B978-0-12-822884-5.00005-2, 00005

Chapter 15

c0025 # Application of big data and artificial intelligence approaches in diagnosis and treatment of neuropsychiatric diseases

Qiurong Song[a], Tianhui Huang[a], Xinyue Wang[a], Jingxiao Niu[a], Wang Zhao[b], Haiqing Xu[c], and Long Lu[a]

[a]*School of Information Management, Wuhan University, Wuhan, PR China,* [b]*Suzhou Zealikon Healthcare Co., Suzhou, PR China,* [c]*Child Health Division, Department of Maternal and Child Health, Maternal and Child Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, PR China*

s0010 ## 1 Introduction

p0010 In recent years, big data technology has been increasingly applied to biomedical and medical information research, and large amounts of biological and clinical data have been generated and collected at an unprecedented speed and scale. For example, the next-generation sequencing (NGS) technologies enable the processing of billions of nucleotides data per day, and the application of electronic health records (EHRs) is documenting huge amounts of patient data. With the emergence of data capture and data generating technologies, and the development of hardware and software for parallel computing, the cost of obtaining and analyzing biomedical data is expected to be significantly reduced.

p0015 The diagnosis and treatment of neuropsychiatric diseases has received increasing attention and has become one of the most difficult challenges in modern medicine. According to a recent report by the World Health Organization (WHO), 50 million people suffer from epilepsy and 24 million from Alzheimer's disease and other dementias. As a discipline to study and prevent mental disorders, brain disorder has long relied on subjective observations for research. Unlike many traditional "physical" diseases, mental illnesses usually have more obvious physiological characteristics and clinical manifestations, but due to the complexity of brain and mind, there are greater difficulties in the assessment and treatment of mental illnesses. Until now, most

Big Data in Psychiatry and Neurology. https://doi.org/10.1016/B978-0-12-822884-5.00005-2

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

of the diagnosis of mental illness has been made on the basis of clinical observations and statistical analysis of symptoms and treatment responses. However, in recent years, with the rise of computational psychiatry, problems such as inaccurate diagnosis (other clinical issues include automatic diagnosis, prediction of treatment outcomes, longitudinal disease progression, and treatment options) may be easier to solve in clinical settings (Qiao, Lin, Cao, & Wang, 2015; Shimada, Shiina, & Saito, 2000). Computational psychiatry uses powerful data analysis, machine learning, and artificial intelligence to detect extreme and abnormal behaviors. On one hand, applying machine learning methods to higher-dimensional data through a data-driven approach can improve disease classification, treatment outcomes prediction, or treatment options (Huys, Maia, & Frank, 2016). On the other hand, the application of artificial intelligence (AI) to the prevention, early screening, diagnosis, treatment, and rehabilitation of mental illnesses can assist doctors, significantly improving the work efficiency of psychiatrists, reducing costs, and complementing the disadvantages of artificial diagnosis and treatment. Furthermore, it can also avoid the subjectivity and bias of psychiatrists. For example, machines can learn directly from medical data to avoid clinical errors caused by human cognitive biases so that a positive impact on patients' diagnosis and treatment could be brought. Today, modern technology and systems enable neurologists to offer appropriate neurological care. Current diagnostic techniques, such as magnetic resonance imaging (MRI), electroencephalography (EEG), generate large amounts of data (size and dimension) for the detection, monitoring, and treatment of neuropsychiatric diseases. A recent study showed that using deep neural networks (DNN) as a computational framework, the AI-driven automatic classification system MNet successfully classified healthy subjects and subjects with two neurological diseases with high accuracy (Aoe et al., 2019). The technique will also be used in the evaluation of the diagnosis, prognosis, and therapeutic effect of various neuropsychiatric diseases.

p0020 The development of AI technology and big data enables us to shift from studying biomarkers at the group level to using brain imaging to predict results at the individual level, which means that the diagnostic methods can be improved by the large amount of studying on single patient's data, different from the extract of covert and common biomarkers in various images. Machine learning can objectively and accurately model massive and multidimensional imaging data, so as to quantify the degree of abnormal brain anatomy and function caused by neuropsychiatric diseases, which is conducive to the diagnosis and prognosis biology of mental diseases and neurological diseases.

p0025 The development of relevant optimization techniques also improves the accuracy of machine learning results. By applying nonlinear hierarchy structures (trees or graphs), some methods involving AI can establish very complex data models. Therefore, in terms of brain image-based disease classification, AI methods can greatly improve the accuracy, sensitivity, and specificity of computer-aided diagnosis (CAD) systems. With the rapid development of

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

medical technology, most diseases can be diagnosed by advanced equipment. Driven by big data technology, this cloud computing-based "AI" will be more useful through the "cloud computing" method of in-depth excavation of medical pathological images and continuous depiction of the pathological regularity "Facebook." Scientific and reasonable results with high feasibility can further promote the accuracy and authority of diagnosis results.

p0030 Moreover, the use of neural network algorithms in the diagnosis of mental illnesses and neurological diseases has also become much more prevalent in the last few years. Neural network algorithms are featured by autonomous learning and complex feature analysis. By using neural network technology to analyze medical data sets, highly abstract diagnostic feature sets can be obtained. In addition, by using multiple data processing techniques in the generated diagnostic feature sets to build neural network models that combined with the patient's vital sign data set including images and other diagnostic criteria, the autonomous learning of the electronic medical record system (EMR) and medical imaging system can be gradually realized, and the diagnostic accuracy of doctors can also be improved.

p0035 This chapter focuses on the introduction of clinical data sources like genomics, EEG signals, eye movement data, and neuroimaging, and some particular devices like wearable equipment would be firstly introduced in the third part. Main algorithms used in the diagnosis and prognosis of different brain disorders would also be referred in the third part, including linear and nonlinear machine learning methods and neural networks. The following section provides an in-depth look at the data types and specific algorithms mentioned before, which are used in areas such as autism spectrum disorders (ASD) and Alzheimer's disease. Finally, we will depict the challenges and promising solutions of big data and AI in neuropsychiatric diseases.

s0015 2 Main data sources

s0020 2.1 Genomics

p0040 Genomic data refer to the genome and DNA data of an organism. They are used in bioinformatics for collecting, storing, and processing the genomes of living things. Genomic data are primarily used in big data processing and analysis techniques, which generally require a large amount of storage and purpose-built software to analyze. Such data are gathered by a bioinformatics system or a genomic data processing software. Typically, genomic data are processed through various data analysis and bioinformatics techniques to find and analyze genome structures and other genomic parameters. Sequencing data analysis techniques and variation analysis are common processes performed on genomic data (Langmead & Nellore, 2018). And genetic counseling is important for the diagnosis of diseases and usually has therapeutic significance. In addition, the related research of genomics can not only be used in the development of related

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

drugs, but also further improve the psychotherapy of mental diseases, including various psychotherapies (such as cognitive behavioral therapy and dialectical behavioral therapy), light therapy, deep-brain stimulation, electroconvulsive therapy, and transcranial magnetic stimulation (TMS), and it will optimize the individualized nursing of patients and improve the prognosis of mental illness (Kumsta, 2019).

2.2 EEG signals

Electroencephalography (EEG) is an electrophysiological monitoring method that records the electrical activity of the brain. It is usually noninvasive, with electrodes placed along the scalp, although invasive electrodes are sometimes used, as in the electrocortical cortex. EEG measures the electrical currents from the generated voltage fluctuations of the ions inside the neuron's brain, which is a spontaneous, rhythmic potential change produced by cortical neurons. Clinically, EEG is the recording of spontaneous electrical activity in the brain over a period of time by means of multiple electrodes placed on the scalp. Nowadays, routine EEG examination in clinic has been very popular. With the rapid development of electronic technology, the application of EEG has become an important index to evaluate the state of brain function and is widely used in the diagnosis of mental diseases (Buettner, Rieg, & Frick, 2020). EEG is most commonly used to diagnose epilepsy, which causes abnormalities in EEG readings. It is also used to diagnose sleep disorders, coma, encephalopathies, and brain death (Sazgar & Young, 2019).

2.3 Eye movement data

Eye movement technology extracts fixation point, fixation time, eye-jumping distance, pupil size, and other data from eye-movement track records. These data can be used to study the internal cognitive processes of individuals. Nowadays, eye movement research has become an important tool for basic neuroscience research. Measures of eye movement have been used for higher brain functions, such as cognition, social behavior, and higher levels of decision-making. With the development of eye tracker, more and more studies have applied eye movement data to mental disorders and found that the basic eye movement characteristics of patients with mental disorders are different from those of healthy controls. Currently, there are many ways to track eye movements, among which the most common noninvasive method is to obtain the position of eyes through video shooting equipment. In the clinical diagnosis of schizophrenia, ASD, and other mental diseases, eye movement data, including the asymmetry of the speed of visually guided saccades Index, plays a significant part (Shiino et al., 2020).

B978-0-12-822884-5.00005-2, 00005

s0035 2.4 Neuroimaging data

p0055 Medical imaging has long been a significant part of medical big data. Among them, neuroimaging plays an important role in the diagnosis and prognosis of patients with brain diseases, providing more data supports for personalized medical treatment. Neuroimaging or brain imaging is the application of various techniques to either directly or indirectly image the structure, function, or pharmacology of the nervous system (Bzdok, Schulz, & Lindquist, 2019). In particular, imaging genetics, which uses imaging as a quantitative biological phenotype, combines genetics, psychiatry, and neuroscience. Neuroimaging genetics can be used to correlate genetic changes with measurable and repeatable results of brain structure or function. Two main methods of neuroimaging genetics are: (1) to identify the imaging changes in the population with definite genetic diseases; (2) to verify the impact of specific genetic changes. A particularly useful aspect of using imaging to define a phenotype is the ability to determine the distance between a diagnosed disease and a subjective self-report inconsistency to track a patient's prognosis. Most importantly, neuroimaging can better understand the pathology of common mental diseases and provide personalized diagnosis and treatment programs by evaluating the relationship between image phenotype and genetic variation (Chiesa, Cavedo, Lista, Thompson, & Hampel, 2017; Uddén et al., 2019).

s0040 2.5 Wearable equipment data

p0060 The Internet of things (IOT) extends the independence of human interaction, contribution, and cooperation in things. The Internet of things has gradually developed heterogeneous technologies with complex protocols and algorithms. In the process of personalized prognosis of mental and neurological disorders, wearable or implanted sensors can continuously track the physiological characteristics of discharged patients, such as blood pressure, heart rate, body temperature, stress rate, electrocardiogram (ECG), EEG, and other physiological states, and it can also continuously provide various data to the cloud, offering more personalized prognosis for mental patients to support patients in emergency decision-making and emergency contact (McGinnis et al., 2018). Physiological information is monitored by human body sensors and biosensors, such as body temperature, diabetes, blood pressure, and heart rate. For example, a smart watch has built-in thermometer, oximeter, accelerometer, and GPS position tracker. Those sensors embedded in clothing track electrocardiograms, EMGs, and patient stress rates. Through the network, the observed data are transmitted to the cloud through network to store and send messages, and then they are transmitted to the client, i.e. doctors' mobile phones, family members, and hospitals (Chen et al., 2017).

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

s0045 3 Main algorithms

p0065 There are a lot of algorithms used in neuropsychiatric diseases, mainly including machine learning, signal time–frequency analysis, feature extraction, and so on. In this paper, the principle of support vector machine, decision tree, random forest, convolution neural network, wavelet transform algorithm in time–frequency analysis, and common space pattern algorithm in feature processing are described in detail.

p0070 Support vector machine (SVM) is a supervised learning algorithm. It is a binary classification algorithm supported by linear classification and nonlinear classification, but after evolution, it also supports multiclassification problem and can also be applied to regression problem. SVM is based on the VC dimension theory of statistical learning theory and the principle of structural risk minimization. According to the limited sample information, it seeks the best compromise between the complexity of the model (that is, the learning accuracy of a specific training sample, accuracy) and the learning ability (that is, the ability to identify any sample without error), in order to obtain the best generalization ability (or generalization ability) (Guyon, Weston, Barnhill, & Vapnik, 2002).

p0075 Decision tree is also a supervised machine learning method, whose generation algorithms include ID3, C4.5, and C5.0. It is a tree structure, in which each internal node represents a judgment on attribute, each branch represents an output of judgment result, and finally each leaf node represents a classification result. Decision tree can be divided into classification tree and regression tree. Classification tree makes decision tree for discrete variables while regression tree makes it for continuous variables (Pradhan, 2013).

p0080 Random forest is a classifier with multiple decision trees, and the output category is determined by the mode of the output category of the individual tree. Each decision tree is a classifier, so for an input sample, n trees will have n classification results. Random forest integrates all the classification voting results and specifies the category with the most voting times as the final output. It is not sensitive to multivariate common linear, and the results are more robust to missing data and unbalanced data. Furthermore, it can predict the role of up to thousands of explanatory variables well and improve the prediction accuracy without significant improvement in the calculation (Diaz-Uriarte & de Andres, 2006).

p0085 Convolution neural network is made up of three parts. The first part is the input layer. And the second one is composed of several convolution layers and pooling layers. Then the third one consists of a fully connected multilayer perceptron classifier. In the convolution layer of the convolution neural network, a neuron is only connected with some neighboring neurons and usually contains several feature planes. Each feature plane is comprised of some neurons arranged in a rectangle. The neurons in the same feature plane share the weight, which is the convolution kernel. Convolution kernel can reduce the connection between all layers of the network and the risk of overfitting (Lessmann et al., 2018).

B978-0-12-822884-5.00005-2, 00005

p0090 In wavelet transform, the infinite trigonometric function basis in Fourier transform is replaced by the finite attenuation wavelet basis. In this way, not only the frequency can be obtained, but also the time can be located. Wavelet transform has two variables: scale a and translation τ. Scale a controls the expansion and contraction of wavelet function, while translation τ controls the translation of wavelet function. The scale corresponds to the frequency (inverse ratio), while the translation τ corresponds to the time (Adeli, Zhou, & Dadmehr, 2003).

p0095 Common spatial pattern (CSP) is a spatial filtering feature extraction algorithm for two classification tasks, which can extract spatial distribution components of each category from multichannel BCI data. The basic principle of the common spatial pattern algorithm is to find a group of optimal spatial filters to project by using the diagonalization of the matrix, so as to maximize the difference between the variance values of the two types of signals, thus obtaining the eigenvectors with high differentiation (Ang, Chin, Zhang, & Guan, 2008).

s0050 4 Applications

p0100 Medical data types and machine learning algorithms commonly used in the field of neuropsychiatry are briefly introduced before. This section will describe the application of big data and AI algorithms in neuropsychiatric diseases from the perspectives of early warning, diagnosis, and prognosis. The major brain disorders involved include ASD, Alzheimer's disease, depressive disorders, and schizophrenia.

s0055 4.1 Early warning

p0105 Most mental diseases are caused by a variety of causes, and their clinical characteristics are diverse. The onset of diseases is often sudden and uncertain, and the effectiveness of treatment varies from person to person. Therefore early warning of diseases through the collection of physiological data can systematically monitor the patient's condition and directly provide a basis for the follow-up nursing decisions, which can buy more time for the treatment of doctors. In that case, the incidence and mortality of mental diseases can be reduced, and the harm caused by them can be greatly moderated. In order to achieve the accurate effect of early warning, we need to rely on a large amount of information collected by various tools. Only through further analysis of big data can we get specific early warning results. Early warning of epilepsy is the most popular application in the early warning through physiological data.

p0110 Epilepsy is a chronic brain dysfunction syndrome caused by various causes. Abnormal discharge of brain nerve cells leads to brain dysfunction and thus epilepsy, which is the clinical manifestation of paroxysmal abnormal hypersynchronous electrical activity of brain nerve cells. It is characterized by repeatability, abruptness, and temporality. Nearly 25% of the world's 50 million

B978-0-12-822884-5.00005-2, 00005

people with epilepsy still have no effective treatments. Due to the sudden and irregular nature of epileptic seizures, early warning of epileptic seizures is the key to break the bottleneck of epileptic treatment, as it can buy patients and doctors time to take appropriate protective measures to avoid accidental injuries caused by sudden seizures.

p0115 For collection of patients with intracranial EEG, cortex EEG signals, such as power spectrum data index, epilepsy early warning system uses the machine learning algorithms such as convolution algorithm of neural network to distinguish between the normal EEG and epileptic seizure EEG, and between epileptic seizure phase and epileptic seizure EEG, allowing doctors to access and track patients' data at any time through the system and then deal with the treatment of late decisions. It not only helps to reduce the rate of seizures and mortality, but also helps to further reveal the regularity and physiological basis of seizures.

s0060 4.2 Diagnosis

s0065 *4.2.1 EEG signals used in the diagnosis of autism*

p0120 Brain electrical activity is a spontaneous and rhythmic potential change generated by cortical neurons. Nowadays, routine clinical electroencephalography has long been pretty popular. And with the rapid development of electronic technology, EEG has become increasingly widely applied. It has become an important index to evaluate the state of brain function and is widely used in the diagnosis of mental diseases. The data output of this kind of data is very large, which provides a sufficient data source for the diagnosis of mental diseases. Among them, it is most commonly used in the field of autism.

p0125 Autism is a highly inherited neurological disorder characterized by psychological symptoms, including repetitive and stereotypical behaviors, as well as a lack of emotional expression, verbal, and social communication skills. It is a mental illness that contains a range of complex neurodevelopmental disorders. Due to the heterogeneity and clinical diversity of autism, the diagnosis method mainly relies on the diagnostic medical staff's judgment of the patient's emotion, language, and social behavior, which is a kind of subjective analysis. However, since its pathophysiology has not yet been established and the existing diagnostic methods are highly subjective, medical staffs will have to work harder to improve the diagnostic accuracy.

p0130 EEG signals can detect the electrical activity of neurons in the brain by traversing the electrodes on the scalp, which has the advantages of no damage, high temporal resolution, and real-time monitoring (Wang et al., 2013). As the real-time data collection of brain electrical signal is so large that it is impossible to accurately analyze all the data according to the subjective judgment of medical staffs, so it is an inevitable trend to apply artificial intelligence technology to data analysis. In addition, existing studies have found that the EEG phase of autistic patients is different from that of normal people. The EEG power

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

spectrum of autistic patients is different between groups in the cerebral hemisphere and between hemispheres. The EEG signals of autistic patients are located in the frontal lobe, significantly decreased in the brain area of the back, reflecting the cognitive decline of autistic patients. The EEG signals of autistic patients are located in the frontal lobe (Daoust, Limoges, Bolduc, Mottron, & Godbout, 2004), which are significantly increased in the parietal lobe area and significantly weakened (Burnette et al., 2011) in the back area of the brain, reflecting the cognitive decline of autistic patients. The characteristic values of children's EEG signals can be extracted by artificial intelligence technology and compared with those of autistic patients. In this way, the high-precision diagnosis of autism can be realized and the work burden of medical workers can be effectively reduced.

p0135 Due to certain behaviors in social situations, it is difficult for autistic patients to keep quiet in the process of data collection. However, the acquisition of resting EEG signals has a high tolerance for the activities of the subjects and can well adapt to the behavioral changes of autistic patients without causing too many errors. Up to now, the diagnostic accuracy of EEG has reached 80%.

s0070 *4.2.2 Electroencephalogram and MRI used in diagnosis of Alzheimer's disease*

p0140 Alzheimer's Disease (AD) is an irreversible degenerative neuropsychiatric disease. Because the etiology of Alzheimer's disease has not been fully elucidated and there is no effective treatment, early diagnosis and intervention of AD quite essential. At present, as a clinical precursory stage, the early diagnosis of Mild Cognitive Impairment (CMI) has also attracted more attention in the academic community. The common clinical diagnosis of AD and MCI on account of neuroimaging is based on, for example, MRI and Positron Emission Computed Tomography (PET). And common analysis methods for neuroimaging are mainly based upon image feature markers (such as hippocampal volume, cortical thickness, etc.), or using diagnostic methods on the basis of machine learning and deep learning to construct diagnostic algorithms.

p0145 At present, the AI diagnosis process for AD includes finding participants, cleaning image data, standardizing classifiers and classification criteria, and extracting as well as recognizing biological features, so as to achieve the classification of different types of AD and MCI. The basic approaches involve SVM, whose aim is to find a predictive model which is able to perform binary group separation. Some researches also use Naive Bayes or Random Forest for pattern recognition models, in terms of Voxel-based diagnosis on functional magnetic resonance imaging (fMRI) (Armananzas, Iglesias, Morales, & Alonso-Nanclares, 2017). Although there are minimal differences when using different voxel subset selection and classification paradigms, the quantitative results suggest that classification models are highly applicable in clinical scenarios.

B978-0-12-822884-5.00005-2, 00005

p0150 In this field, the most widely studied and promising method is prediction based on Convolutional Neural Networks (CNN). CNN is a kind of feed-forward neural network that shows excellent performance in large-scale image processing, including 1D CNN, 2D CNN, and 3D CNN. Traditional machine learning methods usually need to manually check out areas of the brain where AD patients have more obvious changes, such as hippocampus, amygdala, and other regions of interest (ROI). However, CNN and other Deep Learning methods can achieve autonomous learning of feature representations to manage image classification, and then diagnose and classify patients with Alzheimer's disease or mild cognitive impairment. For example, Suk, Lee, and Shen (2014) used pairs of MRI and PET images from ADNI to execute Patch-Level Deep Feature Learning based on Deep Boltzmann Machine (DBM). Increasing evidences have proved that biomarkers from various modalities can provide more complementary information in AD/MCI diagnosis, so multimodal deep feature fusion for different kinds of images is carried out under the joint efforts of Deep Learning methods.

p0155 In addition, deep machine learning algorithms such as Stacked Autoencoders, Noise Reduction Autoencoders, and Deep Belief Networks (DBN) are also gradually applied to the clinical diagnosis of Alzheimer's disease, which usually has high robustness and accuracy (Armananzas et al., 2017). Although most of the related Artificial Intelligence methods require a large amount of data and images for training, they can effectively avoid physicians' personal decision errors and limitations, and multimodal analysis of multiple imaging data can also greatly improve the accuracy and reliability of diagnosis results.

s0075 *4.2.3 Eye movement data used in diagnosis of schizophrenia*

p0160 Eye movement technology is to extract such data as fixation point, fixation time, eye-hop distance, pupil size, and so on from the records of eye movement trajectory, so as to study the internal cognitive process of individuals. Nowadays, eye movement data is increasingly used in detection and classification of mental disorders, combined with AI.

p0165 Schizophrenia is a relatively common psychiatric disease, which is clinically manifested as a syndrome of multiple symptoms. As early as the 1970s, relevant researchers have found that eye movements of patients with schizophrenia are significantly different from normal people. Therefore, in the clinical diagnosis of schizophrenia, eye movement data (including the asymmetry of the speed of visually guided saccades Index, and so on) have become important monitoring indicators. With the widespread application of AI and Big Data technologies in the field of mental illness, eye movement data processing in the diagnosis of schizophrenia has gradually been automatized and intelligentized, and outstanding results have been achieved.

p0170 In the clinical diagnosis of schizophrenia, eye movement data is often directly obtained from medical equipment such as electro-optic eye tracker,

B978-0-12-822884-5.00005-2, 00005

corneal reflection eye tracker, or pupil-corneal reflection vector method medical eye tracker. Data diagnosis mostly plays a role in auxiliary research. Li Yu from Shanghai Jiao Tong University has used principal component analysis in machine learning combined with support vector machine (PCA-SVM) algorithm to establish a diagnostic discrimination model, aiming at using saccade behavior and reverse saccade behavior amplitude in eye movement data to treat schizophrenia patients. The accuracy of classification, diagnosis, and treatment of patients with latency can reach more than 80% (Yu, 2017).

p0175 In addition, the integration of eye movement data and the chronological comparison before and after the illness greatly improves the accuracy of diagnosis. In specific researches, eye movement data can also be combined with EEG data to construct a brain function network based on multilayered complex neural networks, which plays an essential role in the pathological investigation and prognosis of schizophrenia.

p0180 Big Data relevant methods have been widely used in eye movement data analysis. Particular machine-learned computational models (MLCMs) can be used in eye tracking to model reading comprehension (D'Mello & Southwell, 2020), and also, huge amount of eye movement data can assist deep learning models to establish visual sensory–perceptual–cognitive dynamical systems (Assadi et al., 2018). Although big data methods have not yet been applied to diagnosis or early warning of neuropsychiatric diseases detected by eye movement data, eye movement still enjoys significance and intuition as parameter for schizophrenia and depression, etc. It is believed that big data methods would be used in the analysis of eye movement data and give assistance to clinical neuropsychiatric diseases.

p0185 In general, big data or AI which uses eye movement data in the diagnosis of schizophrenia mostly involve saccade research, but do not go deep into smooth tracking or other aspects, so there is still a lot of space for development. For example, the results of exploratory eye movement monitoring can be used as both characteristics and state markers, and the response scores can be used as important biological markers for the diagnosis of schizophrenia.

s0080 *4.2.4 Genomics data used in diagnosis of autism*

p0190 Recent studies have revealed that genetic neurological channelopathies can lead to many different neurological diseases. In general, neurofunctional diseases are inherited in an autosomal dominant manner and cause paroxysmal disorders of neurologic function. These diseases are rare in some cases, but accurate diagnosis is important because they have the function of genetic counseling and usually have therapeutic significance (Spillane, Kullmann, & Hanna, 2016).

p0195 Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder with a strong genetic basis. Evidence accumulated over recent years has shown the correlation of hundreds of gene variants in autism. Yet, only a small fraction of potentially causal genes—about 65 genes out of an estimated several

B978-0-12-822884-5.00005-2, 00005

hundred—are known with strong genetic evidence from sequencing studies. The effects of gene variants are highly variable, and they are usually related to other disorders besides autism. These findings suggest that genetic heterogeneity, mutation penetrance, and genetic pleiotropy are common features of autism.

p0200 Researchers have developed an evidence-weighted machine learning approach on the basis of a human brain-specific gene network to present a genome-wide prediction of autism risk genes and identified likely pathogenic genes within frequent autism-associated copy-number variants (Krishnan et al., 2016).

p0205 In recent years, SNV identification has also made great progress, so that the whole-genome sequencing (WGS) and the whole-exome sequencing (WES) have become feasible alternative methods for selective genotyping. Through the identification of novel and rare variants, it has the significance of risk prediction and diagnosis, and is applied to the treatment of autism and other neuropsychiatric diseases (Carter & Scherer, 2013).

s0085 4.3 Prognosis

p0210 The central theme of personalized medicine is to adopt specific therapies according to the unique physiological characteristics of individuals. Such characteristics mainly include: genetic alterations and epigenetic modifications, clinical symptomatology, biomarker changes, and environmental factors (Ozomaro, Wahlestedt, & Nemeroff, 2013). Therefore the primary objective of personalized medicine is to predict the individual's susceptibility to disease, to achieve accurate diagnosis, and to optimize the most effective and beneficial response to treatment. The realization of personalized medicine in psychiatry can promote a significant reduction in incidence rate and mortality rate.

p0215 Pharmacogenomics involves the customization of therapies through individual genetic makeup, rational drug development, and drug reuse. Several large research teams related to electronic health record (EHR) data have developed gradually. Like new genetic variants that predict drug action have been discovered in many ways, supporting Mendelian randomized trials, showing drug efficacy, and suggesting new indications for existing drugs. Big data approaches may also help identify subtypes of disease, which could suggest differential treatment and prognoses.

p0220 Evidence from recent studies indicates that genetic targets often indicate effective drug targets. The subject of most of these studies was genome-wide association analysis (GWAS), representing drug targets with significant impact on the disease or trait. Electronic health record (EHR) is an effective tool for evaluating the efficacy of drugs. Pharmacogenomics research using EHR data can more efficiently obtain the "actual" situation, complications, related drugs, and long-term therapeutic effect of drug use in patients, so as to better discover the clinical effect of pharmacogenomic interactions and make better use of

B978-0-12-822884-5.00005-2, 00005

patient data. Networks including the eMERGE network have established a strong track record for accurately identifying disease and drug response phenotypes from EHRs, and a published algorithm with a median positive predictive value (PPV) of >95% (Kirby et al., 2016).

p0225 Depressive disorders (DDs) are a common disease in mental disease pathology. About 350 million people worldwide are affected by the condition, according to the World Health Organization. Genetic factors contribute significantly to the risk of DDs and more than 20 genes are linked to episodes of depression (Schosser & Kasper, 2009). Through genome-wide association analysis (GWAS) and electronic case data, a large number of samples of patients and healthy people were used in the algorithm to obtain statistically significant results (Cai et al., 2015; Sullivan et al., 2009). Research has shown that we are able to assess substantial pharmacogenetic traits with accurate performance even if there is not manual review, just as for algorithms of pharmacogenomics. Similarly, the response of antidepressants to major depression also varies with specific genetic changes in certain candidate genes (Shadrina, Bondarenko, & Slominsky, 2018). The associated genetic changes affect not only MDD susceptibility, but also the effectiveness of antidepressant treatment. Thus multiple linear regression was used to generate the algorithm by using the stable dosage of the drugs for depressed disease as the dependent variable and the clinical and genotypic variables as the independent variables. These new applications of biomedical informatics, machine learning, and statistical knowledge have improved the ability to interpret clinical information to identify patients' complex phenotypes and subsets, and to tailor medication and other treatment regimens, accelerating the development of sophisticated medicine.

p0230 This approach has also been shown in personalized medicine for other diseases. For example, coumarins, the most widely used anticoagulants for the treatment and prevention of thromboembolism, vary greatly from individual to individual in the dose required to achieve stable anticoagulation. The multiple linear regression algorithm can be used to predict effectively and improve the dose selection of acenocoumarol (Tong et al., 2016) (Table 1).

s0090 5 Challenges and promising solutions

s0095 5.1 Privacy and security of patient information

p0235 When we get a large amount of data to help patients with personalized prognosis, the privacy of patients including personal details, physiological information, diagnosis reports, and treatment plans will be easily exposed. The storage and use of these data in the cloud may cause the leakage of patient's privacy and bring unnecessary trouble for patients. The security of data is based on the previously mentioned data privacy. When the user's personal privacy is disclosed, the hacker can modify the sensitive information, including health information, so as to cause misdiagnosis or incorrect disease evaluation, further

B978-0-12-822884-5.00005-2, 00005

t0010 **TABLE 1 Briefly summarizes the use of big data and artificial intelligence technologies in early warning, diagnosis, and prognosis, listing the types of data commonly used and the main algorithms used in studies of related diseases, respectively, and the studies of neuropsychiatric diseases listed are presented specifically in the previous paper.**

Disease	Data type	Main algorithm	Application
Epilepsy	Electroencephalogram	Convolutional neural networks	Early warning
Autism	Electroencephalogram	Wavelet transform, common spatial pattern, autoregressive model	Diagnose
Autism	Genomics	Evidence-weighted machine learning approach	Diagnose
Depressive disorders (DDS)	Genomics, electronic health record	Multiple linear regression	Prognosis
Alzheimer's disease (AD)	Electroencephalogram and MRI	Support Vector Machine, Convolutional Neural Network, Deep Belief Network	Diagnose
Schizophrenia	Eye movement	Neural Network, Principal Component Analysis	Diagnose

leading to improper treatment or prognosis, thus increasing the death rate. The data of patients discharged from hospital monitored by Internet of things devices are prone to generate security risks in the transmission process. It can be said that data security is one of the most important challenges in the personalized prognosis of mental diseases or other diseases. In the process of tracking patients' prognosis, hardware and software are also vulnerable to security threats. In order to ensure the secure communication between Internet of things devices, SSL/TLS protocol is widely used. Encryption technology can be used to provide secure communication, but it also needs more storage and processing capacity, which is a major challenge for wireless sensors and computing devices used in healthcare applications. Developing algorithms and solutions to protect health-related information from unauthorized users is a challenging issue, as design and implementation phase rationale errors can also lead to security risks.

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

p0240 To solve this problem, malicious users should not be miscalculated. This requires the establishment of an authentication mechanism for IoT sensor data on the server side to prevent malicious access by third-party users and monitor the application's end-to-end delay in synchronization of data packets. In the case of patients moving independently while wearing sensors, dynamic network topology can quickly adapt to the security switching mechanism from one network to another. At the same time, in order to achieve high throughput of Internet of Things health monitoring and efficient sensing, both communication costs and computation should be cost effective.

s0100 5.2 Information island

p0245 Information island means the problem that information is not shared with each other and information is disconnected from business processes and applications. It is also an important issue that plagues the prognosis of AI and big data in psychiatric disorders. Data islands mainly refer to that medical data cannot be exchanged and shared among hospitals, regions, and departments. Medical activities will generate a large amount of data, but cannot be fully utilized. There are very few data used to guide medical activities, and the overall planning of resources cannot be achieved, which brings great challenges to the integration of medical big data. At present, big data and AI have brought new medical models, but because of the differences in the nature of ownership, the closeness of cooperation, and benefit orientation, among others, it is easy to create new data "islands." On the one hand, different hospitals, medical enterprises, among others, have their own established information systems, where they store massive diagnostic and prognostic information. Therefore, in the process of medical data collection, the lack of unified standards in medical data means that the majority of medical data cannot be shared and used, resulting in information islands; on the other hand, because medical data inevitably contains personal privacy, it is difficult to achieve full sharing in a real sense. At this stage, a large proportion of hospitals maintain a negative attitude toward data sharing, and it is hard to achieve information exchange between hospitals.

s0105 5.3 Storage and analysis capabilities

p0250 In the process of individualized prognosis, storage and analysis capabilities are the major challenges. The storage problem is mainly caused by the number, speed, and various forms of big data analysis. Big data analysis is difficult to use in traditional storage system directly. Therefore cloud storage systems like Amazon S2, Amazon EC2, and elastic block store provide solutions for big data analysis by providing infinite storage systems with fault tolerance. In addition to storage, another challenge is how to move big data to cloud services at lower cost and faster speed. Most of the work in the cloud relies on Amazon Elastic Compute Cloud (Amazon EC2) to provide scalable computing power in the

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

cloud. Nevertheless, cloud platform cannot be effectively used to meet the rapid growth demand for big data analysis and healthcare information.

s0110 5.4 Lack of specialized personnel

p0255 Similarly, based on the vigorous development of personalized medical field, the scarcity of professionals has become one of the challenges facing the construction of medical big data. More specialized operation puts forward more requirements for high-quality compound talents, who not only need to grasp certain information technology ability and data analysis ability, but also need solid medical basic knowledge to better participate in the processing of medical big data. Faced with the huge amount of medical data resources, the value of big medical data cannot be fully realized without professional personnel dealing with it. At present, most hospitals are seriously short of high-quality composite talents, which seriously hinder the development of big medical data. To this end, more support is needed for the training of composite talents, and more excellent talents are guided to join this field.

s0115 6 Conclusions

p0260 In recent years, big data and AI have been widely used in the field of psychiatry. In this paper, the characteristics of biomedical big data itself are analyzed from genetic data, EEG signals, eye movement data, electroencephalogram data, and wearable device data. The main algorithms, such as support vector machine, decision tree, random forest, convolution neural network, wavelet transform algorithm in time–frequency analysis, and common space pattern algorithm in feature processing, are briefly introduced. Due to the rapid development of this field, biomedical big data is widely used in the early warning, diagnosis, and prognosis of mental diseases. And the major brain disorders involved include autism, Alzheimer's disease, and schizophrenia. Finally, we analyzed the challenges encountered in the process of data application, such as possible patient data security issues, information islands, and other issues, and proposed our solutions.

p0265 The future direction of work is mainly to pay more attention to the protection of patients' privacy, while relying on emerging technologies, including Internet of things technology, knowledge mapping technology, etc. to support doctors' diagnosis, treatment, and prognosis. Moreover, we need to accelerate the integration of existing data resources to build a reliable and accurate model for clinical practice; to exchange and share medical data among hospitals, regions, and departments; and to store and use data in a unified way. At the same time, we also need to enhance the ability of data storage and analysis to meet the growing demand for data, as well as to strengthen the training of professionals, which is also the focus of attention in the future.

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

References

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*, *123*(1), 69–87. https://doi.org/10.1016/s0165-0270(02)00340-0.

Ang, K. K., Chin, Z. Y., Zhang, H. H., & Guan, C. T. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In (Vols. 1–8). *2008 IEEE international joint conference on neural networks* (pp. 2390–2397). IEEE: New York.

Aoe, J., Fukuma, R., Yanagisawa, T., Harada, T., Tanaka, M., Kobayashi, M., … Kishima, H. (2019). Automatic diagnosis of neurological diseases using MEG signals with a deep neural network. *Scientific Reports*, *9*(5057). https://doi.org/10.1038/s41598-019-41500-x.

Armananzas, R., Iglesias, M., Morales, D. A., & Alonso-Nanclares, L. (2017). Voxel-based diagnosis of Alzheimer's disease using classifier ensembles. *IEEE Journal of Biomedical and Health Informatics*, *21*(3), 778–784. https://doi.org/10.1109/JBHI.2016.2538559.

Assadi, A. H., et al. (2018). Deep learning models for visual sensory-perceptual cognitive dynamical systems from eye movement data and categories of natural images. *Investigative Ophthalmology & Visual Science*, *59*(9).

Buettner, R., Rieg, T., & Frick, J. (2020). Machine learning based diagnosis of diseases using the unfolded EEG spectra: Towards an intelligent software sensor. In *Paper presented at the information systems and neuroscience, Cham.*

Burnette, C. P., Henderson, H. A., Inge, A. P., Zahka, N. E., Schwartz, C. B., & Mundy, P. C. (2011). Anterior EEG asymmetry and the modifier model of autism. *Journal of Autism and Developmental Disorders*, *41*(8), 1113–1124. https://doi.org/10.1007/s10803-010-1138-0.

Bzdok, D., Schulz, M.-A., & Lindquist, M. (2019). Emerging shifts in neuroimaging data analysis in the era of "big data". In I. C. Passos, B. Mwangi, & F. Kapczinski (Eds.), *Personalized psychiatry: Big data analytics in mental health* (pp. 99–118). Cham: Springer International Publishing.

Cai, N., Bigdeli, T. B., Kretzschmar, W., Li, Y., Liang, J., Song, L., … Consortium, C. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, *523*(7562), 588–591. https://doi.org/10.1038/nature14659.

Carter, M., & Scherer, S. (2013). Autism spectrum disorder in the genetics clinic: A review. *Clinical Genetics*, *83*. https://doi.org/10.1111/cge.12101.

Chen, M., Ma, Y., Li, Y., Wu, D., Zhang, Y., & Youn, C. (2017). Wearable 2.0: Enabling human-cloud integration in next generation healthcare systems. *IEEE Communications Magazine*, *55*(1), 54–61. https://doi.org/10.1109/MCOM.2017.1600410CM.

Chiesa, P. A., Cavedo, E., Lista, S., Thompson, P. M., & Hampel, H. (2017). Revolution of resting-state functional neuroimaging genetics in Alzheimer's disease. *Trends in Neurosciences*, *40*(8), 469–480. https://doi.org/10.1016/j.tins.2017.06.002.

Daoust, A.-M., Limoges, É., Bolduc, C., Mottron, L., & Godbout, R. (2004). EEG spectral analysis of wakefulness and REM sleep in high functioning autistic spectrum disorders. *Clinical Neurophysiology*, *115*(6), 1368–1373. https://doi.org/10.1016/j.clinph.2004.01.011.

Diaz-Uriarte, R., & de Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, *7*, 13. https://doi.org/10.1186/1471-2105-7-3.

D'Mello, S. K., & Southwell, R. (2020). Machine-learned computational models can enhance the study of text and discourse: A case study using eye tracking to model reading comprehension. *Discourse Processes*, *57*(5–6), 420–440.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422. https://doi.org/10.1023/a:1012487302797.

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238.

Kirby, J. C., Speltz, P., Rasmussen, L. V., Basford, M., Gottesman, O., Peissig, P. L., … Denny, J. C. (2016). PheKB: A catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association*, *23*(6), 1046–1052. https://doi.org/10.1093/jamia/ocv202.

Krishnan, A., Zhang, R., Yao, V., Theesfeld, C. L., Wong, A. K., Tadych, A., … Troyanskaya, O. G. (2016). Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nature Neuroscience*, *19*(11), 1454–1462. https://doi.org/10.1038/nn.4353.

Kumsta, R. (2019). The role of epigenetics for understanding mental health difficulties and its implications for psychotherapy research. *Psychology and Psychotherapy: Theory, Research and Practice*, *92*(2), 190–207. https://doi.org/10.1111/papt.12227.

Langmead, B., & Nellore, A. (2018). Cloud computing for genomic data analysis and collaboration. *Nature Reviews Genetics*, *19*(4), 208–219. https://doi.org/10.1038/nrg.2017.113.

Lessmann, N., van Ginneken, B., Zreik, M., de Jong, P. A., de Vos, B. D., Viergever, M. A., & Isgum, I. (2018). Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Transactions on Medical Imaging*, *37*(2), 615–625. https://doi.org/10.1109/tmi.2017.2769839.

McGinnis, R. S., McGinnis, E. W., Hruschak, J., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., & Muzik, M. (2018). Wearable sensors and machine learning diagnose anxiety and depression in young children. In *Paper presented at the 2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*. 4–7 March 2018.

Ozomaro, U., Wahlestedt, C., & Nemeroff, C. B. (2013). Personalized medicine in psychiatry: Problems and promises. *BMC Medicine*, *11*(1), 132. https://doi.org/10.1186/1741-7015-11-132.

Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, *51*, 350–365. https://doi.org/10.1016/j.cageo.2012.08.023.

Qiao, C., Lin, D. D., Cao, S. L., & Wang, Y. P. (2015). The effective diagnosis of schizophrenia by using multi-layer RBMs deep networks. In J. Huan, S. Miyano, A. Shehu, X. Hu, B. Ma, S. Rajasekaran, V. K. Gombar, I. M. Schapranow, I. H. Yoo, J. Y. Zhou, B. Chen, V.Pai & B. Pierce (Eds.), *IEEE international conference on bioinformatics and biomedicine – BIBM* (pp. 603–606).

Sazgar, M., & Young, M. G. (2019). Encephalopathies, brain death, and EEG. In M. Sazgar, & M. G. Young (Eds.), *Absolute epilepsy and EEG rotation review: Essentials for trainees* (pp. 183–198). Cham: Springer International Publishing.

Schosser, A., & Kasper, S. (2009). The role of pharmacogenetics in the treatment of depression and anxiety disorders. *International Clinical Psychopharmacology*, *24*(6), 277–288. https://doi.org/10.1097/yic.0b013e3283306a2f.

Shadrina, M., Bondarenko, E. A., & Slominsky, P. A. (2018). Genetics factors in major depression disease. *Frontiers in Psychiatry*, *9*, 334. https://doi.org/10.3389/fpsyt.2018.00334.

Shiino, T., Miura, K., Fujimoto, M., Kudo, N., Yamamori, H., Yasuda, Y., … Hashimoto, R. (2020). Comparison of eye movements in schizophrenia and autism spectrum disorder. *Neuropsychopharmacology Reports*, *40*(1), 92–95. https://doi.org/10.1002/npr2.12085.

Shimada, T., Shiina, T., & Saito, Y. (2000). Detection of characteristic waves of sleep EEG by neural network analysis. *IEEE Transactions on Biomedical Engineering*, *47*(3), 369–379. https://doi.org/10.1109/10.827301.

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

Spillane, J., Kullmann, D. M., & Hanna, M. G. (2016). Genetic neurological channelopathies: Molecular genetics and clinical phenotypes. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*(1), 37. https://doi.org/10.1136/jnnp-2015-311233.

Suk, H., Lee, S., & Shen, D. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, *101*, 569–582. https://doi.org/10.1016/j.neuroimage.2014.06.077.

Sullivan, P. F., de Geus, E. J. C., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., … Penninx, B. W. J. H. (2009). Genome-wide association for major depressive disorder: A possible role for the presynaptic protein piccolo. *Molecular Psychiatry*, *14*(4), 359–375. https://doi.org/10.1038/mp.2008.125.

Tong, H. Y., Dávila-Fajardo, C. L., Borobia, A. M., Martínez-González, L. J., Lubomirov, R., Perea León, L. M., … Group, P.-A. I. (2016). A new pharmacogenetic algorithm to predict the most appropriate dosage of acenocoumarol for stable anticoagulation in a mixed Spanish population. *PLoS One*, *11*(3), e0150456. https://doi.org/10.1371/journal.pone.0150456.

Uddén, J., Hultén, A., Bendtz, K., Mineroff, Z., Kucera, K. S., Vino, A., … Fisher, S. E. (2019). Toward robust functional neuroimaging genetics of cognition. *The Journal of Neuroscience*, *39*(44), 8778. https://doi.org/10.1523/JNEUROSCI.0888-19.2019.

Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., & Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, *5*.

Yu, L. (2017). *Characteristics of eye-movement in first-episode schizophrenia and first-episode depression and modelling diagnosis and differential diagnosis for schizophrenia*. Master Degree, Shanghai Jiaotong University.

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

Moustafa_Ahmed, 978-0-12-822884-5

B978-0-12-822884-5.00005-2, 00005

Non-Print Items

Abstract:
In recent years, with the rapid development of big data and artificial intelligence, a large number of biological and clinical data have been generated and collected at an unprecedented speed and scale, which have been increasingly used in biomedical and healthcare informatics research, as well as in the field of neuropsychiatry. This paper introduces the application of big data and artificial intelligence in the field of neuropsychiatry and the related data types and algorithms frequently used in recent years. We took autism, schizophrenia, Alzheimer's disease, epilepsy, and other neuropsychiatric disorders as examples for specific analysis, and summarized the characteristics of the application of big data and artificial intelligence in the early warning, diagnosis, and prognosis of these diseases. Then we analyzed the development trend of big data and artificial intelligence in the field of neuropsychiatry. This chapter further points out that the exchange and sharing of medical data and the improvement of data storage and analysis capability are important directions for the future development of this field.

Keywords: Big data; Artificial intelligence; Neuropsychiatric diseases; Diagnose; Early warning; Prognosis