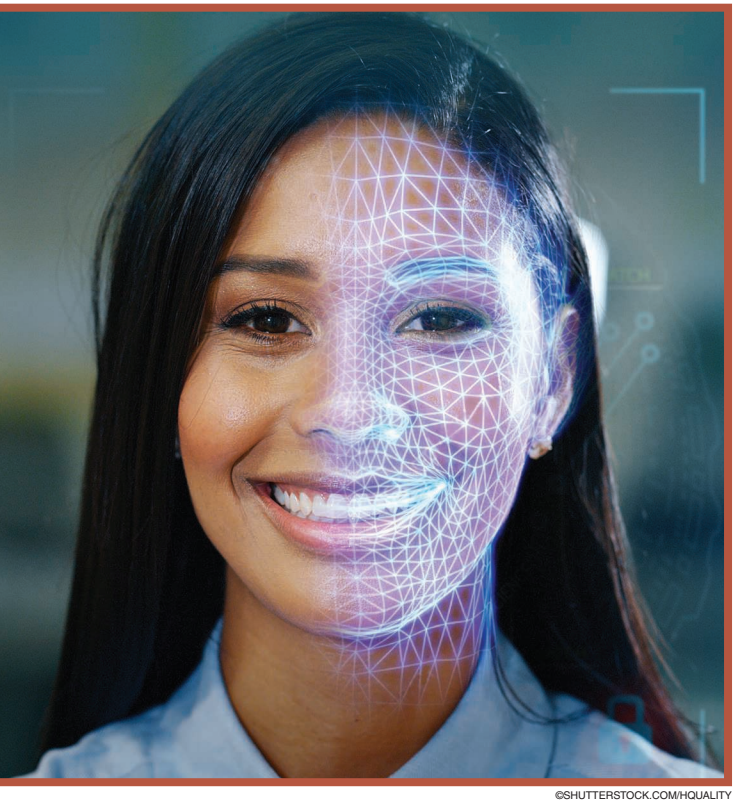
Jing Han, Zixing Zhang, Cecilia Mascolo, Elisabeth André,
Jianhua Tao, Ziping Zhao, and Björn W. Schuller

Deep Learning for Mobile Mental Health

Challenges and recent advances



©SHUTTERSTOCK.COM/HQUALITY

Digital Object Identifier 10.1109/MSP.2021.3099293
Date of current version: 27 October 2021

Mental health plays a key role in everyone's day-to-day lives, impacting our thoughts, behaviors, and emotions. Also, over the past years, given their ubiquitous and affordable characteristics, the use of smartphones and wearable devices has grown rapidly and provided support within all aspects of mental health research and care—from screening and diagnosis to treatment and monitoring—and attained significant progress in improving remote mental health interventions. While there are still many challenges to be tackled in this emerging cross-disciplinary research field, such as data scarcity, lack of personalization, and privacy concerns, it is of primary importance that innovative signal processing and deep learning (DL) techniques are exploited. In particular, recent advances in DL can help provide a key enabling technology for the development of next-generation user-centric mobile mental health applications. In this article, we briefly introduce the basic principles associated with mobile device-based mental health analysis, review the main system components, and highlight the conventional technologies involved. We also describe several major challenges and various DL technologies that have potential for strongly contributing to dealing with these issues, and we discuss other problems to be addressed via research collaboration across multiple disciplines.

Introduction

Mental health is the state of an individual's own ability to control his or her thoughts, feelings, and behaviors, and it helps one determine how to cope with stresses, relationships with others, and challenges in life [1]. It is important to maintain good mental health at every stage of life, from childhood and adolescence through adulthood and old age. It has been reported that 25–50% cases of adult mental illness may be prevented through early intervention in childhood and adolescence [2]. Moreover, graduate students have an eight times higher rate of severe depression and anxiety and are reluctant to seek treatment [3]. Hence, mental well-being is too important to delay trying to improve it. Despite the considerable progress that has been made to promote mental health, much more effort is still required to address the current unmet and underestimated

 1053-5888/21©2021IEEE

mental health needs [1]. In particular, the Lancet Commission on Global Mental Health and Sustainable Development calls for actions to promote mental health for all, to prevent mental disorders among people at high risk, and to reduce the treatment and care gap for people affected by mental disorders [1].

In the last decade, digital techniques have provided new opportunities to reframe the mental health system and alter face-to-face services in a variety of ways, such as delivering prevention messages to educate the public, building online communities to support people with mental health issues, and facilitating remote screening and diagnosis of mental disorders [1]. In particular, the capabilities and functionalities of mobile and wearable technologies to support health care have led to the development of the new interdisciplinary field of mobile health (mHealth). In the following, the term *mobile mental health* (*M²Health*) is used to indicate mHealth systems specifically tailored for mental health.

M²Health has a huge potential to lead the mental health revolution, and it has four key strengths. First, remote measurement technologies (RMTs) based on mobile devices offer M²Health new possibilities for long-term data collection (e.g., heart rate and respiration rate) and continuous monitoring (e.g., sleep duration and sleep quality). In general, traditional assessment largely relies on memory because of a lack of clear and objective biobehavioral markers. By contrast, M²Health can add value to the formal clinical assessment of mental illness by providing clinicians with summaries of the RMT data of the patient collected for a short period before an appointment [4]. Second, a variety of M²health applications has been targeted at different stages, from prevention and assessment to intervention and treatment [5], [6]. Therefore, M²health has a good chance of being integrated with traditional mental health care across entire mental health pathways [1]. Third, the ubiquitous and affordable character of mobile and wearable devices can provide cost-effective M²Health solutions. Currently, 45% percent of the worldwide population lives in countries with fewer than one psychiatrist per 100,000 people [7]. Also, national surveys from China and India revealed that more than 80% of individuals with mental illness did not seek help, for varied reasons, such as being afraid of discrimination and experiencing stigma [1]. M²Health, in this context, can help narrow the treatment and care gaps and reduce inequalities for mental health service. It also offers an alternative choice for users and therefore may encourage higher levels of engagement of those with mental health problems. Fourth, M²Health can use the full capabilities of digital technology to be more effective to meet varied individual needs. As mental health problems can be very complex, even diverse within the same individuals over time, conventional mental health services are not always sufficiently effective, and they largely neglect user experience [1]. Recently, increasing discussions have focused on precision psychiatry and precision mental health care [8]. In this context, M²Health has a better chance to exploit more personal data, enabling it to obtain a better understanding of the problems, provide more accurate diagnoses, and deliver more personalized and user-friendly intervention and treatment. In short, M²Health has a tremendous potential to provide continuous, affordable, and adaptive mental health services and to be involved in clinical pathways.

It has been reported that 25–50% cases of adult mental illness may be prevented through early intervention in childhood and adolescence.

However, as a science in its infancy, there are still barriers and limitations of M²Health that hinder its development and must be addressed [9], such as extracting meaningful features from large-scale heterogeneous data, "green" computing on wearable and mobile devices, the interpretability of the decisions, and data security and privacy issues. In addition, there are other challenges, such as the lack of validation by clinical trials in large cohorts, the public trust of digital devices and tools, the risks of being harmful to mental health, and legal and ethical problems, among others.

In recent years, intelligent signal processing and DL techniques have brought breakthroughs in processing data such as audio, speech, text, and images and video [10]. In particular, by transforming data through multiple layers of nonlinear computational processing units, DL models provide a new paradigm to model complex data [10]. For M²Health systems, DL can be exploited to handle complex data, such as vocal and visual expression and social media data. Moreover, DL can be leveraged to be positioned in varying stages of the care pathways. Take psychiatric counseling as one example. Researchers presented an emotional chatbot, where DL approaches are applied for natural language understanding and continuous emotion monitoring [11]. The aim of this chatbot is to provide sympathetic psychotherapy and treatment services for people with emotional disorders [11].

With all that said, DL techniques are primed to have a major impact on increasing the efficacy and efficiency of mental health applications on mobile devices by tackling some of the aforementioned barriers. Consequently, the aim of this contribution is to discuss recent advances in DL that can help provide a key enabling technology for the development of next-generation user-centric M²Health applications. This article is unique, differing from some previous studies that focused on M²Health but not opportunities for DL [6] and from others that are not centered on mobile health [12], [13]. Another recent work [5] provides a survey through research works related to new techniques (DL advances included) that might be utilized to improve mental well-being. However, the potentials and challenges of DL advances for mobile mental health have not been thoroughly discussed. In the next sections, we provide a systematic introduction of this topic, including a brief overview of a typical M²Health system as well as recent DL technological advances, which may enable new opportunities and possibilities.

A typical mobile mental health analysis system

A mental health spectrum and an intervention spectrum

Mental health is not merely about the absence of mental disorders and illnesses. One may occasionally experience poor

mental health without a mental illness. We all inevitably have days where we feel "little down," anxious, or stressed. These are negative emotional and mental states, while good mental health and well-being is about living and coping well with these challenges and stresses. In particular, similar to physical health, mental health exists on a spectrum [see Figure 1(a)], spanning a continuum ranging from healthy to stressed, to injured/ill, or experiencing a crisis.

Moreover, surrounding the ultimate goal of promoting mental health, mental health-care service should meet a range of needs of individuals, including, but not limited to, the need for assessment and monitoring of general mental well-being, prevention efforts for mental illness development, detection and diagnosis of mental health conditions, and provision of treatment and support for people with mental illness. A variety of mental health interventions is presented in Figure 1(b). It is important to note that M^{2}Health applications can vary widely; some applications might be designed for a general purpose, while others might be devised to support one specific mental health issue.

A typical M^{2}Health system

In general, a typical pipeline to construct an M^{2}Health analysis system (e.g., an automatic depression-detection system) involves gathering raw data, transforming the data into time-sequential feature representations, and feeding the features into neural networks, which then supply a variety of mental health-care applications (see Figure 2). Specifically, first, a wide range of raw data is collected from mobile devices. For instance, these data can be physiological signals via wearable sensors, the current location via smartphones, or the content of users' social media posts. Then, these heterogeneous raw data need to pass through a pipeline of processes to attain representative features in which mental state-related information is extracted. Normally, the collected data are filtered, enhanced, transformed, and integrated into high-level representations. For instance, the tone of voice can be extracted from speech signals during conversation for depression assessment. Likewise, in an interaction therapy for autistic children, the eye-contact frequency can be captured from video signals.

Once these representative features are obtained, DL models are applied to perform learning and inference, aimed at mapping input features to output predictions associated with mental states. Finally, the obtained outputs can be exploited at different stages of the medical-care process (e.g., screening, diagnosis, prevention, treatment, and follow-up). It is worth noting that training a DL model often demands a large amount of data and computing power, and it is generally performed off device (on-device training is out of the scope of this article). Once trained, the inference task of such DL-powered mobile applications can be implemented either on a server where the trained model is stored (also known as *cloud-based deep inference*) or locally on the mobile device (also known as *on-device deep inference*) [14]. These two designs have different pros and cons; cloud-based inference keeps the mobile app simple but requires network access, while on-device inference can be faster; however, it becomes more difficult to update.

Opportunities for DL for M^{2}Health

It can be seen that to develop a good M^{2}Health analysis system, signal processing and DL techniques are both key components as well as challenges in mapping low-level noisy signals to high-level representative features and, further, in obtaining final predictions based on those features. In recent years, DL models have achieved remarkable successes, and the importance of the models has grown to be central in the field of affective and mental health research [13]. One such example is the convolutional neural network, which was originally devised for data-driven extraction of high-level representations from visual data (e.g., facial-expression images). Likewise, the recurrent neural network was designed to make use of contextual information when processing sequence data (e.g., text and speech). These networks have achieved promising performances for solving a range of mental health-related tasks, such as detecting stress from social media posts or predicting mood disorders from speech. Other successful use cases are discussed in [13].

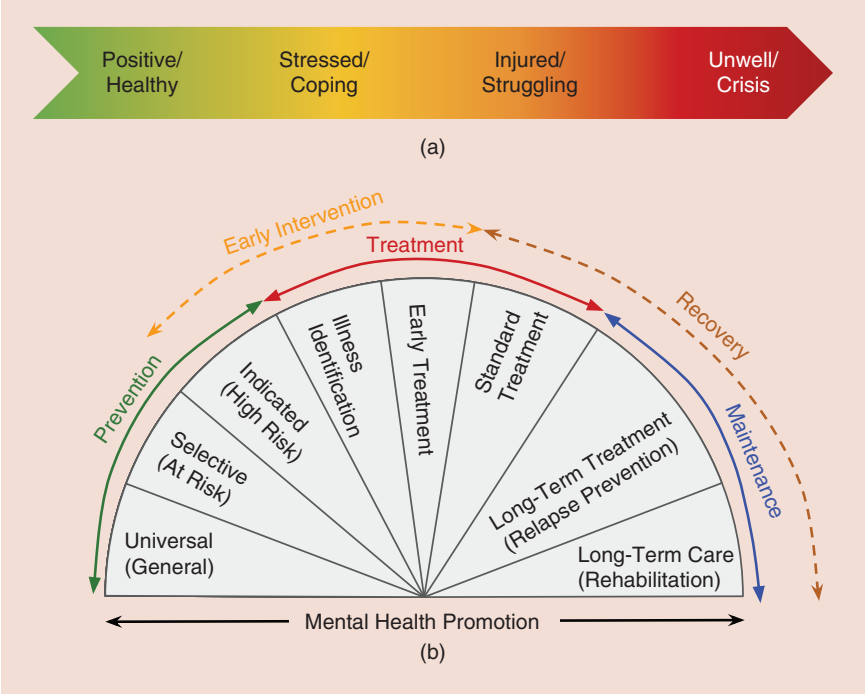


FIGURE 1. (a) The mental health spectrum and (b) its intervention spectrum.

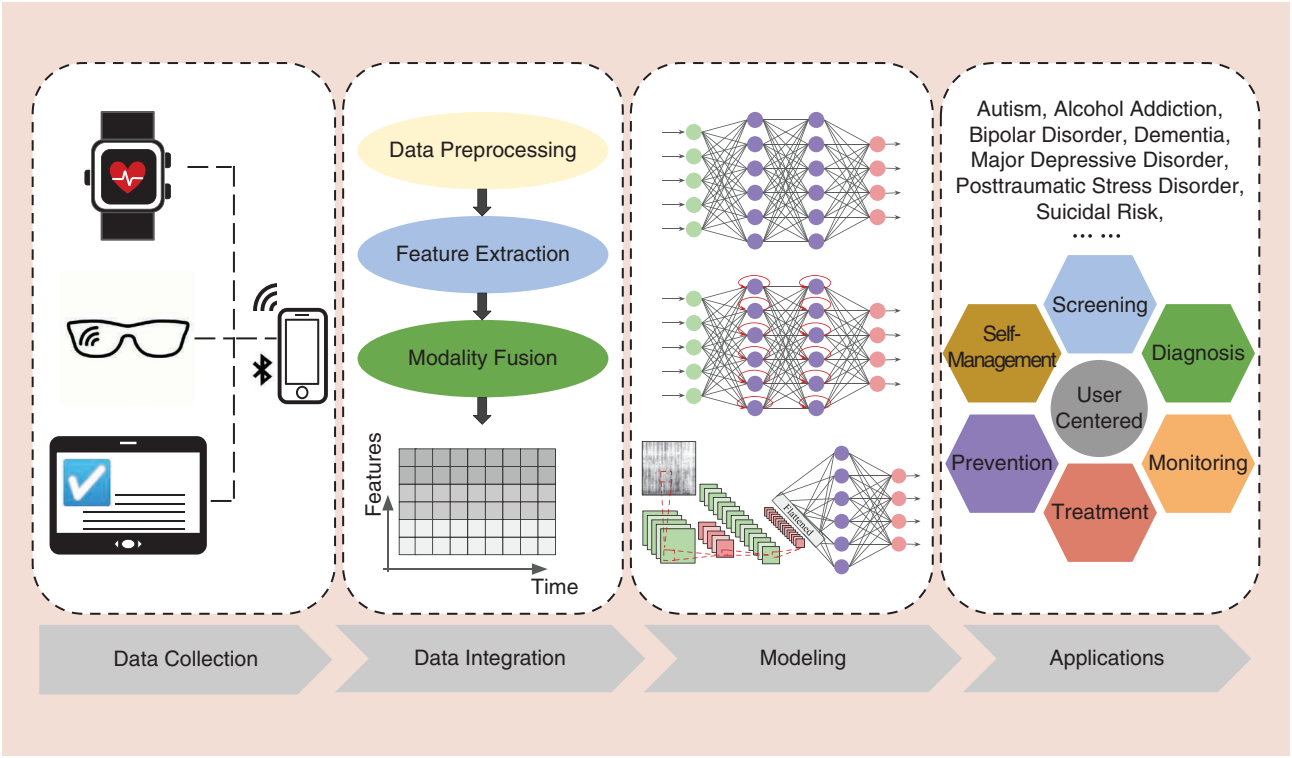


FIGURE 2. A typical mobile device-based mental health analysis system. The system can be partitioned into four key modules: data collection, data integration, deep modeling, and applications. Heterogeneous raw data are collected on which a pipeline of signal processing techniques is performed to produce time-sequential feature representations. Then, deep neural network-based models are learned to estimate mental health states, which power the followed mental health applications.

Taking depression and anxiety disorders (affecting more than 10% of the global population [15]) as an example, DL might be applied in different M^{2}Health applications to address a variety of technical issues. First, in the prevention stage, DL may deal with large-scale RMT data collected via mobile devices as it is helpful to sort out usually occurring asynchronous and heterogeneous data issues [16]. In particular, DL could uncover the changes in the biological, psychological, or social data to predict the mental health status of a person, remotely and around the clock, and thus may facilitate rapid detection and intervention even without the person visiting a hospital. In addition, in the treatment stage, DL may help associate the symptoms and causes of disorders through continuous monitoring over long time periods, and in this manner, caregivers can gain a better understanding and phenotyping of the illness for each patient. As a consequence, rather than one-size-fits-all treatment, individual variability can be considered to select more effective and appropriate treatments, either varied psychological therapies, some specific antidepressant medications, or a combination of the two [8]. Moreover, DL can be useful in responding to individual preferences for mental health service delivery. For instance, DL-based recommendation models can be employed to advocate more personalized emotional improvement services, such as music therapy (music recommended to the user to relax), exercise therapy (exercise guidance given to the user to release stress), and interpersonal talk treatment (status of the user released to close family members or friends) [17]. This tutorial assumes that the readers have a preliminary knowledge of DL and focuses on how to harness the most recent and promising DL algorithms for some arising challenges (see Figure 3) and opportunities in M^{2}Health.

FIGURE 3. Major challenges of M^{2}Health where DL may help.

Limited, heterogeneous, and asynchronous data versus efficient data exploitation

Challenges

Unlike physical illnesses, which can be identified mainly by biomedical markers, mental disorders also closely link to patients' expressive behaviors, such as facial and vocal expressions, head movements and body gestures, and also the reaction of social networks. To determine these behavioral markers, various mobile devices are needed to capture corresponding audio, biological, textual, and video cues. However, efficiently making use of these data is not trivial, and there are several key challenges related to the data. First, it is not sufficiently clear how to explore the complementary semantic information across multiple modalities while discarding individual redundant information. The heterogeneity of multimodal data makes it challenging to construct salient representations and make decisions associated with the mental disorder of interest. For example, texts are often discrete symbols, while audio and video modalities are usually represented as continuous signals. Second, the severe asynchronicity of the data makes it more difficult to identify a direct relationship across modalities. For example, one may want to align patient questionnaire replies with occasionally recorded biological and speech data before and after one week for depression detection. To deal with this, similarity measures are required to determine the possible long-term dependencies and ambiguities. Third, in mental health research, collecting sufficiently fine-grained data with accurate annotations is quite challenging and rare in the real world. In addition to the requirements of data quantity and quality, time-dependent data are also demanding when tracking patient trajectories such as relapse prediction. Last but not least, even if a good outcome can be obtained on the collected data to model the problem, there are other typical issues that need to be addressed for M^{2}Health applications in daily use, such as incomplete data, signal artifacts, and the data mismatch caused by varied hardware and software versions. In consequence, finding ways to efficiently exploit a limited amount of data are of importance in M^{2}Health analysis.

Finding ways to efficiently exploit a limited amount of data are of importance in M^{2}Health analysis.

Opportunities

Despite the heterogeneity and asynchronicity of collected data, previous studies consistently emphasized the benefit of multimodal exploration for mental health analysis due to the complementary information [13]. For M^{2}Health, there are cases such as a health condition being more perceptible in one specific data type, a complementary modality providing additional knowledge of a health problem, or data of one type missing for a period of time. Therefore, it is necessary to take multiple modalities into account to provide a complete and comprehensive profile.

When integrating multiple data sources, learning the salient representations shared across different modalities holds a key position, and this research can be categorized into joint or coordinated representation learning frameworks. Joint representation learning aims to project the heterogeneous data into a shared latent subspace, in which the data in different modalities, but with similar semantics, will be represented by similar vectors. That is, the model is designed to obtain modality-agnostic and semantic-salient properties and simultaneously discard any irrelevant properties. By contrast, coordinated representation learning intends to distill separated but coordinated representations for each modality under some constraints. It is assumed that the separated representations will be helpful in preserving the exclusive and useful modality-specific properties. For example, the learning framework shown in [18] takes two separate networks to extract audio and video representations, which are simultaneously regularized by a similarity loss to learn modality-invariant representations for emotion recognition.

Alternatively, methodologies aiming at integrating multiple data sources could be useful in M^{2}Health. Some typical fusion technologies include 1) feature-level (early) fusion, which simply concatenates the low-level descriptors of each modality into a long vector as network inputs; 2) decision-level (late) fusion, which merges the decisions from each modality-specific model via a fusion mechanism such as averaging; and 3) using an attention mechanism, which automatically controls the contributions from each data source and has the potential to dynamically force-align the asynchronous multimodal data. For instance, by integrating data gathered from smartphone sensors and wearable devices, such as accelerometers, communication logs, and screen interactions, changes in sleep, communication patterns, and activity patterns could be identified, with the ultimate goal of predicting major depressive relapse.

Furthermore, to cope with the data-scarcity issues in M^{2}Health, plenty of existing and emerging DL algorithms can be leveraged, and they can be divided into the following categories: 1) data augmentation, which is a straightforward way to enrich the variety and the size of data based on available annotated data; 2) weakly supervised learning, where incomplete, noisy, or weakly labeled data are efficiently exploited during training rather than being discarded; 3) semisupervised learning, where large-scale unlabeled data are employed to make use of the knowledge from unlabeled data; 4) self-supervised learning, where efficient high-level representations are learned via pretext tasks and then used in downstream tasks; and 5) transfer learning, where knowledge gained from out-of-domain data can be transferred to the task at hand. For instance, a successful example to leverage transfer learning for M^{2}Health is to improve the prediction performance of speech-based depression detection by pretraining the model on a large-scale data set for the speech recognition task.

Generic modeling versus personalized modeling

Challenges

Another challenge of M^{2}Health is the need for precision mental health services. Currently, diagnoses of most mental

disorders are based on clinical symptoms only, and it is impossible to carry out biological tests, like a simple blood test, to accurately diagnose different mental health conditions. One main reason is that clinicians and researchers still lack knowledge of the underlying biological mechanisms of mental health diseases. In addition, signs of a mental health condition may vary from one individual to another. For instance, there is a wide range of manifestations of disorganized speech (one of the core symptoms of schizophrenia), such as a decrease in the amount of speech, empty speech that conveys little information, odd usage of words, and illogical reasoning [19]. Hence, developing methods for personalized diagnosis and monitoring remains a challenging task to date. Besides, individual differences and individual needs and preferences should also be met, especially when health-care providers need to decide which treatment should be given to ensure personalization and precision. In the context of M^{2}Health, signal processing and DL techniques can be exploited to detect subtle anomalies from the wearable sensor data for a specific individual. Innovative methods may enable a future of personalized M^{2}Health, providing an accurate understanding of mental health problems and also delivering individually tailored treatments and interventions.

Data-driven algorithms like autoencoders and clustering methods have the capability to uncover hidden components from the complex data and assign a patient to each of the clusters to different degrees.

Opportunities

First, and most importantly, for practical use, individual-level information (e.g., demographics, social and cultural variables, and personality traits) should be taken into account. Multiple previous studies have demonstrated that a personalized subject-level model can result in a better performance than a general-purpose model, by utilizing knowledge of personal data or learning personalized features [8]. Additionally, contextual data, i.e., the physical and logical environment of the user, are also necessary to better understand external factors that affect the user. The reason is that many mental disorders and corresponding consequences are highly related to the context of the patients. To achieve more personalized and self-adaptive M^{2}Health modeling, four of the most promising, yet not widely implemented, techniques can be investigated: multitask learning, continual learning, reinforcement learning, and zero-/few-shot learning. However, note that deploying these data-driven approaches for successful subject-level outcomes would demand data with rich and fine-grained individual and context information. These types of data are extremely rare; thus, these techniques require further investigation.

Second, another promising avenue is to make use of structured or semistructured data for M^{2}Health. Structured data are ubiquitous, including, but not limited to, electronic health records, location information of the GPS tracker for patients with Alzheimer's disease and the elderly, and online social networks for patients with severe mental illness. However, how to explore the structured and heterogeneous health-care data is still an open, and understudied, question. One possible solution is to make use of graph structures to model the relationship among different data [20]. By using that method, one can also encode the human expert knowledge in the loop of model design. We expect that future research in this direction will lead to more precise decisions for mental health care.

Third, DL may also open up the possibility of deconstructing the traditional symptom-based diagnostic categories into data-derived subgroups that can better predict treatment outcomes [8]. On the one hand, many symptoms are shared among, rather than being unique to, varied mental disorders; on the other hand, the symptom-based diagnostic categories are frequently revised to align with new behavior or biological discoveries [8]. In contrast, data-driven algorithms like autoencoders and clustering methods have the capability to uncover hidden components from the complex data and assign a patient to each of the clusters to different degrees [8]. It is believed that these techniques may reveal and exploit currently unknown interindividual variation and thus improve the effectiveness of precision mental health predictions.

Unexplained decision making versus explainable and trustworthy inference

Challenges

Interpreting how and why the M^{2}Health outcome is achieved in an understandable way for users or caregivers is critical. Hence, for M^{2}Health applications, a trustworthy system has to be transparent and explainable. Such a system needs to maintain capabilities such as pointing out disease signatures from input signals, providing a quantifiable confidence level, associating an unseen case with other similar cases for which decisions are already available, or imitating the inference chain of the physicians. Only with these capabilities will patients and physicians be more likely to accept the involvement of intelligent systems in daily practice. However, efforts toward that direction in mental health-care applications are very limited. Hence, it is crucial to invite signal processing researchers to develop explainable algorithms with the hope of stepping into a new era of responsible systems for mental health.

Opportunities

Although interpretable models, such as linear regression, logistic regression, or decision trees, exist, it is critical to enhance the model interpretability of DL-based M^{2}Health applications. However, most, if not all, DL models are black boxes, the parameters of which are of high complexity and of extremely large scale. This makes the decisions from such models quite difficult to directly explain. Here, we focus on three of the most advanced interpretation approaches: model-agnostic methods, attention mechanisms, and Bayesian neural networks (BNNs) [21].

Model-agnostic methods are flexible enough to be applied to any DL model after it has been trained. In particular, three promising model-agnostic techniques discussed here are individual conditional expectation (ICE), feature importance, and local surrogate models, specifically, local interpretable model-agnostic explanations (LIME). With ICE, a plot is obtained to reveal how the prediction changes when a feature changes. The importance of a feature is measured by calculating the increase of the model error after permuting the feature. Taking detecting depressive symptoms via geolocation data as an example [22], staying at home and visiting fewer locations are identified as the two most important features to indicate depression in bipolar disorder. With LIME, an additional interpretable model (e.g., a linear model) is further trained to understand the deep model by exploring a new data set consisting of permuted samples and the corresponding predictions from the deep model. This approach also demonstrates that traditional statistical methods can be interleaved with DL to enhance the interpretability for M^{2}Health.

Attention-based models pay more importance and greater attention to the most task-relevant parts from input signals when processing the data. Attention mechanisms in DL can thus enhance interpretability and often deliver better performance. For instance, a hierarchical attention transfer mechanism was proposed in [23], aiming at assigning the right amount of attention to each frame-level data item toward a higher-level clinical depression score. This might offer novel insights into the association between artificial attention and human attention for the same task.

Moreover, to obtain uncertainty information from deep models, BNNs were constructed, based on the theoretical foundations of probability theory and Bayesian modeling [24]. In BNNs, model uncertainty can be incorporated by placing distributions over each weight, and the posterior over the weights given one data item can then be deemed as the uncertainty [24]. The posterior, however, is intractable; hence, various approximation methods have been investigated, such as Monte Carlo estimators. Recently, Monte Carlo dropout was proposed to estimate model uncertainty, simply by implementing dropout during inference and analyzing the variance from multiple stochastic forward passes, without changing anything from the existing deep models [24]. For example, this method was applied to obtain the model uncertainty of emotional state predictions [25], and the authors demonstrated how a deep framework improved the interpretability by introducing a confidence threshold.

If an M^{2}Health system can provide human-understandable justifications and explanations along with its predictions, it can disclose the model's inner logic while revealing its strengths and limitations. This will render the diagnostics and intervention outcomes more traceable, transparent, and trustworthy.

Hardware constraints versus mobile device-friendly models

Challenges

Although the users are willing to use M^{2}Health products, the underlying issues associated with the hardware resource constraints remain unresolved [5]. Such issues include power consumption, computational capability, storage and memory size, and transmission bandwidth. Besides, M^{2}Health systems often demand a real-time inference with low latency for reminding patients to take a relevant action or on-device learning for model adaptation and users' privacy protection. These demands further increase the hardware requirement. Therefore, reducing the model size, computational cost, memory footprint, and bandwidth requirement becomes critical. Without addressing the corresponding hardware resource constraints, it will remain impractical to use M^{2}Health to infer and improve mental well-being in real life [5].

Without addressing the corresponding hardware resource constraints, it will remain impractical to use M^{2}Health to infer and improve mental well-being in real life.

Opportunities

Despite the widespread usage of DL in mental health analysis, making corresponding models friendly on mobile devices is still at an early research stage. Reaching this goal will require joint solutions from different disciplines, such as computer architecture, signal processing, and DL. Here, we discuss existing technologies for compressing and accelerating deep neural networks while also retaining their performance. Such technologies include compact network design, parameter sharing and pruning, low-rank decomposition, model quantization, and knowledge distillation, and these solutions can be implemented individually or collaboratively [26].

Compact network design rebuilds part of the conventional network components by replacing them with slimming and compact ones. Such an architecture redesign, however, largely relies on experts' experience and knowledge. To overcome this issue, research projects have increasingly started to employ network architecture search technologies to automatically search for compact network structures, by defining the search space and taking the network size and computational cost as regularization terms.

Parameter pruning removes redundant and noninformative parameters through the evaluation of the importance of model parameters, whereas parameter sharing tends to share weights across different layers to meet the model size demand. It is noted that both parameter pruning and parameter sharing are used for model size reduction; however, the efficiency in inference time may not be improved [26].

Low-rank factorization uses matrix or tensor decomposition to estimate parameters. The decomposition process, however, is often computationally expensive. Moreover, low-rank factorization is normally performed layer by layer and fine-tuned based on a reconstruction error criterion. Therefore, even though the low-rank factorization approaches are

straightforward for model compression, they hardly achieve an optimal compression rate.

Model quantization aims to compress deep models by reducing the number of bits required to represent each model weight, for example, by using a binary neural network, which significantly cuts down the computational cost. To quantify the weights, *k*-means or fixed-point quantization with an optimized bit width can be employed. However, it is worth noting that directly reducing the data precision gives rise to unacceptable performance degradation. Hence, an additional retraining process is often required to maintain the effectiveness of the models.

Knowledge distillation trains a compact neural network (student) by exploiting another larger network (teacher) that has a similar or dissimilar structure. It is supposed that the knowledge learned by the teacher can be distilled to the lighter student network. Recent studies show that a student model trained with the soft labels from a teacher model can achieve much better performance for mental state classification by microexpression [27].

With a compressed DL model implemented on mobile devices, once the streaming data of the user are obtained, the inference can be performed locally in real time. There is less demand for data storage and transfer; thus one can avoid the particular risks that external data handling brings about.

Privacy infringement versus privacy-preserving learning

Challenges

When using smartphone applications for mental health care, user privacy has been a persistent concern, as the gathered data are personal and sensitive [5]. It is a prerequisite to collect and explore users' daily mental and behavioral information with mobile systems for mental health analysis. Nevertheless, this privacy information is exceptionally sensitive. The largest adverse consequence of a data breach is the violation of user/patient privacy. In addition, it often results in other negative consequences, such as ethical and legal issues and an increase in user reluctance. Given these risks, regulating sensitive data acquisition, management, and usage to secure user privacy is highly needed in M^{2}Health.

Opportunities

Privacy-preserving deep modeling attempts to bridge the gap between personal data protection and data usage for clinical routine and research. The privacy-preserving mechanisms can be applied to the whole deep modeling chain, from data aggregation, through model training, to model inference [28]. Typical approaches include deidentification, differential privacy (DP), homomorphic encryption, and secure multiparty computation.

A deidentification mechanism is used to collect and create data sets while protecting the personal information of data contributors. Typical solutions include anonymization and pseudonymization, where anonymization straightforwardly removes the private information from the recorded data, and pseudonymization replaces the sensitive entries with artificially synthesized ones.

Furthermore, DP retains the global statistical distribution of a data set while reducing the individually recognized information [28]. For example, a data set is differentially private if an outside observer is unable to infer whether a specific individual was used for obtaining results from the data set [28]. DP can be implemented together with model inputs, objective functions, gradient updates, and outputs, as well as labels. However, for DP, a tradeoff occurs between the model accuracy and privacy.

Homomorphic encryption allows computation over encrypted data directly. That is, mobile devices send user data in an encrypted way to a server, where data can be utilized without decryption for training or inference. In spite of its promise, this method is computationally intensive.

Secure multiparty computation distributes a computation process across multiple parties, where each single party can access only an encrypted part of the data rather than the entire data. However, the reliability and scalability to a large number of computing parties is a concern.

Apart from these direct privacy-preserving mechanisms, an encouraging execution learning paradigm is federated learning (FL), which can work together with the aforementioned privacy-preserving mechanisms like DP. FL trains a model across multiple decentralized edge devices where data remain locally without disclosing them to other edge devices or the cloud. Specifically, the client downloads the model, updates it by learning the local data, and then sends the updates to a shared model in the cloud using encrypted communication. The shared model averages the updates together with other client updates to improve itself. In FL, all of the training data remain on the edge devices, and no individual information is stored in the cloud [29]. FL as a decentralized learning strategy cannot only boost the efficiency of data utilization by removing data barriers, but, more importantly, can alleviate the privacy risks. FL has become an active research topic today in M^{2}Health because of its promising application potential [30]. For example, an FL framework called *FedHealth* was proposed, aiming at personalizing health care without compromising privacy and security by aggregating heterogeneous data collected from multiple parties [30].

Other outstanding issues, discussions, and conclusions

Although we have covered multiple challenges that provide research opportunities to deploy DL algorithms to address issues in M^{2}Health systems, many other questions and challenges still remain. For instance, note that the reliability and validity of M^{2}Health applications are considerably difficult to verify. Recently, there have been many discussions on the reproducibility issue, raising concerns and questions, such as how to assess the reproducibility of results techniques achieved on small-scale studies when reaching larger populations [5], [6], [9]. In general health care, new treatments are often validated by multiple large-scale randomized clinical trials, but these trials are very limited in mental health. Therefore, it would be worthwhile to

find ways to evaluate the usefulness of new applications in a fast, effective, and pragmatic manner.

Alongside the potential of digital technologies to favor mental health, it is also important to recognize some harms or risks that M^{2}Health might create. For example, the use of mobile devices might lead to or intensify other mental health conditions such as gaming disorders. Also, there are concerns that mobile-based services might lower the effectiveness of mental health care as genuine human interaction is lacking [1]. In addition, M^{2}Health might introduce further inequalities between those who have or do not have access to mobile devices [1]. Moreover, it remains unclear who should be held accountable for the decisions automatically made—in particular, the faults when utilizing M^{2}Health applications.

All of these challenges and barriers need to be overcome before M^{2}Health is more widely applied. Also, a critical step is to promote collaborations among multiple partners, such as clinical professionals; researchers in cognitive science, neuroscience, psychology, psychiatry, signal processing, and computer science; as well as policy makers.

To conclude, in this article, we provided an exploration of the current state of signal processing and DL techniques, with a focus on mental health analysis on mobile devices. We first presented the fundamentals of this research topic, followed by a comprehensive discussion about the most critical challenges and the advanced techniques to tackle them.

Mental health and well-being are important in our day-to-day lives, and DL technologies have been shown to improve or innovate the treatment of mental health care on mobile devices. While the field is still in its early development, there are many open issues to be investigated that could be of interest to the signal processing community. We hope that this article can encourage and inspire researchers to further enhance the analysis of deep mental health and deliver high-impact mobile applications for practical use.

Acknowledgments

This article has been partially funded by the Bavarian Ministry of Science and Arts as part of the Bavarian Research Association ForDigitHealth, the National Natural Science Foundation of China (grants 62071330 and 61702370), and the Key Program of the National Natural Science Foundation of China (grant 61831022).

Authors

Jing Han (jh2298@cam.ac.uk) received her doctoral degree in computer science from the University of Augsburg, Germany, in 2020. She is a postdoctoral researcher in the Department of Computer Science and Technology, University of Cambridge, Cambridge, CB3 0FD, U.K. She has served as a program committee member of the Audio/Visual Emotion Challenge and Workshop in 2018 and a technical program committee member of the Association for Computing Machinery (ACM) Multimedia since 2019. Her research interests are in affective computing and digital health. She (co)authored more than 40 publications in peer-reviewed journals and conference proceedings, such as *IEEE Transactions on Affective Computing, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Computational Intelligence Magazine*, and ACM Multimedia conference proceedings.

Zixing Zhang (zixing.zhang@tum.de) received his Ph.D. degree in computer engineering from the Technical University of Munich, Germany, in 2015. He is a researcher at Cambridge, CB4 0WG, U.K. From 2017 to 2019, he was a research associate with the Department of Computing at the Imperial College London, U.K. Before that, he was a postdoctoral researcher at the University of Passau, Germany. His research mainly focuses on deep learning technologies for speaker-centered state and health computing. He has authored more than 100 publications in peer-reviewed books, journals, and conference proceedings and organized several special sessions and served as a reviewer or program committee member for numerous leading-in-their-field journals and conferences.

Cecilia Mascolo (cm542@cam.ac.uk) received her Ph.D. degree from the University of Bologna. She is a professor of mobile systems in the Department of Computer Science and Technology, University of Cambridge, Cambridge, CB3 0FD, U.K. She is also a fellow of Jesus College Cambridge and the recipient of a European Research Council Advanced Grant. Prior to joining Cambridge in 2008, she was a faculty member in the Department of Computer Science at University College London. Her research interests are in mobile systems and data for health, human mobility modeling, sensor systems and networking, and mobile data analysis. She has published in a number of top-tier conferences and journals in the field.

Elisabeth André (andre@informatik.uni-augsburg.de) received her Ph.D. degree from the University of Saarland, in 1995. She has been a full professor of computer science and a founding chair of Human-Centered Artificial Intelligence at Augsburg University, Augsburg, 86159, Germany, since 2001. In 2010, she was elected a member of the Academy of Europe, and the German Academy of Sciences Leopoldina. In 2017, she was elected to the Computer–Human Interaction Academy, an honorary group of leaders in the field of human–computer interaction. To honor her achievements in bringing artificial intelligence techniques to human–computer interaction, she was awarded a European Association for Artificial Intelligence fellowship in 2013. Since 2019, she has served as the editor-in-chief of *IEEE Transactions on Affective Computing.*

Jianhua Tao (jhtao@nlpr.ia.ac.cn) received his Ph.D. degree in computer science from Tsinghua University, in 2001. He is a professor and deputy director at the National Laboratory of Pattern Recognition, Chinese director of the Sino-European Laboratory of Informatics, Automation, and Applied Mathematics, and assistant president of the Institute of Automation of the Chinese Academy of Sciences, Beijing, 100190, China. He has published more than 300 articles in journals or proceedings, including *IEEE/Association for Computing Machinery Transactions on Audio, Speech, and Language Processing; IEEE Transactions on Affective Computing; IEEE Transactions on Systems, Man, and Cybernetics; IEEE Transactions on Image Processing*; and *IEEE Journal of*

Selected Topics in Signal Processing. His research interests include speech recognition and synthesis, human–computer interaction, affective computing, and big data analysis. He has served as the program/general cochair of numerous international conferences and was a technical chair of Interspeech 2020.

Ziping Zhao (zhaoziping@tjnu.edu.cn) received his Ph.D. degree in automatic prediction of prosodic phrases in 2008 from Nankai University. He is a full professor of computer science at Tianjin Normal University, Tianjin, 300387, China. In 2018, he studied in the Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg, Germany as a visiting scholar. In 2016, he became the vice dean of the college of computer and information engineering at Tianjin Normal University. He has published more than 30 publications in peer-reviewed books, journals, and conference proceedings, including International Conference on Acoustics, Speech, and Signal Processing, Interspeech, and Neural Networks proceedings, and *IEEE Journal of Selected Topics in Signal Processing*. His research fields are affective computing and machine learning.

Björn W. Schuller (bjoern.schuller@imperial.ac.uk) received his Ph.D. degree from the Technical University of Munich, Germany. He is a professor of artificial intelligence in the Department of Computing at the Imperial College London, London, SW7 2AZ, U.K., and a full professor and head of the Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg, Germany. He is the field chief editor of *Frontiers in Digital Health*, president emeritus of the Association for the Advancement of Affective Computing, a Fellow of IEEE, a Golden Core awardee of the IEEE Computer Society, a Fellow of the International Speech Communication Association, and a senior member of the Association for Computing Machinery. He has (co)authored five books and more than 1,000 publications in peer-reviewed books, journals, and conference proceedings.

References

[1] V. Patel et al., "The Lancet Commission on global mental health and sustainable development," *Lancet*, vol. 392, no. 10157, pp. 1553–1598, Oct. 2018. doi: 10.1016/S0140-6736(18)31612-X.

[2] J. Kim-Cohen, A. Caspi, T. E. Moffitt, H. Harrington, B. J. Milne, and R. Poulton, "Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective-longitudinal cohort," *Arch. General Psychiatr.*, vol. 60, no. 7, pp. 709–717, July 2003. doi: 10.1001/archpsyc.60.7.709.

[3] C. Woolston, "PhDs: The tortuous truth," *Nature*, vol. 575, no. 7782, pp. 403–407, Nov. 2019. doi: 10.1038/d41586-019-03459-7.

[4] J. A. Andrews, M. P. Craven, J. Jamnadas-Khoda, A. R. Lang, R. Morriss, C. Hollis, and RADAR-CNS Consortium, "Health care professionals' views on using remote measurement technology in managing central nervous system disorders: Qualitative interview study," *J. Med. Internet Res.*, vol. 22, no. 7, pp. 1–12, July 2020. doi: 10.2196/17414.

[5] K. Woodward, E. Kanjo, D. Brown, T. M. McGinnity, B. Inkster, D. MacIntyre, and T. Tsanas, "Beyond mobile apps: A survey of technologies for mental well-being," *IEEE Trans. Affective Comput.*, early access, July 2020. doi: 10.1109/TAFFC.2020.3015018.

[6] J. E. Bardram and A. Matic, "A decade of ubiquitous computing research in mental health," *IEEE Pervasive Comput.*, vol. 19, no. 1, pp. 62–72, Feb. 2020. doi: 10.1109/MPRV.2019.2925338.

[7] "The WHO special initiative for mental health (2019-2023): Universal health coverage for mental health," World Health Organization, Tech. Rep., 2019. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/310981/WHO-MSD-19.1-eng.pdf

[8] D. Bzdok and A. Meyer-Lindenberg, "Machine learning for precision psychiatry: Opportunities and challenges," *Biol. Psychiatr., Cogn. Neurosci. Neuroimag.*, vol. 3, no. 3, pp. 223–230, Dec. 2018. doi: 10.1016/j.bpsc.2017.11.007.

[9] E. Anthes, "Pocket psychiatry: Mobile mental-health apps have exploded onto the market, but few have been thoroughly tested," *Nature*, vol. 532, no. 7597, pp. 20–24, Apr. 2016. doi: 10.1038/532020a.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. doi: 10.1038/nature14539.

[11] K.-J. Oh, D. Lee, B. Ko, and H.-J. Choi, "A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation," in *Proc. 18th IEEE Int. Conf. Mobile Data Manage. (MDM)*, Daejeon, Korea, 2017, pp. 371–375. doi: 10.1109/MDM.2017.64.

[12] D. Bone, C.-C. Lee, T. Chaspari, J. Gibson, and S. Narayanan, "Signal processing and machine learning for mental health research and clinical applications," *IEEE Signal Process. Mag.*, vol. 34, no. 5, pp. 196–195, Sept. 2017. doi: 10.1109/MSP.2017.2718581.

[13] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: A scoping review," *Transl. Psychiatr.*, vol. 10, no. 1, pp. 1–26, Apr. 2020. doi: 10.1038/s41398-020-0780-3.

[14] T. Guo, "Cloud-based or on-device: An empirical study of mobile deep inference," in *Proc. IEEE Int. Conf. Cloud Eng. (IC2E)*, Orlando, FL, 2018, pp. 184–190. doi: 10.1109/IC2E.2018.00042.

[15] D. Chisholm, K. Sweeny, P. Sheehan, B. Rasmussen, F. Smit, P. Cuijpers, and S. Saxena, "Scaling-up treatment of depression and anxiety: A global return on investment analysis," *Lancet Psychiatr.*, vol. 3, no. 5, pp. 415–424, May 2016. doi: 10.1016/S2215-0366(16)30024-4.

[16] Z. Zhang, N. Cummins, and B. Schuller, "Advanced data exploitation in speech analysis: An overview," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 107–129, July 2017. doi: 10.1109/MSP.2017.2699358.

[17] S. Yang, P. Zhou, K. Duan, M. S. Hossain, and M. F. Alhamid, "emHealth: Towards emotion health through depression prediction and intelligent health recommender system," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 216–226, Sept. 2018. doi: 10.1007/s11036-017-0929-3.

[18] J. Han, Z. Zhang, Z. Ren, and B. Schuller, "EmoBed: Strengthening monomodal emotion recognition via training with crossmodal emotion embeddings," *IEEE Trans. Affective Comput.*, early access, July 2019. doi: 10.1109/TAFFC.2019.2928297.

[19] P. F. Liddle, E. T. Ngan, S. L. Caissie, C. M. Anderson, A. T. Bates, D. J. Quested, R. White, and R. Weg, "Thought and language index: An instrument for assessing thought and language in schizophrenia," *Br. J. Psychiatr.*, vol. 181, no. 4, pp. 326–330, Oct. 2002. doi: 10.1192/bjp.181.4.326.

[20] M. Cheung, J. Shi, O. Wright, L. Y. Jiang, X. Liu, and J. M. Moura, "Graph signal processing and deep learning: Convolution, pooling, and topology," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 139–149, Nov. 2020. doi: 10.1109/MSP.2020.3014594.

[21] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52,138–52,160, Oct. 2018. doi: 10.1109/ACCESS.2018.2870052.

[22] N. Palmius, A. Tsanas, K. Saunders, A. C. Bilderbeck, J. R. Geddes, G. M. Goodwin, and M. De Vos, "Detecting bipolar depression from geographic location data," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1761–1771, Oct. 2016. doi: 10.1109/TBME.2016.2611862.

[23] Z. Zhao, Z. Bao, Z. Zhang, J. Deng, N. Cummins, H. Wang, J. Tao, and B. Schuller, "Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 423–434, Nov. 2019. doi: 10.1109/JSTSP.2019.2955012.

[24] Y. Gal, "Uncertainty in deep learning," Ph.D. thesis, Univ. of Cambridge, Cambridge, U.K., 2016.

[25] R. Harper and J. Southern, "A Bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Trans. Affective Comput.*, early access, Mar. 2020. doi: 10.1109/TAFFC.2020.2981610.

[26] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "Model compression and acceleration for deep neural networks: The principles, progress, and challenges," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 126–136, Jan. 2018. doi: 10.1109/MSP.2017.2765695.

[27] B. Sun, S. Cao, D. Li, J. He, and L. Yu, "Dynamic micro-expression recognition using knowledge distillation," *IEEE Trans. Affective Comput.*, early access, Apr. 2020. doi: 10.1109/TAFFC.2020.2986962.

[28] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Mach. Intell.*, vol. 2, no. 6, pp. 305–311, June 2020. doi: 10.1038/s42256-020-0186-1.

[29] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020. doi: 10.1109/MSP.2020.2975749.

[30] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "Fedhealth: A federated transfer learning framework for wearable healthcare," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, July 2020. doi: 10.1109/MIS.2020.2988604.

SP