

## Review Article

# Data Analytics in Mental Healthcare

**Ayesha Kamran Ul haq,<sup>1</sup> Amira Khattak,<sup>2</sup> Noreen Jamil ,<sup>1</sup> M. Asif Naeem ,<sup>1,3</sup> and Farhaan Mirza<sup>3</sup>**

<sup>1</sup>National University of Computer and Emerging Sciences, Islamabad, Pakistan

<sup>2</sup>Prince Sultan University, Riyadh, Saudi Arabia

<sup>3</sup>Auckland University of Technology, Auckland, New Zealand

Correspondence should be addressed to Noreen Jamil; [noreen.jamil@nu.edu.pk](mailto:noreen.jamil@nu.edu.pk)

Received 20 January 2020; Revised 13 March 2020; Accepted 12 June 2020; Published 4 July 2020

Academic Editor: Shaukat Ali

Copyright © 2020 Ayesha Kamran Ul haq et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Worldwide, about 700 million people are estimated to suffer from mental illnesses. In recent years, due to the extensive growth rate in mental disorders, it is essential to better understand the inadequate outcomes from mental health problems. Mental health research is challenging given the perceived limitations of ethical principles such as the protection of autonomy, consent, threat, and damage. In this survey, we aimed to investigate studies where big data approaches were used in mental illness and treatment. Firstly, different types of mental illness, for instance, bipolar disorder, depression, and personality disorders, are discussed. The effects of mental health on user's behavior such as suicide and drug addiction are highlighted. A description of the methodologies and tools is presented to predict the mental condition of the patient under the supervision of artificial intelligence and machine learning.

## 1. Introduction

Recently the term “big data” has become exceedingly popular all over the world.

Over the last few years, big data has started to set foot in healthcare system. In this context, scientists have been working on improving the public health strategies, medical research, and the care provided to patients by analyzing big datasets related to their health.

Data is coming from different sources like providers (pharmacy and patient's history) and nonproviders (cell phone and internet searches). One of the outstanding possibilities available from huge data utilization is evident inside the healthcare industry. Healthcare organizations have a big quantity of information available to them and a big portion of it is unstructured and clinically applicable. The use of big data is expected to grow in the medical field and it will continue to pose lucrative opportunities for solutions that can help in saving lives of patients. Big data needs to be interpreted correctly in order to predict future data so that

final result can be estimated. To solve this problem, researchers are working on AI algorithms that have a high impact on analysis of huge quantities of raw data and extract useful information from it. There are varieties of AI algorithms that are used to predict patient disease by observing past data. A variety of wearable sensors have been developed to deal with both physical and social interactions practically.

Mental health of a person is measured by a high grade of affective disorder which results in major depression and different anxiety disorders. There are many conditions which are recognized as mental disorders including anxiety disorder, depressive disorder, mood disorder, and personality disorder. There are lots of mobile apps, smart devices like smartwatches, and smart bands which increase healthcare facilities in mobile mental healthcare systems. Personalized psychiatry also plays an important role in predicting bipolar disorder and improving diagnosis and optimized treatment. Most of the smart techniques are not pursued due to lack of resources especially in underdeveloping countries. Like, in Pakistan, 0-1% of the government health budget is being

spent on the mental health system. There is a need for an affordable solution to detect depression in Pakistan so that everyone could be able to pay attention to it.

Researchers are working on many machine learning algorithms to analyze raw data to deduce meaningful information. It is now impossible to manage data in healthcare with traditional database management tools as data is in terabytes and petabytes now. In this survey, we analyzed different issues related to mental healthcare by usage of big data. We analyze different mental disorders like bipolar disease, opioid use disorder, personality disorder, different anxiety disorders, and depression. Social media is one of the biggest and most powerful resources for data collection as every 9 out of 10 people use social networking sites nowadays. Twitter is the main focus of interest for most researchers as people write 500,000 tweets on average per minute. Twitter is being used for sentimental analyses and opinion mining in the business field in order to check the popularity of a product by observing customer tweets. We have a lot of structure and unstructured data in order to reach any decision; data must be processed and stored in such a manner that follows the same structure. We analyzed and compared the working of different storage models under different conditions like mongo DB and Hadoop which are two different approaches to store large amounts of data. Hadoop works on cloud computing that helps to accomplish different operations on distributed data in a systematic manner.

In this survey we discuss the mental health problems with big data into further four sections. The second section describes related work regarding mental healthcare and the latest research on it. The third section describes different types of mental illness and their solutions within the data science. The fourth section describes the different illegal issues faced by the mental patients and early detection of these types of activities. The fifth section describes different approaches of data science towards mental healthcare systems such as different training and testing methods of health data for early prediction like supervised and unsupervised learning methods and artificial neural network (ANN).

## 2. Literature Review

There are a lot of mental disorders like bipolar one, depression, and different forms of anxieties. Bauer et al. [1] conducted a paper-based survey in which 1222 patients from 17 countries were participated to detect bipolar disorder in adults. This survey was translated into 12 different languages with some limitation that it did not contain any question about technology usage in older adults. According to Bauer et al. [1], digital treatment is not suitable for the older adults with bipolar disorder.

Researchers are working on the most interesting and unique method of tremendous interest to check the personality of a person just by looking at the way he or she is using the mobile phone. De Montjoye [2] collected dataset from US Research University and created a framework that analyzed phone call and text messages to check the personality of the user. Participants who did 300 calls or text per

year failed to complete personality measures. They choose optimal sample size that is 69 with mean age = 30.4, S. D. = 6.1, and 1 missing value. Similarly, Bleidorn and Hopwood [3] adopted a comprehensive machine learning approach to test the personality of the user using social media and digital records. Main 9 recommendations for how to amalgamate machine learning techniques provided by the researcher enhance the big five of the personality assessments. Focusing on minor details of the user comprehends and validates the result. Digital mental health has been revolutionized and its innovations are growing at a high rate. The National Health Service (NHS) has recognized its importance in mental healthcare and is looking for innovations to provide services at low cost. Hill et al. [4] presented a study of challenges and considerations in innovations in digital mental healthcare. They also suggested collaboration between clinicians, industry workers, and service users so that these challenges can be overcome and successful innovations of e-therapies and digital apps can be developed.

There are lots of mobile apps, smart devices like smartwatches, smart bands, and shirts which increase healthcare facilities in the mobile healthcare system. A variety of wearable sensors have been developing to deal with both physical and social interactions practically. Combining artificial intelligence with healthcare systems extends the healthcare facilities up to the next level. Dimitrov [5] conducted a systematic survey on mobile internet of things in the devices which allow business to emerge, spread productivity improvements, lock down the cost, and intensify customer experience and change in a positive way. Similarly, Monteith et al. [6] performed a paper-based survey on clinical data mining to analyze different data sources to get psychiatry data and optimized precedence opportunities for psychiatry.

One of the machine learning algorithms named artificial neural network (ANN) is based on three-layer architecture. Kellmeyer [7] introduced a way to secure big brain data from clinical and consumer-directed neurotechnological devices using ANN. But this model needs to be trained on a huge amount of data to get accurate results. Jiang et al. [8] designed and developed a wearable device with multisensing capabilities including audio sensing, behavior monitoring, and environment and physiological sensing that evaluated speech information and automatically deleted raw data. Tested students were split into two groups, those with excessive scores or in excessive score. Participants were required to wear the device to make sure of the authenticity of the data. But one of the major challenges to enable IoT in the device is safe communication.

Yang et al. [9] invented an IoT enabled wearable device for mental well-being and some external equipment to record speech data. This portable device would be able to recognize motion, pressure, monitoring, and physiological status of a person. There are lots of technologies that produce tracking data, such as smartphones, credit cards, websites, social media, and sensors offering benefits. Monteith and Glenn [10] elaborated some kind of generated data using human made algorithm, searching for disease symptoms, hit disease websites, sending/receiving healthcare e-mail, and

sharing health information on social media. Based on perceived data, the system predicted automated decision-making without the involvement of user to maintain security.

Considering all the above issues, there is a need for proper treatment of a disordered person. Mood of the patient is one of the parameters to detect his/her mental health. Public mood is hugely reflected in the social media as almost everyone uses social media in this modern era. Goyal [11] introduced a procedure in which tweets are filtered out for specific keywords from saved databases regarding food price crisis. Data is trained using two algorithms, K nearest neighbor and Naïve Bayes for unsupervised and supervised learning, respectively. Cloud storage is the best option to store huge amounts of unstructured data. Kumar and Bala [12] proposed functionalities of Hadoop for automatic processing and repository of big data. MongoDB is a big data tool for analyzing statistics related to world mental healthcare. Dhaka, P., and Johari [13] presented a way of implementation of big data tool 'MongoDB' for analyzing statistics related to world mental healthcare. The data is further analyzed using genetic algorithms for different mental disorders and deployed again in MongoDB for extracting final data.

But all of the above methods are useless without the user involvement. De Beurs et al. [14] introduced expert-driven method, intervention mapping, and scrum methods which may help to increase the involvement of the users. This approach tried to develop user-focused design strategies for the growth of web-based mental healthcare under finite resources. Turner et al. [15] elaborated in their article that the availability of the big data is increasing twice in size every two year for use in automated decision-making. Passos et al. [16] believed that the long-established connection between doctor and patient will change with the establishment of big data and machine learning models. ML algorithm can allow an affected person to observe his fitness from time to time and can tell the doctor about his current condition if it becomes worst. Early consultation with the doctor could prevent any bigger loss for the patient.

If the psychiatric disease is not predicted or handled earlier, then it enforces the patient to involve into many illegal activities like suicide as most of the suicide attempts are related to mental disorder. Kessler et al. [17] proposed meta-analysis that focused on suicide incidence within 1 year of the self-harm using machine learning algorithm. They analyzed the past reports of suicide patients and concluded that any prediction was impossible to be made due to short duration of psychiatric hospitalizations. Although a number of AI algorithms are used to estimate patient disease by observing past data, the focus of all studies was related to suicide prediction by setting up a threshold. Defining a threshold is a very crucial point or sometimes even impossible to be predicted. Cleland et al. [18] reviewed many studies but were unable to discover principles to clarify threshold. Authors used a random-effects model to generate a meta-analytic ROC. On the basis of correlation results, it is stated that depression prevalence is mediating factor between economic deprivation and antidepressant prescribing.

Another side effect of mental disease is drug addiction. Early drug prediction is possible by analyzing user data. Opioid is a swear type of drug. Hasan et al. [19] explored the Massachusetts All Payer Claim Data (MA APCD) dataset and examined how naïve users develop opioid use disorder. A popular machine learning algorithm is tested to predict the risk of such type of dependency of patent. Perdue et al. [20] predicted ratio of drug abusers by comparing Google trends data with monitoring the future (MTF) data; a well-structured study was made. It is concluded that Google trends and MTF data provided combined support for detecting drug abuse.

### 3. Mental Illness and Its Type

*3.1. Depression and Bipolar Disorder.* Bipolar disorder is also known as the worst form of depression. In Table 1, Bauer et al. [1] conducted a survey to check the bipolar disorder in adults. Data is collected from 187 older adults and 1021 younger adults with excluded missing observations. The survey contained 39 questions which took 20 minutes to complete. Older adults with bipolar disorder were addicted to the internet less regularly than the younger ones. As most of the healthcare services are available only online and most digital tools and devices are evolved, the survey has some limitations that it did not contain any question about technology usage in older adults. There is a need for proper treatment of a disordered person. Mood of the patient is one of the parameters to detect his/her mental health. Table 1 describes another approach of personality assessment using machine learning algorithm that focused on other aspects like systematic fulfillment and argued to enhance the validity of machine learning (ML) approach. Coming with technological advancement in the medical field will promote personalized treatments. A lot of work has been done in the field of depression detection using social networks.

The main goal of personalized psychiatry is to predict bipolar disorder and improve diagnosis and optimized treatment. To achieve these goals, it is necessary to combine the clinical variables of a patient as Figure 1 describes the integration of all these variables. It is now impossible to manage data in mental healthcare with database management traditional tools as data is in terabytes and petabytes now. So, there is a high need to introduce big data analytics tools and techniques to deal with such big data in order to improve the quality of treatment so that overall cost of treatment can be reduced throughout the world.

MongoDB is one of the tools to handle big data. The data is further analyzed using genetic algorithms for different mental disorders and deployed again in MongoDB for extracting final data. This approach of mining data and extracting useful information reduced overall cost of treatment. It provides the best results for clinical decisions. It helps doctors to give more accurate treatment for several mental disorders in less time and at low cost using useful information extracted by big data tool Mongo DB and genetic algorithm.

In Table 1, some of the techniques are handled and stored huge amount of data.

TABLE 1: Types of mental illness and role of big data.

Authors	Discipline(s) reviewed	Keywords used to identify papers for review	Methodology	Number of papers reviewed	Primary findings
Bauer et al. [1]	Bipolar disorder	Bipolar disorder, mental illness, and health literacy	Paper-based survey	68	47% of older adults used the internet versus 87% of younger adults having bipolar disorder
Dhaka and Johari [13]	Mental disorder	Mental health, disorders, and using MongoDB	Genetic algorithm and MongoDB tool	19	Analyzing and storing a large amount of data on MongoDB
Hill et al. [4]	Mental disorder	Mental health, collaborative computing, and e-therapies	(i) Online CBT platform (ii) Collaborative computing	33	(i) Developing smartphone application (ii) For mental disorder (iii) For improving e-therapies
Kumar and Bala, [12]	Depression detection through social media	Big data, Hadoop, sentiment analysis, social networks, and Twitter	Sentimental analysis and save data on Hadoop	14	Analyzing twitter users' view on a particular business product
Kellmeyer [7]	Big brain data	Brain data, neurotechnology, big data, privacy, security, and machine learning	(i) Machine learning (ii) Consumer-directed neurotechnological devices (iii) Combining expert with a bottom-up process	77	(i) Maximizing medical knowledge (ii) Enhancing the security of devices and sheltering the privacy of personal brain data
De Montjoye [2]	Mobile phone and user personality	Personality prediction, big data, big five personality prediction, Carrier's log, and CDR	(i) Entropy: detecting different categories (ii) Interevent time: frequency of call or text between two users (iii) AR coefficients: to convert list of call and text into time series	31	Analyzing phone calls and text messages under a five-factor model
Furnham [21]	Personality disorder	Dark side, big five, facet analysis, dependence, and dutifulness	Hogan 'dark side' measure (HDS) concept of dependent personality disorder (DPD)	34	All of the personality disorders are strongly negatively associated with agreeableness (a type of kind, sympathetic, and cooperative personality)
Bleidorn and Hopwood [3]	Personality assessment	Machine learning, personality assessment, big five, construct validation, and big data	(i) Machine learning (ii) Prediction models (iii) K-fold validation	65	Focusing on other aspects like systematic fulfillment and arguing to enhance the validity of machine learning (ML) approach

Using MongoDB tool, researchers are working to predict mental condition before severe mental stage. So, some devices introduced a complete detection process to tackle the present condition of the user by analyzing his/her daily life routine. There is a need for reasonable solutions that detect disable stage of a mental patient more precisely and quickly.

**3.2. Personality Disorder.** Dutifulness is a type of personality disorder in which patients are overstressed about the disease that is not actually much serious. People with this type of disorder tend to work hard to impress others. A survey was conducted to find the relationship between normal and dutifulness personalities. Other researchers are working on the most interesting and unique method of tremendous interest to

check the personality of a person just by looking at the way he or she is using the mobile phone. This approach provides cost-effective and questionnaire-free personality detection through mobile phone data that performs personality assessment without conducting any digital survey on social media. To perform all nine main aspects of the constructed validation in real time is not easy for the researchers. This examination, like several others, has limitations. This is just a sample that has implications for generalization when it is used in the near-real-time scenario which may be tough for the researchers.

#### 4. Effects of Mental Health on User Behavior

Mental illness is upswing in the feelings of helplessness, danger, fear, and sadness in the people. People do not



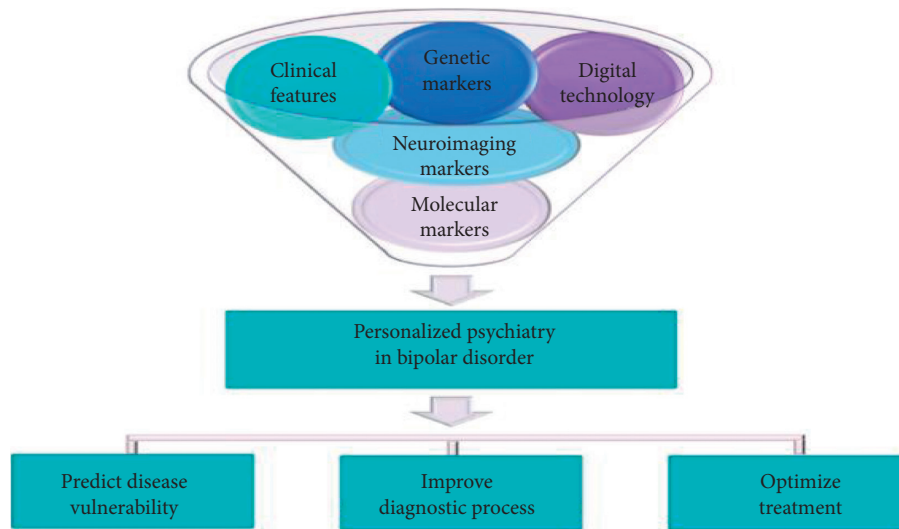


FIGURE 1: Goals of personalized treatment in bipolar disorder [22].

understand the current situation so this thing imposes psychiatric patients to illegal activities. Table 2 described some issues that appear because of mental disorder like suicide, drug abuse, and opioid use as follows.

**4.1. Suicide.** Suicide is very common in underdeveloped countries. According to researchers, someone dies because of suicide in every 40 seconds all over the world. There are some areas in the world where mental disorder and suicide statistics are relatively larger than other areas.

Psychiatrists say that 90% of people who died by suicide faced a mental disorder. Electronic medical records and big data generate suicide through machine learning algorithm. Machine learning algorithms can be used to predict suicides in depressed persons; it is hard to estimate how accurately it performs, but it may help a consultant for pretreating patients based on early prediction. Various studies depict the fact that there are a range of factors such as high level of antidepressant prescribing that caused such prevalence of illness. Some people started antidepressant medicine to overcome mental affliction. In Table 1, Cleland et al. [18] explored three main factors, i.e., economic deprivation, depression prevalence, and antidepressant prescribing and their correlations. Several statistical tools could be used like Jupyter Notebook, Pandas, NumPy, Matplotlib, Seaborn, and ipyleaflet for creation of pipeline. Correlations are analyzed using Pearson's correlation and  $p$  values. The analysis shows strong correlation between economic deprivation and antidepressant prescribing whereas it shows weak correlations between economic deprivation and depression prevalence.

**4.2. Drug Abuse.** People voluntarily take drugs but most of them are addicted to them in order to get rid of all their problems and feel relaxed. Adderall, Divinorum, Snus, synthetic marijuana, and bath salts are the novel drugs. Opioid is a category of drug that includes the illegitimate drug heroin. Hasan et al. [19] compared four machine learning

algorithms: logistic regression, random forest, decision tree, and gradient boosting to predict the risk of opioid use disorder. Random forest is one of the best methods of classification in machine learning algorithms. It is found that in such types of situations random forest models outperform the other three algorithms specially for determining the features. There is another approach to predict drug abusers using the search history of the user. Perdue et al. [20] predicted ratio of drug abusers by comparing Google trends data with monitoring the future (MTF) data; a well-structured study was made. It is concluded that Google trends and MTF data provided combined support for detecting drug abuse.

Google trends appear to be a particularly useful data source regarding novel drugs because Google is the first place where many users especially adults go for information on topics of which they are unfamiliar. Google tends not to predict heroin abuse; the reason may be that heroin is a relatively uniquely dangerous than other drugs. According to Granka [23], internet searches can be understood as behavioral measures of an individual's interest in an issue. Unfortunately, this technique was not going to be very convenient as drug abuse researchers are unable to predict drug abuse successfully because of sparse data.

## 5. How Data Science Helps to Predict Mental Illness?

Currently, there are numerous mobile clinical devices which are established in patients' personal body networks and medical devices. They receive and transmit massive amounts of heterogeneous fitness records to healthcare statistics structures for patient's evaluation. In this context, system learning and data mining strategies have become extremely crucial in many real-life problems. Many of those techniques were developed for health data processing and processing on cellular gadgets.

There is a lot of data in the world of medicine as data is coming from different sources like pharmacy and patient's

TABLE 2: Side effects of mental illness and their solution through data science.

Authors	Side effects of mental disorder	Tools/techniques	Primary findings
Kessler et al. [17]	Suicide and mental illness	Machine learning algorithm	Predicting suicide risk at hospitalization
Cleland et al. [18]	Antidepressant usage	Clustering analysis based on behavior and disease	Identifying the correlation between antidepressant usage and deprivation
Perdue et al. [20]	Drug abuse	(i) Google search history (ii) Monitoring the future (MTF)	Providing real time data that may allow us to predict drug abuse tendency and respond more quickly
Hasan et al. [19]	Opioid use disorder	(iii) Feature engineering (iv) Logistic regression (v) Random forest (vi) Gradient boosting	Suppressing the increasing rate of opioid addiction using machine learning algorithms

history and from nonproviders (cell phone and internet searches). Big data needs to be interpreted in order to predict future data, estimate hypothesis, and conclude results. Psychiatrists should be able to evaluate results from research studies and commercial analytical products that are based on big data.

**5.1. Artificial Intelligence and Big Data.** Big data collected from wearable tracking devices and electronic records help to store accumulating and extensive amounts of data. Smart mobile apps support fitness and health education, predict heart attack, and calculate ECG, emotion detection, symptom tracking, and disease management. Mobile apps can improve connection between patients and doctors. Once a patient's data from different resources is organized into a proper structure, artificial intelligence (AI) algorithm can be used. After all, AI recognizes patterns, finds similarity between them, and makes predictive recommendations about what happened with those in that condition.

Techniques used for healthcare data processing can be widely categorized into two classes: nonartificial intelligence systems and artificial intelligence systems. Although non-AI techniques are less complex, but they are suffering from a lack of convergence that gives inaccurate results as compared to AI techniques. Contrary to that, AI methods are preferable than non-AI techniques. In Table 3, Dimitrov [5] combined artificial intelligence with IoT technology in existing healthcare apps so that connection between doctors and patients remains balanced. Disease prediction is also possible through machine learning. Figure 2 shows hierarchical structure of AI, ML, and neural networks.

One of the machine learning algorithms named artificial neural network (ANN) is based on three-layer architecture. Kellmeyer [7] introduced a way to secure big brain data from neurotechnological devices using ANN. This algorithm was working on a huge amount of data (train data) to predict accurate results. But patients' brain diseases are rare so training models on small data may produce imprecise results. Machine learning models are data hungry. To obtain accurate results as an output, there is a need of training more data with distinct features as an input. These new methods cannot be applicable on clinical data due to the limited economy resources.

**5.2. Prediction through Smart Devices.** Various monitoring wearable devices (Table 3) are available that continuously capture the finer details of behaviors and provide important cues about fear and autism. This information is helpful to recognize mental issues of the user of those devices. Victims were monitored continuously for a month. High level computation performed on the voice requires high complexity data as well as high computational power which leads to a huge pressure on the small chip. In order to overcome power issues, relatively low frequency was chosen.

Yang et al. [9] invented an audio well-being device and conducted a survey in which participants have to speak more than 10 minutes in a quiet room. The first step is to choose the validity of the sample by completing some questions (including STAI, NEO-FFI, and AQ) to the participants. In order to determine whether they are suitable for the experiment or not, a test was conducted based on an AQ question. There was a classification algorithm applied on the AQ data. This type of device has one advantage; it perfectly worked on long-term data instead of low-term one but they used offline data transfer instead of real time.

Although it has different sensors, adding up garbage data to the sensors is a very obvious thing. This is an application that offers on-hand record management using mobile/tablet technology once security and privacy are confirmed. To increase the reliability of IoT devices, there is a need to increase the sample size with different age groups in real time environment to check the validity of the experiment.

There are a lot of technologies that effectuate tracking data like smartphones, credit cards, social media, and sensors. This paper discussed some of the existing work to tackle such data. In Table 3, one of the approaches is human made algorithm; searching for disease symptoms hits disease websites, sending/receiving healthcare e-mail, and sharing health information on social media through this kind of data. These are some examples of activities that perform key rules to produce medical data.

**5.3. Role of Social Media to Predict Mental Illness.** Constant mood of the patient is one of the parameters to detect his/her mental health. According to Lenhart, A. et al. [25] studied almost four out of five internet users of social

TABLE 3: Data analytics and predicting mental health.

Authors	Tool/technology	Methodology	Purpose of finding	Strength	Weakness
Dimitrov [5]	(i) Sensing technology (ii) Artificial intelligence	Emergence of medical internet of things (mIoT) in existing mobile apps	Providing benefits to the customers (i) Avoiding chronic and diet-related illness (ii) improving cognitive function	(i) Achieving improved mental health (ii) improving lifestyles in real time decision-making	Adding up garbage data to the sensors
Monteith et al. [6]	Survey based approach	Clinical data mining	Analyzing different data sources to get psychiatry data	Optimized precedence opportunities for psychiatry	N/A
Kellmeyer [7]	Neurotechnology	(i) Machine learning (ii) Consumer-directed neurotechnological devices (iii) Combining expert with a bottom-up process	Enhancing the security of devices and sheltering the privacy of personal brain	Maximizing medical knowledge	Model needs huge amount of training data as brain disease is rarely captured
Yang et al. [9]	Long-term monitoring wearable device with internet of things	Well-being questionnaires with a group of students	Developing app-based devices linked to android phones and servers for data visualization monitoring and environment sensing	Perfectly working on long-term data	Offline data transfer instead of real time
Monteith and Glenn [10]	Automated decision-making	Hybrid algorithm that combines the statistical focus and data mining	Tracking day-to-day behavior of the user by automatic decision-making	Automatically detecting human decision without any input	How to ignore irrelevant information is a key headache
De Beurs et al. [14]	Online intervention	Expert-driven method Intervention mapping Scrum	Increasing user involvement under limited resources	Standardizing the level of user involvement in the web-based healthcare system	Deciding threshold for user involvement is problematic
Kumar and Bala [12]	Hadoop	Doing sentimental analysis and saving data on Hadoop	Analyzing twitter users' view on a particular business product	Checking out popularity of a particular service	Usage of two programming languages needs experts
Goyal [11]	KNN and Naïve Bayes classifier	Text mining and hybrid approach combining KNN and Naïve Bayes	Opinion mining of tweets related to food price crisis	Cost-effective way to predict prizes	Data needs to be cleaned before training

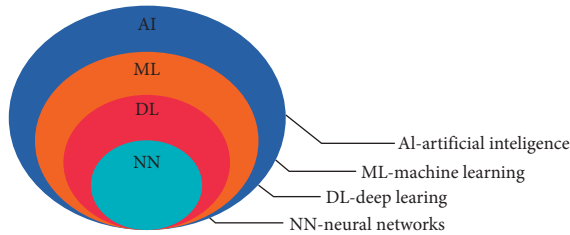


FIGURE 2: AI and ML [24].

media. In Table 3, researchers used twitter data to get online user review that helps the seeker to check out popularity of a particular service or purchase a product. In order to collect opinion of people on Airtel, they did analysis on it. Filter of the keyword is done using *Filter by content* and *Filter by location*. First of all, special character, URL, spam, and short

words are removed from the tweets. Secondly, remaining words from the tweets are then tokenized and TF-IDF score is calculated for all the keywords. After cleaning of data, classification algorithm named K nearest neighbor and Naïve Bayes algorithm were applied on the text in order to extract feature. Location filters work on specific bounding filter. Although hybrid recommendation system is providing 76.31% accuracy of the result, then Naïve Bayes is 66.66%. At the end, automated system is designed for opinion mining.

There is another point of consideration that Tweeter has unstructured data so handling such a huge amount of unstructured data is a tedious task to take up. Due to lack of schema structure, it is difficult to handle and store unstructured data. There is a need for storage devices to store an insignificant amount of data for processing. Cloud storage is the best option for such a material. The entire program is designed in Python so that it could be able to

catch all possible outcomes. Hadoop works on cloud computing that helps to accomplish different operations on distributed data in a systematic manner. Success rate of the above approach was around 70% but authors have done these tasks using two programming languages. Python code for extraction tweets and Java is used to train the data which required expert programmers on each language. It will help doctors to give more accurate treatment for several mental disorders in less time and at low cost. Infusing this approach provides predetection of depression that may preserve the patient to face the worst stage of mental illness.

#### 5.4. Key Challenges to Big Data Approach

- (i) Big data has many ethical issues related to privacy, reusability without permission, and involvement of the rival organization.
- (ii) To work in diverse areas, big data requires collaboration with expert people in the relative field including physicians, biologists, and developers that is crucial part of it. Data mining algorithms can be used to observe or predict data more precisely than traditional clinical trials.
- (iii) People may feel hesitant to describe all things to the doctors. One of the solutions to estimate the bad mental illness before time is automated decision-making without human input as shown in Table 3. It collects data from our behavior that is unsophisticated to the digital economy. Key role of digital providence must be inferred in order to understand the difficulties that technology may be responsible for people with mental illness.
- (iv) There are many security issues while discussing sensitive information online as data may be revealed so a new approach to provide privacy protections as well as decision-making from the big data through new technologies needs to be introduced.
- (v) Also, if online data is used to predict user personality, then keeping data secured and protected from hacker is a big challenge. A lot of cheap solutions exist but they are not reliable from a user's perspective especially.
- (vi) Major challenges for enabling IoT in the device is communication; all of the above methods are useless without the user involvement. User is one of the main parts of the experiment especially if the user's personal or live data is required. Although many web-based inventions related to mental health are being released, the actual problem of active participation by end users is limited. In Table 3, an expert-driven method is introduced that is based on intervention mapping and scrum methods. It may help to increase the involvement of the users. But if all the users are actively involved in the web-based healthcare system, then it becomes problematic.
- (vii) When deciding on the level of user involvement, there is a need to decide about user input with the

accessibility of resources. It required an active role of technological companies and efficient time consumption. Further research should provide direction on how to select the best and optimized user-focused design strategies for the development of web-based mental health under limited resources.

## 6. Conclusions

Big data are being used for mental health research in many parts of the world and for many different purposes. Data science is a rapidly evolving field that offers many valuable applications to mental health research, examples of which we have outlined in this perspective.

We discussed different types of mental disorders and their reasonable, affordable, and possible solution to enhance the mental healthcare facilities. Currently, the digital mental health revolution is amplifying beyond the pace of scientific evaluation and it is very clear that clinical communities need to catch up. Various smart healthcare systems and devices developed that reduce the death rate of mental patients and avert the patient to associate in any illegal activities by early prediction.

This paper examines different prediction methods. Various machine learning algorithms are popular to train data in order to predict future data. Random forest model, Naïve Bayes, and k-mean clustering are popular ML algorithms. Social media is one of the best sources of data gathering as the mood of the user also reveals his/her psychological behavior. In this survey, various advances in data science and its impact on the smart healthcare system are points of consideration. It is concluded that there is a need for a cost-effective way to predict intellectual condition instead of grabbing costly devices. Twitter data is utilized for the saved and live tweets accessible through application program interface (API). In the future, connecting twitter API with python, then applying sentimental analysis on 'posts,' 'liked pages,' 'followed pages,' and 'comments' of the twitter user will provide a cost-effective way to detect depression for target patients.

## Data Availability

The authors will provide the data used for the experiments, if requested.

## Conflicts of Interest

There are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors are thankful to Prince Sultan University for the financial support towards the publication of this paper.

## References

- [1] R. Bauer, T. Glenn, S. Strejilevich et al., "Internet use by older adults with bipolar disorder: international survey results,"



- International Journal of Bipolar Disorders*, vol. 6, no. 1, p. 20, 2018.
- [2] Y.-A. De Montjoye, J. Quoidbach, F. Robic, and A. Pentland, "Predicting personality using novel mobile phone-based metrics," in *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pp. 48–55, Berlin, Heidelberg, April 2013.
  - [3] W. Bauer and C. J. Hopwood, "Using machine learning to advance personality assessment and theory," *Personality and Social Psychology Review*, vol. 23, no. 2, pp. 190–203, 2019.
  - [4] C. Hill, J. L. Martin, S. Thomson, N. Scott-Ram, H. Penfold, and C. Creswell, "Navigating the challenges of digital health innovation: considerations and solutions in developing online and smartphone-application-based interventions for mental health disorders," *British Journal of Psychiatry*, vol. 211, no. 2, pp. 65–69, 2017.
  - [5] D. V. Dimitrov, "Medical internet of things and big data in healthcare," *Healthcare Informatics Research*, vol. 22, no. 3, pp. 156–163, 2016.
  - [6] S. Monteith, T. Glenn, J. Geddes, and M. Bauer, "Big data are coming to psychiatry: a general introduction," *International Journal of Bipolar Disorders*, vol. 3, no. 1, p. 21, 2015.
  - [7] P. Kellmeyer, "Big brain data: on the responsible use of brain data from clinical and consumer-directed neurotechnological devices," *Neuroethics*, vol. 11, pp. 1–16, 2018.
  - [8] L. Jiang, B. Gao, J. Gu et al., "Wearable long-term social sensing for mental wellbeing," *IEEE Sensors Journal*, vol. 19, no. 19, 2019.
  - [9] S. Yang, B. Gao, L. Jiang et al., "IoT structured long-term wearable social sensing for mental wellbeing," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3652–3662, 2018.
  - [10] S. Monteith and T. Glenn, "Automated decision-making and big data: concerns for people with mental illness," *Current Psychiatry Reports*, vol. 18, no. 12, p. 112, 2016.
  - [11] S. Goyal, "Sentimental analysis of twitter data using text mining and hybrid classification approach," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 2, no. 5, pp. 2454–132X, 2016.
  - [12] M. Kumar and A. Bala, "Analyzing twitter sentiments through big data," in *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 2628–2631, New Delhi, India, March 2016.
  - [13] P. Dhaka and R. Johari, "Big data application: study and archival of mental health data, using MongoDB," in *Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 3228–3232, Chennai, India, March 2016.
  - [14] D. De Beurs, I. Van Bruinessen, J. Noordman, R. Friele, and S. Van Dulmen, "Active involvement of end users when developing web-based mental health interventions," *Frontiers in Psychiatry*, vol. 8, p. 72, 2017.
  - [15] V. Turner, J. F. Gantz, D. Reinsel, and S. Minton, "The digital universe of opportunities: rich data and the increasing value of the internet of things," *IDC Analyze the Future*, vol. 16, 2014.
  - [16] I. C. Passos, P. Ballester, J. V. Pinto, B. Mwangi, and F. Kapczynski, "Big data and machine learning meet the health sciences," in *Personalized Psychiatry*, pp. 1–13, Springer, Cham, Switzerland, 2019.
  - [17] R. C. Kessler, S. L. Bernecker, R. M. Bossarte et al., "The role of big data analytics in predicting suicide," in *Personalized Psychiatry*, pp. 77–98, Springer, Cham, Switzerland, 2019.
  - [18] B. Cleland, J. Wallace, R. Bond et al., "Insights into antidepressant prescribing using open health data," *Big Data Research*, vol. 12, pp. 41–48, 2018.
  - [19] Hasan M. M., M. Noor-E-Alam, Patel M. R., Modestino A. S., Young G. Sanchez L. D., A Novel Big Data Analytics Framework to Predict the Risk of Opioid Use Disorder. 2019.
  - [20] R. T. Perdue, J. Hawdon, and K. M. Thames, "Can big data predict the rise of novel drug abuse?" *Journal of Drug Issues*, vol. 48, no. 4, pp. 508–518, 2018.
  - [21] A. Furnham, "A big five facet analysis of sub-clinical dependent personality disorder (dutifulness)," *Psychiatry Research*, vol. 270, pp. 622–626, 2018.
  - [22] E. Salagre, E. Vieta, and I. Grande, "Personalized treatment in bipolar disorder," in *Personalized Psychiatry*, pp. 423–436, Academic Press, Cambridge, MA, USA, 2020.
  - [23] L. Granka, "Inferring the public agenda from implicit query data," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, MA, USA, July 2009.
  - [24] V. Sinha, 2019, <https://www.quora.com/What-are-the-main-differences-between-artificial-intelligence-and-machine-learning-Is-machine-learning-a-part-of-artificial-intelligence>.
  - [25] Lenhart A., Purcell K., Smith A., Zickuhr K., Social media & mobile internet use among teens and young adults. Millennials, Pew Internet & American Life Project, Washington, DC, USA, 2010.