for Medicare & Medicaid Services. Community-Based Care Transitions Program. https://innovation.cms.gov/initiatives/CCTP/ (accessed March 16, 2017).

73. Centers for Medicare & Medicaid Services. Chronic Care Management Services. In: Services CfMM, ed., 2015.

74. Centers for Medicare & Medicaid Services. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models. https://www.gpo.gov/fdsys/pkg/FR-2016-11-04/pdf/2016-25240.pdf (accessed March 16, 2017).

75. Department of Human and Health Services. Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA. In: HealthIT.gov, ed. Washington, D.C., 2016.

76. National Institute of Mental Health. Depression. [Web Page] https://www.nimh.nih.gov/health/topics/depression/index.shtml (accessed September 1, 2016).

7. Appendix

Appendix A. List of Recognized Depression Symptoms

According to DSM-IV and NIMH, the following symptoms are consistent with depression:

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide. [39]

According to the National Institute of Mental Health (NIMH), other symptoms may include:

- Feelings of hopelessness, or pessimism
- Difficulty concentrating, remembering, or making decisions
- Aches or pains, headaches, cramps, or digestive problems without a clear physical cause and/or that do not ease even with treatment[76]

Appendix B. PHQ-9 Items:

1. Little interest or pleasure in doing things
2. Feeling down, depressed, or hopeless
3. Trouble falling or staying asleep, or sleeping too much
4. Feeling tired or having little energy
5. Feeling bad about yourself or that you are a failure or have let yourself or your family down
6. Poor appetite or overeating
7. Trouble concentrating on things, such as reading the newspaper or watching television
8. Moving or speaking so slowly that other people could have noticed. Or the opposite being so fidgety or restless that you have been moving around a lot more than usual
9. Thoughts that you would be better off dead, or of hurting yourself
10. If you checked off any problems, how difficult/not difficult at all have these problems made it for you to do your work, take care of things at home, or get along with other people?

The scores are totaled to create a composite score. The composite score is categorized as follows: [21]

Interpretation of Total Score

Total Score	Depression Severity
1-4	Minimal depression
5-9	Mild depression
10-14	Moderate depression
15-19	Moderately severe depression
20-27	Severe depression

Appendix C. BlūBot prediction model data flow legend

1. Sensor data are recorded by PDK Data Generators.
2. PHQ-9 surveys completed by the user are recorded by the custom BlūBot PHQ-9 Data Generator.
3. The predictive model queries the data generators for sensor data an survey responses to train a statistical model relating the two.
4. The predictive model logs PHQ-9 predictions derived from sensor data.
5. PDK data generators log predictions and sensor data to the Generator Manager.
6. The Generator Manager relays incoming data to the app display to update the BlūBot display and the HTTPS Transmitter to prepare the data for transmission.

A. The HTTPS Transmitter sends the data to the cloud server over an SSL encrypted data connection.
B. The Data Repository stores the incoming data for analysis and dissemination.
C. The Data Repository stores a file containing the PHQ-9 data within the OpenEMR MySQL Database. It also creates alert messages if the composite score exceeds a threshold of 10.
D. The OpenEMR Application Server stores and retrieves data from the MySQL Database, including data inserted by the Data Repository.
G. The OpenEMR Application Server retrieves PHQ-9 visualizations to embed in the main clinician display.