Chapter 8

Concluding Remarks and Future Directions

In this chapter, we have highlighted the main contribution of our work.

8.1. Contribution Summary

Search space optimization has been studied for efficient exploration of large solution space and exploitation of optimal but better solutions. In this thesis, we develop five optimization methods to improve the global performance of the Brain-Computer Interface. In this context, we categorized the popular baseline optimization methods into three categories: (1) Filter, (2) Wrapper, and (3) Hybrid methods. We reviewed the advantages and disadvantages of each category in terms of solution quality, effect on classification accuracy, and execution speed. Considering the limitations of the above-mentioned categories, we performed multiple experiments to improve the overall performance of the filter, wrapper, and hybrid methods.

In Chapter 3, we implement an effective curve smoothing method to minimize the noise and outliers in the early phase of multiclass MI classification. In this experiment, we used the 3rd-order Savitzky-Golay filtering approach on the signal window size of 1800 points. The final classification results showed that the applied signal smoothing strategy improves overall classification results significantly compared to different state-of-the-art models. The key limitation of this experiment was high time complexity because of the involvement of a decision tree-based XgBoost classification scheme. Conventionally, height management and selecting optimal node-splitting conditions is a time-consuming procedure and reduces the execution rate of the XgBoost classifier. This limitation was reduced in Chapter 4 where we propose a filter-based channel selection method using a novel Dynamic Channel Relevance (DCR) technique. This approach uses information-theoretic paradigms to determine the best set of MI-specific channels or electrodes. Since only a relevant number of channels were selected using the DCR method, the overall temporal complexity, as well as execution rate, were also reduced. Based on the spatial distribution of the selected channels, we conclude that the frontal and parietal cortex also play a significant role in the regulation of MI tasks. This finding was strongly validated in chapters 5 and 6 where metaheuristic algorithms were used to determine the optimal channel subset.

In Chapter 5, we develop a wrapper method to deal limitations of the proposed filter approach. Here, we use the local search strategy of the Firefly Algorithm (FA) to update the position of observations and sort the channels using the associated Fisher Information score. The rank of each channel is decided in the neighborhood of candidate solutions. This approach improves the overall performance significantly compared to different filter and wrapper methods. However, this approach was limited by some key challenges such as (1) local optima entrapment, (2) premature convergence, and (3) poor solution exploitation when used with a large number of channels. These limitations were resolved in Chapter 6 using global and local search strategies of the Butterfly Optimization Algorithm (BOA). It enjoys the advantages of two mutual complementary sigmoid functions to compute better-quality solutions. Similar to Chapter 6, the solution quality is referred to in terms of low classification accuracy error rate and high channel reduction rate. The optimization scheme guarantees to improve the solution quality because it selects the best solution between the sigmoid solution and the previous solution. The global performance of the proposed algorithm indicates that this method suits well for the large-size dataset. Similar to previously proposed methods, this algorithm realizes superior performance to various state-of-the-art methods.

Finally, in Chapter 7, we develop a hybrid feature selection using Dynamic BOA (DBOA) and Feature Interaction Maximization (FIM). DBOA is an improved variant of BOA that uses a local search strategy using mutation operation. In each iteration, DBOA updates the position of instances and FIM selects the features that are highly associated with the target variable. Moreover, the proposed methods showed three noticeable improvements over conventional filter and wrapper methods: (1) better trade-off between the exploration and exploitation phase, (2) effective dealing with local optima problems, and (3) avoiding irrelevancy and redundancy of selected features. The validation results obtained from twenty publically available datasets indicate that the proposed hybrid method significantly improves classification accuracy with minimum information loss compared to ten wrapper-based filter selection algorithms. In addition, the proposed method obtains superior specificity and sensitivity measures compared to baseline models.

8.2. Scope for Future Work

This section illustrates various new research dimensions that can be employed to enhance the performance of the experiments discussed in this thesis.

8.2.1. Graph-based Similarity Search Techniques for Channel Selection

Graphs are one of the effective data structures that can be used to explore the relationship between normal and candidate channels in an agile way. To transform channel selection into a graph-based similarity search problem, a common approach is to use the one-to-one mapping between channels and nodes and edges connecting pairs of nodes as the similarity relationship between pairs of channels [152]. The similarity measures can be either distance or difference measures that can be computed between pairs of nodes. Here, the set of known solutions works as a candidate channel set while the remaining functions as a normal channel set. Further, the application of clustering-specific metaheuristic algorithms can be used to update the instance or feature vector of each channel iteratively and relative distance is computed. After a set of iterations, the clusters with the maximum number of candidate channels and the minimum number of normal channels can be used in the classification process.

In other words, this set represents the group of strongly coupled normal channels with candidate channels. This iterative mapping strategy has been effectively used to solve community detection problems where a set of interactive nodes form communities based on their structural and functional properties. As per our knowledge, this technique is not implemented in the channel selection problem, therefore, in the future, it can be used to determine the most effective set of channels that can provide the most discriminatory MI- MI-specific features and improve the overall performance of the BCI model.

8.2.2. Enhancing Causal Inference for Channel Selection

Causality defines the influence of one event, process, or state on a different event where the cause is partially responsible for the effect, and the effect is partly dependent on the cause [153]. In machine learning, causality performs better in pattern recognition than correlation because correlation limits to determine the effect of the selected channel on the classification accuracy. The idea of implementing a causality paradigm in channel selection is fundamentally based on the greater significance of some special EEG channels while measuring the causality of these channels onto others, during the MI actions. A time-series data collected from channel X is said to Granger-cause Y if X provides predictive information about (future) values of Y over and above what may be predicted from past values of Y. Based on the computed predictive information the channels can be sorted and their effect can be seen on the classification accuracy. Based on the performance, the most effective channel subset can be selected and used for the BCI experiment.

8.2.3. Multi-view Deep Learning with Channel Aware Attention Scheme

In multi-channel data analysis, interpreting the learned representation is important to understand the significance of each channel in the original set. Here, we focus on discovering the contribution scores of the channel and identifying which ones are crucial to identifying MI patterns. The main theme is to derive a global context vector that captures relevant channel information to help detect the performed MI tasks. In multi-view learning [154], the channel attention mechanism explores the global context vector as a function of two different latent representations: (1) channel-specific, and (2) global views of each channel. The first representation calculates the energy score of each channel without capturing its relationship with other channels while the latter representation computes the relationship among all possible channel pairs. Combining both information vectors highly associated channels are filtered to compute the global classification accuracy.

8.2.4. Feature Selection Using Graph Mining

Graph mining is an interesting concept that can be used to mitigate the CoD problem with minimum information loss [155]. In addition, it can be used in unsupervised learning where the label information is absent therefore, direct application of any label-guided strategy fails without computing the optimal feature subset. Some graph theoretic concepts such as vertex cover, graph cut, and independent set can be used to compute the significance and interdependency of features in the classification process. One prominent example in the category is the Feature Association Map (FAM) which aims to visualize the inter-feature similarity or potential redundancy present in the target dataset. Similar to our proposed feature selection method, FAM also maintains an effective balance between relevance and redundancy measures. In the future, FAM can be extended as Intuitionistic Fuzzy FAM by involving feature uncertainty management. Here, the selection or rejection of a feature depends on its membership value which can be computed by a unique membership function. Moreover, the effectiveness of the graph-based feature selection method requires a comprehensive study of set merging and splitting techniques and their validation on multiple datasets.

8.2.5. Enhancing the Scope of Proposed Optimization Methods

In the future, the application of our developed methods can be tested on various industrial projects such as task variant allocation in robotics where different software processes are assigned to multiple processors in a distributed environment to maximize the efficiency of a robot. The selection procedure is optimized in such a way that the group of allotted tasks is minimally independent of each other otherwise they can not be simultaneously executed. Another application of our methods is query optimization where the query optimizer attempts to determine the most efficient way to execute a given query by considering the possible query plans without altering the target. In this problem, the optimizer selects the most optimal path in the query tree where each branch of the tree refers to a specific query plan.

Moreover, the selection of a path depends on various factors such as minimum join operations, null value handling techniques, and ordering of selected data items. Based on historical data, some candidate operations and their suitability with other filtering operations such as selection, projection, and join can be grouped to execute new queries. These advanced techniques can be used to develop a minimal set of global search operations that can be used to execute most of the queries.

Our proposed optimization techniques can be extended in related medical applications such as multi-channel neurodegenerative pattern analysis in Alzheimer's disease, and epilepsy classification and investigate the role of Diffusion Tensor Imaging (DTI) in the differentiation of different cognitive state patterns.