Functional and Process Model on Big Data, Machine Learning, and Digital Phenotyping in Clinical Psychiatry

Arunkumar Annamalai[1] iD

Advances in information technology and the advent of big data (BD) and automated evaluation have changed how we acquire, analyze, and present data in the clinical care continuum.[1] The use of technology is increasingly seen as a solution for addressing the skewed doctor–patient ratio in the field of mental health (MH), albeit with a number of challenges to be addressed.[2] Subsequently, the role of data science and technology in psychiatric practice has been actively explored in recent years.[3] Machine learning (ML) techniques use complex multimodal data in psychiatry for clinical decision-making, otherwise unexplored by psychiatrists.[4,5] This technology deployment is shifting the clinical psychiatry practice from theory- and experience-driven to be data- and technology-driven.[6–8] The important aspect of this technology-enabled clinical psychiatry is the generation of BD and its analysis and presentation for clinical decision-making.[9,10] This paradigm shift, if not accompanied by a corresponding increase in the skills of the psychiatrists

to handle new technology, will nullify any advance in the technology.[11]

However, BD in psychiatry has been published with less straightforward

implications for psychiatrists dealing with the patients directly. BD in MH has been predominantly discussed from a computational and engineering perspective.[12] There are no reviews that broaden the knowledge landscape in clinical psychiatry that underpin the basic science of artificial intelligence (AI) for MH, psychiatry epidemiology, and MH policymaking.

This article aims to create an easy-to-understand functional and process model for psychiatrists with little to no knowledge about BD, ML, and AI. The best way to realize this aim is to synthesize knowledge from the existing literature and personal experiences in dealing with clinical data. Therefore, this article was written to provide familiarity with upcoming concepts and trends in BD, AI, and their potential applications for psychiatrists, without delving into the technical jargon. This viewpoint article will also help psychiatrists collaborate with technology counterparts to translate these ideas and concepts for effective clinical applications. Discussion on complex techniques and technology has been

[1] Clinical Data Science and AI, Deep Medicine Labs, Chennai, Tamil Nadu, India.

Address for correspondence: Arunkumar Annamalai, Deep Medicine Labs, #6/3, Annai Mary Third Street, Kamarajapuram, Anakaputhur, Chennai, Tamil Nadu 600070, India. E-mail: drarundatascientist@gmail.com

intentionally avoided. Rather, emphasis has been laid upon the actual applications in MH and psychiatry.

Subjects and Methods

Rationale for Choosing the Methodology

Considering the aim of the paper, it was necessary to summarize the existing literature on complex questions using mixed-method reviews.[13–15] It is impossible to do a meta-analysis as different studies have used different types of data and ML techniques to address a variety of MH disorders. A systematic review would help gain ideas and concepts but would still lack the knowledge synthesis that can be understood by psychiatrists who have no background in data science or ML. Therefore, a meta-narrative review approach has been used to provide direct, actionable insights to enable the translation of ideas from bench to bedside. It will also help them collaborate with experts/teams of other disciplines. This approach will help offset the methodological weaknesses of various studies and create a comprehensive knowledge of the concepts being studied.[16]

Search Strategy and Study Selection

To meet the study's objectives, several research questions were framed in December 2019 (Box 1). Subsequently, the search strategy (**Table 1**) and study selection criteria were predefined. The criterion for the study selection was that the information

BOX 1.
Research Questions

1. What is BD in MH and psychiatry?
2. What are the different sources, types, varieties, and methods of acquiring BD in psychiatry?
3. What are the characteristics of this data?
4. How to leverage the potential of BD?
5. How to analyze this data?
6. What is the role of ML and AI in analyzing this data?
7. How can a clinical psychiatrist use BD and related technology for delivering primary care?
8. What are the various applications of AI and BD in clinical care?

TABLE 1.
Search Strategy

Name of Database	Type of Database	Search Fields	Search Terms
PsycINFO	Health care	Abstract, title, and keywords	Big data and health care
PubMed			Big data and mental health
Cochrane			Big data and psychiatry
IEEE Xplore	Information technology		Big data and machine learning
AIM Digital Library			Artificial intelligence and psychiatry
Springer	Both		Artificial intelligence and mental health
Scopus			Data science and psychiatry
ScienceDirect			Data science and mental health

from the studies should have a direct or indirect impact on clinical decision-making. A systematic review was done between January 2020 and June 2021, using the predefined criteria. Articles published in English and peer-reviewed from 2010 to 2021 were selected for review. A title search was done to find the correct papers. Then those screened in the title search were selected for reading the abstracts. Those that fulfilled the eligibility criteria were selected for reading the entire article. The papers were then studied to extract ideas. The articles provide insight on BD in psychiatry, ML in psychiatry, and applications of ML/AI in MH (according to the World Health Organization definition of MH).

Knowledge Extraction, Analysis, and Synthesis

Initially, three broad themes were framed (data in MH/psychiatry, analysis, and applications in MH/psychiatry). Full papers were read thoroughly, and knowledge was extracted and classified under these themes. Three guiding questions in the analysis of the information extracted were; (a) Is this information relevant for clinical decision making (directly or indirectly)? (b) What is the straightforward implication of this knowledge in various applications of AI/ML in MH and psychiatry, and how do different components interact with each other in the medical-technology spectrum? (c) How can a clinician use this information to contribute to AI/ML in MH and psychiatry?

Within each theme, the content was explored using focused coding and the constant comparative method. Codes and categories were identified. Irrelevant codes were then excluded from the analysis.[17] The information was analyzed using meta-ethnography techniques, and an interpretative qualitative synthesis was done using the best-fit framework synthesis approach.[18] **Figure 1** shows the summary of the method.[19] Subsequently, functional and process models from a clinical psychiatry perspective were developed. Realist and meta-narrative evidence syntheses (RAMESES) method for meta-narrative reviews has been followed in reporting this study.

Results of the Review

Analysis of the content revealed codes and categories summarized in **Table 2**. The findings and salient features of the functional and process model are discussed in detail under discussion.

Discussion

Some reviews and studies dealt with the technical aspects of AI in psychiatry and MH. The studies detailed the design of the algorithm, the architecture required to implement them, and the function of the algorithm.[20,21] However, when a psychiatrist tries to understand this concept, it becomes overwhelming as these concepts are written from a nonclinical perspective.[22] The need for describing these concepts from a clinical perspective is essential to ensure that psychiatrists can collaborate with technology experts in using AI to solve common problems in clinical practice. Additionally, this paper will serve as a starting point for psychiatrists trying to get an overview of data science and AI in MH without a technical conundrum.

Therefore, the discussion section has been divided into two parts.

Part A: A functional model that discusses the individual components of the

TABLE 2.

Themes, Categories, and Nodes from the Analysis

Themes	Data in Mental Health/Psychiatry			Analysis				Applications in Mental Health/Psychiatry			
Categories	Data Acquisition	Data management		ML Techniques	DPMH	CDSS		Detection and Diagnosis	Prognosis, Treatment, and Support	Public Health Applications	Research and Clinical Administration
Nodes	Sources and methods	Storage	Data Retrieval	Supervised	Affect analytics	Basic /Advanced	Presentation of Findings	Screening	Personalize treatment	Surveillance	Improve resource allocation
		Classification		Semi-supervised	Cognitive analytics			Diagnosis	Ensure adherence to treatment	Clinical Epidemiology	Improve research methods
		Verification		Unsupervised	Behavioral analytics			Risk Prediction	Detect Behavioral changes		Use existing secondary data for research
	Types	Validation			Social analytics				Predict long term outcome		
	Characteristics	Update			Biomarker analytics						
	Barriers to acquisition										

CDSS, clinical decision support system; DPMH, digital phenotyping in mental health.

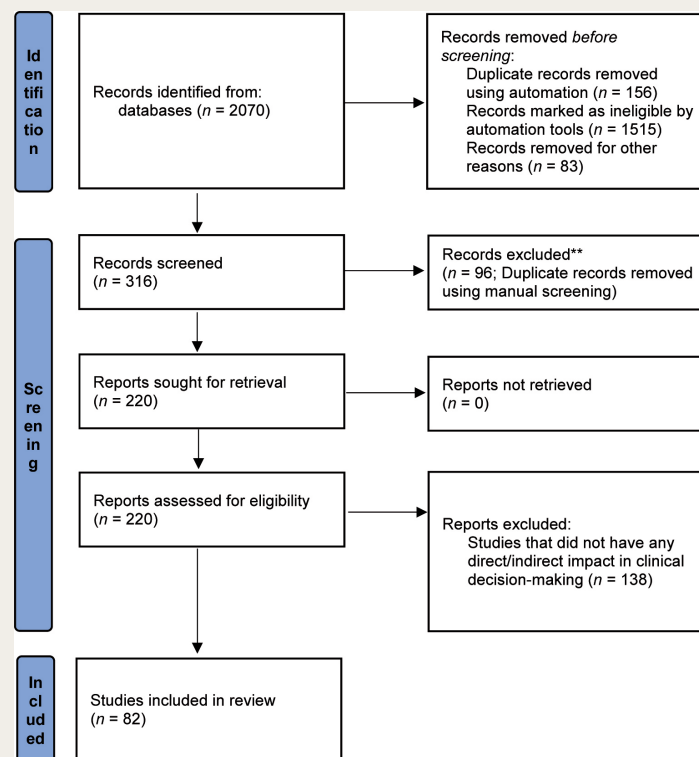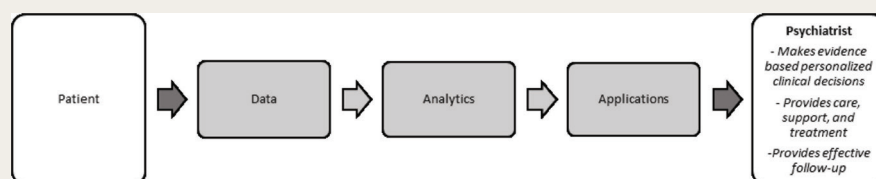**FIGURE 1.**

Summary of the Research Process

FIGURE 2.

Functional Model From a Clinical Psychiatry Perspective

entire bench-to-bedside pathway of data science and AI in psychiatry

Part B: A process model on how these components interact and what is expected out of the psychiatrist, the essential knowledge required, and how to enable meaningful innovation in psychiatry

Part A: Functional Model

The functional model was derived from breaking the complex concepts into an easy-to-understand framework for direct clinical implications in psychiatry. Therefore, only insights that are relevant to clinical psychiatry have been discussed. Hence, concepts have been populated and discussed without definitions or detailed descriptions. Most of the insights apply to the entire health care sector, although specific implications in psychiatry have been discussed. **Figure 2** shows the functional model from a clinical psychiatry perspective. The ultimate beneficiary is the patient, who is also the model's starting point.

This section is further organized under the major themes of the review; data in MH/psychiatry, analysis, and applications in MH/psychiatry.

Data in Mental Health/Psychiatry

The major headings in this subsection are data acquisition (where, what, and how data is collected), data management

TABLE 3.
Sources and Methods of Data Collection (Traditional Perspective)

Source Where It Is Generated (Traditional Approach)	Names of the Sources
Patient	Case notes, electronic medical records (EMR), electronic health records (EHR), discharge notes, pharmacy, genomics, social media, imaging, claims, lab/biomarkers, clinical trials.
Physician	Prescriptions, research experience, survey data, reference data.
Institutions	Claims, sales consumption, research experience, reference data.
Countries	Registries

TABLE 4.
Sources and Methods of Data Collection (Big Data Perspective)

Based on the Source of Collection (Big Data Approach)	Description	Objective
Biomedical data	Data from hospitals and scientific communities, research and development data from pharmaceutical companies, electronic medical records (EMR), electronic health records (EHR), and clinical trials data.	To develop tools for personalized medicine
Internet	Navigation history, search history, and shopping history.	To understand real-time trends in health care
Social networks	Social media like Twitter, Facebook, LinkedIn, etc.	To analyze the impact of disease To under people's perception
Wearables	Data from smart watches, medical devices	To understand human mobility and interaction patterns
Mobile phones	Embedded sensors like GPS, accelerometer, gyroscope, camera, microphone, apps.	To understand human mobility and interaction patterns

TABLE 5.
Characteristics of BD in Mental Health/Psychiatry

Domain-specific/General	Characteristics	Description
Specific to clinical BD in psychiatry	Data Fragmentation	Data is stored in different formats in different places
	Unstructured	The data doesn't have a specific form
	Lack of interoperability	Data from different sources can't be integrated
	Lack of standardization	There is no uniform way of collecting data
Applicable to all BD in health care	Volume	Data explosion because of exponential growth of data that is generated each day
	Variety	Data is generated in different types from different sources
	Velocity	Continuous and massive inflow of real-time data
	Veracity	Data has a lot of noise, abnormality, and biases
	Validity	Some data is irrelevant in most the cases
	Volatility	All data cannot be perpetually stored and managed

BD: big data.

(when, where, and how data is kept), and data retrieval.

Data Acquisition[23–25]

Data acquisition can be understood under four common headings:

Source and Methods of Data Collection

It is essential to understand how and from where the data has been generated. There are two ways to view this. The traditional way of seeing this is to identify who generates this data (**Table 3**). The other way is the big data perspective based on how and where this data is collected from (**Table 4**). There will be considerable overlap between these two approaches and between categories within these approaches.

Types of Clinical Data in Mental Health

The data is any of the following types: tabular data, images, voices, speech, video, text, and time series.

Characteristics of Data in Psychiatry/ MH

Please see **Table 5**.

Barriers to Efficient Data Acquisition in Clinical Psychiatry

The most common barriers to data acquisition are[25]:
1. Technical barriers (lack of interoperability).
2. Political barriers (lack of effective data-sharing policies).
3. Ethical barriers (lack of consensus on privacy and confidentiality).
4. Administrative barriers (inadequate human resources to handle clinical data).

Data Management and Retrieval

This is purely a nonclinical aspect of BD, consisting of verification, validation, storage, classification, update, standardization, and review. The process of extracting data from a database for analysis or visualization is called retrieval. This is beyond the scope of this paper to discuss in detail. However, psychiatrists need to be familiar with the vocabulary, terminologies, and various data standards in health care.[26]

BD Analysis in Psychiatry[27]

Machine Learning Techniques (ML Techniques)

ML techniques can be supervised, semi-supervised, or unsupervised. There are many reviews that detail ML techniques in MH. An important concept to remember is that ML techniques are a way to use computers to analyze BD. This vast domain is not for psychiatrists, although basic knowledge is necessary to collaborate with ML engineers. Each technique has a unique purpose and function. The choice of ML technique depends on the need, available data, and available resources. At any point, the choice of an ML technique is always taken by the nonclinical counterpart; therefore, a general understanding of the concept is enough.

Digital Phenotyping of MH (DPMH)[28,29]

With such voluminous data and multiple correlations with different variables that can be spurious at times, it is essential to identify those data points that are clinically significant and serve as a surrogate marker for any underlying psychiatric disorder. Also, it is essential to find variables that represent many other variables or variables with higher weightage so that the less significant variables can be ignored from the overall decision-making model. The process of extracting clinically significant digital biomarkers that provide meaningful insight for decision-making in psychiatry is called as digital phenotyping of MH (DPMH). It is of the following types;

1. Affect recognition (emotion recognition and sentiment analysis)
2. Cognitive analytics (acoustic analysis, linguistic analysis, and emotion processing)
3. Behavioral anomaly detection (sleep monitoring, substance abuse, suicide, and eating disorder)
4. Social analytics (social influence, social dynamics, and social participation)
5. Biomarker analytics (genome and neurological imaging)

Clinical Decision Support Systems (CDSS)[30]

Clinical decision support systems (CDSS) are computerized and noncomputerized tools and interventions that aid clinical decision-making for care, support, and treatment. Depending on the complexity of the problem to which support is provided, CDSS can be basic (simple alerts for taking drugs, prompts for better data collection in health management information systems (HMIS), etc.) or advanced (checks for drug-disease interactions, providing treatment recommendations, etc.). Basic CDSS is part of the regular clinical care continuum; however, advanced CDSS is still a theory.

Applications in Mental Health/Psychiatry[31]

Data science and AI in MH/psychiatry can be used for the following:

1. Detection and diagnosis of clinical conditions (*screening, diagnosis, and risk prediction*).
2. Prognosis, treatment, and support (*personalizing treatment, ensuring adherence to treatment, detecting behavioral changes, and predicting long-term outcome*).
3. Public health applications (*surveillance and clinical epidemiology*).
4. Research and clinical administration (*improving resource allocation and research methods, using existing secondary data for research*).

Part B: A Process Model with Implications in Clinical Psychiatry

The challenge with translational research is that the knowledge of the application of concepts is different from the actual understanding of the concepts. This section is drafted considering that psychiatrists have little to no understanding of data science and AI concepts. However, this should not limit them in providing meaningful contributions to technology experts who require psychiatrists' input and validation. Therefore, this section deals with a simple process model that any psychiatrist is likely to encounter

FIGURE 3.

Process Model From a Clinical Psychiatry Perspective

when dealing with data science and AI in MH. The process can be broadly divided into three phases (**Figure 3**).

Ideation

This is the most important stage in any innovation or invention. Technology experts lack this insight and rely on psychiatrists to look for gaps where technology can play an efficient role. There is a need to identify a problem that needs a solution and define the problem statement in terms that technology experts can understand. Also, it is essential to determine if an AI-based solution is necessary or will it be another tool that will soon be forgotten and lost.

Design

The design phase is purely for the technology experts to work on, although the choice of tools for data collection, sources, and methods are largely determined by the psychiatrist in line with the clinical practice requirements. This phase is where active collaboration takes place between various stakeholders. Hence, holistic knowledge of various stakeholders and activities is necessary.

Execution

This phase is important for a psychiatrist only from the perspective of verification and clinical validation of the technology and ensuring that every procedure is in accordance with good clinical practices.

Table 6 summarizes the process model with examples, along with the role and essential skills for a psychiatrist.[32]

Limitations of the Study

The paper is limited by a single author perspective. However, it is challenging to find psychiatrists who can understand the jargon in engineering and simplify the article. Therefore, a single author perspective is good to start with, considering the nascent

TABLE 6.

Role of a Psychiatrist and Essential Skills and Competencies

Stage	Step	Role of a Psychiatrist	Essential Skills/Competencies	Example (Build an Automated Data Collection)
Ideation	Defining the problem	Provide evidence that an AI-based solution is the best way forward	Clear communication skills	Psychiatrists spend time doing clerical work that hinders the time spent with the patients. Patients are unable to provide an accurate history of their illness.
	Exploring the need for AI-based solutions and also checking for alternate solutions		Brainstorming and making efficient decisions on the type of solution to focus on. Conduct scoping/systematic reviews	Can an AI-based app be built that can collect automated information from the user?
Design	Data acquisition	Actively design the tools for data collection	Define tools for data collection. Finalize the sources and methods of data collection. Define the type of data to be collected	–
	Data management	Collaborate with the data management team to prepare the right dataset	Basic understanding of data cleaning process. A fair understanding of data standards	–
	Data analysis	Collaborate with analysts to help make better choices of statistical techniques to use	Basic understanding of biostatistics	–
Execution	Deployment and upgrade	Collaborate with the technical team to improve the product and do verification and validation. Collaborate with regulatory experts for necessary regulatory and ethical approvals.	Do quality control and provide the necessary feedback for improvement. Have a complete understanding of the regulatory approval process. Have professional ethics	–

stage of clinical data science and clinical AI in MH and psychiatry. The paper doesn't delve in-depth into the data management and ML concepts. There are a lot of studies that discuss them in detail, and there is a necessity for consolidating the knowledge into understandable domains.

Conclusion

Psychiatrists have a huge responsibility in the ideation and execution of AI-based applications. Hence, it is necessary to understand the important concepts of data science, digital phenotyping, CDSS, and analytics from the process model perspective to make it easier for psychiatrists to understand and collaborate with different stakeholders wherever necessary. Future studies should focus on maintaining ethics, privacy, security, and confidentiality of the data collected through these novel modes of technology.

ORCID iD

Arunkumar Annamalai (iD) https://orcid.org/0000-0003-3195-6374

References

1. Obermeyer Z and Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* September 29, 2016; 375(13): 1216.
2. Bauer M, Monteith S, Geddes J, et al. Automation to optimise physician treatment of individual patients: Examples in psychiatry. *Lancet Psychiatry* April 1, 2019; 6(4): 338–349.
3. Torous J, Staples P, and Onnela JP. Realizing the potential of mobile mental health: New methods for new data in psychiatry. *Curr Psychiatry Rep* August 2015; 17(8): 1–7.
4. Fakhoury M. Artificial intelligence in psychiatry. In: Kim YK. (ed) *Frontiers in psychiatry. advances in experimental medicine and biology,* vol 1192. Singapore: Springer.
5. Kalanderian H and Nasrallah HA. Artificial intelligence in psychiatry. *Curr Psychiatry* August 2019; 18(8): 33–38.
6. Hsin H, Fromer M, Peterson B, et al. Transforming psychiatry into data-driven medicine with digital measurement tools. *NPJ Digit Med* August 22, 2018; 1(1): 1–4.
7. Huys QJ. Advancing clinical improvements for patients using the theory-driven and data-driven branches of computational psychiatry. *JAMA Psychiatry* March 1, 2018; 75(3): 225–226.
8. Rutledge RB, Chekroud AM, and Huys QJ. Machine learning and big data in psychiatry: Toward clinical applications. *Curr Opin Neurobiol* April 1, 2019; 55: 152–159.
9. Wang Y, Kung L, and Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol Forecast Soc Change* January 1, 2018; 126: 3–13.

10. Sanchez-Pinto LN, Luo Y, and Churpek MM. Big data and data science in critical care. *Chest* November 1, 2018; 154(5): 1239–1248.

11. Wykes T. Racing towards a digital paradise or a digital hell? *J Mental Health* 2019; 28(1): 1–3.

12. Wang XJ and Krystal JH. Computational psychiatry. *Neuron* November 5, 2014; 84(3): 638–654.

13. Barnett E and Thomas J. Methods for the synthesis of qualitative research: A critical review. *BMC Med Res Methodol* 2009; 1: 1–1.

14. Gough D, Oliver S, Thomas J, editors. *An introduction to systematic reviews*. SAGE Publications, March 28, 2017, 181–211.

15. Sandelowski M, Voils CI, Leeman J, et al. Mapping the mixed methods–mixed research synthesis terrain. *J Mix Methods Res* October 2012; 6(4): 317–331.

16. Mays N, Pope C, and Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policymaking in the health field. *J Health Serv Res Policy* July 2005; 10(Suppl 1): 6–20.

17. Gibson W and Brown A. *Working with qualitative data*. SAGE Publications, May 18, 2009, 65–74.

18. Carroll C, Booth A, Leaviss J, et al. "Best fit" framework synthesis: Refining the method. *BMC Med Res Methodol* December 2013; 13(1): 1–6.

19. Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: Realist syntheses. *BMC Med* December 2013; 11(1): 1–4.

20. Bone D, Lee CC, Chaspari T, et al. Signal processing and machine learning for mental health research and clinical applications [perspectives]. *IEEE Signal Process Mag* September 6, 2017; 34(5): 196–195.

21. Thieme A, Belgrave D, and Doherty G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans Comput-Hum Interact* August 17, 2020; 27(5): 1–53.

22. Thenral M and Annamalai A. Challenges of building, deploying, and using AI-enabled telepsychiatry platforms for clinical practice among urban Indians: A qualitative study. *Indian J Psychol Med* July 2021; 43(4): 336–342.

23. Garcia-Ceja E, Riegler M, Nordgreen T, et al. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive Mob Comput* December 1, 2018; 51: 1–26.

24. Rutledge RB, Chekroud AM, and Huys QJ. Machine learning and big data in psychiatry: Toward clinical applications. *Curr Opin Neurobiol* April 1, 2019; 55: 152–159.

25. Kubben P, Dumontier M, and Dekker A. Fundamentals of clinical data science. 2019. DOI: 219.10.1007/978-3-319-99713-1.

26. Schulz S, Stegwee R, and Chronaki C. Standards in healthcare data. *Fundamentals of clinical data science*. 2019, 19–36.

27. Shatte AB, Hutchinson DM, and Teague SJ. Machine learning in mental health: A scoping review of methods and applications. *Psychol Med* July 2019; 49(9): 1426–1448.

28. Liang Y, Zheng X, and Zeng DD. A survey on big data-driven digital phenotyping of mental health. *Inf Fus* December 1, 2019; 52: 290–307.

29. Onnela JP and Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* June 2016; 41(7): 1691–1696.

30. Wasylewicz AT and Scheepers-Hoeks AM. Clinical decision support systems. *Fundamentals of clinical data science*. 2019, 153–169.

31. Saleem TJ and Chishti MA. Exploring the applications of machine learning in healthcare. *International J Sens Wirel Commun Control* August 1, 2020; 10(4): 458–472.

32. Federer L, Foster ED, Glusker A, et al. The Medical Library Association Data Services Competency: A framework for data science and open science skills development. *J Med Libr Assoc* April 2020; 108(2): 304.