# Cognitive Fatigue Detection from EEG Signals using Topological Signal Processing

Arup Kumar Das, Kriti Kumar, Rahul D. Gavas, Dibyanshu Jaiswal, Debatri Chatterjee,
Ramesh Kumar Ramakrishnan, M Girish Chandra, Arpan Pal
TCS Research and Innovation, Tata Consultancy Services Ltd., India.
Email: {arup.kdas, kriti.kumar, rahul.gavas, dibyanshu.jaiswal, debatri.chatterjee, ramesh.kumar, m.gchandra, arpan.pal}@tcs.com

***Abstract*—Topological signal processing has attracted substantial interest lately owing to its attribute of multi-scale tracking of simplicial complexes. This distinctive aspect is exploited to study the topological evolution of time series/signals. Specifically, EEG signals analysis is considered here for the challenging task of cognitive fatigue detection. This work utilizes the topological attributes like Betti numbers, and persistent homology of dimension 0 and 1 extracted from EEG signals to study the cognitive state of an individual. Using the CogBeacon dataset, a comparison of the topological features with the *conventional* time and frequency domain features is presented. Random forest classifier is used to classify the fatigue state. Results show that the performance of topological features is at par with the *conventional* features even when significantly less number of topological features are used. Also, enhancement in classification accuracy is observed by appropriately combining both *conventional* and topological features which outperforms the state-of-the-art method for fatigue detection. Additionally, recursive feature elimination is applied on combined features to reduce redundancy by selecting a subset consisting of prominent features. Analysis indicates that all topological features derived from EEG signals contribute to the best performing subset, which also increases the overall accuracy.**

***Index Terms*—Topological Signal Processing, Electroencephalogram, Cognitive Fatigue, Persistent Homology, Classification, Recursive Feature Elimination**

## I. Introduction

Cognitive Fatigue (CF) fundamentally refers to mental tiredness. It can cause decreased productivity in executing routine procedures that can have adverse affect in diverse fields such as medicine [1], traffic [2], defense [3], and construction industry [4] among many others. Gradual accumulation of fatigue in individuals result in lack of attention which in turn hampers the overall performance and efficiency of the task being carried out. Thus, timely detection of CF becomes crucial specially for reducing imparity in critical decision making abilities.

Detection and prediction of CF is quite challenging owing to the high level of uncertainty and variability associated with it across individuals [5]. Cognitive states can be assessed in one of 3 ways [6]: (i) using subjective feedback or questionnaire, (ii) using task performance metrics, and (iii) using physiological sensing. The work in [7] uses questionnaire or subjective feedback to assess mental fatigue. The main challenge in such systems is the bias associated with the participants while providing the feedback. In addition, the assessment can be done either in the beginning or at the end of the task. If it is carried out at multiple intervals during the task, it interrupts the participant from performing the task. The second measurement technique using task performance metrics or test scores can be used only at the end of the task to determine the fatigue levels and are more task-specific and lack generalization ability across tasks. Both these methods lack the ability to monitor the participant continuously and hence, physiological sensing-based measurements are now a days used to assess the cognitive state of the individuals during task execution. However, this usually requires high resolution sensor devices which are costly and difficult to handle. On the other hand, affordable low cost sensors usually provide low resolution data and the corresponding signal processing, noise reduction, and so on are challenging [8]. This work exploits the use of a low cost and easy to deploy electroencephalogram (EEG) device to compute effective features for determining the CF of an individual while performing routine tasks in an enterprise scenario. The focus is not on clinical applications that require medical grade EEG devices. It is more targeted towards mass deployment for CF detection.

EEG signals are used to examine neuron potentials at targeted positions for identification of possible CF conditions [5], [9]. Both these methods utilize the *conventional* EEG features for classifying the CF state. The *conventional* features comprise of frequency and time domain features that describe the morphology of the EEG signal. Recently, topological aspects of the EEG signals have been explored which helps in capturing the crucial markers pertaining to the signal [10]. The topology-based signal characteristics can be obtained by using concepts from Topological Signal Processing (TSP) [11]. TSP aims to reflect upon properties of a signal by inspecting different topological shapes embedded in the signal. The various mathematical tools employed in the TSP framework can distinguish the intrinsic shapes of the signal from those due to noisy offshoots [12]. This makes TSP robust to noise. Owing to such an attribute, TSP is being increasingly used in a multitude of applications over *conventional* signal processing techniques. The work in [12] used TSP features for classifying EEG data for motor-intention based brain-computer interface systems. The TSP features were reported to be more effective than the *conventional* band power based features. The work in [10] used TSP features in combination with the *conventional* features for epileptic seizure detection in EEG signals. However, to the best of our knowledge, the TSP based features have not been used for cognition related tasks. Hence, we aim to explore the applicability of these features in the domain of cognitive state classification where the cognitive state under test is fatigue.

In this work, TSP-based features like Betti numbers and persistent homology of dimension 0 and 1 are used to study the cognitive state of the person using EEG signals. Using the CogBeacon dataset, it is observed that with significantly less number of features, TSP based method is able to achieve performance equivalent to the state-of-the-art *conventional* EEG features based method. The classification accuracy is shown to further increase by combining the *conventional* and TSP based features, thereby, outperforming the state-of-the-art method for cognitive state assessment. The main contributions of the paper are: $(i)$ exploration of TSP features for the classification of cognitive state like fatigue, $(ii)$ comparison of TSP features with *conventional* EEG features, and $(iii)$ combining TSP and *conventional* EEG features for better classification accuracy.

Towards providing the necessary details of the proposed work, the rest of the paper is organized as follows. A brief background on TSP is given in Section II. Section III presents the relevant TSP features used for fatigue classification. This is followed by Section IV which presents the classification results obtained with TSP and *conventional* features. Finally, Section V concludes the paper.

## II. Brief Background on TSP

TSP focuses on the characterization of different algebraic structures or homology observed in the data. TSP is being increasingly used in various applications as it can analyze the structures of the associated topological spaces at multiple scales. Some applications of TSP include forecasting [13], classification [14], and clustering [15]. A comprehensive review on the performance of topological features for a diverse range of use cases is detailed in [16]. A brief description of the different processing blocks involved in TSP analysis is presented below.

The scalar time series signal is divided into appropriate windows or segments chosen based on the properties of the signal under observation. Subsequently, a high dimensional point cloud is constructed by time delay embedding. Takens' time-delay embedding technique [17] is a popular technique that is used for this task and is described as follows. Given a scalar sequence $\{x[n]\}$, a $m$-dimensional sequence $\{X[n]\}$ is computed for time epoch $n$ where $m \geq 2$. The $m$-dimensional observation $X[n] \in \mathbb{R}^m$ can be expressed as:

$$X[n] = \Big(x_1[n], x_2[n], \cdots, x_m[n]\Big), \tag{1}$$

where, $x_j[n] = x[n + (j-1)d]$ for $j = \{1,..,m\}$ with $d$ being the delay parameter. The values of $m$ and $d$ are chosen appropriately based on the application.

With the multi-dimensional point cloud, the topological shape of the data at a particular scale can be studied using the Vietoris-Rips complex (or simply, Rips complex). Given a Rips complex, Betti numbers are calculated to characterize the underlying topological space. The $i$-th Betti number or $\beta_i$ represents the number of structures observed in dimension $i$ of the topological space. In any topological space, $\beta_0$ represents the number of connected components, $\beta_1$ represents the number of holes, $\beta_2$ represents the number of voids (or 2-D holes), and so on. For example, the topological space of a sphere is characterized by $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 1$, and $\beta_i = 0$ for all $i \geq 3$. More details on Betti numbers for common topological spaces can be found in [11].

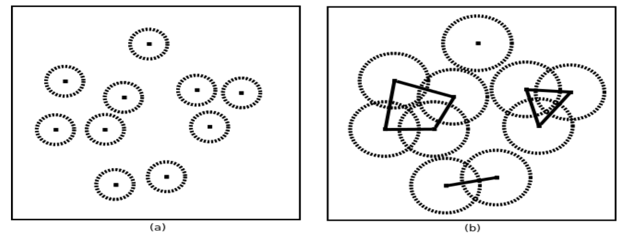


Fig. 1. Evolution of Rips complex at two different radii - (a) $r_1$ and (b) $r_2$, with $r_1 < r_2$.

The topological space obtained at a chosen scale may not be sufficient to reveal all underlying topological structures pertaining to the data. Thus, instead of studying topological shape only at a particular scale, it is necessary to track these shapes across all possible scales. This can be done by successively increasing the scale parameter, starting from zero, in small steps and observing the corresponding Rips complex for necessary shape information. At any scale $r_\epsilon$, Rips complex is constructed on the point cloud and the corresponding Betti numbers are calculated. Subsequently, the scale is incremented by a small value $\epsilon$ and the corresponding Rips complex is studied. The process is continued till all possible scales are considered, i.e. when the underlying topological space does not change with further increase in the scale. This method of analyzing homology groups at different scales is known as the Rips filtration technique [18].

In the Rips filtration technique, the Rips complex at any scale $r$ can be constructed by connecting the pair of vertices (or pair of sample points of the point cloud) by an edge if the distance between them is $\leq 2r$. Fig. 1 presents a pictorial description of Rips filtration for two different scales $r_1$ and $r_2$ for a 2-dimensional point cloud consisting of 10 sample points. At radius $r_1$, since no pairs of vertices are within the required distance threshold, there are no edges. The Betti numbers for this scale are $\beta_0 = 10$ and $\beta_i = 0$ for all $i \geq 1$. When the scale is increased to $r_2$, a number of vertices pairs lie within the distance threshold and so are connected with an edge (as shown in the Fig. 1). The Betti numbers for scale $r_2$ are $\beta_0 = 4$, $\beta_1 = 2$ and $\beta_i = 0$ for all $i \geq 2$. This process of incrementing the scale and connecting relevant vertex pairs is continued until every vertex has an edge with every other vertex in the point cloud and hence, the structure does not change with further increase in the scale.

It can be observed from Fig. 1 that the Betti numbers

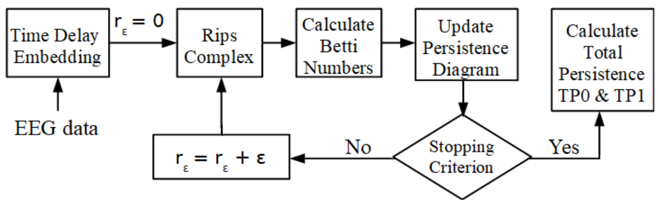
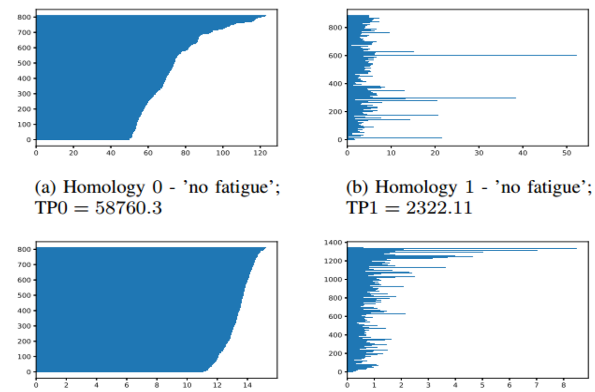



Fig. 2. Flow diagram for computation of TSP features.



(a) Homology 0 - 'no fatigue'; TP0 = 58760.3

(b) Homology 1 - 'no fatigue'; TP1 = 2322.11

(c) Homology 0 - 'fatigue'; TP0 = 10798.1

(d) Homology 1 - 'fatigue'; TP1 = 655.85

Fig. 3. Persistence levels of different homology groups observed through Rips filtration. The total persistence levels for different homology groups for 'no-fatigue' and 'fatigue' classes is indicated. The independent axis indicates the lifespan or persistence level, and the dependent axis refers to the indices of different homology groups.

change with the change in scale as the associated topological space also changes. During the evolution of the Rips complex, different topological structures or homology groups get created and destroyed (i.e. merged with other topological structures). For a particular homology group that gets created at scale $r_b$ and destroyed at scale $r_d$, the persistence is defined by $r_d - r_b$, where $r_b$ and $r_d$ denote the birth and death time of that homology group. Topological signatures such as Persistent Diagram (PD) and barcode graphs give knowledge about persistence levels of different homology groups. Specifically, greater interest lies in knowing the persistence of different homology groups that exist in the data. This is due to the fact that the structures which are persistent for large intervals can be attributed to the inherent structure of the data, whereas the smaller persistent structures generally imply noise.

## III. Methodology for Cognitive Fatigue Detection

This work presents the use of topological features for classifying the cognitive state of the individual using the relevant EEG channels [19]. In this section, a detailed explanation of the different topological features is provided along with a brief description of the dataset considered for CF detection.

### A. Dataset

The publicly available CogBeacon [5] data is considered here for CF detection. This data contains 4 channel EEG data collected from 19 participants while performing the original Wisconsin Card Sorting Test (WCST) [20], and its modified versions [5] for assessing the mental tiredness of the individual. EEG data is collected from frontal and temporal lobes of the human brain, by placing electrodes at locations TP9, AF7, AF8, and TP10 (channels 1 through 4 respectively),as per the International Standard 10-20 system of EEG electrode scalp locations. There are a total of 76 sessions of EEG data that is collected using the Muse headset [21] with a sampling frequency of 220 Hz. Two modified versions of Wisconsin Card Sorting Test (WCST) are designed to generate increased user interaction and focus, resulting in higher CF. While performing the tasks, the participants reported fatigue state by a button press. This is considered as the ground truth to learn models for CF detection.

### B. Signal Processing and Feature Extraction

The raw EEG signals from the CogBeacon data is pre-processed to remove the eye-blink artifacts using the Independent Component Analysis (ICA) technique [22]. The average duration of each trial of the cognitive tests is found to be $4.53 \pm 1.43$ seconds. Thus, non-overlapping windows of 4 seconds duration are considered to compute the TSP and *conventional* features. Each window comprising of 880 instances is labeled as 'no fatigue' or 'fatigue' based on the cognitive state associated with the last instance of that window. It is assumed that once a participant reports a 'fatigue' state, the participant remains in that state till the end of the session.

*TSP features:* TSP features are computed for each of the 4 EEG channels using the TSP pipeline presented in Fig. 2. Time-delay embedding with $m = 70$ and $d = 1$ is performed on each EEG channel to build a high dimensional point cloud of 811 points in $\mathbb{R}^{70}$ space for each window. The value of the embedding dimension is chosen using grid search with values ranging from $5 - 110$ in steps of 5. For this data, it is observed that the performance of the TSP features has little effect on the value of $m$ chosen for experimentation. The best results are obtained with $m = 70$, and so it is considered here.

Rips filtration (described in Section II) is subsequently used to calculate Betti numbers corresponding to the commonly used homology groups of dimension 0 and 1. PD is used to ascertain persistence levels of these homological groups. The total persistence TP0 and TP1 corresponding to 0-dimensional and 1-dimensional persistent homology respectively are considered for fatigue detection. TP0 denotes the sum of persistence levels of all 0-dimensional homological structures. On the other hand, TP1 denotes the sum of persistence levels of all 1-dimensional homological structures. These features are chosen as they are able to discriminate between the cognitive state of the individual. Fig. 3 shows the persistent levels for homology groups of dimensions 0 and 1 for a particular participant computed using channel 2 of the Muse EEG headset. It can be seen that the lifespan of topological structures are much longer for both homology groups when the participant is in 'no fatigue' state. The values of TP0 and TP1 for both 'no fatigue' and 'fatigue' are also presented in Fig. 3. A good separation is observed in these feature values for both the cognitive states, thus, demonstrating discriminative ability of TSP features in detecting CF. A similar trend is observed for other participants. Thus, in this work, TP0 and TP1 are calculated for each EEG channel, and a total of 8 features are used for fatigue classification. The TSP features are computed using the persistent homology library *ripser* [23].

*Conventional EEG features:* As in [9], the *conventional* EEG features calculated comprise of both frequency and time domain features. Nine frequency domain features consisting of 5 EEG band powers (delta, theta, alpha, beta and gamma) and 4 derived features based on the band power ratios are extracted. In addition, 11 time-domain features like mean, skewness, kurtosis, standard deviation, variance, minimum, maximum and three Hjorth parameters (activity, mobility and complexity) are extracted. Thus, in total, 80 *conventional* features (20 per EEG channel) are extracted for assessing CF.

### C. Learning Models for Classification

The CogBeacon dataset has high class imbalance with less number of 'fatigue' instances reported by the participants. Considering 4 second windows, there are a total of 3807 'no fatigue' and 1805 'fatigue' instances. The models learned using this data can be biased and inaccurate. To address this issue, Synthetic Minority Over-sampling Technique (SMOTE) [24] is used on the topological and *conventional* feature sets individually to generate the class balanced data. SMOTE interpolates on the minority class to restore the class balance. Post SMOTE application, 3807 number of instances are obtained for both 'no fatigue' and 'fatigue' classes. Random Forest classifier is used to learn 4 models with different sets of features: $(i)$ $TSP_F$ - topological features, $(ii)$ $Con_F$ - *conventional* features, $(iii)$ $Joint_F$ - combination of topological and *conventional* features, and $(iv)$ $Joint_{FS}$ - features obtained by selecting best performing features of $(iii)$. Recursive Feature Elimination (RFE) [25] is performed on the learning model in $(iv)$ to select the top features for classification. The performance of the different models are evaluated using the standard metrics like accuracy, precision, recall and F1-score. Results of the proposed method are compared against the state-of-the-art method employing *conventional* EEG features [9].

## IV. Results

The performance of the TSP and *conventional* features on the CogBeacon dataset is presented in this section. A 10-fold cross-validation averaged over 10 times is used to report the corresponding performance metrics of the learning models $(i) - (iv)$, described in Section III-C, for assessing CF. For the learning model in $(iv)$, RFE [25] is used on class balanced data to perform feature selection on the set of 88 TSP and *conventional* features. To do this, the data is split randomly into train and test sets in a $90 - 10$ ratio. RFE is used to rank these features individually according to their performance obtained with the random forest classifier on the training data. Next, these features are progressively grouped together into 88 different collections in order to train the learning models, and ascertain the corresponding accuracies of each group of

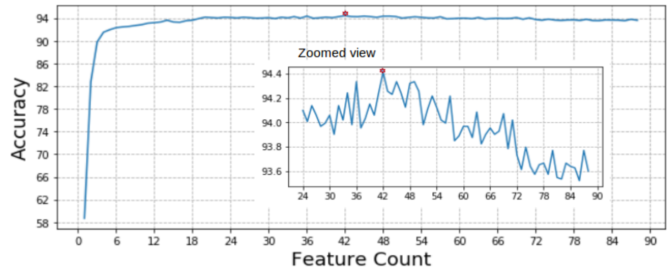



Fig. 4. Average accuracy (in %) of selected number of features using RFE. A zoomed in view of the the accuracies for feature sets containing 24 to 88 most relevant features is presented in the inset box.

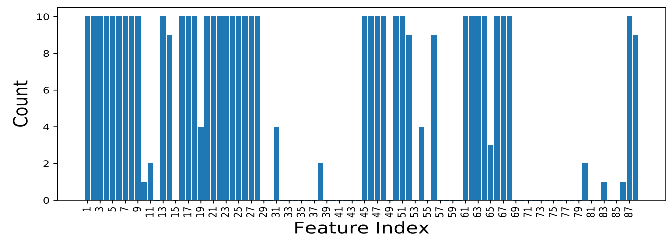



Fig. 5. Feature indices $1 - 8$ represent TSP features, while indices $9 - 88$ represent conventional features. The figure shows the count of different features occurring in the best performing RFE feature set over 10 iterations.

features. The grouping is done in steps of 1, starting from the most relevant feature, and including the next best feature, as dictated by RFE, to the feature set. Thus, a total of 88 accuracy values are obtained, one corresponding to each of the feature collections. This entire process is iterated 10 times in order to negate the randomness of train-test sets selection. Fig. 4 shows the accuracy values corresponding to 88 different feature collections averaged over 10 iterations. It is observed that the collection containing 42 features performs the best with an accuracy of 94.41% as opposed to 93.6% obtained with total 88 features. Thus, it can be concluded that the use of all 88 features is not required to achieve the best accuracy.

The top 42 features for the classification task considered in this work are obtained as follows. Since the train and test instances are generated randomly, the 42 top features are not the same across iterations. Hence, those 42 features which emerge most number of times across all 10 iterations are considered as the top features. Fig. 5 depicts the count of all 88 features appearing in the first 42 positions across the different train and test sets. The first 8 feature indices in Fig. 5 represents TSP features, and the *conventional* frequency and time features are represented in the remaining feature indices. It is observed that the TSP features from all 4 EEG channels consistently appear in the collection of 42 features across all iterations. This indicates the suitability of topology based features for CF detection. It is observed that, among the top 42 features, theta, alpha, beta, gamma and activity features from all 4 EEG channels are present. On the other hand, features like variance, standard deviation, etc. appear for few of the channels in this best performing feature set. $Joint_{FS}$ considers these 42 features selected by RFE technique.

Table I presents the values of the performance metrics

TABLE I
RESULTS OF COGNITIVE FATIGUE DETECTION WITH DIFFERENT MODELS

| Model | Accuracy (%) | Precision | Recall | F1-score |
|---|---|---|---|---|
| **Class balanced data** | | | | |
| **Model** | **Accuracy (%)** | **Precision** | **Recall** | **F1-score** |
| $TSP_F(\#8)$ | $92.5 \pm 0.1$ | $0.91 \pm 0.001$ | $0.94 \pm 0.001$ | $0.93 \pm 0.001$ |
| $Con_F(\#80)$ | $92.7 \pm 0.19$ | $0.91 \pm 0.002$ | $0.95 \pm 0.001$ | $0.93 \pm 0.001$ |
| $Joint_F(\#88)$ | $93.3 \pm 0.11$ | $0.91 \pm 0.002$ | $0.96 \pm 0.002$ | $0.93 \pm 0.001$ |
| $Joint_{FS}(\#42)$ | $93.83 \pm 0.11$ | $0.92 \pm 0.001$ | $0.96 \pm 0.001$ | $0.94 \pm 0.001$ |
| **Class imbalanced data** | | | | |
| **Model** | **Accuracy (%)** | **Precision** | **Recall** | **F1-score** |
| $TSP_F(\#8)$ | $89.25 \pm 0.2$ | $0.85 \pm 0.002$ | $0.8 \pm 0.006$ | $0.83 \pm 0.003$ |
| $Con_F(\#80)$ | $89.52 \pm 0.15$ | $0.85 \pm 0.003$ | $0.81 \pm 0.004$ | $0.83 \pm 0.003$ |
| $Joint_F(\#88)$ | $90.18 \pm 0.25$ | $0.86 \pm 0.004$ | $0.83 \pm 0.005$ | $0.84 \pm 0.004$ |
| $Joint_{FS}(\#42)$ | $90.96 \pm 0.22$ | $0.87 \pm 0.003$ | $0.84 \pm 0.004$ | $0.86 \pm 0.003$ |

obtained with the different models for detecting CF. The numbers succeeding model names in this table correspond to the number of features used for that model. In addition to the class balanced data, results with class imbalanced data are also provided here for reference. It can be observed that the accuracy of both topological and *conventional* features is high with class balanced data. Further, the performance of the model using 8 topological features is similar and at par with the model using 80 *conventional* features. This indicates that the transient temporal variations of the EEG signals caused by CF are effectively captured in the evolution of different homology groups considered here. It is also observed that the combination of the topological and *conventional* features outperforms the state-of-the-art method. Additionally, feature selection on the combined features further increases the accuracy and reduces the feature count from 88 to 42 incorporating all the TSP features. It is also observed that the F1-score of the learning model $Joint_{FS}$ is better than the other considered models for both the class balanced and imbalanced data.

Further, correlation analysis is performed to study the inter-relationships among the top 42 features obtained from RFE. It is observed that the TSP features of frontal channels are strongly correlated ($\geq$ 0.95) with the *conventional* features of the same channel. Since frontal channels represent cognition, cognitive fatigue shows up here, which is being picked up by TSP and *conventional* features. This demonstrates the usefulness of TSP features for the task of cognitive state detection with EEG data. Eliminating as many as 10 such correlated *conventional* features and 1 TSP feature, an accuracy of $93.1\% \pm 0.07\%$ was obtained using the feature set of reduced size 31. This result is comparable with the $Joint_{FS}$ model.

## V. CONCLUSION

This paper presents the use of topological signal processing for studying the time evolution of EEG signals for cognitive fatigue detection. Literature suggests that energies of various low frequency EEG sub bands like alpha (4-8 Hz), theta (8-16 Hz) and beta (16-30 Hz) are important in indicating mental fatigue state. These low frequency variations are effectively captured in persistent homology through systematic non-linear transformations of TSP. The topological features based on 0-dimensional and 1-dimensional homology groups are used as features for assessing the cognitive state of the individuals. Experimental results show that the topological features perform at par with the *conventional* features with relatively less number of features. Moreover, the combination of topological and *conventional* features perform better than the state-of-the-art method. The use of feature selection technique aids in identifying the disposable features, thereby reducing redundancy and further increasing the accuracy. The initial results obtained with TSP are very promising. TSP is known to be robust to noise which makes it suitable for cognitive assessments. In future, the TSP features can be tested on other datasets and further exploited for studying other physiological signals for different applications.

## REFERENCES


[1] Marilyn Sue Bogner, *Human error in medicine*, CRC Press, 2018.

[2] Luna and et al, "A study of usaf air traffic controller shiftwork: sleep, fatigue, activity, and mood analyses.," *Aviation, space, and environmental medicine*, vol. 68, no. 1, pp. 18–23, 1997.

[3] Weeks et al, "Physiological and psychological fatigue in extreme conditions: the military example," *PM&R*, vol. 2, no. 5, 2010.

[4] Hallowell et al, "Worker fatigue: Managing concerns in rapid renewal highway construction projects," *Professional safety*, vol. 55, no. 12, pp. 18–26, 2010.

[5] Papakostas Akilesh and Fillia, "Cogbeacon: A multi-modal dataset and data-collection platform for modeling cognitive fatigue," *Technologies*, vol. 7, no. 2, pp. 46, 2019.

[6] Gavas and et al, "Estimation of cognitive load based on the pupil size dilation," in *2017 IEEE SMC conference*. IEEE, 2017, pp. 1499–1504.

[7] Bentall et al, "A brief mental fatigue questionnaire," *British journal of clinical psychology*, vol. 32, no. 3, pp. 375–377, 1993.

[8] Rahul et al, "Enhancing the usability of low-cost eye trackers for rehabilitation applications," *PloS one*, vol. 13, no. 6, 2018.

[9] Gavas et al, "Multivariate variational mode decomposition based approach for blink removal from eeg signal," in *2020 IEEE PerCom Workshops. Accepted for publication*. IEEE, 2020.

[10] Ghadyali, "Applications of topological data analysis and slidingwindow embeddings for learning on novel features of time-varying dynamical systems," *PhD thesis*, 2017.

[11] Elizabeth Munch, "A users guide to topological data analysis," *Journal of Learning Analytics*, vol. 4, no. 2, pp. 47–61, 2017.

[12] Altindis et al, "Use of topological data analysis in motor intention based brain-computer interfaces," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1695–1699.

[13] Gidea and Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018.

[14] Khasawneh, Munch, and Perea, "Chatter classification in turning using machine learning and topological data analysis," *IFAC-PapersOnLine*, vol. 51, no. 14, pp. 195–200, 2018.

[15] Chazal et al, "Persistence-based clustering in riemannian manifolds," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–38, 2013.

[16] Seversky et al, "On time-series topological data analysis: New data and opportunities," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 59–67.

[17] Floris Takens, "In dynamical systems of turbulence," *Lecture notes in mathematics*, vol. 898, pp. 366, 1981.

[18] Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[19] Michal Teplan et al., "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[20] Monchi et al, "Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging," *Journal of Neuroscience*, vol. 21, no. 19, pp. 7733–7741, 2001.

[21] MUSE EEG, "https://choosemuse.com/," 2020.

[22] EEGLAB, "https://sccn.ucsd.edu/eeglab/index.php," 2019.

[23] Christopher et al, "Ripser.py: A lean persistent homology library for python," *The Journal of Open Source Software*, vol. 3, no. 29, 2018.

[24] Chawla and et al, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[25] Guyonand et al, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.