The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders

Aalto University,

Bachelor's Thesis

10. Sept 2020

Samuel Sihvonen

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

KANDIDAATINTYÖN TIIVISTELMÄ

Tekijä: Samuel Sihvonen

Työn Nimi: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders

Päiväys: 10.09.2020

Sivumäärä: 31

Pääaine: Tietotekniikka

Vastuuopettaja: Professori Eero Hyvönen

Työn ohjaaja: Tohtori Talayeh Aledavood

Digitaalisuuden kehittyminen on mahdollistanut monimuotoisen, henkilökohtaisen datan keräämisen sosiaalisesta mediasta, älypuhelimista, älykelloista ja muista vastaavista henkilökohtaisista laitteista. Datan määrän kasvu ja laskentatehon kasaantuminen ovat johtaneet koneoppimisteknologian kehittymiseen olemassa olevien järjestelmien päälle. Koneoppiminen on päässyt mukaan myös terveydenhuoltoon, jossa sen tehtävinä on esimerkiksi hoidon laadun parantaminen tai diagnoosin tarkempi arviointi.

Tämän tutkielman tarkoituksena on kartoittaa mielialahäiriöden oireiden ilmentyvyyttä koneoppimisen metodien avulla. Kandidaatintyössä vertaillaan koneoppimisessa tarvittavia datan lähteitä, joita tarvitaan mallien tarkkuuden parantamiseksi ja tarkkuuden testaamiseksi. Datan lähteet ovat tutkimuksessani henkilökohtaisia älylaitteita. Lisäksi tutkielmassani kyseenalaistetaan vastaavien koneoppimismetodien tarpeellisuus mielenterveydenhoidon tulevaisuudessa.

Kandidaatintyöni on kirjallisuustutkimus, jossa arvioidaan vuosina 2014-2020 ilmestyneiden 29 artikkelin menetelmiä ja tuloksia. Erityisesti masennukseen ja kaksisuuntainen mielialahäiriö ovat suurena tutkimuksenkohteena valittujen artikkeleiden joukossa. Datan lähteinä hyödynnettiin eniten älypuhelinsovelluksia, joilla on kerätty muun muassa dataa unesta, liikkumisesta ja sosiaalisista kanssakäymisistä. Koneoppimismenetelminä tärkeimmiksi muodostuivat ‘random forest’ ja ‘support vector machine’. Koneoppimismallien ennustuustarkkuus sairauksille ja oireille on julkaisuissa merkittävä ja vakavasti otettava: 68% - 96%

Tekoälyn ja koneoppimisen hyödyntäminen mielenterveyden hoitamisessa on kandidaatintyöni perusteella kasvava ja jalansijaansa vahvistava ala. Tulevaisuudessa voidaan olettaa, että metodeita hyödynnetään yhdessä kliinisen hoitotiimin kanssa. Sitä ennen tulee valmistella isompia tutkimuksia, standardisoida menetelmiä ja vahvistaa datan turvallisuutta.

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

ABSTRACT OF BACHELOR'S THESIS

Author: Samuel Sihvonen

Title of thesis: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders

Date: 10.09.2020

Pages: 31

Pääaine: Computer Science

Vastuuopettaja: Professor Eero Hyvönen

Työn ohjaaja: Dr. Talayeh Aledavood

The development of digitalization has made it possible to collect diverse, personal data from social media, smartphones, smartwatches, and other similar personal devices. The increase in the amount of data and the accumulation of computing power have led to the development of machine learning technology on top of existing systems. Machine learning has also become part of health care, where its tasks include improving the quality of care or a more accurate assessment of the diagnosis.

The purpose of this thesis is to map symptoms of mood disorders using machine learning methods. The bachelor's thesis compares the data sources needed in machine learning to improve the accuracy of the models and to test the accuracy. The data sources in my research are personal smart devices. Also, my dissertation questions the need for similar machine learning methods in the future of mental health care.

My bachelor thesis is a literature review that evaluates the methods and results of 29 articles published in 2014-2020. Depression and bipolar disorder in particular are major research topics among the selected articles. The most widely used data sources were smartphone applications, which collected data on sleep, movement, and social interactions, among other things. The most important machine learning methods were 'random forest' and 'support vector machine'. The predictive accuracy of machine learning models for diseases and symptoms is a significant 68% - 96%.

The utilization of artificial intelligence and machine learning in the treatment of mental health is a growing and strengthening field based on my bachelor's thesis. In the future, it can be assumed that the methods will be utilized in conjunction with the clinical care team. Before that, larger studies need to be prepared, methods standardized and data security strengthened.

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

Aalto University

Author: Samuel Sihvonen

Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders

Date: 10.09.2020

Major: Computer Science

Instructor: Dr. Talayeh Aledavood

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

Introduction

This bachelor's thesis is based on the methods of machine learning in the prediction of symptoms of mood disorders. The work is a literature review in which 21st-century literature sources related to my topic are used. The focus is on studying research papers that have implemented machine learning methods for predicting symptoms and causes of mood disorders, where the data used in the algorithms is gathered from wearables and other data sources personal to the subjects. This type of data collection is being used in other mental health research context as well[1]. Background of the state of mental health, mood disorders, and artificial intelligence and machine learning is provided before introducing the methods used in this thesis. After the methods section, the results of composing a review of relevant papers including present methods, and a section for discussion about the selected articles and materials.

There are two relevant items, to understand machine learning models. First, is the data sources the machine learning model is being trained and tested with. The question of which data sources are the ones providing the best predictions. The data from the wearable devices can be from e.g. a GPS[2–18] or a heart rate monitor[2,6,11], or sleep [2–4,6,9,11,14,15,17,19,20]. Second are the machine learning models, which provide quantitative prediction scores for mood disorders. The question of which methods stand out from the multiple possible machine learning algorithms developed over the years, and which ones are being implemented by the studies and why.

Ultimately, the goal is predicting mood would presumably be to further clinical care by making diagnosis easier to access and developing more lean processes in health care. Therefore, the last point of exploration in this thesis is how much value do the models bring to the table and can they possibly better treatment results by smarter and more lenient ways of predicting moods and mood disorders.

With the reasons mentioned above, the research questions for my thesis are the following:

1. What data sources are the best for collecting data for mood disorders?
2. What kind of machine learning methods are being used in mood prediction?
3. Are machine learning methods necessary to boost the efficacy of treatment, where are they beneficial and where should we consider other methods?

Background

Introduction to Artificial Intelligence in Mental Health

The rise of digitalization has provided various applications of information and communication technology, which now have also started to be applied in healthcare services. Humanity is coming to a critical point in reaching the Fourth Industrial Revolution[21]. The revolution consists of the lines between 'physical, digital, and biological'[21] worlds mixing with the implementation and use of e.g. artificial intelligence and machine learning. This phenomenon and movement has also been called as the 'Industry 4.0'. One of the concepts strongly associated with today's digitalization and one that is gaining traction in both academic and business circles is artificial intelligence, more specifically machine learning and its various applications. In mental health, machine learning can potentially be used in many different applications, such as the identification and diagnosis of disease risk factors.[22]

Due to technological development in areas such as social media, smartphones, wearables, and neuroimaging[23], where health data can be collected, machine learning algorithms allow academics and software developers to make personalized treatment suggestions and predict the course and outcome of an illness based on individual characteristics and background[24].

This new ability to take into account multiple variables concerning a single patient opens state-of-the-art predictive possibilities for mental health and mood disorders.

State of Mental Health and Mood Disorders

In 2013, the World Health Organization [25] released a report called 'Mental Health Action Plan 2013-2020". In this report, the statement *'Mental well-being is a fundamental component of WHO's the definition of health.',* hardly leaves room for inference on the importance of mental health in the eyes of the World Health Organization. Mental health disorders affect 10,7% of the global population [26], and mood disorders affect 9,7% of all US citizens, 18 or older, studied by Harvard Medical School[27].

Mood affective disorders are defined as health disorders, which affect an individual's mood. A person is a 'good mood', is psychiatrically characterized [28] as having better verbal ability and implicit memory, and also, the studied person may feel a sense of ease upon themselves. Common mood disorders include conditions such as depression, bipolar disorder, type 1, and 2. Type 1 is characterized by the presence of at least one manic episode, with or without a history of major depressive episodes, and type 2 requires at least one hypomanic and one major depressive episode, and also seasonal affective disorder. Harvard Medical School evaluated that 9,7% of adults in the United States have had some type of mood affective disorder in the past year [27]. Also, Kessler RC et al. [29] estimated in their research that from all those adults, who had suffered from a mood disorder of any kind during the last year, as much as 45% had suffered from a serious impairment, 40% from a moderate impairment, and 15% from a mild impairment.

The diagnosis of these mood disorders is largely done by interviewing the patient about their mood, depending on the disorder that the clinician is screening for. The problem with questionnaires such as the Patient Health Questionnaire (PHQ-9), is that many of the mental health assessments have a low positive predictive value. Positive predictive value is the

probability that the patient in question with a positive test result has the disease[30], i.e. is not false positive.

Utilization of Wearables for Collection of Data

A wearable is defined as being a technological device, that collects non-invasive psychophysiological signals from the wearer's body and ultimately transmits and analyzes the results for the wearer's sake, giving them some type of insight[31]. Wearable technology, a more prevalent megatrend year-by-year, has doubled in size from 2014 to 2018 in the United States, by gaining 27 million more users in the course of the four years as described by [32], with the population of the United States being 328,9 million. Wearables, integrated with IoT (Internet of Things), will be an important piece in establishing a new paradigm: 'Healthcare 4.0'[33], which can be linked to the already mentioned 'Industry 4.0'. Healthcare 4.0 will further combine the biological human anatomy with technologies such as machine learning.

Machine Learning Methods Used in Health Care

Artificial intelligence, machine learning, and deep learning processing are techniques that are relevant to healthcare. Machine learning can be used to analyze structured data contained in electronic health records, such as blood pressure, weight, and x-rays. Natural language processing can be utilized to process unstructured text data and thereby also bring patient reports to machine learning algorithms in a comprehensible form[34]. Different applications of artificial intelligence have a wide range of potential applications at different levels of disease prevention[35]

Under experimental conditions, artificial intelligence has shown its potential to be effective in preventing disease, but in practice, experiments are still needed to confirm the results. Obstacles to the use of this practice include the difficult human comprehensibility of certain artificial intelligence techniques and open ethical and legal issues[33].

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

Methods

The work has been carried out in its entirety as a literature review and does not include empirical research. Literature sources were collected using databases from Scopus and Google Scholar. The development and research of artificial intelligence in the use of mood prediction and mental health have been abundant especially in recent years, and most of the studies on the applications of artificial intelligence were published in the 21st century. Some of the different technologies of artificial intelligence have been developed before the 21st century.

To collect these materials, a selection of 'keywords' was established that best combined the different possibilities of mood prediction with artificial intelligence technologies and the gathering of subject data done by personal tracking devices. The combinations were done by combining 'keywords' in the same category with boolean OR operator and stringing the categories together with boolean AND operator, as is a *de facto* way in obtaining relevant articles for a review.

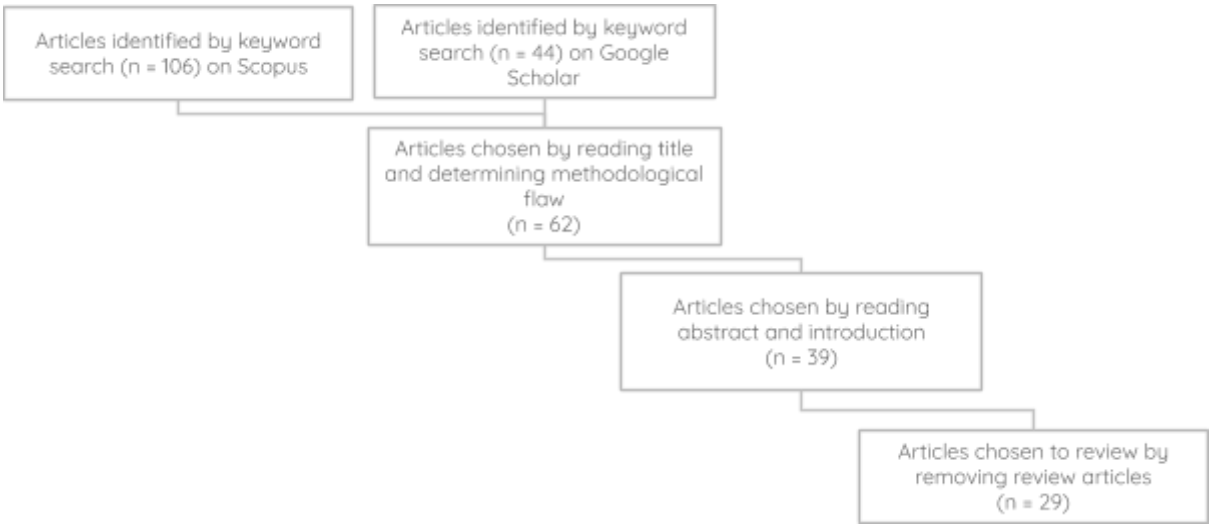
Search terms:

- *"Machine learning" OR "artificial intelligence" OR "deep learning" OR AI OR "neural network" OR "unsupervised learning" OR "supervised learning"*

- *"psychiatry" OR "psychiatric" OR "depression" OR "depressive" OR "suicidality" OR "suicide" OR "mood disorder" OR ("mood" AND "disorder") OR "major depressive episode" OR "major depression" OR bipolar OR "bipolar disorder"*

- *"smartphone" OR "smartwatch" OR "fitness tracker" OR "tracker" OR "sleep tracker" OR "sleeptracker"*

Article Selection for Literature Review



The search was done by searching only in the *Title, Keywords,* and *Abstract* of publications on Scopus and Google Scholar. The mentioned search terms found a total of exactly 106 articles on Scopus and 44 on Google Scholar. After the initial search results, the title was evaluated, and fitting titles were included in the next round of selection. During this round, methodological flaw was also determined, which included issues e.g. a publication studying an irrelevant disorder to this literature review. After this initial filtering process, an article's abstract and introduction was read through to further assess it's fitness to the study. From these articles, 29 were ultimately chosen to be a part of the literature review.

Results

All articles concluded in the literature review were published during the years 2014 to 2020, where the publication count had a growing trend. In figure 1, a display of articles released on a particular year is visualized. The figure demonstrates an increase in the amount of research done in the field over the last 6 years. The first papers appear in 2014 and an increasing trend can be noticed in the number of publications from 2016 onward. In 2020, the year this thesis is written, 8 articles were published. The article collection was gathered during mid-July and therefore, it is most likely that the article collection of 2020 will not remain under 10 publications. This exemplifies the growing relevance of the field year by year.

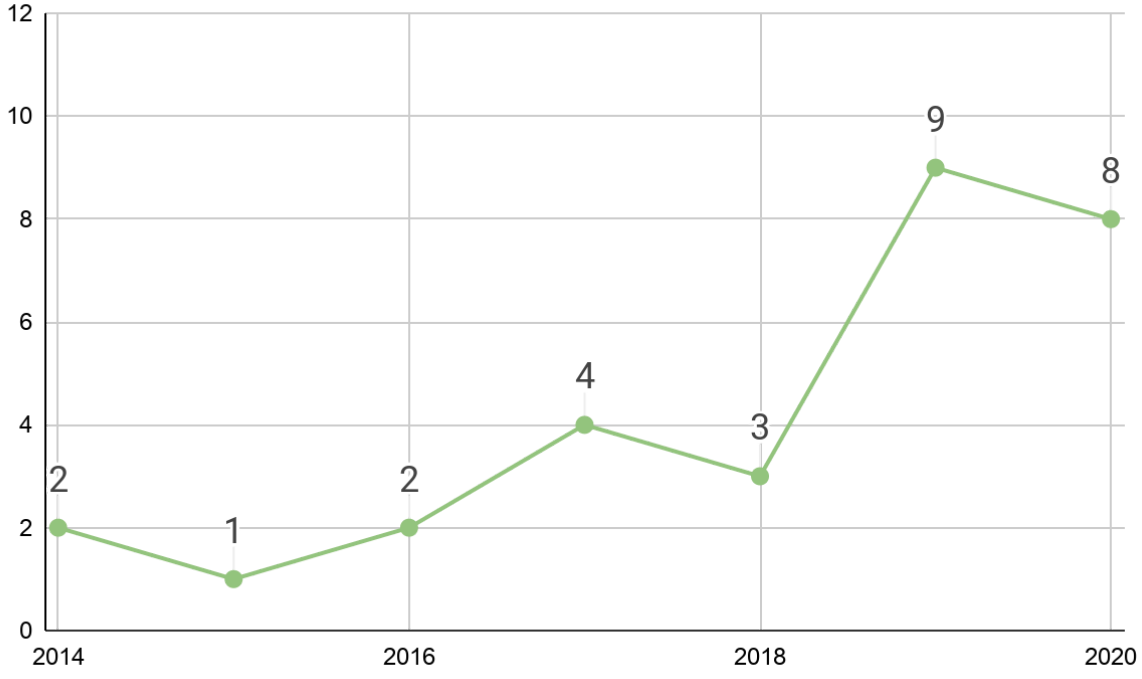


Figure 1

To help better grasp how these studies are performed, the common structure of a study close to the following:

1. Gather data with wearables
2. Train the machine learning model with the gathered data
3. Test the newly trained model against e.g. a clinical score (PHQ-9)[36], EMA[37], or mood questionnaire and record its predictive power

In the reviewed collection of articles, 27/29 used phones as the type of wearable. Only 1/29[38] used a fitness tracker (Fitbit[39]), and the remaining study (1/29) had pre-existing medical data from hospital visits to train their models with[40].

Major Themes in Results

Disorders

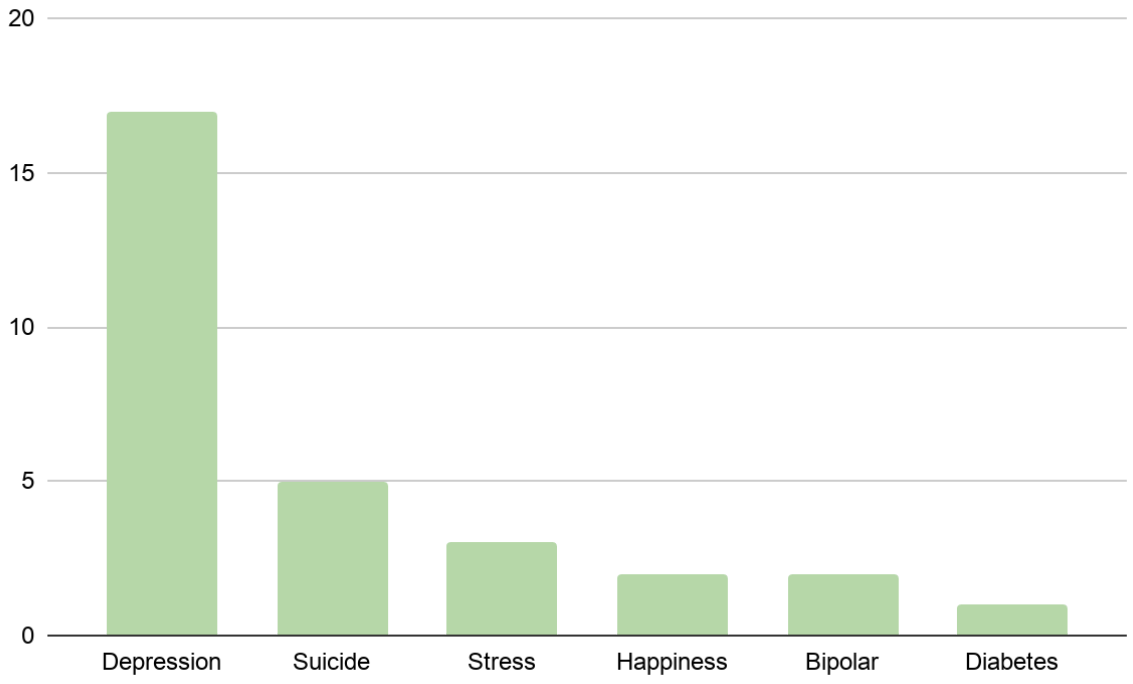


Figure 2

Depression & Suicidality

17/29 of the reviewed articles [2,3,5–9,11,13,14,16,17,19,41–44] studied predicting depression from different angles. As seen from figure 2, depression is the most prevalent topic or disorder

that was studied. The publications tried to predict the probability of an upcoming depressive period or forecast a developing depressive period. The sample sizes in these themes ranged from, ranging from 3 to 335, and the median being 47 participants. 4/17 of these articles [2,8,16,17] were published on undergraduate students, all of the respective universities located in the United States. Accuracy for the studies ranged from 52% [41] to 92.6% [5], with the median being 91%. Also in the collection of selected articles, 5 featured [4,10,40,41,43] predicting suicide risk. 6/17 depression publications compared the scores predicted by the machine learning models to widely used questionnaires in clinical care: PHQ-9[5,8,9,16,42] and PHQ-2[7]. The rest of the 17 depression & suicidality articles that didn't use the PHQ-9 or PHQ-2 used mood questionnaires or EMAs.

Bipolar

A small number of studies, 2/29, featured Bipolar Disorder, both type 1 and 2[45]. One paper[6] had a goal to predict mood for the next three days via prognosis, that a disturbance in circadian rhythm would cause disarray into the status of a patient's mood. The study concluded a notable 2 years of data collection into its publication, with the accuracy of predicting manic episodes of 94% and depressive episodes by 87%. The article did not only confine data collection to sleep with a Fitbit[38], but also collected Activity and Sleep data through the same device. The study had a sample size of 55. The other study [46] aimed to predict mood for the following day with a sample size of 130. The study tried to predict six different emotions, more accurately: anxious, elated, sad, angry, irritable, energetic. A comparison of a patient's mood questionnaire and predicted emotion level was done to determine the accuracy of the machine learning method. It succeeded to predict the mentioned emotions with accuracy ranging from 82% to 90% for patients suffering from bipolar disorder and for healthy subjects with an accuracy of 89% to 97%.

Data Sources

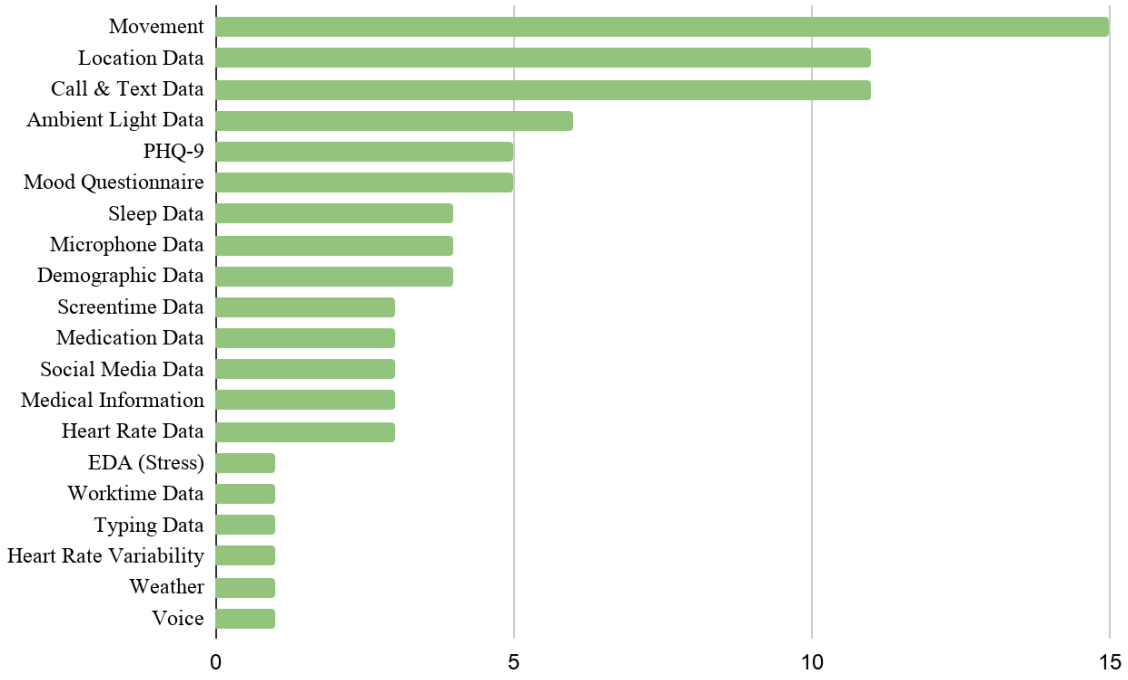


Figure 3

Sleep

Sleep was measured in 10 total papers [2–4,6,9,11,14,15,17,19], with ambient light [2,6,9,11,14,15], heart rate [2,6,11], accelerometer [6,14], or microphone data [4,6,14] utilized to determine quality of sleep, which the lack or disturbance of it [47] has been shown to be a sign of depression. 9/10 of the publications studied predicting depression or suicidality, which indicates that predicting depression with sleep analysis is a recognized field. All the studies regarding sleep had a rather small sample size from 3 [14] to 68 [17], except one larger with a sample size of 335[3].

Location or Movement

Location or movement tracking and analysis were used by 17/29 [2–18] of reviewed articles. This meant measuring traveled distance, daily routine, time spent at home, etc. The hypothesis was that by analyzing if people have a weekly routine and or spend too much time at home, this would signal a potentially starting depressive or manic period or rising risk of suicide. The success of the studies was varied. As an example, two studies conducted in 2020 that represent the smallest and largest sample size, with the best accuracy to the worst, respectively: Ware et al. [18] and also Masud et al. [5] gathered data on e.g. location variance, circadian movement, and entropy of movement. Masud et al. were able to reach the best accuracy of 92,6% for healthy individuals and 87,2% for severely depressed participants with n = 33. Ware et al. managed to acquire an F1 score ranges from 61% to 83% for all participants, with the sample size being a larger 182 college students. Both used smartphones to gather participant data.

Heart Rate & Heart Rate Variability

3/29 articles featured heart rate as a data source [2,6,11] and 1/3 of these articles featured heart rate variability[2]. Jacobson and Chung were able to calculate a correlation coefficient r = 0.587 with 95% CI across participants. In the study, they predicted mood for hourly intervals, which was the shortest time interval of prediction in all 29 papers. The two other publications [6,11] featuring used heart rate to analyze sleep. Jacobson and Chung also used HRV outside of sleep.

Typing

Typing data was gathered in only one article[42]. The article looked at sequences of keypresses and releases. Also the deletion rate, the number of characters typed, and typing session duration. It only gathered this typing data from participants and compared it to PHQ-9

results. The study claimed it reached 82%/86%, sensitivity/specificity respectively. Also, a correlation coefficient of >0.60 was calculated when compared to PHQ-9 scores. This coefficient signals a significant correlation.

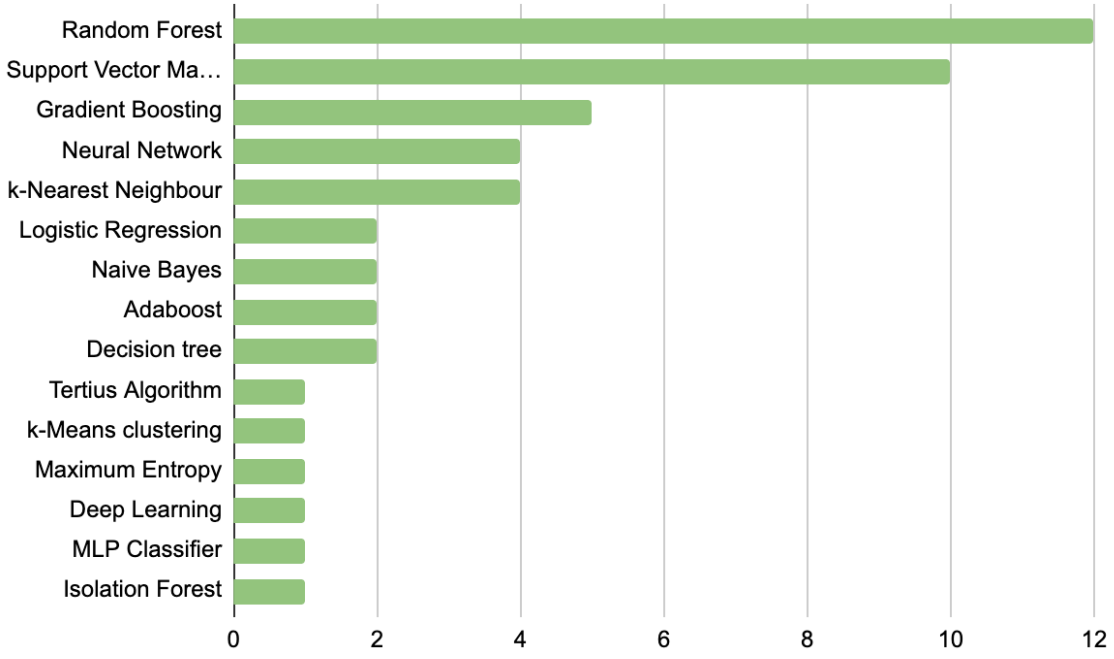


Figure 4

Predominant Machine Learning Methods

Random Forests

Figure 4 visualizes how the Random Forest method[48] has been used the most, 12/29 in total [3,4,6–9,11,15,17,40,42,46]. 6/12 of these publications [3,8,9,11,17,42] also used support vector machines together with the random forest method. Accuracy of results ranging from 68% to 96%, median 80.54%, with the best accuracy selected. Only four articles [6,7,40,46] used random forest

method on its own, the rest of the twelve compared it with others. The articles that only used random forests had an accuracy of ranging from 82% to 94%. Random forest articles had a sample size starting from 20[11] and a maximum of 335[3]

Support Vector Machines

10/29 articles used Support Vector Machines[3–5,8,9,11,16,17,42,44]. Only one article[16] used Support Vector Machines by themselves and didn't use any other machine learning methods. The best accuracy of the results ranged from, 68% to 96%, median 85%. Articles featuring support vector machines had a sample size ranging from 33[5] to 335[3], median 48 and the second-largest sample was 79.

Aalto University
Author: Samuel Sihvonen
Name: The Methods of Machine Learning in the Prediction of Symptoms for Mood Disorders
Date: 10.09.2020
Major: Computer Science
Instructor: Dr. Talayeh Aledavood

Discussion

Overall Study Quality

Studies especially after 2018 try to replicate other similar results and develop existing methodologies further. The gross outlook of the studies is engineering-focused, not focused on the clinical side as they could be. Also, because mental health and psychiatry is a very human-centric field, the design of the studies are sometimes overly technical and seem to reveal a lack of designers and clinicians in the research groups. This can be noticed from the concentration on the application of machine learning over treatment results, or that many of the studies developed they're own mobile app for data collection, without addressing the user experience aspect of a 'less-polished' mobile app. One could predict with adequate confidence, that a bad mobile app would discourage users to for example their mood of the day. Also, developing mobile apps in the United States to follow HIPAA[49], the healthcare privacy legislation in the US, is difficult and a timesink. The studies as mentioned, focused on machine learning and technicalities, instead of clinical applications but this might be because the articles were published on journals for computer science & engineering.

Realization of a New Study Field

The selected studies try to search for new ways in predicting mood. Through either different data sources, machine learning methods, analyzing symptoms of mood disorders, they try to attain the best predictive scores with varying results. The field is only coming out of its inception. There are a few noticeable signs of this:

1. A shift in the quality and quantity of articles starting from 2019. Articles during and after this year especially start to bring about more varied methods and have developed a principled study routine to gather relevant data
2. Sample sizes in all studies that gathered their own data, and didn't use already existing medical information, were small. Only 2/17 publications studying depression had a sample size larger than 100[3,7].
3. The machine learning methods used are elementary, such as random forests or support vector machines, which are not arduous to set up and are rather easy to test.
4. Discussion points in the articles such as privacy of data and relevance of these new applications, i.e. the studies question their own practical use.

Privacy and Ethics of Mood Prediction

Privacy and security are discussed mostly at the end of the studies. However, developing rigorous methods to keep privacy secure was not a large part of the methods in the articles. Mental health user data is extremely sensitive because it describes the patient's inner, family, economic, and other delicate data[50]. Even though full automation and independence from the clinical side can be a significant development in mental health services, acknowledgment of ethical, privacy and legal aspects in psychiatric prediction is important and vital for practical application of these methods. As an example, the United States already has existing and distinct health data regulation (HIPAA)[49] which tries to mitigate risks regarding health data security. This legislation guides healthcare organizations and companies to anonymize the data that they collect and process.

Wearables as a Data Source

In the materials collected for this bachelor's thesis, wearables are used because they are an easy and reliable source of continuous data. Wearables are personal for the user and are most often used throughout the daily lives of study participants. This eases the strain of

research studies on the participants since the wearables are products designed by large and professional teams of designers, developers, and business people like iOS or Android phones from Apple and Alphabet respectively. In possible clinical use, wearables can be a major data collection device, since the devices themselves have a sense of familiarity with their users. Wearables seem to have a significant role in mood prediction techniques in the future by the wide outlook drawn by these 29 research articles.

Challenges with wearables include things like varied apps used by the studies, user experience issues, and consent to give personal data to be processed. All the studies used or had developed a different application to track mood and gather data. This poses a challenge from a point of generalization, since using a new, different, or multiple apps for individual prediction techniques will be exhausting to users. At the same time, user experience is an industry by itself, and to bring such prediction techniques and data collection to customers, excellent user design will need to be performed, which the studies lack. Ultimately, consent to give your data away is a challenge, that is most likely the most prevalent out of all modern challenges. Somehow, the technologies will need to be able to convince the users that giving their data away is beneficial, but this will be an arduous task for the mood prediction field and market.

Application of Machine Learning for Predicting Mood

Predicting mood, possible symptoms, and upcoming episodes of mood disorders with machine learning is an interesting opportunity. The article collection gathered for this thesis presents machine learning models with respectable accuracy & sensitivity scores. This implies that the models used in these 29 articles could be used somewhat of a clinical setting to reinforce already existing care. What needs to be remembered with the prediction models, is that they do not supply clinical care, but more or less direct physicians and therapists to modify and iterate their care plans for individuals patients. Individuality and personalization is a big advantage that the machine learning models bring in, also called digital phenotyping[51].

The small sample sizes lift the question of overfitting of the models on the table. The traditional tradeoff in machine learning with bias over variance[52] cannot be fulfilled without a proper sample size. When models overfit, they tend to have high accuracy for the already existing population but have weak accuracy when a new input is introduced[53]. This can be examined by drawing a scatter plot like Figure 5, where sample sizes and accuracies of studies are compared. Only 10 studies listed both accuracy and sample size[2–6,8,9,11,17,46], other publications listing sensitivity/specificity, or F1 scores in results. In the scatterplot, a negative trend can be distinguished, with a correlation coefficient of -0.40. If we remove sample sizes of over 100, the correlation coefficient decreases to -0.80. This is done because the sample of ten has few data points with sample size >75, and the larger sample sizes might distort the 'real' trend. The coefficients and the trendline on the figure imply that smaller studies might overfit their results and therefore acquire better accuracy.

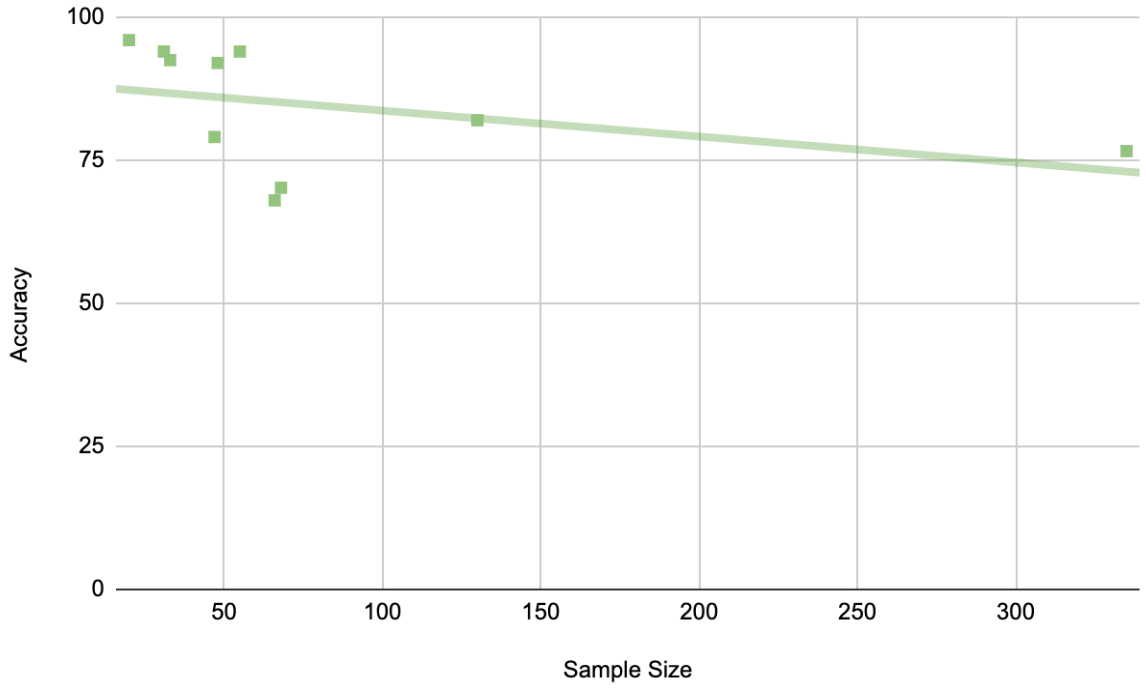


Figure 5

Conclusions

In this bachelor's thesis, a literature review of 29 papers was put together, evaluated, and analyzed. In this section, a conclusion of the composed materials and literature is done to give an overview of the future of artificial intelligence methods for mood prediction. This is done by evaluating the answers in results and discussion sections to the research questions posed in the introductory section at the beginning of the thesis. In the end, I will examine the success of the thesis itself, it's methods, results, and extents drawn from those.

The first research question consisted of asking which data sources are the best in the context of predicting mood, while the second one asked which machine learning methods provide the best accuracy. Location, movement, sleep were the most important data sources that were used in the studies, which also provided the best accuracies and sensitivities for predicting mood and mood disorders. These are also easy to derive from GPS, accelerometers, and ambient light sensors, which explains their prevalence in an emerging field, where easy-to-access data will be used primarily. Data sources combined with support vector machines and random forest algorithms were the most commonly mixed in the selection of articles.

The last research question of three asked if methods of artificial intelligence and machine learning are needed in mood disorder prediction & diagnosis. The articles give varied angles to this topic. On one hand, the models and methods used in the studies have impressive accuracy and their use in clinical context can be established quickly and the time required for maintenance is low. They show the strengths of machine learning by its ability to personalize care and predict with a relatively low previous familiarity of a patient. On the other hand, while the accuracies and the results on their own are impressive, transferring the technology to real applications and further studies will be challenging for three things: one, industry standards will have to be set in order to smooth out the larger future studies. Two, data security and privacy will need to be addressed with proper seriousness. And finally, three,

larger studies will need to be done to find out whether smaller sample sizes are overfitting or not.

References

1. Aledavood, T. *et al.* Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. *JMIR Res. Protoc.* **6**, e110 (2017).
2. Jacobson, N. C. & Chung, Y. J. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors* **20**, 1–16 (2020).
3. Dogrucu, A. *et al.* Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. *Smart Health* **17**, (2020).
4. Haines, A. *et al.* Testing out suicide risk prediction algorithms using phone measurements with patients in acute mental health settings: a feasibility study. *JMIR mHealth and uHealth* (2020) doi:10.2196/15901.
5. Masud, M. T. *et al.* Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. *J. Biomed. Inform.* **103**, (2020).
6. Cho, C.-H. *et al.* Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: Prospective observational cohort study. *J. Med. Internet Res.* **21**, (2019).
7. Pratap, A. *et al.* The accuracy of passive phone sensors in predicting daily mood. *Depress. Anxiety* **36**, 72–81 (2019).
8. Gerych, W., Agu, E. & Rundensteiner, E. Classifying Depression in Imbalanced Datasets Using an

Autoencoder- Based Anomaly Detection Approach. in *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019* 124–127 (2019). doi:10.1109/ICOSC.2019.8665535.

9. Sarda, A., Munuswamy, S., Sarda, S. & Subramanian, V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: Cross-sectional observational study. *JMIR mHealth and uHealth* **7**, (2019).
10. Berrouiguet, S. *et al.* Combining mobile-health (mHealth) and artificial intelligence (AI) methods to avoid suicide attempts: The Smartcrises study protocol. *BMC Psychiatry* **19**, (2019).
11. Narziev, N. *et al.* STDD: Short-term depression detection with passive sensing. *Sensors* **20**, (2020).
12. Mehrotra, A. & Musolesi, M. Designing effective movement digital biomarkers for unobtrusive emotional state mobile monitoring. in *DigitalBiomarkers 2017 - Proceedings of the 1st Workshop on Digital Biomarkers, co-located with MobiSys 2017* 3–8 (2017). doi:10.1145/3089341.3089342.
13. Egilmez, B. *et al.* UStress: Understanding college student subjective stress using wrist-based passive sensing. in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017* 673–678 (2017). doi:10.1109/PERCOMW.2017.7917644.
14. Doryab, A., Min, J. K., Wiese, J., Zimmerman, J. & Hong, J. I. Detection of behavior change in people with depression. (2014).
15. Gosling, S. D. *et al.* Using Smartphones to Collect Behavioral Data in Psychological Science. *Perspect. Psychol. Sci.* **11**, 838–854 (2016).
16. Farhan, A. A. *et al.* Behavior vs. introspection: Refining prediction of clinical depression via smartphone sensing data. in *2016 IEEE Wireless Health, WH 2016* 30–37 (2016). doi:10.1109/WH.2016.7764553.

17. Jaques, N. *et al.* Predicting students' happiness from physiology, phone, mobility, and behavioral data. in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015* 222–228 (2015). doi:10.1109/ACII.2015.7344575.
18. Ware, S. *et al.* Predicting depressive symptoms using smartphone data. *Smart Health* **15**, (2020).
19. Hagerer, G., Cummins, N., Eyben, F. & Schuller, B. 'did you laugh enough today?' -deep neural networks for mobile andwearable laughter trackers. in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* vol. 2017-Augus 2044–2045 (2017).
20. Aledavood, T. *et al.* Smartphone-Based Tracking of Sleep in Depression, Anxiety, and Psychotic Disorders. *Curr. Psychiatry Rep.* **21**, 49 (2019).
21. Schwab, K. *The Fourth Industrial Revolution*. (Crown, 2017).
22. Shatte, A. B. R., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: a scoping review of methods and applications. *Psychol. Med.* **49**, 1426–1448 (2019).
23. Chen, M., Mao, S. & Liu, Y. Big Data: A Survey. *Mobile Networks and Applications* vol. 19 171–209 (2014).
24. Luo, J., Wu, M., Gopukumar, D. & Zhao, Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed. Inform. Insights* **8**, 1–10 (2016).
25. Organization, W. H. & Others. Mental health action plan 2013-2020. (2013).
26. Ritchie, H. & Roser, M. Mental Health. 2018.
27. Kessler, R. C. & Walters, E. The National Comorbidity Survey. *Textbook in Psychiatric Epidemiology* 343–362 doi:10.1002/0471234311.ch14.
28. Biss, R. K., Hasher, L. & Thomas, R. C. Positive mood is associated with the implicit use of distraction. *Motiv. Emot.* **34**, 73–77 (2010).
29. Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication.

Arch. Gen. Psychiatry **62**, 617–627 (2005).

30. Levis, B., Benedetti, A., Thombs, B. D. & DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* **365**, l1476 (2019).
31. Düking, P., Achtzehn, S., Holmberg, H.-C. & Sperlich, B. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. *Sensors* **18**, 1632 (2018).
32. Meeker, M. BOND Internet Trends 2019. *Teknisk rapport* (2019).
33. Chanchaichujit, J., Tan, A., Meng, F., Eaimkhong, S. & Others. Healthcare 4.0. *Springer Books* (2019).
34. Manning, C. & Schutze, H. *Foundations of Statistical Natural Language Processing*. (MIT Press, 1999).
35. Jiang, F. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* **2**, 230–243 (2017).
36. Maurer, D. M. Screening for depression. *Am. Fam. Physician* **85**, 139–144 (2012).
37. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* **4**, 1–32 (2008).
38. Feehan, L. M. *et al.* Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. *JMIR Mhealth Uhealth* **6**, e10527 (2018).
39. Kwon, J. Smart watch. *US Patent* (2016).
40. Walsh, C. G., Ribeiro, J. D. & Franklin, J. C. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clin. Psychol. Sci.* **5**, 457–469 (2017).
41. Mdhaffar, A. *et al. DL4DED: Deep Learning for Depressive Episode Detection on Mobile Devices*. vol. 11862 LNCS 109–121 (2019).

42. Mastoras, R.-E. *et al.* Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Sci. Rep.* **9**, (2019).
43. Helbich, M. Dynamic Urban environmental exposures on Depression and Suicide (NEEDS) in the Netherlands: A protocol for a cross-sectional smartphone tracking study and a longitudinal population register study. *BMJ Open* **9**, (2019).
44. Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M. & Lee, S. Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. in *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017* vol. 2017-Decem 138–140 (2017).
45. Sommers, M. *Bipolar Disorder and Manic Depressive Illness*. (Rosen Publishing Group, 2003).
46. Perez Arribas, I., Goodwin, G. M., Geddes, J. R., Lyons, T. & Saunders, K. E. A. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Transl. Psychiatry* **8**, (2018).
47. Tsuno, N., Besset, A. & Ritchie, K. Sleep and depression. *J. Clin. Psychiatry* **66**, 1254–1269 (2005).
48. Ho, T. K. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition* doi:10.1109/icdar.1995.598994.
49. Cohen, I. G., Glenn Cohen, I. & Mello, M. M. HIPAA and Protecting Health Information in the 21st Century. *JAMA* vol. 320 231 (2018).
50. Bennett, K., Bennett, A. J. & Griffiths, K. M. Security considerations for e-mental health interventions. *J. Med. Internet Res.* **12**, e61 (2010).
51. Onnela, J.-P. & Rauch, S. L. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* **41**, 1691–1696 (2016).
52. Geurts, P. Bias vs Variance Decomposition for Regression and Classification. in *Data Mining and*

Knowledge Discovery Handbook (eds. Maimon, O. & Rokach, L.) 733–746 (Springer US, 2010). doi:10.1007/978-0-387-09823-4_37.

53. Frost, J. The danger of overfitting regression models. *Minitab Blog* (2015).