

Chapter 13

Brain Imaging Methods in Social and Affective Neuroscience: A Machine Learning Perspective



Lucas R. Trambaiolli, Claudinei E. Biazoli Jr, and João R. Sato

Abstract Machine learning (ML) is a subarea of artificial intelligence which uses the induction approach to learn based on previous experiences and make conclusions about new inputs (Mitchell, Machine learning. McGraw Hill, 1997). In the last decades, the use of ML approaches to analyze neuroimaging data has attracted widening attention (Pereira et al., Neuroimage 45(1):S199–S209, 2009; Lemm et al., Neuroimage 56(2):387–399, 2011). Particularly interesting recent applications to affective and social neuroscience include affective state decoding, exploring potential biomarkers of neurological and psychiatric disorders, predicting treatment response, and developing real-time neurofeedback and brain-computer interface protocols. In this chapter, we review the bases of the most common neuroimaging techniques, the basic concepts of ML, and how it can be applied to neuroimaging data. We also describe some recent examples of applications of ML-based analysis of neuroimaging data to social and affective neuroscience issues. Finally, we discuss the main ethical aspects and future perspectives for these emerging approaches.

Keywords Brain imaging methods · Machine learning · Neuroscience machine learning · Emotion/affective decoding · Neurofeedback

L. R. Trambaiolli (✉)

Basic Neuroscience Division, Mclean Hospital – Harvard Medical School,
Belmont, MA, USA

e-mail: ltrambaiolli@mclean.harvard.edu

C. E. Biazoli Jr · J. R. Sato

Center for Mathematics, Computing, and Cognition, Federal University of ABC,
São Bernardo do Campo, Brazil

Introduction

Machine learning (ML) is a subarea of artificial intelligence which uses the induction approach to learn based on previous experiences and make conclusions about new inputs (Mitchell, 1997). In the last decades, the use of ML approaches to analyze neuroimaging data has attracted widening attention (Pereira et al., 2009; Lemm et al., 2011). Particularly interesting recent applications to affective and social neuroscience include affective state decoding, exploring potential biomarkers of neurological and psychiatric disorders, predicting treatment response, and developing real-time neurofeedback and brain-computer interface protocols. In this chapter, we review the bases of the most common neuroimaging techniques, the basic concepts of ML, and how it can be applied to neuroimaging data. We also describe some recent examples of applications of ML-based analysis of neuroimaging data to social and affective neuroscience issues. Finally, we discuss the main ethical aspects and future perspectives for these emerging approaches.

Brain Imaging Methods

Most neuroimaging experiments in human social and affective neuroscience are based on two groups of techniques (Fig. 13.1) (Min et al., 2010). The first group comprises measurements of either electrical or magnetic features associated with the electrophysiological activity of neuronal assemblies. This group includes the electroencephalography (EEG) and the magnetoencephalography (MEG) data acquisitions. On the other hand, the second group comprises measurements of metabolic or hemodynamic features that are indirectly associated with neural activity. This second group of neuroimaging techniques includes functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), and positron emission tomography (PET).

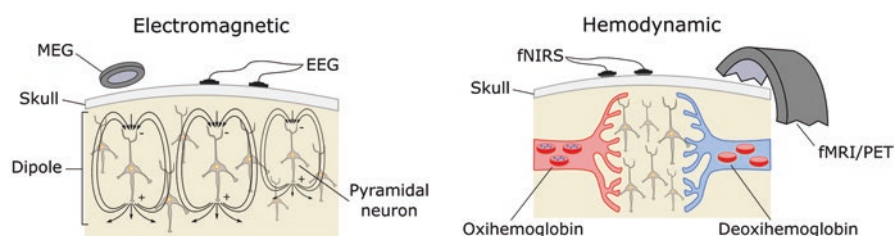


Fig. 13.1 Electromagnetic-based imaging approaches (left) use electric or magnetic sensors to capture the electromagnetic resultants from the neuronal and synaptic activity. Hemodynamic-based procedures (right) use light or magnetic sensors to measure the cerebral blood flow and oxygen consumption levels

Among the electromagnetic approaches, the EEG uses electrodes positioned over the scalp to record the sum of excitatory and inhibitory postsynaptic potentials in which the resulting dipoles are perpendicularly aligned to the scalp (Niedermeyer & da Silva, 2005). In consequence, its spatial resolution is limited and further compromised by volume conduction effects. However, its simplicity, low cost, and high temporal resolution (reaching the order of kilohertz in modern systems) make it one of the most common techniques in social and affective experiments. Similarly, MEG signals are resultant from the magnetic field generated by postsynaptic currents in apical dendrites (mainly those tangential to the skull) (Hansen et al., 2010). Despite presenting some mapping limitations similar to the EEG, MEG has a better spatial resolution, though restricted to superficial cortical sulci activity. Moreover, its higher cost and less availability when compared to EEG result in relatively fewer studies in human affective neuroscience using this technique (Min et al., 2010).

PET scanning is the pioneering metabolic and hemodynamic imaging approach. This technique uses an injected radioactive tracer to track brain tissue variations on blood flow and metabolic features associated with local neural activity (Maquet, 2000). However, with the emergence of noninvasive fMRI protocols, which did not depend on exogenous tracers, PET experiments became relatively less common in current research. The fMRI uses the paramagnetic properties of the deoxyhemoglobin molecules, which work as an endogenous tracer, to measure the blood-oxygen-level-dependent (BOLD) contrast effect (Ogawa et al., 1990). Both PET and fMRI acquisitions provide the highest spatial resolution among the brain imaging approaches, allowing the evaluation of both cortical and subcortical structures associated with social behavior and affective states (Liu et al., 2015). The worldwide availability of MRI scanners in clinical settings made it the most used neuroimaging technique in the last two decades. Among fMRI, main limitations in affective and social process research when compared with other approaches are its lower temporal resolution, scanner noise, and the setup that restrict movement (Doi et al., 2013). Hence, fMRI acquisition does not allow more naturalistic, out-of-the-laboratory protocols. As a complementary hemodynamics-based technique for more naturalistic settings, the fNIRS has the advances of portability, low cost, and a relatively good temporal-spatial ratio (Doi et al., 2013). This technique measures the absorption of near-infrared light by oxyhemoglobin and deoxyhemoglobin molecules in superficial layers of the brain tissue, during local neural activity (Ferrari & Quaresima, 2012). However, fNIRS acquisitions only cover brain layers close to the scalp, as is the case with MEG (Min et al., 2010), and with a sparse representation limited by the optodes arrangement.

In sum, each neuroimaging modality has advances and disadvantages, and the choice for a particular technique should be based on the specific research question. More recently, the use of multimodal setups emerged as a promising approach in the neuroimaging field. These approaches use two or more neuroimaging techniques aiming to combine its advantages and provide complementary and convergent information regarding the underlying neural phenomena (Liu et al., 2015). The most common combination involves at least one electromagnetic and one hemodynamic approach, such as EEG-fMRI, EEG-fNIRS, or EEG-fNIRS-fMRI. However,

combinations into the same group of techniques are possible, such as EEG-MEG and fNIRS-fMRI.

Basic Concepts of Machine Learning

The primary aim of a machine learning algorithm is to *learn* (i.e., extract knowledge) from an original dataset (training set), validate its ability to make predictions in an independent dataset (validation set), and then make decisions or predictions in new samples (test set) (Mitchell, 1997). During the learning process, the decision model bases its conclusions on patterns observed on the features of the examples in the training set. Such features might include, for example, frequencies of neural activity during specific tasks, event-specific potentials for a particular set of stimuli, or the connectivity level between different brain areas (Rubinov & Sporns, 2010; Sakkalis, 2011).

Learning Process

Three main approaches might be used to guide the learning process, according to the presence or absence of labels for each example (i.e., instance or subject in the dataset) (Fig. 13.2). The first approach (which will be the focus of this chapter) is the supervised learning, where each instance has a corresponding label (e.g., patient

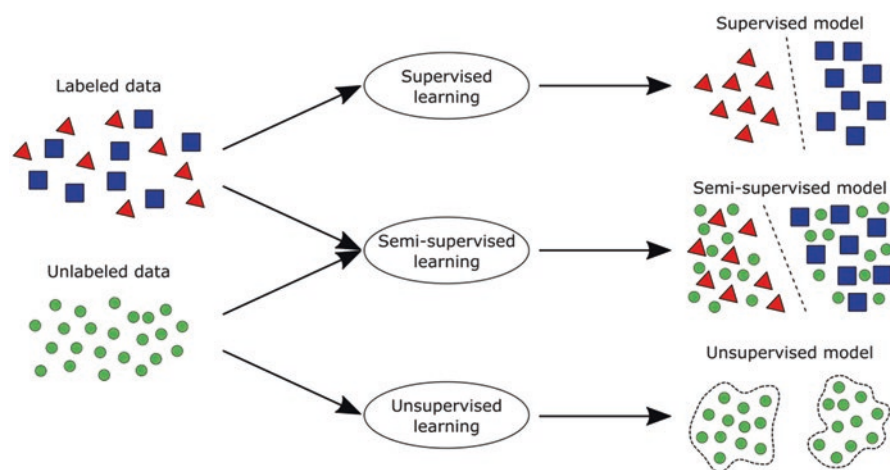


Fig. 13.2 Supervised learning methods use labeled examples to learn from data, while unsupervised learning methods extract patterns from data using unlabeled inputs. The recently proposed semi-supervised approach, otherwise, combines both labeled and unlabeled inputs during the learning process

or healthy subject). In this case, the objective is to develop models which can predict the desired labels with minimal error (Larranaga et al., 2006). Thus, during the learning process, the algorithm continually evaluates and adjusts the decision model until it reaches a near-to-optimal performance (Kuhn & Johnson, 2013). In unsupervised procedures, on the other hand, labels are not provided during the learning process. Here, the aim is to extract patterns exclusively based on similarities among groups of features (usually grouping examples according to these measures) (Larranaga et al., 2006). Finally, the third approach merges the characteristics from both previous methods. In this so-called semi-supervised approach, both labeled and unlabeled examples are used during the learning process. This approach takes advantage of the higher precision from the labeled training, as well as the lower computational cost from the non-labeled training (Cohen et al., 2004).

Validation Procedures

To converge in optimal decisions during the learning process, the decision model is continuously tested with a second dataset (i.e., the validation set) and, if necessary, remodeled using the training set (Kuhn & Johnson, 2013). The best approach for this procedure would be to train and validate the model with as much data as possible. However, due to experimental design constraints or to limited sample sizes, this task is commonly performed using somewhat suboptimal datasets (Lemm et al., 2011). Critically, to avoid variance and bias, the processes of training, validating, and testing the model *should not* be performed on the same data (Pereira et al., 2009). Different validation approaches have been proposed to overcome the issues raised by using limited datasets. One popular strategy for experiments using supervised learning is the cross-validation method (Lemm et al., 2011). In this approach, a small sample of the dataset is first split to be used as the test set, while the remaining part is further splitted into the train and the validation sets (Fig. 13.3a). This partitioning procedure is repeated several times to create different samples for each iteration (Lemm et al., 2011).

Different partitioning schemes might be used for this division. For example, in k-fold cross-validation (Fig. 13.3b), the dataset is divided into k disjoint subsets with equal size. Then, k-1 folds are used to train the model, and the remaining one is used for validation. This last step is repeated k times until all subsets are used as the validation set (Pereira et al., 2009). Another popular approach is the leave-one-out cross-validation (Fig. 13.3c), which is a particular case of k-fold cross-validation where k is equal to the number of examples.

Finally, in Monte Carlo cross-validation (Kuhn & Johnson, 2013) (Fig. 13.3d), the train and validation sets are composed by a fixed number of examples (e.g., X% for training and 100-X% for validating). Then, samples are randomly selected to form each set. This procedure might be repeated until all combinations are tested (high computational cost) or up to a predetermined number of permutations.

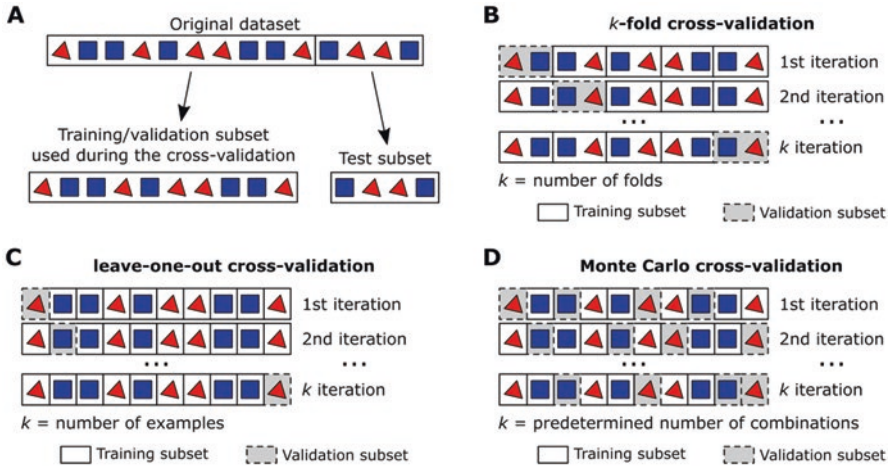


Fig. 13.3 Different steps and approaches for data splitting. **(a)** The first step of the validation process is to select a sample subset for testing purposes. Then, cross-validation approaches are used to split the remaining data into training and validation subsets. **(b)** During the *k*-fold cross-validation, data is split into *k*-folds of similar lengths. Then, the algorithm is validated *k* times, until all folds were used as the validation subset. **(c)** The leave-one-out cross-validation is a particular case of *k*-fold cross-validation, where each fold corresponds to a single example. **(d)** The Monte Carlo cross-validation performs a predetermined number of combinations, where the validation subset is composed of a fixed quantity of randomly selected samples

Dimensionality Reduction

In contrast to a limited number of examples, supervised models usually have a wide range of features associated with them. This growing abundance of assessed features relates to the improvement of brain imaging technologies and the development of new feature extraction methods. However, contrary to a common belief, the increasing high dimensionality of neuroimaging datasets does not necessarily lead to improved ML models. Indeed, much of these new features are redundant or irrelevant to the model design and might even cause a decrease in performance (Guyon & Elisseeff, 2003). With this in mind, dimensionality reduction strategies became a fundamental step for model building (Lemm et al., 2011).

As the learning approaches, feature selection (FS) methods can be grouped into unsupervised and supervised categories. The common spatial pattern (CSP), an example of supervised method, uses the class label to search for an optimal and reduced subset of features, where the maximum of relevant information is held (Lemm et al., 2011). On the other hand, unsupervised methods, such as the principal component analysis (PCA) and the independent component analysis (ICA), are mainly used for dimensionality and noise reduction based on projections to the more relevant factors or based on grouping of effects (Lemm et al., 2011). However, unlike the supervised category, unsupervised methods often require manual selection of relevant factors or groups.

Over the last decades, supervised FS methods have become popular in neuroscience (Huang, 2015). To select these optimal subsets of features, some topics should be established, such as the search strategy and the level of interaction with the ML algorithm.

Regarding the search strategy, two main approaches are possible, according to the subset composition. For the first strategy, all features are sorted according to some relevance criteria. Then, only those features with higher positions are selected to compose the subset (Huang, 2015). On the other strategy, subgroups are created with random features from the original feature set. Then, these subsets are evaluated according to its capacity to describe the whole dataset (Huang, 2015). The ideal FS algorithm would explore all combinations available to compose the feature subsets (i.e., to perform an exhaustive search) (Guyon & Elisseeff, 2003). However, due to the complexity of the problem and to computational limitations, it is common to establish a stop criterion that defines when the algorithm decides for one subset of features (e.g., when the model reaches a specific performance threshold or when the subset reaches a particular amount of features) (Guyon & Elisseeff, 2003).

According to the level of interaction with the ML model, feature selection algorithms might also be grouped into three approaches (Kohavi & John, 1997) (Fig. 13.4): filter, wrapper, and embedded. The filter approach is the most commonly used procedure. In this, the feature selection is performed before and independently to the model induction (Fig. 13.4a). For the wrapper approach, every feature set is submitted to the ML algorithm, and the model performance is used to evaluate the selected subset (Fig. 13.4b). Finally, embedded approaches merge the feature selection and the model induction steps, with the subsets being created internally by the ML model (Fig. 13.4c).

Types of Classifiers

Different types of classifiers are defined according to the specific assumptions made during the learning process (Pereira et al., 2009). For example, logic-based algorithms create successive layers in which instances are classified according to the values of a single feature. These algorithms might be described as a decision tree which is composed by nodes and branches (Fig. 13.5a). Each node has a particular rule that divides the instance into different branches according to the corresponding feature value (Murthy, 1998). The first node of the tree is the feature that best separates the training data, followed by nodes ordered by a decreasing predictive power until no more rules become necessary to classify the dataset correctly. This kind of algorithm tends to perform better when dealing with categorical features (Kotsiantis, 2007).

In perceptron-based algorithms, the perceptron calculates a linear combination of the input features and, further, sum all weighted inputs to make a decision. When the result is higher than a specified threshold, the instance is labeled as class A or marked as class B otherwise (Mitchell, 1997). These weights are randomly

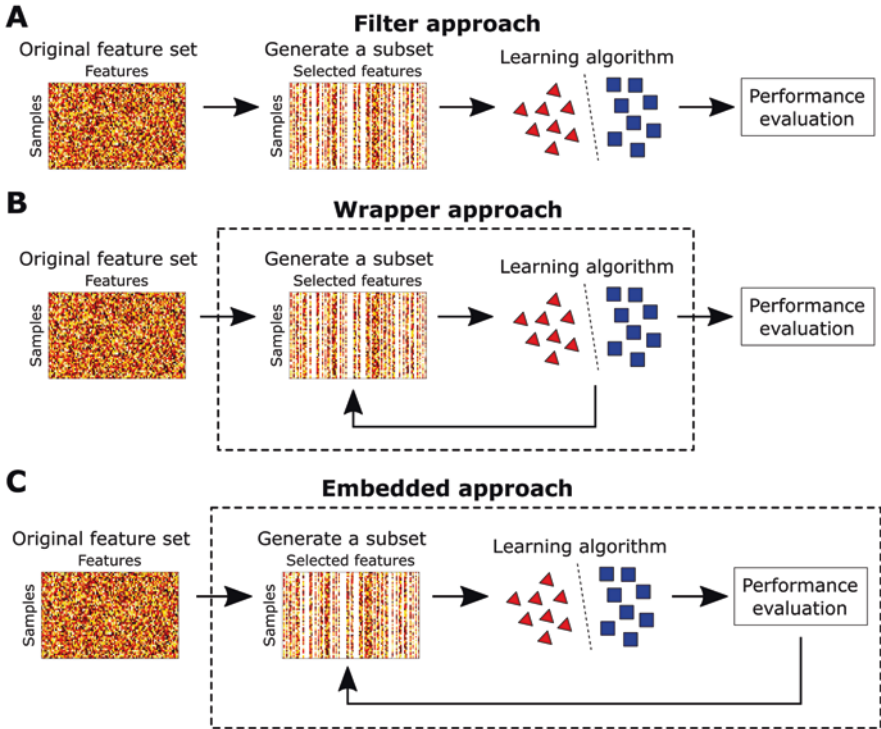


Fig. 13.4 Level of interaction between the feature selection algorithm and the classifier. (a) During the filter approach, the feature selection is performed before and apart from the classifier. (b) During the wrapper approach, every single feature subset is submitted to the classifier, and the classification performance is used to evaluate the sample. (c) During the embedded procedure, both the feature selection and the classifier algorithms are merged and happen simultaneously

established at first but optimized during the learning process until they reach near-to-optimal predictions (Mitchell, 1997). The perceptron approach, however, can only classify linearly separable inputs (Kotsiantis, 2007). To perform nonlinear discrimination, the use of artificial neural networks (ANN) was proposed. In this, multiple perceptrons are combined creating a complex network where the output from one single perceptron might be used as an input for several other perceptrons (Fig. 13.5b) (Zhang, 2000).

Unlike other classifiers, statistical-based algorithms provide the probability of the evaluated instance belonging to any given class (Kotsiantis, 2007). A classic example of this group of algorithms is the linear discriminant analysis (LDA) which explores linear combinations of features that best label instances into the desired classes (Fig. 13.5c) (Balakrishnama & Ganapathiraju, 1998).

Finally, support vector machines (SVM) compose a non-probabilistic method inspired by statistically based approaches. In this case, data is separated into two classes by a hyperplane (Vapnik, 1995). This hyperplane is defined trying to

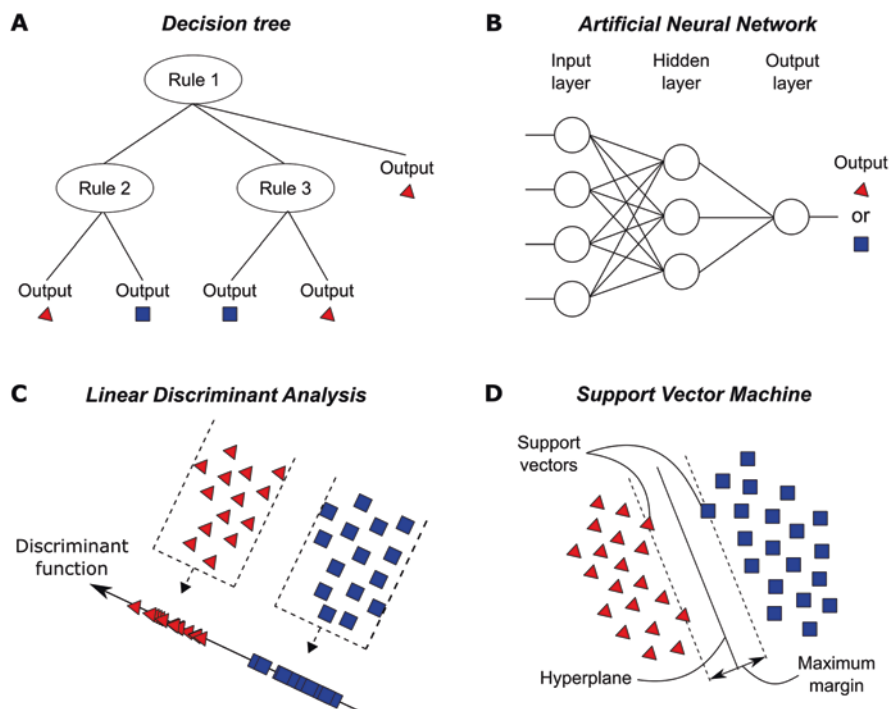


Fig. 13.5 Examples of classifiers commonly applied to neuroimaging studies. (a) A decision tree, (b) artificial neural networks, (c) linear discrimination analysis, (d) support vector machines

maximize its distance (margin) to the instances on either category (Fig. 13.5d) and, consequently, reducing the expected generalization error (Cristianini & Shawe-Taylor, 2000). For the classification of non-separable data, the dataset might be translated onto a higher-dimensional space using kernel methods, to apply the SVM-designed hyperplane (for more details about kernel methods, please refer to Cristianini & Shawe-Taylor, 2000).

Although multiclass classification approaches have been architected for the previously listed classifiers, binary classification (e.g., task vs. control group, task A vs task B, etc.) is most commonly applied in social and affective neuroscience studies.

Evaluating and Interpreting a Machine Learning Model

One easy way to evaluate the performance of a binary classifier is the use of a confusion matrix (or error matrix) (Sokolova & Lapalme, 2009). This matrix represents the relation between the actual and the predicted classes (Fig. 13.6a). Four main measures might be extracted from this matrix (Sokolova & Lapalme, 2009). The first measure, named accuracy, is the ratio between the number of examples

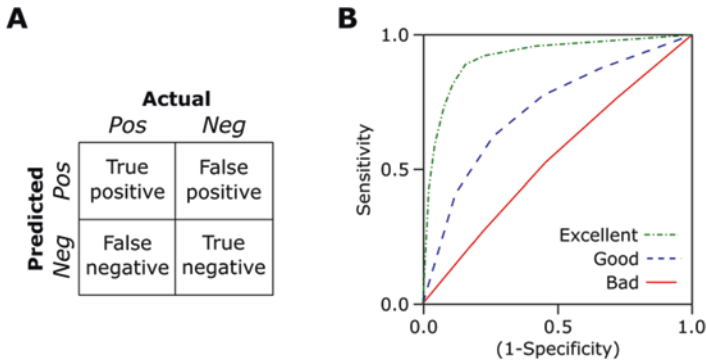


Fig. 13.6 Illustrative example of (a) a confusion matrix and (b) three different examples of ROC curves representing classifiers with excellent (dotted line), good (dashed line), and bad (continuous line) performances

correctly predicted (true positives and true negatives) by the total of samples available. The second is named precision, which is the ratio between the number of true positives by the total of examples predicted as positive (true and false positives). Sensitivity is the ratio between the number of true positives by the total of positive examples (true positives and false negatives), while specificity is the ratio between the number of true negatives by the total of negative samples (true negatives and false positives).

In general, an optimal model should present high sensitivity and specificity. However, real-world datasets tend to show an unbalance between these measures. To evaluate this aspect, the receiver operating characteristic curve (ROC curve) presents an illustrative plot of the discriminant ability of the binary classifier for different thresholds (Fawcett, 2006). This curve is plotted using the sensitivity of the classifier as the y-axis and the fall-out (i.e., 1-specificity) as the x-axis (Fig. 13.6b). Thus, the area under the ROC curve (AUC) describes the probability that the classifier will rank a random positive instance higher than a random negative example (Fawcett, 2006). In other words, when comparing the AUC of different classifiers, the higher the AUC, the better is the classifier average discriminative power.

Finally, linear classifiers such as the LDA and the linear SVM present weights relative to each variable. These weights describe how relevant each variable is to identify each class (Sato et al., 2009). In addition to performance measures, this information adds valuable clues regarding the neural basis of the studied mental process. For example, that specific frequencies in some brain areas are more related to one affective state than the other or that the volume of a subcortical structure might be a predictor of a given psychiatric disease.

Besides the evaluation methods listed in this chapter, other performance metrics might be used according to the characteristics of the ML algorithm and the experimental design. For a comparative review, please refer to Sokolova and Lapalme (2009).

ML Applications in Social and Affective Neuroscience

Computer-Aided Diagnosis

Psychiatric disorders are defined by the presence of specific set of symptoms. However, some symptoms are shared across disorders and a single patient might satisfy criteria for multiple disorders, or do not fit the requirements for any precise diagnosis (Huys et al., 2016). In this context, an increasingly popular application of ML in social and affective neuroscience is in the quest for imaging biomarkers of psychiatric disorders. This popularity is due to a recent focus on individualized medicine. Although classical statistical approaches provide biomarker descriptions at the group level, physicians should make clinical decisions about individuals (Orri et al., 2012). Thus, ML has been an active area of research to the development of potential computer-aided individualized diagnosis methods.

From this perspective, the use of structural MRI data combined with ML approaches is presenting promising results for the better comprehension of the obsessive-compulsive disorder (OCD). For example, Soriano-Mas et al. (2007) successfully classified patients with OCD from healthy control with more than 90% of accuracy based on brain structural features. Also, these data were used to predict the severity of obsessive-compulsive symptoms (Hoexter et al., 2013), as well as to list potential biomarkers using dimensionality reduction approaches (Trambaiolli et al., 2017).

In depressive spectrum disorders, structural MRI also achieved accuracies around the 90% threshold when classifying patients and controls (Mwangi et al., 2012), while functional MRI successfully discriminated between bipolar and unipolar depression with similar performances (Grotegerd et al., 2013). Also, structural and functional variations in affective-related brain regions, such as the amygdala, the insula, and the cingulate cortex, predicted symptom severity and treatment response (Siegle et al., 2006; Chen et al., 2007). Similarly, ML predictive approaches efficiently predicted the treatment response from patients with anxiety disorder for both pharmacological (Whalen et al., 2008) and cognitive behavioral (Doehrmann et al., 2013) therapies. However, it is important to emphasize that such findings had not yet reached clinical significance and are not currently incorporated in psychiatric practice.

Emotion/Affective Decoding

Brain decoding is the identification of someone's mental states based exclusively on measurements of their brain activity (Haynes & Rees, 2006). This stands on the idea that different neural activity patterns are associated with different mental states. Thus, decoding these patterns might be fundamental for our understanding of the neural basis of human cognition (Haynes & Rees, 2006). In this context, the ability

from ML methods to identify and learn from patterns makes it a quite suitable approach for affective brain decoding.

A spectral power asymmetry over the frontal regions during emotion elicitation is a classical effect reported from EEG data analysis (Balconi et al., 2015). Applying an ML approach, Wang et al. (2014) reached more than 80% of predictive accuracy when distinguishing between positive and negative affective valences. Similar classification results were reported using fNIRS recordings over the prefrontal cortex when comparing positive or negative affective states with neutral states (Trambaiolli et al., 2018a). Also, the prefrontal activity even during resting state seems to be related with the emotional processing, since resting state frontal asymmetry predicts responsiveness to affective elicitation (Balconi et al., 2015).

However, human emotions involve complex networks comprising areas not accessed by the EEG or fNIRS spatial sampling and resolution. Using fMRI data, Baucom et al. (2012) achieved up to 90% accuracy in single participant classification between positive and negative valences using voxels from the medial and the ventrolateral prefrontal cortex, anterior cingulate, and amygdala, among other regions. Later, Lindquist et al. (2016) developed a meta-analytic study compiling data from 397 functional studies and different ML learning methods to investigate different hypotheses of network organization during the elicitation of affective valence. Their evidence suggests a single network composed by areas such as the dorsomedial prefrontal cortex, ventrolateral prefrontal cortex, supplementary motor area, anterior insula, amygdala, ventral striatum, and thalamus, which respond both for positive and negative valence, but with different patterns of activation depending on the affective state (Lindquist et al., 2016).

Neurofeedback

Due to the recent success of ML in decoding different mental states, this approach was also used to develop therapeutic applications, such as neurofeedback. Neurofeedback is a real-time procedure where a feedback of the neural activity in specific neural substrates is provided to the volunteer aiming to achieve the self-regulation of these areas or networks (Sitaram et al., 2017). Specifically, affective neurofeedback targets substrates related to emotional processing (Trambaiolli et al., 2018b) and might be useful as a nonpharmacological treatment for psychiatric symptoms or disorders, such as schizophrenia, major depressive disorder, attention-deficit/hyperactivity disorder, and obsessive-compulsive disorder (Fovet et al., 2015).

Different imaging methods allow different approaches to control affective networks. On the one hand, electrophysiological methods usually aim to control specific frequency bands in particular subsets of electrodes (Begemann et al., 2016; Enriquez-Geppert et al., 2017). For example, EEG alpha asymmetry in frontal electrodes was tested to reduce depressive symptoms, while central beta suppression and theta enhancement were applied to minimize inattention and impulsivity symptoms (Begemann et al., 2016).

On the other hand, hemodynamic methods use the upregulation or downregulation of the local blood flow in specific targets (Sulzer et al., 2013). For example, depressive patients who achieved self-control of the amygdala through fMRI-based neurofeedback showed reduced indices of anxiety and increased indices of happiness (Young et al., 2014), as well as a positive correlation between the symptom improvement and the reorganization of amygdala functional connectivity after the neurofeedback training (Young et al., 2018).

Social Neuroscience

Despite the indisputable importance of living in a structured society for human affective and cognitive processes, how the human brain works throughout simple to complex social contexts remains largely elusive (Babiloni & Astolfi, 2014).

In current social neuroscience, the possibility of simultaneously recording brain activity of two or more people interacting (i.e., hyperscanning) and of conceptualizing the connectivity emerging from such interactions (i.e., hyperconnectivity) has gained momentum (Montague et al., 2002). In this context, ML algorithms could be applied to modeling some level of a causal relation in social interactions mediated by interactions in brain activities (Konvalinka & Roepstorff, 2012). Anders et al. (2011) used fMRI records to predict the level of neural activity in romantic partners while experiencing the same emotional feelings. For this, the model was trained using data from one partner and used to successfully estimate the brain functional activation pattern of the other partner.

Another appealing field of research questions using ML approaches is the investigation of the neural correlates complexing social preferences and behaviors, such as friendship or engagement with political ideologies. For example, Kanai et al. (2011) applied a classifier to differentiate between participants with self-declared conservative or liberal political ideologies. Using the gray matter volume of the anterior cingulate and the right amygdala as inputs, the classifier reached near to 70% accuracy (Kanai et al., 2011). In another study, liberal and conservative participants were classified using functional MRI data with remarkable AUC values of more than 98% (Ahn et al., 2014).

Future Perspectives and Ethical Aspects

During the last decade, the neuroimaging community is making a continuous effort to create structured and standardized publicly available datasets, covering a wide range of samples and experiments (Poldrack & Gorgolewski, 2014). This action is fundamental to the development of optimized models for computer-aided diagnosis, for example. With larger samples, population heterogeneity, and standardized protocols, new ML models will be less susceptible to outliers and noise influence and

will present higher generalization power (Schnack & Kahn, 2016). The extensive information resulting from these datasets will allow the use of ML approaches to confirm or to explore new aspects regarding the neural basis of affect and social interactions.

A promising instrumental evolution is the development of portable imaging devices, such as wearable EEG and fNIRS systems (Piper et al., 2014; von Lüthmann et al., 2017). This technology allows studies outside the laboratory environment, leading to the observation of how the social brain acts in real-life situations (Balardin et al., 2017). Although ML algorithms should be adapted to deal with new levels of physiological (e.g., movement-related artifacts) and environmental (e.g., diverse magnetic fields) noises, a new range of naturalistic responses will be available for analysis. Neurofeedback applications would also be benefited by portable devices, with the possibility of location-independent training or the passive control of affect-driven software or equipment.

Another exciting prospect is the use of ML to develop new concepts of social interaction, such as the named collaborative brain-computer interfaces (BCI) (Wang & Jung, 2011). Following the idea of neurofeedback, in BCI, the user intends to control a computer exclusively based on their brain activity (Sitaram et al., 2017). Thus, collaborative BCI uses brain waves from multiple users to control one single machine, leading to increased task performances as high is the number of participants (Wang & Jung, 2011). Still, in the context of BCI, other social environments were created with the assistance of ML algorithms. For example, Rao et al. (2014) proposed the brain-to-brain interface in humans, where the EEG signals from one user were used to stimulate the brain of a second subject through transcranial magnetic stimulation (TMS). Later, this concept was expanded for the idea of a “brain-net,” where the signals of some users (senders) were collaboratively merged to stimulate the brain of an independent participant (receiver) (Jiang et al., 2018).

The advance of ML applications in affective and social neuroscience also raises some ethical concerns. In clinical settings, for instance, the use of ML algorithms will only be possible after careful evaluation and when proper evidence for improvement in either diagnosis accuracy or treatment efficacy is in place. To date, no conclusion or clinical decision should be taken exclusively based on the ML output, and future applications surely will depend on the integration of ML procedures to expert knowledge (Fu & Costafreda, 2013). Decoding affective states is an essential tool for the understanding of the brain basis of the human mind, as well as for the development of therapeutic approaches such as neurofeedback. However, an essential ethical and legal aspect regarding brain decoding applications is ensuring privacy or non-consented commercial use of data or decoding results (Haynes, 2011).

Final Considerations

In this chapter, we introduced concepts of brain imaging and ML methods. Aside from describing learning and validation methods, dimensionality reduction and feature selection approaches, performance estimations, and currently popular

classifiers, we purposefully focused on supervised methods. This choice was based on the facts that these are the best examples for an initial overview of the ML topic and the most popular approach in neuroimaging studies. We also described some uses of ML to social and affective neuroscience problems, from basic investigations to clinical and therapeutic applications. Promising prospects were also mentioned to contextualize the reader to cutting-edge advances in this area. Finally, we also highlighted some ethical aspects that might be carefully considered when developing applications of ML in social and affective neuroscience.

Acknowledgments JRS is grateful to Sao Paulo Research Foundation (FAPESP, Grants #2018/04654-9, #2018/21934-5 and #2021/05332-8).

References

- Ahn, W. Y., Kishida, K. T., Gu, X., Lohrenz, T., Harvey, A., Alford, J. R., ... Montague, P. R. (2014). Nonpolitical images evoke neural predictors of political ideology. *Current Biology*, 24(22), 2693–2699.
- Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *NeuroImage*, 54(1), 439–446.
- Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience and Biobehavioral Reviews*, 44, 76–93.
- Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. *Institute for Signal and Information Processing*, 18, 1–8.
- Balardin, J. B., Zimeo Morais, G. A., Furuch, R. A., Trambaiolli, L., Vanzella, P., Biazoli, C., Jr., & Sato, J. R. (2017). Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. *Frontiers in Human Neuroscience*, 11, 258.
- Balconi, M., Grippo, E., & Vanutelli, M. E. (2015). Resting lateralized activity predicts the cortical response and appraisal of emotions: An fNIRS study. *Social Cognitive and Affective Neuroscience*, 10(12), 1607–1614.
- Baucom, L. B., Wedell, D. H., Wang, J., Blitzer, D. N., & Shinkareva, S. V. (2012). Decoding the neural representation of affective states. *NeuroImage*, 59(1), 718–727.
- Begemann, M. J., Florisse, E. J., Van Lutterveld, R., Kooyman, M., & Sommer, I. E. (2016). Efficacy of EEG neurofeedback in psychiatry: A comprehensive overview and meta-analysis. *Translational Brain Rhythmicity*, 1(1), 19–29.
- Chen, C. H., Ridler, K., Suckling, J., Williams, S., Fu, C. H., Merlo-Pich, E., & Bullmore, E. (2007). Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biological Psychiatry*, 62(5), 407–414.
- Cohen, I., Cozman, F. G., Sebe, N., Cirelo, M. C., & Huang, T. S. (2004). Semisupervised learning of classifiers: Theory, algorithms, and their application to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12), 1553–1566.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., ... Pollack, M. (2013). Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry*, 70(1), 87–97.
- Doi, H., Nishitani, S., & Shinohara, K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in Human Neuroscience*, 7, 770.
- Enriquez-Geppert, S., Huster, R. J., & Herrmann, C. S. (2017). EEG-neurofeedback as a tool to modulate cognition and behavior: A review tutorial. *Frontiers in Human Neuroscience*, 11, 51.

- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2), 921–935.
- Fovet, T., Jardri, R., & Linden, D. (2015). Current issues in the use of fMRI-based neurofeedback to relieve psychiatric symptoms. *Current Pharmaceutical Design*, 21(23), 3384–3394.
- Fu, C. H., & Costafreda, S. G. (2013). Neuroimaging-based biomarkers in psychiatry: Clinical opportunities of a paradigm shift. *Canadian Journal of Psychiatry*, 58(9), 499–508.
- Grotegerd, D., Suslow, T., Bauer, J., Ohrmann, P., Arolt, V., Stuhmann, A., ... Dannlowski, U. (2013). Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: A pilot study. *European Archives of Psychiatry and Clinical Neuroscience*, 263(2), 119–131.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar), 1157–1182.
- Hansen, P., Kringelbach, M., & Salmelin, R. (Eds.). (2010). *MEG: An introduction to methods*. Oxford University Press.
- Haynes, J. D. (2011). Brain reading: Decoding mental states from brain activity in humans. In Judy Illes and Barbara J. Sahakian (Eds.), *The Oxford handbook of neuroethics*, Oxford University Press, Oxford. pp. 3–13.
- Haynes, J. D., & Rees, G. (2006). Neuroimaging: Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523.
- Hoexter, M. Q., Miguel, E. C., Diniz, J. B., Shavitt, R. G., Busatto, G. F., & Sato, J. R. (2013). Predicting obsessive-compulsive disorder severity combining neuroimaging and machine learning methods. *Journal of Affective Disorders*, 150(3), 1213–1216.
- Huang, S. H. (2015). Supervised feature selection: A tutorial. *Artificial Intelligence Research*, 4(2), 22.
- Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404.
- Jiang, L., Stocco, A., Losey, D. M., Abernethy, J. A., Prat, C. S., & Rao, R. P. (2018). BrainNet: A multi-person brain-to-brain interface for direct collaboration between brains. *arXiv*, 1809.08632.
- Kanai, R., Feilden, T., Firth, C., & Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, 21(8), 677–680.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324.
- Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6, 215.
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. In I. G. Maglogiannis (Ed.), *Emerging artificial intelligence applications in computer engineering*, IOS Press, Amsterdam, Netherland. (Vol. 160, pp. 3–24).
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
- Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., ... Robles, V. (2006). Machine learning in bioinformatics. *Briefings in Bioinformatics*, 7(1), 86–112.
- Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, 56(2), 387–399.
- Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex*, 26(5), 1910–1922.
- Liu, S., Cai, W., Liu, S., Zhang, F., Fullham, M., Feng, D., ... Kikinis, R. (2015). Multimodal neuroimaging computing: A review of the applications in neuropsychiatric disorders. *Brain Informatics*, 2(3), 167.
- Maquet, P. (2000). Functional neuroimaging of normal human sleep by positron emission tomography. *Journal of Sleep Research*, 9(3), 207–232.

- Min, B. K., Marzelli, M. J., & Yoo, S. S. (2010). Neuroimaging-based approaches in the brain-computer interface. *Trends in Biotechnology*, 28(11), 552–560.
- Mitchell, T. M. (1997). *Machine learning*. McGraw Hill.
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., ... Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, 16, 1159–1164.
- Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2(4), 345–389.
- Mwangi, B., Ebmeier, K. P., Matthews, K., & Douglas Steele, J. (2012). Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain*, 135(5), 1508–1521.
- Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24), 9868–9872.
- Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and Biobehavioral Reviews*, 36(4), 1140–1152.
- Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, 45(1), S199–S209.
- Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., ... Schmitz, C. H. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *NeuroImage*, 85, 64–71.
- Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, 17(11), 1510.
- Rao, R. P., Stocco, A., Bryan, M., Sarma, D., Youngquist, T. M., Wu, J., & Prat, C. S. (2014). A direct brain-to-brain interface in humans. *PLoS One*, 9(11), e111332.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3), 1059–1069.
- Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine*, 41(12), 1110–1117.
- Sato, J. R., Fujita, A., Thomaz, C. E., Martin, M. D. G. M., Mourão-Miranda, J., Brammer, M. J., & Junior, E. A. (2009). Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage*, 46(1), 105–114.
- Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Frontiers in Psychiatry*, 7, 50.
- Siegle, G. J., Carter, C. S., & Thase, M. E. (2006). Use of fMRI to predict recovery from unipolar depression with cognitive behavior therapy. *The American Journal of Psychiatry*, 163(4), 735–738.
- Sitaram, R., Ros, T., Stoekel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., ... Birbaumer, N. (2017). Closed-loop brain training: The science of neurofeedback. *Nature Reviews Neuroscience*, 18(2), 86.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4), 427–437.
- Soriano-Mas, C., Pujol, J., Alonso, P., Cardoner, N., Menchón, J. M., Harrison, B. J., ... Gaser, C. (2007). Identifying patients with obsessive-compulsive disorder using whole-brain anatomy. *NeuroImage*, 35(3), 1028–1037.
- Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M. L., ... Herwig, U. (2013). Real-time fMRI neurofeedback: Progress and challenges. *NeuroImage*, 76, 386–399.
- Trambaiolli, L. R., Biazoli, C. E., Jr., Balardin, J. B., Hoexter, M. Q., & Sato, J. R. (2017). The relevance of feature selection methods to the classification of obsessive-compulsive disorder based on volumetric measures. *Journal of Affective Disorders*, 222, 49–56.

- Trambaiolli, L. R., Biazoli, C. E., Cravo, A. M., & Sato, J. R. (2018a). Predicting affective valence using cortical hemodynamic signals. *Scientific Reports*, 8(1), 5406.
- Trambaiolli, L. R., Biazoli, C. E., Cravo, A. M., Falk, T. H., & Sato, J. R. (2018b). Functional near-infrared spectroscopy-based affective neurofeedback: Feedback effect, illiteracy phenomena, and whole-connectivity profiles. *Neurophotonics*, 5(3), 035009.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.
- von Lüthmann, A., Wabnitz, H., Sander, T., & Müller, K. R. (2017). M3BA: A mobile, modular, multimodal biosignal acquisition architecture for miniaturized EEG-NIRS-based hybrid BCI and monitoring. *IEEE Transactions on Biomedical Engineering*, 64(6), 1199–1210.
- Wang, Y., & Jung, T. P. (2011). A collaborative brain-computer interface for improving human performance. *PLoS One*, 6(5), e20422.
- Wang, X. W., Nie, D., & Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing*, 129, 94–106.
- Whalen, P. J., Johnstone, T., Somerville, L. H., Nitschke, J. B., Polis, S., Alexander, A. L., ... Kalin, N. H. (2008). A functional magnetic resonance imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. *Biological Psychiatry*, 63(9), 858–863.
- Young, K. D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W. C., & Bodurka, J. (2014). Real-time fMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PLoS One*, 9(2), e88785.
- Young, K. D., Siegle, G. J., Misaki, M., Zotev, V., Phillips, R., Drevets, W. C., & Bodurka, J. (2018). Altered task-based and resting-state amygdala functional connectivity following real-time fMRI amygdala neurofeedback training in major depressive disorder. *Neuroimage: Clinical*, 17, 691–703.
- Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451–462.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

