CHAPTER 2

2 Literature Review

2.1 Background

In recent years, advancements in medical and computational sciences have developed a communication pathway between the human brain and external devices; such methods are popularly referred as Brain-computer interfaces (BCIs) [20]. As a medical application, BCI devices are widely used to assist people with neuromuscular disorders. Various control signals, such as Visually Evoked Potentials (VEP), Slow Cortical Potentials (SCP), p300 evoked potentials, Sensorimotor rhythms, etc., have been used to design BCIs [21].

Among many types of BCIs [22–24], motor imagery (MI) based BCI uses brain signals associated with the imagination of motor movement-related tasks [9]. Various studies have suggested that when a subject thinks about a specific motor movement, there are significant relative power changes that occur in the mu (8-13 Hz) and beta (13-30 Hz) rhythms of EEG acquired over the sensorimotor cortex area of the brain [14,25,26]. Subsequently, these power changes in EEG are processed and classified using pattern recognition methods to control external devices [27,28]. The power changes in EEG occur due to imagination of limb movements, are referred as event-related desynchronization (ERD)/ event-related synchronization (ERS), which can be further processed to control an external device [14]. However, EEG time series are highly contaminated by body motion artifacts and environmental noises, due to which distinguishing between different motor movements is a challenging exercise to perform [29]. Hence it is essential to employ preprocessing methods to suppress artifacts and noises before extracting useful information from the EEG signals [30].

In BCI systems, another popularly used control signal is Steady-State Visually Evoked Potentials (SSVEP). SSVEPs are highly used due to high signal-to-noise ratio and robustness [31]. SSVEP is a resonance phenomenon that is primarily observed in the occipital lobe of brain when the subject is focusing on a light source flickering at a constant frequency [17].

This chapter of the thesis presents a brief survey of the state-of-the-art signal processing methods used in various studies for MI and SSVEP-based BCIs. Section 2.2 describes popular brain signal recording methods. Section 2.3 gives a literature survey on signal processing methods applied to EEG for noise and artifact reduction. Section 2.4 and 2.5 literature reviews signal processing methods in MI-BCI and SSVEP-BCI, respectively.

2.2 Brain Signals Recording

The neural oscillations or brain waves are continuous time-varying patterns that originate due to the mental activities in the central nervous system (CNS). During a mental activity, a neuron interacts with another neuron in the brain, which induces an electrical current. The naturally induced brain current produces electric and magnetic fields, which can be recorded using different modalities. The commonly used modalities to map brain activities are magnetoencephalography (MEG) [32], Electroencephalography (EEG) [33], positron emission tomography (PET) [34], Electrocorticography (ECoG) [35], and Functional Near-Infrared Spectroscopy (fNIRS) [36]. A comparison of these acquisitions

Table 2.1 Comparison of Different modalities used for the recording of the Brain activities.

	Year	**Portability**	**Temporal Resolution**	**Spatial resolution**	**Invasive/ non-invasive**
EEG	1924	Yes	High	Low	Non-invasive
MEG	1968	No	High	Low	Non-invasive
PET	1977	No	Low	High	Non-invasive
fNIRS	1985	Yes	Low	High	Non-invasive
ECoG	early 1950s	No	High	Very high	invasive

devices are drawn in Table 2.1.

2.3 Signal processing in BCI

EEG time series is a nonstationary and random signal; thus, its study requires various mathematical and signal analysis tools. It is also highly prone to noises and body artifacts. The types of noises mostly found in acquired EEG are power line 50 Hz noise, environment electromagnetic waves, and thermal noise of the electronic components present in the EEG acquisition device. Most of the noises are rejected by the hardware module of the EEG recorder. Moreover, digital filters are employed at the software level to eliminate the remaining noises. On the other hand, body artifacts are the muscle signals that contaminate the raw EEG data. Some of these body artifacts originate due to the motion of various body parts and can be avoided by instructing the subject not to make any motions while recording EEG. However, artifacts from non-voluntary motions cannot be avoided, such as cardiac signals. Blinking and movement of eyes, popularly referred as ocular artifacts, highly affect EEG, and the subject cannot completely avoid eyeball movements during the acquisition of EEG. For these artifacts, signal processing of the contaminated EEG is done. Another need for signal processing is to extract only the useful information from EEG and eliminate redundant data points. In general, there are three types of analysis required to get useful information from the EEG signal, namely temporal, spectral and spatial analysis. The temporal resolution of EEG is high and provides good features [37]. However, the frequency and the spatial resolution of EEG are poor. Since EEG signal carries vital information in both frequency and time domain, there is a requirement of signal analysis tools that can extract time and frequency information at the same time. The Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), and Wavelet transform have been introduced in most of the studies to analyze EEG [38–40]. The FFT represents the spectral component present in the frequency domain but does not provide any time-domain information about the signal. On the other hand, STFT overcomes the problem of FFT and provides information in both time and frequency domain. But the shortcoming of STFT is that it does not provide multi-resolution information of the signal. Wavelet transform shows high multi-resolution properties and is considered as one of the most potent tools for time-frequency analysis of complex signals like EEG.

2.3.1 Preprocessing: Removal of artifacts and noises

The brain activities generate EEG in the frequency range of 0.5-30 Hz. As explained earlier, noises are the unwanted signals added with the raw brain activities during EEG acquisition and need to be eliminated at the preprocessing stage. Digital bandpass filter designed with a passband frequency between 0.5 and 30 Hz suppresses the dc values and noises with higher frequency components. Also, a 50 Hz notch filter removes the power line noise. However, the body movement artifacts have frequency ranges that coincide with that of EEG. For instance, ECG and EMG frequency range is 0-100 Hz, and ocular artifacts frequency range lies between 0 and 10 Hz. Thus, motion artifacts cannot be removed by directly applying the bandpass filtering on the contaminated EEG.

To serve this purpose spatial filtering methods for example, Independent component analysis (ICA), surface Laplacian derivation (SLD), and common average reference (CAR) have been proposed. Spatial Filtering requires multi-channel EEG for denoising. ICA is proposed for the separation of eye motion artifacts from the EEG signal in [41,42]. To enhance the performance, an adaptive filter was used in combination with the ICA [43]. SLD is generally called Laplacian filter or surface Laplacian [44–47]. SLD is derived by subtracting the target channel from the weighted mean of four neighbouring channels. Small Laplacian derivation (SLAP) and Large Laplacian derivation (LLAP) are the two subdivisions of SLD. SLAP takes the average of the closest four neighbouring channels whereas, LLAP takes the average of the next four close channels. CAR is equal to denoising channel minus average of all other channels [48,49]. These techniques require multi-channel EEG signals for processing. However, most BCI systems are based on MI and Event-related potential; for such systems, single-channel EEG signal processing is essential. In the work of M. Davies et al. [50], ICA algorithm was used on single-channel signals for the first time. In this work, sources were assumed as stationary and discontinuous in their frequency spectrum. As all biomedical signals are nonstationary, these assumptions may not hold. Therefore, noise suppression in single-channel EEG time series is an important issue for detecting accurate brain activities.

Decomposition methods such as singular spectrum analysis (SSA), and wavelet transform decomposition have been used for denoising of Single EEG channel. The singular spectrum analysis (SSA) is an extensive method for non-linear time series analysis. SSA decomposed the EEG into distinct spectrums and applied thresholding to remove noises [51]. Compared with conventional methods, as an effective analysis technique, wavelet transform is a frequently used tool for nonstationary signal processing in many fields. It is notable that for denoising non-linear and nonstationary signals, wavelet decomposition [52–54] and signal separation algorithm [55,56] have been recognized as powerful tools. In many works, multi-resolution property of wavelet transform has been used to eliminate the OAs [57–60]. In general, wavelet functions are used to decompose contaminated EEG signal into different levels. Then a wavelet shrinkage function is used to adjust the threshold value, which is compared with the decomposed wavelet coefficients and eliminates the OA components. After that, the clean EEG time series data is reconstructed using inverse wavelet transform. Some of the popular shrinkage strategies in the wavelet domain are VisuShrink, SureShrink, BayesShrink, and NeighShrink. Dual-tree complex wavelet transform (DTCWT) with NeighCoeff shrinkage has been proposed in [61] as a denoising algorithm for nonstationary mechanical vibration signals. The work of Miao et al. [62] indicates that the DTCWT based decomposition method for EEG signals is effective compared to those DWT-based. Decomposition of the signal using discrete wavelet transform (DWT) results in energy losses at the transition band of adjacent scales [63]. At the same time, the DWT shows a large aliasing phenomenon [64]. DTCWT can resolve the energy losses and aliasing problem and has some advantages, for instance, nearly shift-invariance, multidimensional direction selectivity, and perfect reconstruction [65]. Table 2.2 provides a brief description and references of signal processing methods in preprocessing.

2.4 Signal processing in MI-based BCI

MI-BCI systems have been used in medical applications, for instance, neuro-rehabilitation of stroke patients. In the work of Jessica Cantillo-Negrete et al. [66], a scheme has been proposed to couple MI-BCI with a robotic orthosis device to rehabilitate human upper extremity post-stroke. In general, the raw EEG data of the patients are analyzed using computational models to classify various mental tasks. This EEG pattern classification can be performed using pattern recognition methods such as feature extraction, feature selection, and classification or regression [9,67].

In the case of MI EEG feature extraction, the selection of an appropriate frequency band is a very crucial step as most of the information lies in a particular band of frequency. In [14], it has been shown that the motor-related mental tasks generate event-related desynchronization (ERD) and event-related synchronization (ERS) patterns in two frequency bands, namely mu (8-13 Hz) and beta (13-30 Hz). It was reported that when the subject thinks about the movement of his limbs, average power in the mu rhythms attenuates and that in the beta rhythms increases. Therefore, for the classification of MI tasks, the EEG signal must be bandpass filtered in mu and beta rhythms frequency ranges.

Table 2.2 Signal processing algorithms in pre-processing.

Algorithms	Descriptions	References
Band-pass filtering and Notch filtering	Band-pass filters are applied during the signal processing of EEG to suppress the dc values and noises with higher frequency components. Hence, it is important to employ band pass filtering before spatial filtering. The disadvantage is that some useful information gets eliminated. Notch filtering is usually implemented to remove the power line noise (50/60 Hz).	[31–33,49–54,54,54–76]
ICA	ICA method separates the mixed signal into independent components. It is capable of removing body artifacts such as EOG, EMG artifacts from EEG signal. Advance variant of ICA, namely, Fast ICA is popular because it converges faster than ICA.	[32,49,63,74, 77–81]
CAR	CAR considers the average activity of other channels as noise and thus subtract it from the target channel. CAR improves the signal to noise ration of EEG.	[28,32,33,56, 58,60,63,65, 78,82–86]
Surface Laplacian derivation (SLD)	SLD finds the difference between the target channel and weighted mean of four neighbouring channels. It is divided into SLAP and LLAP. SLAP uses average of nearest channels and LLAP uses average of next four nearest channels.	[28–31,61,65, 74,84,85,87–93]
Singular Spectrum Analysis	SSA decomposed the EEG into distinct spectrums. Then, it groups the decomposed coefficients into EEG and noises. Only EEG coefficients are used during the reconstruction phase.	
Wavelet transformation with thresholding	DWT decomposed the EEG signal and then a threshold on the coefficients of kth scale, i.e., the scale which has a frequency range 0-10 Hz as the strength of ocular artifacts is strong in this range of frequency. Finally, regenerate the artifact-free EEG signal using inverse-DWT.	[94]
CSP	CSP projects the EEG into a subspace where variance of one class is maximized and that of another class is minimized. Although it is used mainly for feature extraction, but some studies used it at preprocessing stage.	[66,80]

Some of the bandpass filters used by researchers are elliptical filter, Butterworth, Chebyshev, and finite impulse response (FIR) filter [68,69]. A few have reported the use of wavelet transform [70] and wavelet packet decomposition [71] for bandpass filtering of the EEG.

2.4.1 Feature Extraction in MI-BCI

Once the EEG signal is filtered, various time, frequency, and phase features are calculated. Some of the features most generally used in MI BCI are explained below.

Time features: Statistical analysis of EEG in time domain provides the main information that exists in the signal that can improve the discrimination scope between the different motor movements related to mental tasks [72]. In various studies, commonly extracted statistical features are wavelet domain energy, sample entropy, variance, root mean square, and maximum of the EEG signal [73]. Band power (BP) is commonly used for extracting ERD/ERS patterns [74]. It computes the average power of the bandpass filtered EEG signal. Basically, BP is evaluated by squaring the filtered EEG and then averaging it over all the trials.

Frequency Domain features: Power spectral density (PSD) is commonly used to extract features from EEG in the frequency domain. ERD/ERS is the relative power of the EEG from the channel locations C3 and C4, which provide discriminatory information for two-class motor imagery [14]. To evaluate EEG signal power in the frequency domain, Fourier transform is used.

Phase Features: In order to extract phase features, the phase relation between the EEGs associated with the different tasks is studied. In MI-BCI literature, mainly used phase-based features are phase-locking value (PLV), instantaneous phase difference (IPD), and mean phase difference (MPD). PLV computes the phase synchronization between the two signals using the Hilbert transform [75]. IPD is the instantaneous phase difference between each pair of EEG channels [76,77]. MPD is defined as the mean of the IPD between a pair of EEG signals over the time window [75].

Time-frequency analysis: Most common time-frequency analysis methods applied in BCI research are Short–Time Fourier Transform (STFT), Morlet Wavelet Transform (MWT), Wavelet Filter Bank (WFB), and Wavelet Packet Decomposition (WPD). STFT divides the time-series signal into smaller windows and then computes Fourier transforms of the windowed signal [78]. MWT computes the magnitude and phase coefficients of a time-varying signal [79]. It is a continuous wavelet transform. Multi-resolution property of wavelet transform has been widely used in BCI research to analyze time-frequency relationship in EEG signals [70]. Wavelet functions are used to decompose EEG signals into different frequency coefficients of details and approximations. Then, coefficients of useful frequency bands are used to reconstruct the EEG signal. Moreover, there are a variety of mother wavelet basis functions used for decomposition. Selection of a mother wavelet is important as the outcome of different mother wavelets may vary for the same problem. Researchers select mother wavelet on an experiment basis. In WPD method, coefficients of specific frequency bands are used as features for classification. Table 2.3 describes feature extraction methods in MI signal processing.

2.4.2 Feature Selection

Although many computational algorithms have been proposed to classify the mental tasks in different BCI applications [27,28,109–113], obtaining a higher classification performance is still a challenging task. High dimensions of features extracted from MI data make EEG a very complex signal to analyze. To address these issues, feature selection techniques are in use to reject the redundant features [114]. Numerous feature selection techniques are being employed in various studies,

Table 2.3 Feature extraction methods in MI signal processing.

Methods	Description	References
Statistical features	Various statistical features such as wavelet domain energy, sample entropy, variance, root mean square value, and maximum of the time series EEG signal are evaluated to represent main information in the signal.	[73]
Band power (BP)	BP is computed by band-pass filtering, squaring, and averaging over all the trials in time domain. In frequency domain it is sometimes calculated using Fourier Transform. It contains significant information and has been widely used in BCI research.	[31,51,52,63, 65,66,68,76, 77,81,84,88, 89,109–113]
Power spectrum density (PSD)	Parametric estimation methods such as autoregression (AR) mode and non-parametric are used for calculating PSD. Other PSD methods such as Spectrogram, Welch, and Periodogram are calculated using Fast Fourier transform. PSD shows significant features but consumes relatively more time.	[54,58,69,78, 82,83,85,87, 90,114]
Phase features: PLV, IPD, and MPD	Phase locking value (PLV) evaluates the phase synchronization between the two signals using the Hilbert transform [107]. The PLV range is [0-1], with value 0 indicating no synchrony between the signals, whereas 1 indicates the relative phase between the signals is identical.	[75] [76,77]
MWT and STFT	STFT computes the Fourier transform on shifted time windowed signal. MWT, as continuous wavelet transform, computes the magnitude and phase coefficients of a time-varying signal.	[78] [79]
Wavelet packet decomposition	WPD evaluates coefficients from sensorimotor rhythm frequency bands and uses them for classification.	[71,108]

including principal component analysis (PCA) [115], Independent component analysis [116], and Evolutionary algorithms such as Firefly Algorithm (FA) [117], Differential Evolution (DE) [118] Particle swarm optimization (PSO), and Artificial Bee Colony (ABC) Optimization [119,120]. However, in various literature, efforts are made to find the passable combination of feature selection methods and classifiers, a few compared an assortment of algorithms applied to the same MI dataset to analyze their performances in different classifiers. In the work of Ramos et al. [121], a comparison of five filter methods

(Correlation-based Feature Selection (CFS), ReliefF, Consistency, mRmR, and C4.5) and one wrapper method (Genetic Algorithm(GA)) as a feature selection technique is drawn in terms of classification accuracy and kappa value in different classifiers, declaring that GA with linear discriminant analysis (LDA) classifier outperforms all the other combinations. In a study [122], a hybrid model, GA-PSO, is implemented for a two-class MI problem. It is reported that GA-PSO outperformed GA and PSO.

Based on the framework, feature selection methods are categorized into three groups: the wrapper approach, the filter approach, and the embedded approach. The following paragraph defines the three approaches in detail.

- Wrapper methods use a fitness evaluation model to select feature subsets by assigning a score. The subset with the best score is chosen for the classification task. The score of each subset is evaluated using a learning classifier. Each subset trains the classifier and tests using a hold-out test set. The generalization error of the outcome provides the score for that subset. Due to the separate learning of all the subsets, wrapper methods are time-consuming and computationally expensive but generally select the best subset of features and enhance the classification performance for a particular classification problem.
- On the contrary, filter methods calculate the best subset of features without using any classification algorithms. There are two stages involved in filter algorithm. In the first stage, features are assigned a rank based on certain performance measuring methods. In the second stage, the higher-performing features, i.e., features with ranks above a manually chosen threshold rank, are selected for classification. In recent studies, numerous performance measuring methods have been proposed, such as Fisher score [123], the mutual information between the features [124], and ReliefF and its advanced versions [125].

- Filter models are computationally more efficient than the wrapper methods since they evaluate the subset without utilizing the classifier and cross-validation approaches. However, the performance of filter methods is low for some of the applications due to avoidance of the biases of the classifier. For instance, the outcome of ReliefF algorithm would not provide relevant subset of features for Naive-Bayes because it is observed that in most of the problems, the generalization performance of Naive-Bayes classifier enhances with the elimination of relevant features [126]. On the other hand, wrapper model quantifies the features using a predefined classifier and avoids the representational biases of the classifier but takes a longer execution time due to the involvement of cross-validation in its structure that makes it computationally expensive algorithm. Embedded Models incorporate the advantages of (1) wrapper models - they include the interaction with the classification model and (2) filter models - they are far less computationally intensive than wrapper methods. Examples of embedded methods include Lasso Regularization [127], neighborhood component analysis [128], recursive feature elimination using support vector machine (SVM) [129] , ID3 [130], and C4.5 [131].

2.4.3 Classification

In the last stage of MI-BCI system, a decision is made by a classifier to distinguish between the different motor imagery classes correctly. A classifier is an algorithm that learns from the previous data to predict the true class of future data. In motor imagery classification, supervised learning is most commonly used as the training data is labeled. Numerous studies have experimented with the classification of MI data using different classifiers, but the most commonly used classifiers are Linear Discriminant analysis (LDA), support vector machine (SVM), Bayesian classifier (BSC), and logistic regression (LR).

In LDA framework, a hyperplane is created to separate different classes. For classification of two-class problem, it assumes that the data is linearly separable. SVM is very simple yet effective to solve the problem of classification associated with a small sample size of the dataset, non-linear relationship, and multi-class classification. The working of SVM is based on the building of an optimal hyperplane as the decision-making surface to discriminate between the different classes. The SVM model is a representation of the data values as points in space, mapped so that the data values of the different classes are distributed by a clear margin that is globally maximum. More than half of the BCI researches have used either LDA or SVM for classification [132]. Another commonly used classifier in MI tasks, BSC, is based on Bayes theorem. It is a statistical method that evaluates posteriori probability based on priori probability. BSC is relatively more time-consuming. Other commonly used classifiers in MI BCI research are Neural network (NN), K-nearest neighbours (KNN), Mahalanobis distance (MD), and Hidden Markov model (HMM). NN is a combination of artificial neurons arranged in input layer, hidden layers, and output layer. NNs such as Linear vector quantization and backpropagation NN are used for MI task classification in some studies [133–136]. KNN is generally used for multi-class classification problems. It is prone to the curse of dimensionality and hence, rarely used for MI tasks classification [137,138]. MD is suitable for two-class and multi-class and has been used to develop asynchronous BCI [139,140]. In a research [141], an online BCI system is developed using HMM classifier. However, the use of HMM is very limited in BCI field. Other rarely used classifiers are Gaussian classifier and random forest [142,143]. Table 2.4 provides a brief description and references of classification methods in MI signal processing.

Table 2.4 Classification methods in MI signal processing.

Methods	**Descriptions**	**References**
LDA	LDA creates a hyperplane to separate different classes. It assumes that the data is linearly separable to solve a two-class problem. It is used frequently in BCI.	[30,31,33,50, 51,53,54,54, 55,59,61, 65–67,69–73,77,81,84, 87,88,109,112, 113,140–142]
SVM	SVM builds an optimal hyperplane as the decision-making surface to discriminate between the different classes. In SVM, the data values of the different classes are distributed by a clear margin that is globally maximum. SVM gives good classification performance in BCI field and is less prone to curse of dimensionality.	[56,58,60, 64,65,68,73, 75,78,86,91, 92,111,113, 141,143–145]
BSC	BSC is a statistical method that calculates posteriori probability based on priori probability. This algorithm takes relatively more time.	[52,63,72,85, 93,146]
NN	NNs are suitable to train almost all kind of datasets. It uses an assortment of artificial neurons arranged in layers. For MI tasks classification, Linear vector quantization NN and backpropagation NN are used in some studies.	[133–136]
KNN	KNN is seldomly used for MI classification because it is suitable for multi-class problems.	[137,138]

2.4.4 Spatial filtering-based models in MI-BCI:

In recent years, various computational models based on pattern recognition techniques and machine learning are proposed for MI task classification [27,28,169–171]. Among them, common spatial pattern (CSP) is the most widely used algorithm for feature extraction from the raw MI EEG dataset [172]. CSP projects the MI EEG data recorded for two classes (left-hand and right-hand MI) into a subspace such that the data in new subspace become more discriminative between the two classes [172]. The variance of the projected data from one class gets maximized while that from the other class gets minimized. With these variance values, a feature set is generated to train a classifier. However, the performance of CSP is dependent on the choice of the frequency band. In general, CSP is applied on a wide frequency band of range 4-40 Hz. This manual selection of a fixed wide frequency band for different subjects leads to insignificant classification rate because the neural responses are subject-specific. To alleviate this issue, advancements have been made in conventional CSP algorithm by filtering EEG using different narrow frequency band filters and then applying frequency band selection algorithms to select subject-specific frequency bands automatically [68,173–175]. In the work of [173], a method, namely, common spatio-spectral pattern (CSSP), was proposed to enhance the CSP performance by optimizing a finite impulse response filter within CSP. An improved CSSP method, termed as Common sparse spectral-spatial pattern (CSSSP), was proposed by [176], which finds spectral patterns common to all the channels instead of finding different spectral patterns for each channel as in CSSP.

Further, sub-band common spatial patterns (SBCSP) proposed in [177] used multiple narrow bandpass filters to filter the EEG before extracting the CSP features separately from each sub-band. After that, linear discriminant analysis (LDA) reduced the dimensionality of SBCSP extracted features. However, SBCSP attained a higher classification performance in comparison with CSSSP, CSSP, and CSP, but the mutual information between the CSP features from different sub-bands was ignored. The work proposed in [68] presented a filter bank common spatial pattern (FBCSP) to select optimal frequency band by measuring the correlation among the CSP features from multiple sub-bands. Recently, discriminative filter bank common spatial pattern (DFBCSP), an advanced variant of FBCSP algorithm, was introduced [178], where only the most discriminative sub bands were chosen by exploiting a fisher score. Table 2.5 gives brief description and references of spatial filtering-based models in MI signal processing.

Table 2.5 Spatial filtering-based models in MI signal processing.

Methods	**Descriptions**	**References**
CSP	CSP enhances the discriminating capabilities between the classes by maximizing the variance of one class while minimizing that of the other class. A bandpass filter with a wider frequency band (4-40Hz) is applied to the EEG to extract CSP features. However, a fixed wide frequency band fails to provide optimal classification results because ERD/ERS patterns occur at different frequency bands in different subjects [179]. Hence, the performance of CSP is highly dependent on the choice of the frequency band in which MI-related EEG is filtered.	CSP [50,51,53, 54,71,75,146]; CSP-log [28,30,33,54, 55,67,70,72, 73,77,86,164, 165]
CSSP	CSSP enhanced the CSP performance by optimizing a finite impulse response filter within CSP. An improved CSSP method, termed as Common sparse spectral-spatial pattern (CSSSP), was proposed, which finds spectral patterns common to all the channels instead of finding different spectral patterns for each channel as in CSSP.	CSSP [173] CSSSP [176]
SBCSP	In SBCSP, filters were utilized to filter the EEG, and thus CSP features were extracted. However, SBCSP achieved significant classification results but ignored the mutual information between the CSP features from various sub-bands.	[177]
FBCSP	FBCSP used mutual information-based feature selection to optimize CSP features at different frequency bands.	[68]
DFBCSP	As an advanced version of FBCSP, Discriminative FBCSP (DFBCSP) applied overlapped shifted filters on the EEG before extracting CSP features and selected the optimal sub-bands evaluating a fisher score	[178]

2.4.5 Time- window optimization models in MI-BCI:

Apart from optimizing the frequency bands, another critically important but very rarely investigated research area is the optimization of time windows of MI-related EEG. In most of the previous studies [68,182], the time segment used by feature extraction methods for MI classification is usually fixed (i.e., between 0.5 and 2.5 s after the cue), but the brain of different subjects shows different time latencies to given MI tasks [183]. MI paradigm generally has an imagination preparation period (0 to 1 s) and post imagination period (3.5 to 4 s) [183]. These intervals may not show the presence of MI-related patterns in all subjects. Moreover, a significant time window that captures discriminative features varies subject to subject [184]. Therefore, for each subject, it is important to select a time window that captures discriminative features. Hence, simultaneous optimization of time windows and frequency bands is crucial to further improve the MI classification. Basically, EEG is segmented into multiple time windows using sliding time window approach, then each time window is filtered at multiple frequency bands, and finally, discriminative features are extracted. With time windows and frequency bands, the extracted feature space becomes multi-view tensor data. Algorithms [184–186], which have used time window optimization and frequency band optimization for MI classification, converted the multi-view tensor EEG into a single large matrix by unfolding and concatenating the multiple matrices and then applying a feature selection method. However, this approach results in loss of internal structure of multi-view EEG data and degrades the BCI classification accuracy [187].

2.5 Signal processing in SSVEP-based BCI

Numerous target identification techniques have been designed to identify SSVEPs in BCI [17,188,189]. The Power Spectrum Density Analysis (PSDA) – based methods such as Fast Fourier Transform (FFT) were primarily used for this purpose [17]. Feature extraction method in SSVEP based BCI detects the frequency in EEG time series. PSDA uses the FFT spectral calculation method to identify the SSVEP frequency. The disadvantage of PSDA is prone to noise. To overcome that, superior methods are introduced. In the work of Lin et al. [190], canonical correlation analysis (CCA) is proposed for feature extraction in SSVEP. The CCA method evaluates the maximal correlation coefficient between the predefined sinusoidal reference signals and multi-channel EEG signals filtered at simulation frequency. Thus, it recognizes the target frequency using the evaluated canonical correlation values. The CCA method has found wide applications in online SSVEP BCI research because of its advantages like superiority in performance, ease of execution, and no calibration[18,191–193]. CCA outperformed PSDA methods. Hence advanced variants of CCA were introduced aiming to further improve the classification accuracy [194–196]. Although CCA method is effective in identifying SSVEPs, its performance is limited by involvement of spontaneous EEG signals [197]. In some studies[194,198], Effects of spontaneous EEG signals on the performance of CCA-based SSVEPs detection are reduced by incorporating phase and latency information in CCA method. Some of the advanced versions of CCA are phase constrained CCA (PCCA), individual template-based CCA (IT-CCA), multi-way CCA (MwayCCA), L1-regularized multi-way CCA (L1-MCCA), and multi-set CCA (MsetCCA). Table 2.6 describes the advanced variants of CCA method.

Kalunga et al. [188] presented the use of Riemannian geometry in online SSVEP-based BCI. However, all these methods do not incorporate learning algorithms, which have been used in various BCIs other than SSVEP-based BCI [199,200]. Also, these methods have shown good accuracy in target identification when the subject is actually focusing on the target, but they do not identify idle state, i.e., when subject is not targeting any frequency, with the same accuracy. Hence increased number of False Positive outputs are often observed in these SSVEP-based BCIs.

Table 2.6 Advance CCA methods in SSVEP signal processing.

Method	**Description**	**References**
PCCA	In PCCA approach, the phases of the sinusoidal reference signals were fixed according to the visual latency estimated from the calibration data.	[194]
CCA (IT-CCA)	IT-CCA method, the reference signals were VEP templates obtained by averaging across multiple EEG trials in the calibration data from each individual.	[198]
MwayCCA	MwayCCA method finds appropriate reference signals for SSVEP detection based on multiple standard CCA processes with the calibration data.	[195]
L1-MCCA	L1-MCCA method optimizes the reference signals in SSVEP recognition.	[201]
MsetCCA	MsetCCA method optimizes the reference signals from common features in multiple calibration trials.	[196]