Biomedical Signal Processing and Control 86 (2023) 105274

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Detecting mild traumatic brain injury for athletes using SSVEP classification: A case study

Quang Thien Hoang [a,c,d,*], Ken-Tye Yong [c,d], Xiaochen Liu [d], Dylan Mahony [a], Alissa Chaitarvornkit [a], Adrian Cohen [a,1], Tijl Grootswagers [b,1]

[a] Headsafe MFG Pty Ltd, Research & Development, Sydney 2010, Australia
[b] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney 2214, Australia
[c] School of Biomedical Engineering, The University of Sydney, Sydney 2006, Australia
[d] The Biophotonics and Mechano-Bioengineering Lab, The University of Sydney, Sydney 2006, Australia

ARTICLE INFO

Keywords:
Mild traumatic brain injury classification
Steady-state visual-evoked potential
Machine learning

ABSTRACT

Mild traumatic brain injury (mTBI) can have detrimental impacts on the well-being of individuals, especially athletes with millions of injury cases reported per year. Nevertheless, the current assessment and diagnostic tools for mTBI have limitations due to their subjectivity and the lack of accessibility. This study aimed to evaluate the potential of machine learning algorithms in combination with steady-state visual evoked potentials (SSVEP) to provide mTBI diagnoses. The participants of this study included 36 athletes diagnosed with mTBI, aged 17–54, and 400 matched healthy controls without mTBI. Altogether, we extracted 51 SSVEP-based features from the collected observations and transformed them via principal component analysis (PCA) for feature reduction. Several machine learning algorithms were trained and validated using the transformed features for further analysis and comparison. Linear Discriminant Analysis (LDA) was found to be the best-performing classifier with 62 % balanced accuracy and has the potential to improve further with additional data. Overall, the findings of this study indicate that machine learning models have the potentials to be utilized as a diagnostic tool for mTBI when used with SSVEP data.

1. Introduction

Traumatic brain injury (TBI), a widespread injury among the general population, has been reported to be over 50 million annually worldwide [1,2]. Within this broad category, Mild traumatic brain injury (mTBI), which is frequently observed in contact sports participants such as rugby and football players, constitutes a significant subset [3–5]. Previous research has estimated the annual incidence of mTBI in the USA to be in the range of 1.5 to 3.8 million [6,7], with more than 2 million cases reported in Europe annually [8]. In addition, the estimated hospital cost for sport-related mTBI in Victoria, Australia between 2002 and 2011, was reported to be approximately A$18 million. This commercial burden is considered to be underestimated as it fails to take into account undiagnosed and unreported mTBI, which were theorised to be higher than the reported cases [9]. With the rising trend of mTBI incidence with each passing year [10,11], the hospital cost for mTBI is only expected to grow. The aetiology of mTBI is largely attributed to mechanical impacts to the head or body which result in differential movement of the brain within the skull and consequent damage to brain tissue [12,13]. Although the impact of mTBI may initially present as minimal, without adequate medical intervention and management, it may result in chronic neurodegenerative conditions including cognitive, sensory and psychological dysfunctions [14–16]. Furthermore, chronic traumatic encephalopathy (CTE), a fatal brain condition, was found to associate with exposure to repeated head trauma [17]. This finding further signifies the importance of detecting mTBI in athletes in order to provide adequate medical care and reduce their exposure to repetitive head trauma. Despite the debilitating consequences, mTBI often remains undiagnosed and untreated [18,19]. It is challenging to isolate the reasons behind the under-diagnosis of mTBI as they vary from patients' subjectivity such as symptoms underreporting to the lack of sufficient biomarkers and diagnostic tools.

Underreporting of mTBI in athletic populations can be attributed, in part, to the absence of interdisciplinary consensus and education

* Corresponding author at: Headsafe MFG Pty Ltd, Research & Development, Sydney 2010, Australia.
E-mail address: quang.hoang@sydney.edu.au (Q.T. Hoang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bspc.2023.105274
Received 3 April 2023; Received in revised form 29 June 2023; Accepted 16 July 2023

regarding the injury [18,20]. It has been reported that the lack of awareness and understanding among athletes themselves can lead to a failure to seek appropriate medical attention [21]. Additionally, athletes may be inclined to minimize their symptoms during mTBI assessments to avoid removal from play [22,23] or due to external pressures [24], leading to subjectivity and bias in the diagnostic process. Moreover, the symptoms associated with mTBI are often reported to be subtle and rapidly resolving, with adverse effects observed within 24 to 72 h post-injury [25–27]. Therefore, these factors would greatly affect the reliability of current standard assessment tools such as The Sports Concussion Assessment Tool (SCAT), which diagnoses sports-related mTBI on the sideline through a multimodal assessment of common symptoms [25,28]. Consequentially, SCAT is regarded as insufficient due to inconsistent and potentially subjective assessments [18,29], highlighting the need for new approaches.

Currently, the diagnosis of mTBI is usually made through clinical evaluation by a physician based on the reported symptoms from the patients [29,30]. As this approach is inherently subjective, supplementary diagnostic aids have been created to provide more objective measurements. The most commonly used measurements include neuroimaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) [31,32] and neuronal activity recordings such as electroencephalography (EEG) [33–35]. CT and MRI, as the current standard modalities for neurostructural trauma assessment, offer an excellent spatial resolution for diagnostic purposes [33,36]. However, these techniques are not ideal for the diagnosis of mTBI, as the condition is generally regarded as functional, instead of structural [37]. Furthermore, they are not feasible for rapid field-based assessments due to the need for specialized equipment, facilities, and trained personnel [31,33], limiting their use primarily to the hospital or clinical settings. Therefore, in recent years, the necessity for non-imaging biomarkers with sufficient sensitivity and specificity for mTBI detection has been emphasized [13]. The need for these biomarkers is more apparent in sports field assessments as medical facilities are limited and physicians are infrequently present [29,33,38]. It is imperative that more objective, reliable, and readily available diagnostic methods be developed to address the growing burden of mTBI in athletic populations and mitigate the limitations of current approaches.

In contrast to the aforementioned diagnostic aids, electroencephalography (EEG) has been put forward as a promising alternative, offering several advantages over CT and MRI including its superior temporal resolution, portability, low cost, and minimal training requirements [36,39]. While CT and MRI scans provide a static snapshot of the brain, EEG provides a direct and ongoing measurement of functional brain activity [33,39], making it a valuable diagnostic aid. For the use of EEG-based biomarkers, most recent studies employed resting state EEG (rs-EEG) for mTBI classification [39–41]. One methodological variable in these studies is the time of recording with some studies requiring 4 min while others are as long as 10 min. The time required, albeit only minutes, could be argued to be lengthy, limiting its appeal to patients and in particular athletes who are not required nor compensated for their time to conduct an assessment by their governing body. Further to this limitation is the concern that detectable changes in EEG signals may not occur during rest, but rather require excitation or stress to be amplified to the point of detection. Taking both limitations in combination further highlights the advantage of using event-related potentials instead of standard EEG, which was reported to be sensitive to cognitive dysfunction [42].

In particular, Steady-State Visual Evoked Potentials (SSVEP), which quantify neural population responses to a repeating visual stimulus [43], have demonstrated potential as an alternative EEG-based biomarker for mTBI with less than a minute required for data collection [29]. SSVEP signal was selected for this study over standard EEG signal as SSVEP allows the measurement of brain response to visual stimuli, which was reported to show abnormalities in mTBI patients [44,45]. Moreover, SSVEP has several advantages over traditional VEP such as simpler equipment requirements, relative noise artefact resilience and enhanced ability to withstand changes in contact impedance [46,47]. Thus, these advantages make SSVEP a more favourable biomarker in a non-clinical environment. In light of this, the present preliminary study seeks to expand the understanding of SSVEP as a novel potential diagnostic aid for mTBI by exploring its viability as a biomarker for machine learning classification using a data-driven approach.

As the current understanding of the underlying mechanisms is limited [48,49], mTBI poses a complicated challenge for the development of effective classification solutions. However, machine learning algorithms have the ability to analyse complex patterns in high-dimensional data [50] and thus, offering an attractive solution for addressing the diagnostic challenges associated with mTBI. Positively, there is an increase in opportunities for medical practitioners to explore neurophysiological data with recent machine learning advancements as clinical tools [51,52]. Moreover, there exist various EEG signal processing techniques, including Fast Fourier Transformation (FFT), and Wavelet, that broaden the scope of EEG feature engineering in both time and frequency domains [53,54]. Some studies took advantage of these techniques and employed traditional machine learning algorithms, such as Support Vector Machines, K-nearest neighbours, and Linear Discriminant Analysis [40,55–57], to develop classification models. Whereas other studies have capitalized on the strengths of deep learning (DL) algorithms, such as Neural Networks on EEG signals to develop prediction models without the need for feature engineering [39,58]. Though there are numerous studies on typical EEG signals, the literature remains sparse on SSVEP application with machine learning for mTBI classification.

In this study, our objective was to evaluate the effectiveness of machine learning in classifying mild traumatic brain injury (mTBI) through the analysis of steady-state visual evoked potential (SSVEP) signals. These signals were obtained using a specialized portable electroencephalogram (EEG) device. We used both raw and filtered EEG data to preserve data integrity and extract SSVEP-based features for training the machine learning algorithms. Our primary aim was to determine the feasibility of combining machine learning with SSVEP signals to create an objective diagnostic tool for mTBI. To this end, we conducted a systematic comparison of several widely used classifiers. Our findings revealed that multiple classifiers demonstrated performance comparable to the current mTBI assessment standard, the Sports Concussion Assessment Tool (SCAT). This study represents a significant contribution toward the development of an innovative, portable, and objective system for mTBI detection, which has the potential to revolutionize diagnostics in this field.

2. Methods

2.1. Participants

Initially, we recruited a total of 468 participants from sports settings where individuals are frequently subjected to head injury. After filtering for poor data, we ended up with 436 participants for the study as outlined in Table 1. The participants were divided into two groups: a non-mTBI control group (n = 400) and a clinically diagnosed mTBI group (n = 36). The age of the participants ranged from 17 to 54 years, and the cohort included males (n = 299, mean age = 23 years) and females (n = 137, mean age = 26 years). SSVEP signals were recorded from athletes

Table 1
Breakdown of study participants by sports cohorts and mTBI status.

Cohort	Non-mTBI (control) participants	Injured (mTBI) participants
Rugby	332	35
Combat sport	62	1
Other sport	6	0

diagnosed with mTBI within four days of diagnosis (mean [SD], 2 [1.29] days). All clinical data were collected under protocols approved by a Human Research Ethics Committee (Bellberry Limited, HREC 180SSVEP). Participants were given detailed information regarding the experiment and provided informed consent to participate in the study.

The study's inclusion criteria consisted of athletic individuals aged 14 years and above who were proficient in English. Exclusion criteria were created around limitations to electrode placement or complications that could arise from the flickering visual stimulus. These included any open head wound, epilepsy, seizure history, sensitivity to flashing lights, legal blindness, or any structural brain injury and/or condition. A control participant was defined as someone who had not experienced any injury that raised the possibility of mTBI in the month before assessment and was not receiving treatment for a previous mTBI. Furthermore, observations were classified as injured (mTBI) if the subject had been diagnosed with mTBI by a qualified medical practitioner. The period between the date of injury and SSVEP recording was documented to select the data for this study.

2.2. Data acquisition

The prototype SSVEP acquisition device (Fig. 1A) was employed to collect SSVEP data. The prototype device is composed of a rear section that houses the unipolar electrodes for recording the signals, and a front visor section that delivers a 15 Hertz (Hz) flickering visual stimulus using white light-emitting diodes (LEDs). The system collects continuous SSVEP signals through three occipital electrodes (O1, O2, and Oz) and employs two parietal electrodes (P1 and P2) for the reference and ground/bias, respectively, as illustrated in Fig. 1B. Common noise measured across the electrode channels was utilized for common-mode rejection. The SSVEP signals were recorded for a fixed period of 30 s at a sampling rate of 250 Hz [29].

All measurements were conducted in non-laboratory settings including office spaces, sporting grounds change rooms or medical clinics. To minimize potential distractions and noise levels during data collection, the research team took measures to create a calm and quiet environment. There was little to no noise during data collection and the lighting was controlled to be soft with no abrupt changes in brightness so no visual artefacts would present in the signals. Participants were instructed to remain seated, silent, and relaxed throughout the testing period, with their gaze fixed on the visual stimulus. Once the participants were seated, the research team then removed any potential source of artefacts from the participants' hair by wiping down the hair with an alcohol wipe and applying the saline solution to the electrodes for better contact. Before recording SSVEP signals, the rear section of the prototype headset was positioned above the inion to ensure proper placement of the unipolar electrodes on the occipital lobe, as directed by the manufacturer. Subsequently, the front section of the headset was placed snugly on the face over the nasal bridge, to avoid any external light entering the visor. The straps on both sides of the headset were tightened to ensure an appropriate fit. The contact impedances of the electrodes were monitored through a mobile application to ensure that SSVEP readings were obtained with an impedance of less than 25 kOhms. To further ensure the quality of collected data, the electrodes were replaced after each use.

2.3. Data pre-processing and representation

The SSVEP signals were processed using MATLAB (version R2021a). Prior to signal processing, any collected data with higher impedance than the criteria were rejected from the study to minimize potential noises being introduced to the dataset. The accepted signals were filtered with a Butterworth bandpass filter from 5 to 35 Hz to minimize the effects of low-frequency noise, main power artefacts and DC voltage offset [59]. Following that, Fast Fourier Transformation (FFT) was performed to convert filtered time-domain SSVEP data to the frequency domain which allows the generation of an FFT plot for further screening as shown in Fig. 2. The screening procedure involved examining for common EEG artefacts such as muscle activity, electrode and harmonic artefacts. Our acceptance criteria require the samples to have a minimum 15 Hz response of 1 μV (based on pilot testing) and no mentioned artefacts to be included in the machine learning analysis. Any failed observation based on the criteria was removed from the study.

Fig. 2 shows examples of clean observation accepted for analysis and poor observation which were excluded from the study due to these data quality criteria. A total of 51 unique features were extracted from the datasets; with 30 features based on the filtered frequency series, including the band power of conventional EEG frequency bands; and the remaining 21 features were obtained from filtered time series, including first-order statistics. This resulted in a 436 by 51 matrix of observations by features which were used as input to the classification analysis.

2.4. Classification setup and analysis

The feature selection method utilised principal component analysis (PCA) to produce a new set of orthogonal principal components by linearly transforming the original feature set. This approach had a low computational complexity, which is a clear advantage over other conventional techniques [60]. The principal components were selected based on their ability to retain the most informative and discriminatory

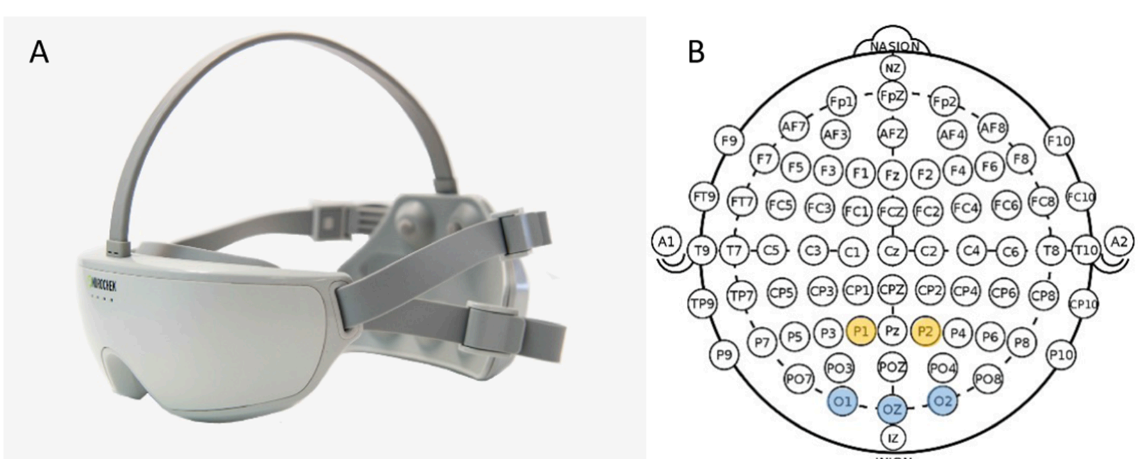

Fig. 1. Electrode overview of prototype SSVEP acquisition device (A). Electrodes are positioned according to the international standard 10–20 EEG system (B). Reference (P_1) and ground/bias (P_2) electrodes are yellow. Signal electrodes are blue (O_1, O_2, and O_z).

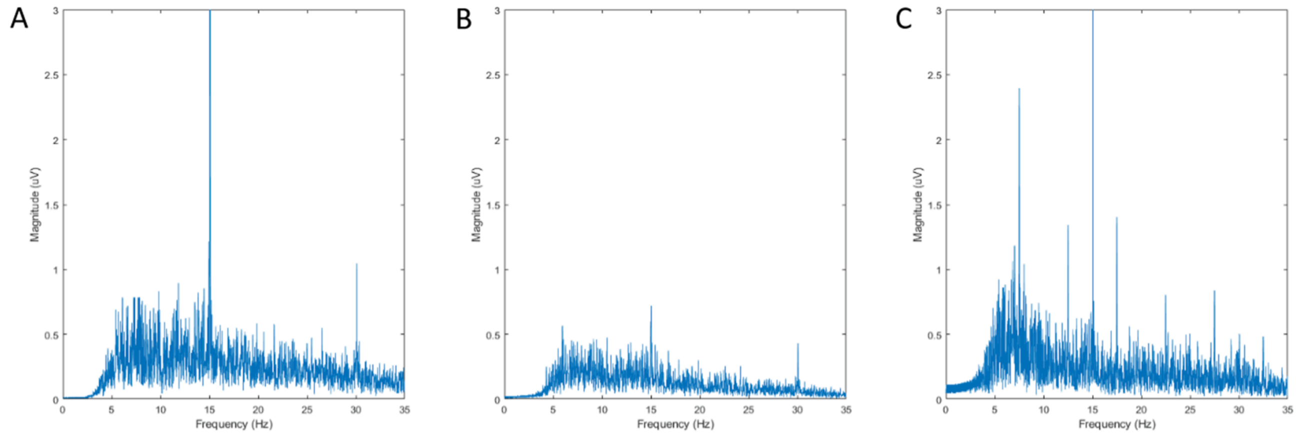

Fig. 2. Quality assurance standards for SSVEP readings screened for inclusion in the training dataset. (A) Accepted sample with strong 15 Hz response and minimal frequency noise, (B) Rejected sample due to low 15 Hz response, (C) Rejected sample due to harmonics present throughout the frequency range.

information while minimizing the presence of noise and reducing the risk of overfitting. Given the data size available for this study, deep learning algorithms such as Neural Networks are excluded as it is suggested to have a sample size of at least 50 times the number of weights [61]. As such, with only over 400 samples, the degree of freedom in designing the deep learning models would be greatly limited and there would be a risk of overfitting. Four classification algorithms, namely Naïve Bayes (NB), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Support Vector Machine (SVM), were employed in the study as they were utilized in past studies for mTBI classification and were reported to show positive outcomes [41,62,63]. The performance of the classifiers was evaluated using stratified k-fold cross-validation [50,56], which partitioned the dataset into k non-overlapping folds where the class proportion is maintained across each fold. For each algorithm, the model was trained using all but one fold and tested using the left-out fold. This pipeline was repeated 10,000 times to obtain a more accurate and reliable estimate of classification performance on the limited dataset [64], where different combinations of healthy and mTBI observations were used for training and testing in each repetition. The number of folds used in this process was equal to the number of mTBI participants, which was 36. The performance metrics to evaluate the classifiers were sensitivity, specificity and f1 score. To consider both sensitivity and specificity as a single measure of ease of comparison, classifier performance was also measured as balanced accuracy, computed as:

$$Balanced\ \ Accuracy(\%) = \frac{sensitivity + specificity}{2}$$

The optimal number of principal components necessary for each classifier to achieve peak performance would vary due to the unique underlying mechanics of each algorithm. SVM, for instance, is predicated on geometric properties, while the other classifiers rely on statistical properties and disparate assumptions. Thus, the classification pipeline was performed on all algorithms with varying numbers of principal components in a systematic manner. Beginning with the first principal component, which accounted for the highest variance present in all original features, the pipeline was run to completion once all repetitions per pipeline were achieved. Then, additional principal components, each explaining the next highest variance, were added in subsequently until all principal components were employed. The optimal number of principal components utilized by each algorithm was thus determined by selecting the configuration that produced the best-performing classification results.

This study postulated that incorporating more informative data into the training dataset would allow a more accurate overall classifier performance estimation. Specifically, it is proposed that increasing the number of data points, represented by mTBI samples, could potentially enable the classifiers to attain a higher performance ceiling. In order to obtain the estimated trajectory, the performance of the classifiers was evaluated while progressively increasing the number of mTBI observations in the dataset. The classification started with a non-stratified independent random sampling of three randomly selected mTBI observations. Subsequently, the training data was augmented with an additional mTBI observation, and the classification process was iterated until all mTBI observations were utilized. For comparative purposes, classifier performance was assessed using the optimal principal components (as determined previously), as well as four commonly employed principal component cut-offs that explain 90 %, 95 %, 99 %, and 99.9 % of the variance. Fig. 3 summarizes the experimental process flow of this study for visual representation.

2.5. Statistical assessment

The study conducted a permutation test on all classifiers to assess the reliability of classification results [65,66]. First, the labels for each group were shuffled 10,000 times, with the mTBI patient and control labels randomly assigned to the 436 observations. The same classification procedure was then applied to the shuffled data, across all classifiers, which allowed further statistical computation for the test. The p-values were calculated by comparing the observed test statistic to the randomized distribution of test statistics. To control the familywise error rate in multiple comparisons, the maximum statistic correction method was utilized to obtain the corrected p-values [67]. The significance of the classification result was determined using a confidence threshold of $\alpha = 0.05$ [68,69]. Consequently, we established which numbers of principal components were significant for each classifier. The number of components among the established with the highest balanced accuracy was chosen for each classifier for further comparison and analysis.

3. Results

The optimal number of principal components required for best-performing classification was first determined. Fig. 4A illustrated the variance explained as a function of principal components. The evaluation of classifier performance as a function of the principal component size used was executed for four different classifiers, as depicted in Fig. 4B. The peak of each data series in this Fig. 4B identified the optimal number of principal components for the respective classifiers. From Fig. 4A and 4B, the LDA classifier achieved its best performance with 24 principal components, which explained 97 % variance, whereas LR reached its optimal performance with 23 components (96.5 % variance). The SVM classifier, on the other hand, performed best with 19

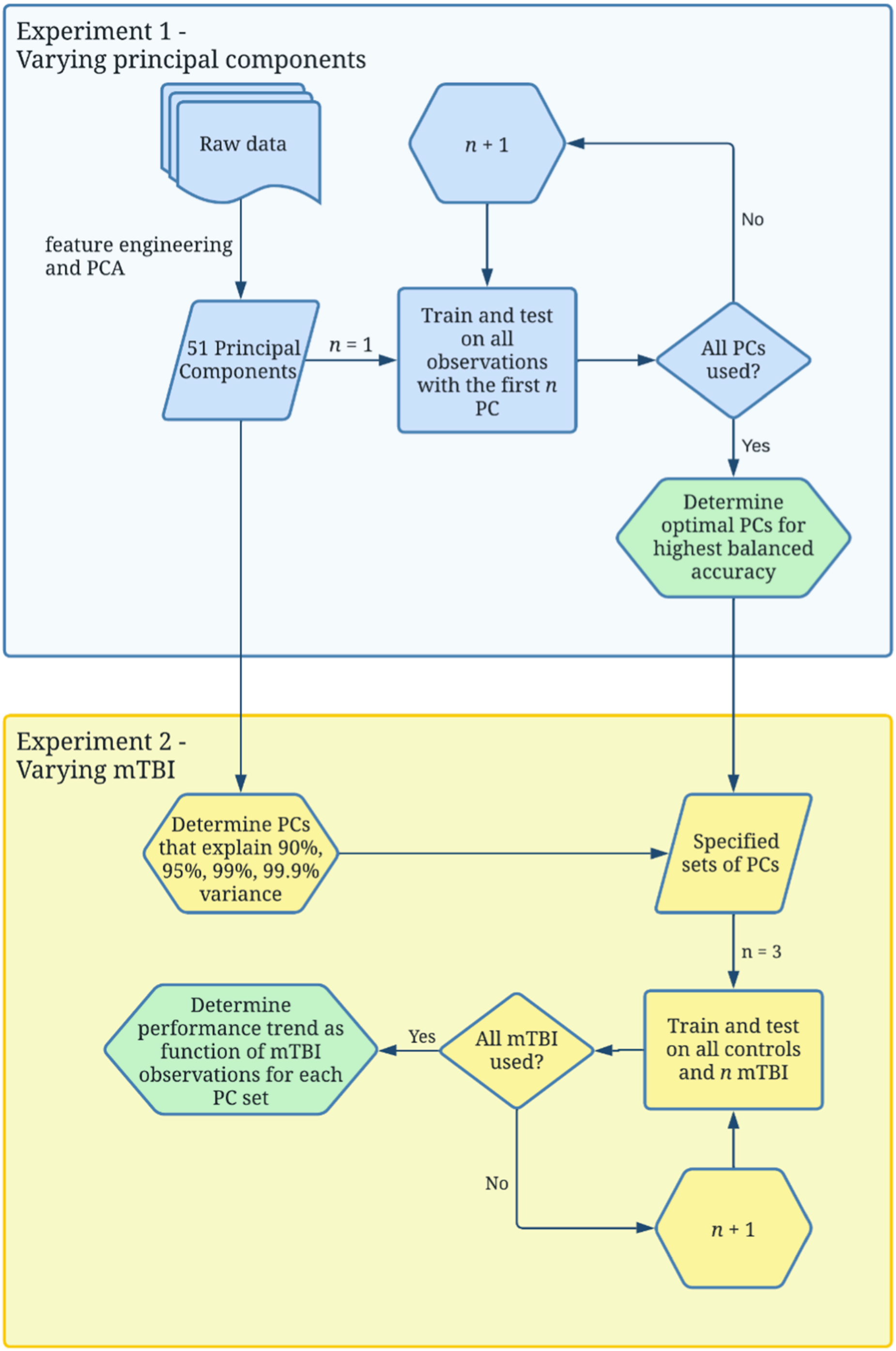

Fig. 3. Experimental Methodology Process Flowchart. Experiment 1 began by testing various numbers of principal components for training to identify the optimal combination for each algorithm. Once the ideal principal components combination was determined, Experiment 2 proceeded to train the algorithms using a variable mTBI database. Finally, the performance of the algorithms, when trained with the established optimal principal components combination, was compared to their performance when trained with commonly used variance percentages as inputs, rather than principal components.

components (93.7 % variance), while NB achieved its highest performance with only five components (62.4 % variance). Most classifiers in this study used over 90 % variance to obtain their best performance, except for NB. Additionally, the classifiers' trends varied considerably, indicating the need for further reliability testing. Among the classifiers, the SVM classifier achieved the best performance (~64 % balanced accuracy) using the aforementioned number of principal components, followed by the LDA and LR classifiers (~62 % balanced accuracy), while the NB classifier performed the worst (~56 % balanced accuracy). The permutation test was used to identify statistically significant principal components for each classifier, using the p-values threshold of less than 0.05 as depicted in Fig. 4C. The hollow circles in this figure represented the statistically significant principal components for each classifier. LDA had the most significant components, followed by LR and SVM respectively, and NB had no components that achieved a p-value less than 0.05. After maximum statistic correction, LDA, LR, and SVM classifiers had, respectively, one, six, and two statistically corrected significant principal components, as shown by the dots within the hollow circles. The principal components that yield the highest balanced accuracy for most classifiers were statistically significant.

All relevant mean performance metrics of the classifiers were listed in Table 2. Except for NB, which was skewed towards sensitivity, the remaining classifiers had a better balance between sensitivity and specificity. Particularly, SVM and LDA had better specificity than sensitivity whereas it is vice versa for LR. NB achieved the highest sensitivity (~65 %) with the trade-off being its specificity (~47 %).

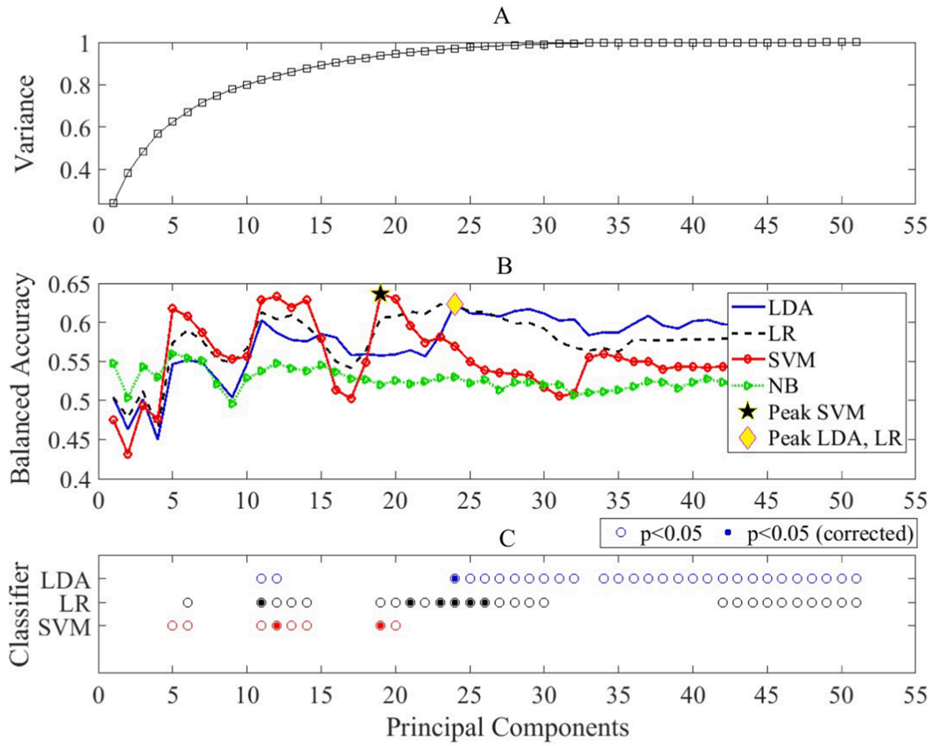

Fig. 4. (A) Cumulative proportion of variance explained in the dataset as a function of the number of principal components. (B) Classifier performance as a function of the number of principal components. (C) Statistical significance of the classification results, where the hollow circles represent classifier performance significantly higher than chance, and the dots represent significant results after using correcting for the multiple tests.

Table 2
Performance metrics of all classifiers, Support Vector Machine (SVM), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and Logistic Regression (LR), including mean balanced accuracy, mean sensitivity, mean specificity and mean f1 score.

Classifier	Balanced Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 Score
SVM	63.64	62.78	64.50	0.6445
NB	56.00	65.28	46.63	0.4839
LDA	62.28	60.00	64.55	0.6422
LR	62.43	63.61	61.25	0.6147

Whereas both SVM and LDA achieved the highest specificity (~64.5 %) without any significant trade-off. In addition, both SVM and LDA had similar f1 scores (~0.64), followed by LR (~0.61) and then NB (~0.48).

With the optimal principal components determined for each classifier, the corresponding balanced accuracy averaged across 10,000 repetitions, was plotted as a function of mTBI data size as shown in Fig. 5. Overall, the findings revealed that performance improved across all classifiers as more mTBI observations were added to the training data, although the levels of improvement varied among the classifiers. Notably, only NB's performance had a linearly proportional relationship with training data, where each addition of mTBI observation led to an approximately 0.2 % increase in accuracy. The LR classifier's performance, on the other hand, exhibited a sharp improvement after a small number of mTBI observations were added, which was followed by a decline in the accuracy rate from 0.6 % per observation down to 0.2 % per observation as the mTBI data size continued to grow. In contrast, the LDA classifier showed an increasing rate of performance improvement as the mTBI observations used in training approached the exhaustion of the available dataset. This was remarkably different from the trend observed in the LR classifier, which suggests that the LDA classifier might outperform the LR classifier as more observations are added. Additionally, the SVM classifier shared a similar trajectory to the LDA classifier, with a slightly better final performance. Notably, there was a spike in SVM's performance towards the end of the available mTBI dataset, increasing by approximately 1 % per mTBI observation added, which was distinct from the trends observed in the other classifiers.

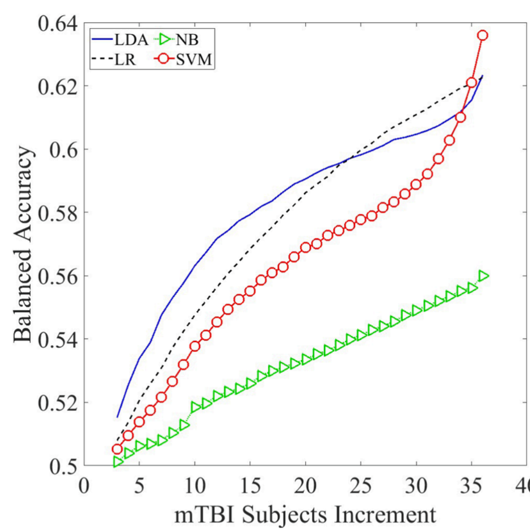

Fig. 5. A comparison in balanced accuracy performance as a function of mild traumatic brain injury (mTBI) observations for classifiers, Support Vector Machine (SVM), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and Logistic Regression (LR), using their optimal amount of variance as estimated via principal component analysis (PCA).

To enable a more direct comparison between classifiers, Fig. 6 presents their performance relative to mTBI data size using principal components that account for the same variance. LDA (Fig. 6A) showed a distinct performance trend, with only 95 % variance experiencing a

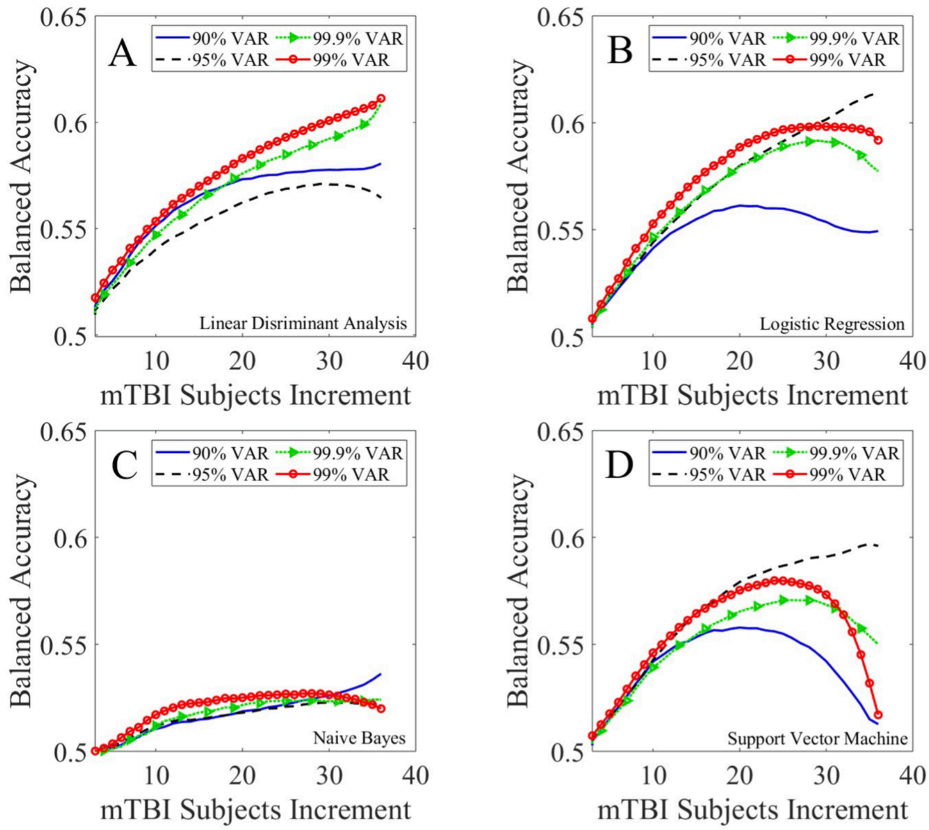

Fig. 6. A comparison of common variance settings of 90 %, 95 %, 99 % and 99.9 % for classifier performance as a trajectory of mild traumatic brain injury (mTBI) observations increment (A) Linear Discriminant Analysis, (B) Logistic Regression, (C) Naïve Bayes and (D) Support Vector Machine. The trend of each plot would be used to determine how the statistical significance of variance input would affect the potential of the classifier performance.

decline after reaching its peak, while the other variance thresholds exhibited an increasing trend with a steeper slope of performance increase. In contrast, LR and SVM (Fig. 6B and 6D) exhibited similar performance behaviour for most variance thresholds, where the balanced accuracy reached a certain peak before declining as the dataset size increases. For both classifiers, the 95 % variance plot was an exception, which had a persistent increasing trend. On the other hand, NB (Fig. 6C) exhibited relatively similar performance regardless of variance threshold.

With the statistical significance of the principal components taken into account (Fig. 4C), it was observed that the mTBI increment trajectory of the LR and SVM classifiers for 95 % variance was the only significant trajectory with an increasing trend. On the other hand, LDA had a wider range of significant principal components, including its variance settings that returned an increasing trend in the mTBI increment trajectory. NB was not considered as it had no significant principal component.

4. Discussion

We investigated the potential of machine learning models developed using SSVEP-based features for the classification of mTBI, considering both the current performance and the possibility for further development of the models. The capability and reliability of the machine learning models were evaluated by validating them with different principal components extracted from the available dataset. The results demonstrated reliable classifiers with comparable performance to the existing standard assessment tool. Additionally, the study found that the performance ceiling for the classifiers had not been reached, as evidenced by the increasing trend in balanced accuracy with increasing mTBI observation size. The study included participants diagnosed with mTBI within three days (with two exceptions) and controls, indicating the potential of the classifiers as a diagnostic aid. The protocol for data collection required 30 s of SSVEP recordings to be collected, using a wireless device controlled by a tablet or phone, making it a convenient tool for use on the sideline, locker room, or in limited spaces with minimal preparation required.

The initial experiment in this study demonstrates the potential of machine learning models to achieve performance comparable to the standard assessment tool, SCAT. By systematically increasing the principal components used for model training and testing, the study identified the optimal principal components for each algorithm. Notably, most classifiers examined, namely Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Support Vector Machines (SVM), achieved a balanced accuracy of 62 % or higher (over 60 % sensitivity, over 61 % specificity, over 0.61 f1 score). In comparison, a recent study that utilized SCAT to diagnose 36 participants, consisting of 19 individuals with mTBI and 17 controls, achieved a balanced accuracy of 60 % (100 % sensitivity, 20 % specificity) [70]. In this study, the mTBI assessment method used to train and classify against the machine learning models, EEG, was demonstrated to be quick (30 s of data). This is important when considering its ability to provide similar diagnostic aiding performance to SCAT, while overcoming SCAT's major disadvantages, including testing time and objectivity. There is a consensus that SCAT required a minimum of 10 min to complete [28]. This led to the test being reported as lengthy and often interrupted [70]. Furthermore, SCAT was designed to take advantage of mTBI symptoms via a series of neurological and functional brain tests [25]. However, this design is detrimental to the result reliability as the symptom endorsement was reported to be subjective [71].

The second experiment of this study suggests that the classifiers examined have the potential for further improvement. Overall, with optimal principal components, all classifiers demonstrated an increasing trend in balanced accuracy as the mTBI data size increased, with LDA and SVM displaying the highest increase rates. Although, the increasing trend was not observed in other variance setups, with only LDA demonstrating a consistent trend in different trials. Notably, except for NB, the remaining classifiers' optimal components are all statistically significant according to the adjusted p-values. Based on the availability of significant components, it could be argued that LDA and LR are more reliable than SVM, although both classifiers were 2 % less accurate than SVM. When considering both accuracy and reliability, LDA was deemed to be the leading classifier. However, this finding is limited to the applied dataset and may be subject to change as more observations result in different variances. Therefore, while the study provides evidence to support LDA as the most suitable classifier among those investigated, it does not imply that other classifiers should be disregarded for this application.

A limitation of our study was its relatively small sample size, which restricted the findings' statistical power and the classifiers' optimization due to the risk of overfitting [72]. It was reported that up to 560 annotated samples could be needed to reach the performance target with acceptable error [73]. This limitation was attributed to the logistical challenges encountered during data collection, particularly concerning mTBI data. The process of identifying and assessing mTBI is complex and time-consuming, requiring incident reports by team coaches, diagnosis by a medical practitioner, and SSVEP data collection on-site by trained personnel, all within a limited timeframe. Another limitation was subject compliance in completing the study and its associated follow-up assessments. To manage these variables, the study design focused on controlled research sites of high concussion occurrence, primarily contact sporting clubs in Australia. However, this approach may affect the generalizability of the findings, which is a common issue for an application intended for the general population. The dataset used in this study was collected from a single city and mostly from one sport discipline, rugby union. Although contacts in this sport can be compared to those in other football codes such as rugby league, Australian Rules Football, and American Football. On the other hand, traditional football ("soccer") does not have the same quantum of direct person-to-person physical contact but has the additional element of "heading" the ball, which should be considered applicable to some extent as both elements were reported to correlate with traumatic brain injury in the sport [74,75]. By expanding and diversifying the data collection sites and other sporting disciplines, the machine learning approach could be broadened, resulting in a more robust mTBI classification model and increased reliability.

Ongoing research efforts are dedicated to improving the performance of the investigated classifiers by expanding the dataset size, exploring additional SSVEP-based features, and optimising the models. Refinements in the prototype SSVEP device since the collection of analysed data are indicative that the model can be improved in future studies. Consequently, the future direction for the research would involve investigating more SSVEP features such as connectivity and complexity features. Additionally, the machine learning algorithms researched in this study would be validated on a larger dataset to ensure the generalizability of the study's findings.

5. Conclusion

This study demonstrates that machine learning classifiers, utilizing SSVEP-derived data, show promise in distinguishing between mTBI patients and healthy individuals. Overall, the classifiers employed in this research exhibited a gradual increase in classification performance as more mTBI observations were integrated into the model. This finding suggests that the classifiers have not yet reached their performance peak and have further room for improvement. Upon closer examination, Linear Discriminant Analysis (LDA) proved to be the most reliable and overall best-performing model. Additionally, LDA achieved a 62 % balanced accuracy, which is comparable to SCAT, the current standard assessment tool for sports-related mTBI. These findings carry significant implications for the clinical diagnosis of mTBI, a form of brain injury that often goes undetected using traditional diagnostic methods. The implementation of an SSVEP-based classification approach could provide clinicians with an additional tool to aid in the identification of mTBI, potentially enhancing patient outcomes through early intervention and treatment.

Author contributions

Q.T.H and D.M designed study; Q.T.H, D.M, A.C and A.J.C collected data; D.M, A.J.C, K.Y and T.G conducted review and editing; Q.T.H. X.L and A.C wrote the paper; A.J.C provided funding acquisition; Q.T.H conducted study; Q.T.H and T.G conducted analysis.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work contains a commercial conflict of interest by employees (Q.T. H, D.M, A.C, A.J.C) of Headsafe. A.J.C has filed a patent for the technology in this general area of mTBI diagnosis aids. T.G, K.Y, X.L declare no conflict of interest.

Data availability

I can share the processed data in form of features. Raw data is not available for sharing.

Study materials are available upon request, please contact quang.hoang@sydney.edu.au.

Acknowledgements

We would like to acknowledge and thank the participants, researchers, and support staff, including those of the various sporting associations and human research ethics committee, for their contributions to this work.

This research has been funded by Headsafe MFG through the Industrial PhD Sponsorship (No. 216229) and Start up Funding from University of Sydney (No. H3165).

References

[1] M.C. Dewan, et al., Estimating the global incidence of traumatic brain injury, J. Neurosurg. 130 (4) (Apr. 2019) 1080–1097, https://doi.org/10.3171/2017.10.JNS17352.

[2] A.I.R. Maas, et al., Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research, Lancet Neurol. 16 (12) (Dec. 2017) 987–1048, https://doi.org/10.1016/S1474-4422(17)30371-X.

[3] A.J. Gardner, G.L. Iverson, W.H. Williams, S. Baker, P. Stanwell, A Systematic Review and Meta-Analysis of Concussion in Rugby Union, Sports Med. 44 (12) (Dec. 2014) 1717–1731, https://doi.org/10.1007/s40279-014-0233-3.

[4] K.G. Kilcoyne, et al., Reported Concussion Rates for Three Division I Football Programs: An Evaluation of the New NCAA Concussion Policy, Sports Health Multidiscip. Approach 6 (5) (Sep. 2014) 402–405, https://doi.org/10.1177/1941738113491545.

[5] C.W. Fuller, A. Taylor, M. Raftery, Epidemiology of concussion in men's elite Rugby-7s (Sevens World Series) and Rugby-15s (Rugby World Cup, Junior World Championship and Rugby Trophy, Pacific Nations Cup and English Premiership), Br. J. Sports Med. 49 (7) (Apr. 2015) 478–483, https://doi.org/10.1136/bjsports-2013-093381.

[6] B. Ferry and A. DeCastro, "Concussion," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Jan. 18, 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK537017/.

[7] V.G. Khurana, A.H. Kaye, An overview of concussion in sport, J. Clin. Neurosci. 19 (1) (Jan. 2012) 1–11, https://doi.org/10.1016/j.jocn.2011.08.002.

[8] T. Skandsen, et al., The epidemiology of mild traumatic brain injury: the Trondheim MTBI follow-up study, Scand. J. Trauma Resusc. Emerg. Med. 26 (1) (Dec. 2018) 34, https://doi.org/10.1186/s13049-018-0495-0.

[9] C.F. Finch, A.J. Clapperton, P. McCrory, Increasing incidence of hospitalisation for sport-related concussion in Victoria, Australia, Med. J. Aust. 198 (8) (May 2013) 427–430, https://doi.org/10.5694/mja12.11217.

[10] L. Langer, C. Levy, M. Bayley, Increasing Incidence of Concussion: True Epidemic or Better Recognition? J. Head Trauma Rehabil. 35 (1) (Jan. 2020) E60–E66, https://doi.org/10.1097/HTR.0000000000000503.

[11] P. Veliz, S.E. McCabe, J.T. Eckner, J.E. Schulenberg, Trends in the Prevalence of Concussion Reported by US Adolescents, 2016–2020, JAMA 325 (17) (May 2021) 1789, https://doi.org/10.1001/jama.2021.1538.

[12] C. Cao, S. Slobounov, Application of a novel measure of EEG non-stationarity as 'Shannon- entropy of the peak frequency shifting' for detecting residual abnormalities in concussed individuals, Clin. Neurophysiol. 122 (7) (Jul. 2011) 1314–1321, https://doi.org/10.1016/j.clinph.2010.12.042.

[13] H.S. Levin, R.R. Diaz-Arrastia, Diagnosis, prognosis, and clinical management of mild traumatic brain injury, Lancet Neurol. 14 (5) (May 2015) 506–517, https://doi.org/10.1016/S1474-4422(15)00002-2.

[14] S.T. DeKosky, M.D. Ikonomovic, S. Gandy, Traumatic Brain Injury — Football, Warfare, and Long-Term Effects, N. Engl. J. Med. 363 (14) (Sep. 2010) 1293–1296, https://doi.org/10.1056/NEJMp1007051.

[15] S.M. Rice, A.G. Parker, S. Rosenbaum, A. Bailey, D. Mawren, R. Purcell, Sport-Related Concussion and Mental Health Outcomes in Elite Athletes: A Systematic Review, Sports Med. 48 (2) (Feb. 2018) 447–465, https://doi.org/10.1007/s40279-017-0810-3.

[16] G. Manley, et al., A systematic review of potential long-term effects of sport-related concussion, Br. J. Sports Med. 51 (12) (Jun. 2017) 969–977, https://doi.org/10.1136/bjsports-2017-097791.

[17] T.D. Stein, V.E. Alvarez, A.C. McKee, Concussion in Chronic Traumatic Encephalopathy, Curr. Pain Headache Rep. 19 (10) (Oct. 2015) 47, https://doi.org/10.1007/s11916-015-0522-z.

[18] C. Prince, M. Bruhns, Evaluation and Treatment of Mild Traumatic Brain Injury: The Role of Neuropsychology, Brain Sci. 7 (12) (Aug. 2017) 105, https://doi.org/10.3390/brainsci7080105.

[19] W.P. Meehan, R.C. Mannix, M.J. O'Brien, M.W. Collins, The Prevalence of Undiagnosed Concussions in Athletes, Clin. J. Sport Med. 23 (5) (Sep. 2013) 339–342, https://doi.org/10.1097/JSM.0b013e318291d3b3.

[20] T. Rice, R. Curtis, Parental knowledge of concussion: Evaluation of the CDC's 'Heads up to parents' educational initiative, J. Safety Res. 69 (Jun. 2019) 85–93, https://doi.org/10.1016/j.jsr.2019.02.007.

[21] J.K. Register-Mihalik, K.M. Guskiewicz, T.C.V. McLeod, L.A. Linnan, F.O. Mueller, S.W. Marshall, Knowledge, Attitude, and Concussion-Reporting Behaviors Among High School Athletes: A Preliminary Study, J. Athl. Train. 48 (5) (Oct. 2013) 645–653, https://doi.org/10.4085/1062-6050-48.3.20.

[22] A. Ackery, C. Provvidenza, C.H. Tutor, Concussion in Hockey: Compliance with Return to Play Advice and Follow-up Status, Can. J. Neurol. Sci. J. Can. Sci. Neurol. 36 (2) (Mar. 2009) 207–212, https://doi.org/10.1017/S0317167100120281.

[23] P.S. Echlin, et al., A prospective study of physician-observed concussions during junior ice hockey: implications for incidence rates, Neurosurg. Focus 29 (5) (Nov. 2010) E4, https://doi.org/10.3171/2010.9.FOCUS10186.

[24] E. Kroshus, B. Garnett, M. Hawrilenko, C.M. Baugh, J.P. Calzo, Concussion under-reporting and pressure from coaches, teammates, fans, and parents, Soc. Sci. Med. 134 (Jun. 2015) 66–75, https://doi.org/10.1016/j.socscimed.2015.04.011.

[25] P. McCrory *et al.*, "Consensus statement on concussion in sport—the 5 th international conference on concussion in sport held in Berlin, October 2016," *Br. J. Sports Med.*, p. bjsports-2017-097699, Apr. 2017, doi: 10.1136/bjsports-2017-097699.

[26] I. Pozzato, et al., Challenges in the acute identification of mild traumatic brain injuries: results from an emergency department surveillance study, BMJ Open 10 (2) (Feb. 2020) e034494.

[27] R. M. Ruff, G. L. Iverson, J. T. Barth, S. S. Bush, D. K. Broshek, and the NAN Policy and Planning Committee, "Recommendations for Diagnosing a Mild Traumatic Brain Injury: A National Academy of Neuropsychology Education Paper," *Arch. Clin. Neuropsychol.*, vol. 24, no. 1, pp. 3–10, Feb. 2009, doi: 10.1093/arclin/acp006.

[28] R. J. Echemendia *et al.*, "The Sport Concussion Assessment Tool 5th Edition (SCAT5)," *Br. J. Sports Med.*, p. bjsports-2017-097506, Apr. 2017, doi: 10.1136/bjsports-2017-097506.

[29] D. H. Fong *et al.*, "Objectively Assessing Sports Concussion utilizing Visual Evoked Potentials," *J. Vis. Exp.*, no. 170, p. 62082, Apr. 2021, doi: 10.3791/62082.

[30] C.H. Tator, Concussions and their consequences: current diagnosis, management and prevention, Can. Med. Assoc. J. 185 (11) (Aug. 2013) 975–979, https://doi.org/10.1503/cmaj.120039.

[31] B. Lee, A. Newberg, Neuroimaging in traumatic brain imaging, NeuroRx 2 (2) (Apr. 2005) 372–383, https://doi.org/10.1602/neurorx.2.2.372.

[32] T.A. West, D.W. Marion, Current Recommendations for the Diagnosis and Treatment of Concussion in Sport: A Comparison of Three New Guidelines, J. Neurotrauma 31 (2) (Jan. 2014) 159–168, https://doi.org/10.1089/neu.2013.3031.

[33] P.E. Rapp, et al., Traumatic Brain Injury Detection Using Electrophysiological Methods, Front. Hum. Neurosci. 9 (Feb. 2015), https://doi.org/10.3389/fnhum.2015.00011.

[34] M. Gay, W. Ray, B. Johnson, E. Teel, A. Geronimo, S. Slobounov, Feasibility of EEG Measures in Conjunction With Light Exercise for Return-to-Play Evaluation After Sports-Related Concussion, Dev. Neuropsychol. 40 (4) (May 2015) 248–253, https://doi.org/10.1080/87565641.2015.1014486.

[35] A.C. Conley, et al., Resting State Electroencephalography and Sports-Related Concussion: A Systematic Review, J. Neurotrauma 36 (1) (Jan. 2019) 1–13, https://doi.org/10.1089/neu.2018.5761.

[36] F. Amyot, et al., A Review of the Effectiveness of Neuroimaging Modalities for the Detection of Traumatic Brain Injury, J. Neurotrauma 32 (22) (Nov. 2015) 1693–1721, https://doi.org/10.1089/neu.2013.3306.

[37] S.A. Dambinova, J.C. Maroon, A.M. Sufrinko, J.D. Mullins, E.V. Alexandrova, A. A. Potapov, Functional, Structural, and Neurotoxicity Biomarkers in Integrative Assessment of Concussions, Front. Neurol. 7 (Oct. 2016), https://doi.org/10.3389/fneur.2016.00172.

[38] M. Vishwanath, et al., Investigation of Machine Learning Approaches for Traumatic Brain Injury Classification via EEG Assessment in Mice, Sensors 20 (7) (Apr. 2020) 2027, https://doi.org/10.3390/s20072027.

[39] K. Thanjavur, et al., Recurrent neural network-based acute concussion classifier using raw resting state EEG data, Sci. Rep. 11 (1) (Dec. 2021) 12353, https://doi.org/10.1038/s41598-021-91614-4.

[40] N. Vivaldi, M. Caiola, K. Solarana, M. Ye, Evaluating Performance of EEG Data-Driven Machine Learning for Traumatic Brain Injury Classification, IEEE Trans. Biomed. Eng. 68 (11) (Nov. 2021) 3205–3216, https://doi.org/10.1109/TBME.2021.3062502.

[41] A.E. Jacquin, et al., Concussion assessment potentially aided by use of an objective multimodal concussion index, J. Concussion 5 (Jan. 2021), https://doi.org/10.1177/20597002211004333.

[42] S.D. Fickling, et al., Brain vital signs detect concussion-related neurophysiological impairments in ice hockey, Brain 142 (2) (Feb. 2019) 255–262, https://doi.org/10.1093/brain/awy317.

[43] R. Kothari, P. Bokariya, S. Singh, R. Singh, A Comprehensive Review on Methodologies Employed for Visual Evoked Potentials, Scientifica 2016 (2016) 1–9, https://doi.org/10.1155/2016/9852194.

[44] N.K. Yadav, K.J. Ciuffreda, Objective assessment of visual attention in mild traumatic brain injury (mTBI) using visual-evoked potentials (VEP), Brain Inj. 29 (3) (Feb. 2015) 352–365, https://doi.org/10.3109/02699052.2014.979229.

[45] D. Poltavski, P. Lederer, L.K. Cox, Visually Evoked Potential Markers of Concussion History in Patients with Convergence Insufficiency, Optom. Vis. Sci. 94 (7) (Jul. 2017) 742–750, https://doi.org/10.1097/OPX.0000000000001094.

[46] A.M. Dreyer, C.S. Herrmann, Frequency-modulated steady-state visual evoked potentials: A new stimulation method for brain–computer interfaces, J. Neurosci. Methods 241 (Feb. 2015) 1–9, https://doi.org/10.1016/j.jneumeth.2014.12.004.

[47] A.M. Norcia, L.G. Appelbaum, J.M. Ales, B.R. Cottereau, B. Rossion, The steady-state visual evoked potential in vision research: A review, J. Vis. 15 (6) (May 2015) 4, https://doi.org/10.1167/15.6.4.

[48] C. Cao, R.L. Tutwiler, S. Slobounov, Automatic Classification of Athletes With Residual Functional Deficits Following Concussion by Means of EEG Signal Using Support Vector Machine, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (4) (Aug. 2008) 327–335, https://doi.org/10.1109/TNSRE.2008.918422.

[49] M. Vishwanath *et al.*, "Investigation of Machine Learning and Deep Learning Approaches for Detection of Mild Traumatic Brain Injury from Human Sleep Electroencephalogram," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico: IEEE, Nov. 2021, pp. 6134–6137. doi: 10.1109/EMBC46164.2021.9630423.

[50] C.M. Bishop, Pattern recognition and machine learning. Information science and statistics, Springer, New York, 2006.

[51] R.T. Schirrmeister, et al., Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis, Hum. Brain Mapp. 38 (11) (Nov. 2017) 5391–5420, https://doi.org/10.1002/hbm.23730.

[52] G. Erus, M. Habes, and C. Davatzikos, "Machine learning based imaging biomarkers in large scale population studies: A neuroimaging perspective," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, Elsevier, 2020, pp. 379–399. doi: 10.1016/B978-0-12-816176-0.00021-1.

[53] J.S. Kumar, P. Bhuvaneswari, Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study, Procedia Eng. 38 (2012) 2525–2536, https://doi.org/10.1016/j.proeng.2012.06.298.

[54] A.S. Al-Fahoum, A.A. Al-Fraihat, Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains, ISRN Neurosci. 2014 (Feb. 2014) 1–7, https://doi.org/10.1155/2014/730218.

[55] Y. Höller, et al., Comparison of EEG-Features and Classification Methods for Motor Imagery in Patients with Disorders of Consciousness, PLoS One 8 (11) (Nov. 2013) e80479.

[56] H.U. Amin, W. Mumtaz, A.R. Subhani, M.N.M. Saad, A.S. Malik, Classification of EEG Signals Based on Pattern Recognition Approach, Front. Comput. Neurosci. 11 (Nov. 2017) 103, https://doi.org/10.3389/fncom.2017.00103.

[57] M.C. Guerrero, J.S. Parada, H.E. Espitia, EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks, Heliyon 7 (6) (Jun. 2021) e07258.

[58] S.D. Wickramaratne, M.S. Mahmud, R.S. Ross, in: Use of Brain Electrical Activity to Classify People with Concussion: A Deep Learning Approach, IEEE, Jun, Dublin, Ireland, 2020, pp. 1–6, https://doi.org/10.1109/ICC40277.2020.9149393.

[59] J.T. Moyer, et al., Standards for data acquisition and software-based analysis of in vivo electroencephalography recordings from animals. A TASK 1- WG 5 report of the AES/ ILAE Translational Task Force of the ILAE, Epilepsia 58 (S4) (Nov. 2017) 53–67, https://doi.org/10.1111/epi.13909.

[60] F. Song, Z. Guo, and D. Mei, "Feature Selection Using Principal Component Analysis," in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, Yichang, China: IEEE, Nov. 2010, pp. 27–30. doi: 10.1109/ICSEM.2010.14.

[61] A. Alwosheel, S. Van Cranenburgh, C.G. Chorus, Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis, J. Choice Model. 28 (Sep. 2018) 167–182, https://doi.org/10.1016/j.jocm.2018.07.002.
[62] J.D. Lewine, et al., Quantitative EEG Biomarkers for Mild Traumatic Brain Injury, J. Clin. Neurophysiol. 36 (4) (Jul. 2019) 298–305, https://doi.org/10.1097/WNP.0000000000000588.
[63] R. Boshra, et al., From Group-Level Statistics to Single-Subject Prediction: Machine Learning Detection of Concussion in Retired Athletes, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (7) (Jul. 2019) 1492–1501, https://doi.org/10.1109/TNSRE.2019.2922553.
[64] C. Beleites, R. Salzer, Assessing and improving the stability of chemometric models in small sample size situations, Anal. Bioanal. Chem. 390 (5) (Mar. 2008) 1261–1271, https://doi.org/10.1007/s00216-007-1818-6.
[65] M. Ojala and G. C. Garriga, "Permutation Tests for Studying Classifier Performance," in *2009 Ninth IEEE International Conference on Data Mining*, Miami Beach, FL, USA: IEEE, Dec. 2009, pp. 908–913. doi: 10.1109/ICDM.2009.108.
[66] J. Ludbrook, H. Dudley, Why Permutation Tests are Superior to *t* and *F* Tests in Biomedical Research, Am. Stat. 52 (2) (May 1998) 127–132, https://doi.org/10.1080/00031305.1998.10480551.
[67] P.H. Westfall, J.F. Troendle, Multiple Testing with Minimal Assumptions, Biom. J. 50 (5) (Oct. 2008) 745–755, https://doi.org/10.1002/bimj.200710456.
[68] J.W.L. Hooton, Randomization tests: statistics for experimenters, Comput. Methods Programs Biomed. 35 (1) (May 1991) 43–51, https://doi.org/10.1016/0169-2607(91)90103-Z.
[69] R.L. Nuzzo, Randomization Test: An Alternative Analysis for the Difference of Two Means, PM&R 9 (3) (Mar. 2017) 306–310, https://doi.org/10.1016/j.pmrj.2017.02.001.
[70] A. Sik, B. Kaveney-Gibb, J. Cooper, J. Pearson, P. Larsen, and A. Rogan, "Is the Sport Concussion Assessment Tool 5th Edition a useful concussion assessment tool in head-injured patients presenting to the emergency department?," *Emerg. Med. Australas.*, pp. 1742-6723.14144, Dec. 2022, doi: 10.1111/1742-6723.14144.
[71] Z.Y. Kerr, J.K. Register-Mihalik, S.W. Marshall, K.R. Evenson, J.P. Mihalik, K. M. Guskiewicz, Disclosure and non-disclosure of concussion and concussion symptoms in athletes: Review and application of the socio-ecological framework, Brain Inj. 28 (8) (Jul. 2014) 1009–1021, https://doi.org/10.3109/02699052.2014.904049.
[72] X. Ying, An Overview of Overfitting and its Solutions, J. Phys. Conf. Ser. 1168 (Feb. 2019), 022022, https://doi.org/10.1088/1742-6596/1168/2/022022.
[73] R.L. Figueroa, Q. Zeng-Treitler, S. Kandula, L.H. Ngo, Predicting sample size required for classification performance, BMC Med. Inform. Decis. Mak. 12 (1) (Dec. 2012) 8, https://doi.org/10.1186/1472-6947-12-8.
[74] G. Bunc, J. Ravnik, T. Velnar, May Heading in Soccer Result in Traumatic Brain Injury? A Review of Literature, Med. Arch. 71 (5) (2017) 356, https://doi.org/10.5455/medarh.2017.71.356-359.
[75] J. Mooney *et al.*, "Concussion in soccer: a comprehensive review of the literature," *Concussion*, vol. 5, no. 3, p. CNC76, Sep. 2020, doi: 10.2217/cnc-2020-0004.