



Article

A Novel Method to Assist Clinical Management of Mild Traumatic Brain Injury by Classifying Patient Subgroups Using Wearable Sensors and Exertion Testing: A Pilot Study

Joshua P. McGeown^{1,2,3,*} , Mangor Pedersen⁴, Patria A. Hume^{2,3,5,6} , Alice Theadom^{2,3}, Stephen Kara⁷ and Brian Russell^{3,5,8}

¹ Mātai Medical Research Institute, Gisborne 4010, New Zealand

² Traumatic Brain Injury Network, Auckland University of Technology, Auckland 1142, New Zealand; patria.hume@aut.ac.nz (P.A.H.); alice.theadom@aut.ac.nz (A.T.)

³ Sports Performance Research Institute New Zealand (SPRINZ), Auckland University of Technology, Auckland 0632, New Zealand; brian@ambientcognition.com

⁴ Department of Psychology and Neuroscience, Auckland University of Technology, Auckland 1142, New Zealand; mangor.pedersen@aut.ac.nz

⁵ Auckland Bioengineering Institute, The University of Auckland, Auckland 1010, New Zealand

⁶ WA Law School, The University of Western Australia, Perth, WA 6009, Australia

⁷ Axis Sports Medicine, Auckland 1072, New Zealand; stephenkara1965@gmail.com

⁸ Ambient Cognition, Nelson 7173, New Zealand

* Correspondence: j.mcgeown@matai.org.nz; Tel.: +64-027-580-2659

Abstract: Although injury mechanisms of mild traumatic brain injury (mTBI) may be similar across patients, it is becoming increasingly clear that patients cannot be treated as one homogenous group. Several predominant symptom clusters (PSC) have been identified, each requiring specific and individualised treatment plans. However, objective methods to support these clinical decisions are lacking. This pilot study explored whether wearable sensor data collected during the Buffalo Concussion Treadmill Test (BCTT) combined with a deep learning approach could accurately classify mTBI patients with physiological PSC versus vestibulo-ocular PSC. A cross-sectional design evaluated a convolutional neural network model trained with electrocardiography (ECG) and accelerometry data. With a leave-one-out approach, this model classified 11 of 12 (92%) patients with physiological PSC and 3 of 5 (60%) patients with vestibulo-ocular PSC. The same classification accuracy was observed in a model only using accelerometry data. Our pilot results suggest that adding wearable sensors during clinical tests like the BCTT, combined with deep learning models, may have the utility to assist management decisions for mTBI patients in the future. We reiterate that more validation is needed to replicate the current results.

Keywords: mTBI; convolutional neural network; ecological validity; neurophysiology; biomechanics; symptom clusters



Citation: McGeown, J.P.; Pedersen, M.; Hume, P.A.; Theadom, A.; Kara, S.; Russell, B. A Novel Method to Assist Clinical Management of Mild Traumatic Brain Injury by Classifying Patient Subgroups Using Wearable Sensors and Exertion Testing: A Pilot Study. *Biomechanics* **2023**, *3*, 231–249. <https://doi.org/10.3390/biomechanics3020020>

Academic Editor: Tibor Hortobagyi

Received: 13 April 2023

Revised: 12 May 2023

Accepted: 16 May 2023

Published: 26 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Traumatic brain injuries (TBIs) occur as a consequence of exposure to sudden physical trauma from external forces resulting in deformation and damage of neuronal and vascular tissues within the brain [1]. An estimated 95% of TBI patients have suffered mild TBI (mTBI) determined by a Glasgow Coma Scale score ≥ 13 , absence of skull fracture or positive neuroimaging, and associated with symptoms resulting from impaired neurological function [2–5]. Approximately 20% of TBIs result from a sport/physical activity related mechanism of injury, with 98% of these sport-originated brain injuries considered mild [6].

Until recently, recommendations for managing mTBI followed a one-size-fits-all model, where all patients received the same care regardless of their specific symptom burden [7]. It was thought that 85–90% of these mTBI patients would experience spontaneous clinical

recovery and resolution of symptoms within 10–14 days post-injury; thus a “wait and see” approach was adopted. Access to proactive interventions was consequently limited [8]. However, updated reports from large epidemiological studies of the general population have indicated that up to half of mTBI patients can experience prolonged symptom burden beyond this window [2,9–12]. Moreover, incomplete recovery at six and 12 months has been observed in 50% of mTBI patients who presented to an emergency department within 24 h post-injury [11,12]. These poor outcomes, even in early-presenters, suggest that improved assessment and management pathways are needed to reduce the burden of mTBI. Although injury mechanisms of mTBI may be similar across patients, it is becoming increasingly clear that patients cannot be treated as one homogenous group as several predominant symptom clusters (PSCs) have been identified, each requiring specific and individualised treatment plans [13–18]. Criteria have been developed to identify different PSCs to determine whether a predominantly physiological, vestibulo-ocular, or cervicogenic origin appears to contribute to unresolved symptoms at ≥ 21 days post-injury [13]. The criteria to determine the PSC utilise the patient’s clinical history and examination along with results of a provocative exercise test, commonly referred to as the Buffalo Concussion Treadmill Test (BCTT) [13,19–21]. A framework has been provided to help understand the pathophysiology responsible for each PSC, and how to approach the prescription of individualised exercise-based interventions to address and resolve these underlying mechanisms [13,18].

It has been hypothesised that symptoms consistent with physiological PSC are manifestations caused by uncoupling of the autonomic nervous and cardiovascular systems [13,22–24]. Previous laboratory research has identified differences in autonomic regulation during exercise, but not at rest, between athletes who had recently sustained a mTBI, and healthy controls, as measured by heart rate variability using electrocardiography (ECG) [23,24]. Conversely, symptoms that characterise vestibulo-ocular PSC are thought to be due to underlying issues with sensorimotor integration and vestibular function [13,22,25,26]. Static balance and gait differences between healthy controls and participants with mTBI have been found, using force plates and accelerometers [27–29]. Individuals with vestibulo-ocular PSC may display greater impairment in balance/gait than those with physiological PSC, and vice versa for autonomic regulation. Presently, even when the BCTT is used, identifying PSCs relies on athlete’s honest symptom reporting, which is an issue when managing athletes with mTBI [30–33], and the training/experience of the supervising clinician [34–36]. Previous observations with laboratory grade equipment have highlighted that technological methods may be sensitive to impairments not detected by clinical tests [23,24,27–29]. However, restricted accessibility to laboratory grade equipment in clinical environments has limited translation of these findings.

The development of wearable sensors validated against laboratory grade equipment has enabled research to be transitioned from the lab to the real-world [37]. Integrating the BCTT within clinical practice presents an opportunity to use a wearable sensor to collect ECG and accelerometry data under ecologically valid conditions. Adding objective measures to this testing protocol might assist accurate PSC classification when symptom reporting may be untrustworthy and/or when clinicians have limited experience working with mTBI patients. Machine learning models are an increasingly popular approach for analysing time series data such as ECG and accelerometry [37–41]. There is potential for machine learning algorithms, specifically deep learning techniques, to support medical decision making by automatically detecting the most important features related to patient outcomes i.e., high versus low risk for a given condition [38,42]. The capability of a deep learning model trained using time series signals from ECG (as a measure of autonomic control) and accelerometry during gait (as a surrogate of sensorimotor integration) to classify PSC subgroups in mTBI patients has not been previously explored.

Therefore, this pilot study aimed to investigate whether a deep learning approach could accurately classify mTBI patients with physiological PSC versus vestibulo-ocular PSC using wearable sensor data collected during the BCTT. A secondary aim was to compare classification performance using ECG and accelerometry data in combination

versus accelerometry alone. If accurate PSC classification was observed, a final aim was to perform a post-hoc signal exploration to provide preliminary insight into features that may differ between PSCs, thus contributing to classification.

2. Methods

2.1. Study Design and Participants

A cross-sectional design was adopted to evaluate the accuracy of a deep learning model trained with ECG and accelerometry data in order to classify patients with physiological PSC versus vestibulo-ocular PSC. Data collection took place at a single dedicated sport-related mTBI clinic between April and July 2019. A total of 43 consenting patients diagnosed with mTBI by a sport and exercise medicine physician completed a BCTT during the data collection period for this study. Diagnosis of sport-related mTBI was made in line with the 2017 Concussion in Sport Group Consensus Statement [8]. The sample was inclusive for age and sex. Institutional (AUTEC 18/374) and health and disability committee (HDEC 18/NTA/108) ethical approvals were obtained, and this study was conducted according to the ethical standards of the Declaration of Helsinki. Participants provided written consent (child assent and parental consent for participants <16 years old) for their data to be used for research and publication. All participant data were de-identified prior to extraction/data analysis to ensure confidentiality.

2.2. Procedures

2.2.1. Timing of BCTT

Participants received mTBI management as part of usual clinical care; the details of this clinical pathway have been previously described [9]. Clinical scheduling constraints meant that not all patients completed a BCTT. Testing was completed if requested by the supervising physician to confirm suspected PSC based on patient history and clinical examination and/or to inform an individualised treatment plan. The appointment during which the BCTT was administered depended on the proximity of the participant's initial clinical assessment to the date of injury. Efforts were made to administer a BCTT approximately two weeks post-injury to aid in developing a targeted and individualised exercise treatment plan for participants with unresolved symptoms based on their PSC. Participants presenting for initial clinical assessment ≥ 10 days post-injury completed a BCTT on their first visit if requested by the physician. Treadmill testing was performed during the first follow-up appointment if participants initially presented within 10 days of their injury and had not reported improvement in their symptoms between the first and second visits.

2.2.2. Treadmill Testing Protocol

The BCTT protocol is a safe and reliable method to assist patient prognosis, PSC identification, and management decisions for mTBI patients with unresolved symptoms [13,19–21,43]. Contraindications, including inability to exercise because of orthopedic injury or increased cardiorespiratory risk, were considered when deciding whether a participant had completed a BCTT [20]. Pre-testing symptom burden was established by participants rating their current symptom burden on a 1–10 scale (1 = minimal/no symptoms; 10 = severe symptoms), and specific symptoms were recorded [20].

Participants walked on a treadmill at 5 km/h at 0% incline for one minute, then incline was increased to 2% and a further 1% every minute thereafter, with the velocity remaining at 5 km/h [20]. Participants were asked to walk normally, without resting their hands on the treadmill rails. After every minute of the protocol, participants were asked to report any changes in their symptom burden and rating of perceived exertion (scale of 0–10; 10 representing maximal effort) [44]. Heart rate was monitored and recorded at the end of each minute. Testing proceeded in this way until participants experienced symptom exacerbation (increase of +3 from their initial reports) or volitional fatigue (rating of perceived exertion ≥ 9) [20]. If the participant did not experience symptom exacerbation or fatigue by the maximum incline of 15%, velocity was increased by 1 km/h while

maintaining the 15% incline, to a maximum testing duration of 20 min. All BCTTs were delivered by a clinician and/or researcher with training in exercise physiology.

2.2.3. Expert Labelling

Physicians with training and experience managing mTBI athletes identified the PSC using published criteria [13]. Given that physiological and vestibulo-ocular PSCs were the most prevalent in our previous research [45] combined with the novel nature of the current study, we elected to focus specifically on these two groups. Data from 26 participants (18 physiological PSC; eight vestibulo-ocular PSC) who completed a BCTT were eligible for the current study. For analysis purposes, vestibulo-ocular PSC was considered the positive class, since it is less prevalent and a predictor of worse recovery outcomes.

2.3. Instrumentation

During the BCTT, heart rate was recorded using a Zephyr BioHarness™ (Medtronic, Dublin, Ireland) that contained three-lead ECG (sampling rate 250 Hz) and a tri-axial accelerometer (sampling rate: 100 Hz; x-axis = vertical, y = lateral, z-axis = sagittal). The BioHarness™ has been shown to be a valid and reliable tool for physiological monitoring in field-based applications with military, first responders, and athletes [46–48]. Data were transmitted in real-time via Bluetooth to a smartphone application (IoTool sensor platform from SenLab Ljubljana, Slovenia). Raw data were stored within local memory and downloaded after recording. The BioHarness™ was worn in the same way as a typical heart rate monitor, where the electrodes were lightly dampened with cold water and the strap was placed snugly around the torso at the level of the xiphoid process. Two BioHarness™ of different sizes were used for all testing based on subject chest size (sizes XS-M and M-XL). Preservation of clinical flow and ecological validity of findings was a priority of this study. Once a consistent heart rate was registered on the smartphone application, normal procedures for explaining and administering the BCTT were initiated without further adjustment of the BioHarness™. All BCTTs were performed on a Life Fitness Engage 95T treadmill (Hamilton, New Zealand).

2.4. Deep Learning Pipeline

While there are several deep learning methods, previous findings that indicate a convolutional neural network (CNN) trained with ECG or accelerometry time series data can accurately classify patients with/without cardiac dysfunction [41] and different human activities [37,39,40], respectively. Notable advantages of CNNs include computational efficiency and little to no manual feature engineering [39,49]. For these reasons, a CNN appeared to be a good starting point to explore if deep learning could automatically identify features in ECG and/or accelerometry signals that differentiate between mTBI PSCs. Figure 1 shows the data pipeline from patient identification through the experiments executed in this study. All data pre-processing and deep learning models were completed in Python (v3.8.5) using the *Sklearn* (v0.23.2), *Tensorflow* (v2.3.1), and *Keras* (v2.4.0) packages.

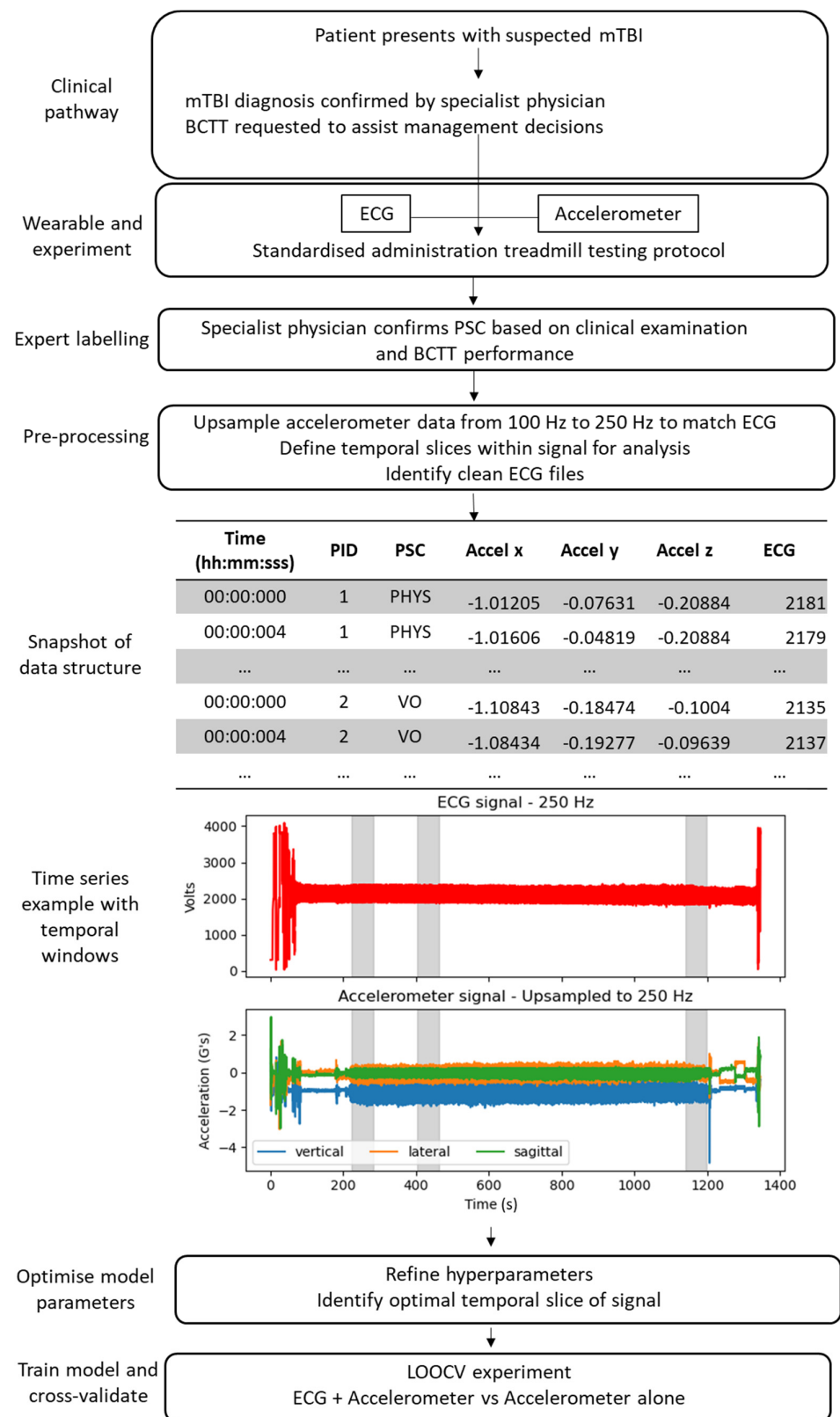


Figure 1. Data pipeline from mTBI to deep learning model. Abbreviations: BCTT—Buffalo Concussion Treadmill Test; ECG—electrocardiography; LOOCV—leave-one-out-cross-validation; PID—participant ID; PHYS—physiological; PSC—predominant symptom cluster; VO—vestibulo-ocular.

2.4.1. Dataset Preparation and Pre-Processing

Prior to analysis, accelerometry time series data were time aligned, labelled and up-sampled from 100 Hz to 250 Hz to match the sampling frequency of ECG data using linear interpolation. After up-sampling, three temporal slices of the ECG and accelerometry signals were defined to standardise data inputs to the CNN across participants, due to the symptom-dependent variation in BCTT duration. These slices included the first 60 s of testing, the third minute, and the final 60 s before the BCTT was ended. The first minute was chosen to explore if different PSCs could be classified using sensor data collected early in the protocol without the need to push the patient to symptom exacerbation, whereas the final 60 s should coincide with the onset of symptom exacerbation or volitional fatigue. The third-minute slice was selected as the latest common time for all subjects, because the shortest BCTT collected for this study was four minutes. The third minute provided a standardised timepoint to evaluate all participants while under physiological load (5 km/h, 4% incline), but prior to symptom exacerbation or fatigue. When administering the BCTT, the velocity of the treadmill gradually increased to 5 km/h, and the first minute of testing began once the treadmill was up to the required velocity and the participant settled into steady state walking. The beginning of testing was identified in each participant's time series file by plotting their accelerometry data and manually identifying the timestamp when the participant settled into a repeatable gait pattern, defined by 5+ consecutive gait cycles with similar waveform profiles (see Figure 2B).

The first-minute slice was the next 60 s from the manually defined point, and the third minute was from 180 to 240 s. A similar approach was used to determine the final 60 s slice. The time series accelerometry plots were manually inspected to identify when there was a sudden change in the gait pattern, representing when the participant stepped onto the treadmill rails, decreased the velocity or stopped the treadmill (Figure 2C). The 60 s of signal leading up to the timestamp of this disruption in gait served as the final slice of data to be explored.

ECG files were visually inspected to assess if a clean ECG signal with distinct and repeatable QRS complexes had been recorded within the temporal slices of interest. Clean ECG time series were observed in 17/26 participants (12 physiological PSC; five vestibulo-ocular PSC). The remaining nine participants' files were deemed unfit for further analysis because they contained large amounts of noise, making it difficult to discern the QRS pattern. Noise was likely due to poor coupling between the ECG nodes and the skin, due to the quick deployment the BioHarness™ to prioritise clinical flow. An example of the final dataset columns are provided in Figure 1.

Time series data were segmented for training and cross-validation of CNN models, where n samples were divided into windows of W samples wide, with S overlap producing D rows of data (Equation (1)) [37].

$$D = 1 + \frac{n - W}{S} \quad (1)$$

Data were transformed into an array (D, W, F) with D rows, width of W samples, and F features ($F = 4$ for ECG + accelerometry (x, y, z); $F = 3$ for accelerometry (x, y, z) only) so that they were compatible with the CNN [37]. A separate one-dimensional array contained PSC labels corresponding to each row within the signal data array based on which participant the data belonged to.

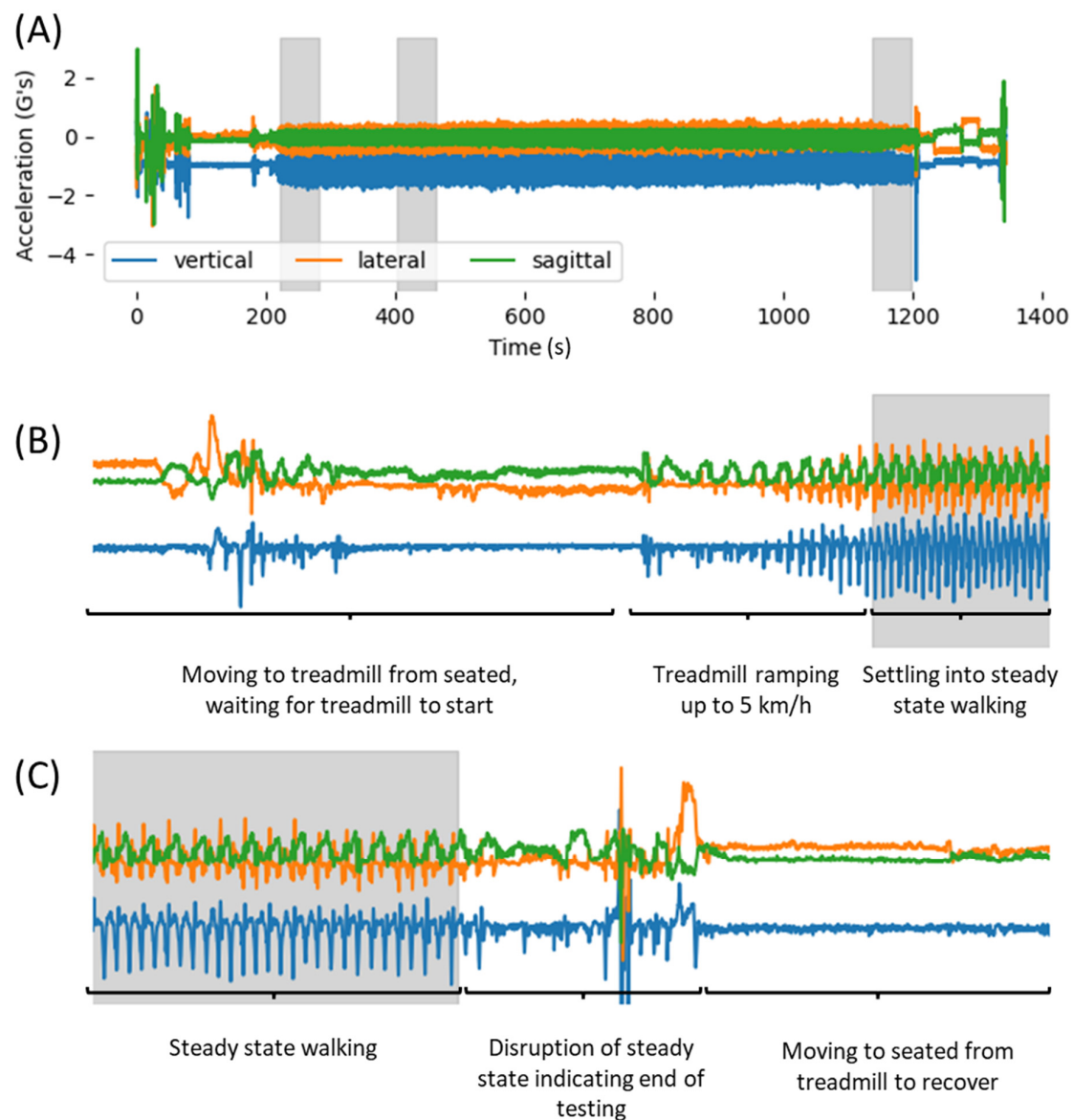


Figure 2. Process showing how temporal slices were determined. (A) Example of complete signal from time BioHarness™ was turned on until turned off, with temporal slices shown in grey. (B) How accelerometry waveforms were used to determine the start of the first minute. (C) How accelerometry waveforms were used to determine the end of BCTT.

2.4.2. Deep Learning Model

The CNN consisted of four separate 1D convolutional networks for each axis of the accelerometry signal (x, y, z) and ECG signal (see Figure 3 for CNN topology) combined into a single classifier output stage. Each channel used two convolutional 1D layers and a ReLU activation function with filter and kernel sizes of 64 and 3, respectively. Convolutional layers were followed by a drop out layer of 50% for generalizability and a max pool layer with pool size of 0.5. A flatten layer was used to combine separate channels, followed by a dense layer with ReLU activation. Output weights from previous layers were transformed into probabilities for each class using a SoftMax activation function. Learning used an epoch of 50 and batch size of 100. Classification metrics, including accuracy, Cohen's kappa, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), were calculated to evaluate the CNN's performance when classifying windows with physiological labels versus vestibulo-ocular labels.

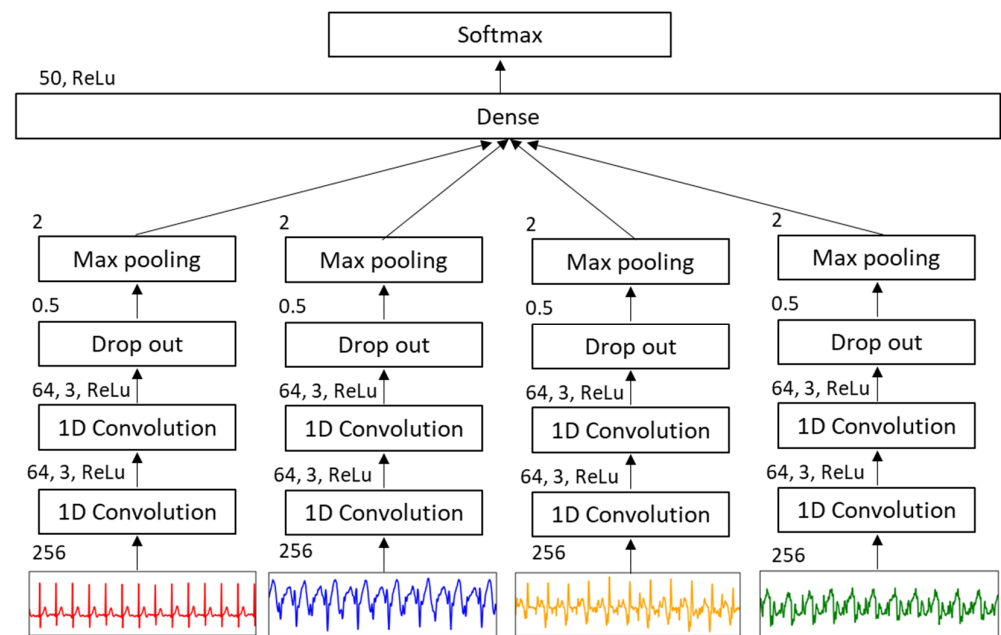


Figure 3. Structure of CNN.

2.4.3. Hyperparameter Tuning

Window size tuning was performed to optimise PSC classification accuracy. Windows ranging in size from 16, 32, 64, 128, and 256, with a 50% overlap between windows, were evaluated. For each set of hyperparameters, a randomised train test split of 0.66/0.33 was selected using *Sklearn* in Python. During hyperparameter tuning, classification metrics were calculated based on performance of the CNN to correctly classify the label associated with each window in the test set, irrespective of which participant the window came from. Priority was placed on hyperparameters that maximised Cohen's kappa and PPV due to the imbalance in PSC classes. This process was applied for each temporal slice (first minute, third minute, final minute) to cross-reference optimal hyperparameters within each slice and to evaluate if differences in classification metrics were evident.

2.4.4. LOOCV Experiment

Once optimum hyperparameters were determined the following experiment was conducted to explore whether the CNN could correctly classify a given participant's PSC. Due to a limited sample size, a leave-one-out-cross-validation (LOOCV) methodology was implemented to train the CNN on data from $n - 1$ participants, followed by testing the model's capacity to correctly identify the PSC of the left-out case using their sensor data. This process was repeated for each participant and allowed exploration of how well the model trained on current data might perform at classifying physiological versus vestibulo-ocular PSC following a BCTT in a new mTBI patient.

Use of time series data in this study meant that 15,000 rows ($250 \text{ Hz} \times 60 \text{ s}$) corresponded to each participant per temporal slice. To effectively accomplish LOOCV for each participant, *Sklearn's* stratified k-folds method was used, wherein all data were split into 17 equal folds of 15,000 rows of data without shuffling. The first iteration held out rows 0–14,999 from participant 1 for LOOCV and trained on rows 15,000–254,999 (participants 2–17), the second iteration held out rows 15,000–29,999 from participant 2 while training on 0–14,999 and 30,000–254,999 (participants 1, 3–17), and so on. The CNN was trained on $k - 1$ folds using a randomised train test split of 0.66/0.33. A probability for each class was calculated corresponding to each of the D windows in the validation fold and these probabilities were stored in an array. For a given participant, some windows may have had a higher probability of belonging to the physiological PSC class, while the rest indicated that vestibulo-ocular PSC was more probable. Therefore, the mean probability across D

windows was calculated for each class and the class with the greater mean probability was recorded as the predicted class for the participant to which the windows belonged. The predicted class was saved after each iteration. Once all iterations were complete, a confusion matrix was produced, from which sensitivity, specificity, PPV, NPV, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve were calculated. This process is visualised in Figure 4.

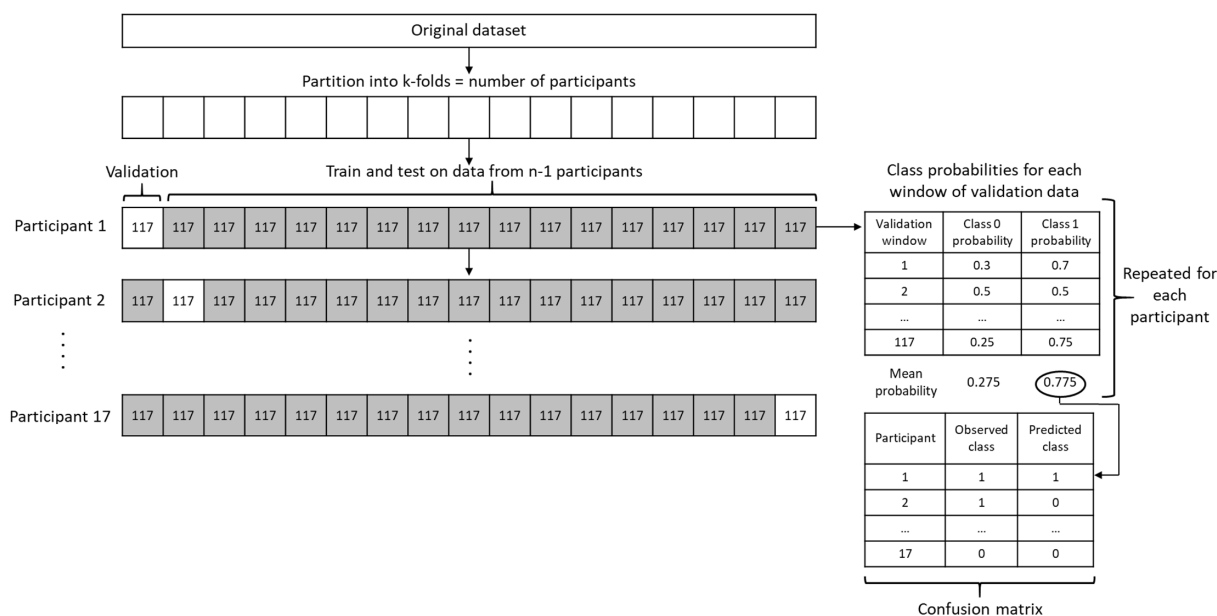


Figure 4. Diagram of how LOOCV was implemented to evaluate potential utility of a CNN to predict the class of ‘new’ sensor data collected during a BCTT.

This methodology was first used to explore the potential of a CNN trained with both ECG and accelerometry time series data to accurately classify the PSC of 17 participants with clean ECG data. Since noisy ECG data led to exclusion of nearly 1/3 of eligible participants (9/26), this experiment was repeated using only accelerometry data to see the level of classification accuracy that might be obtained, without the need to collect ECG data.

2.5. Post-Hoc Feature Exploration: PSD and Entropy Analysis

An advantage of deep learning models such as CNN is that they automatically detect features in the data that maximise classification accuracy which may not be apparent to a human analyst. This significantly reduces the time and expertise required to develop the model, compared to machine learning techniques that rely on manually engineered features [42]. Such an approach was appropriate for the current study, since our primary aim was to evaluate if a model trained using wearable sensor data collected during the BCTT *could* classify different PSCs in mTBI patients [38]. This approach did not allow inferences to be made as to *how* the model differentiated between groups, leading to deep learning techniques such as CNNs being commonly described as a ‘black box’ [38].

To provide a degree of explainability to our LOOCV experiment results, we performed post-hoc analysis on two common time series features: power spectral density and multiscale entropy. Power spectral density (PSD) provides a method to estimate the power or energy of a temporal signal at different frequencies [50] and PSD-based analyses have demonstrated utility in neuroscientific investigations [51–53]. Multiscale entropy (MSE) quantifies the complexity (i.e., repeatability or randomness) within physiological/physical time series signals while accounting for different time scales [54,55] and has proven a useful feature in the analysis of gait and ECG data [56,57]. For reference, periodic signals have low entropy and random signals have high entropy.

The maximum PSD of accelerometry and ECG signals were calculated using the `signal.welch()` function in Python's *SciPy* (v1.5.4) package. The MSE complexity index of signals were determined using the Python *EntropyHub* (v0.2) package with coarse graining and a scale factor of 10. Given the limited sample in this study, Hedges' *g* effect sizes with 95% confidence intervals were calculated to evaluate group differences in max PSD and MSE complexity index between physiological and vestibulo-ocular PSC using the R package *effsize* (v0.8.1). Finally, the magnitude of differences between PSCs for PSD and MSE were presented in an effect measure plot using Python's *zEpid* package (v0.9.0).

3. Results

3.1. Demographics

Of the 17 participants whose data were analysed, 11 were male (eight physiological PSC; three vestibulo-ocular PSC) and six were female (4 physiological PSC; 2 vestibulo-ocular PSC), with a median age of 21.0 (IQR: 16.0; 28.0) years. Participants completed a BCTT with a median of 17.0 (IQR: 13.0; 18.5) days post-injury, and with a median BCTT duration (mm:ss) of 14:12 (IQR: 9:12; 17:55).

3.2. Hyperparameters and Temporal Slices

Results of hyperparameter tuning are shown in Table 1. The best classification performance of windows corresponding to physiological PSC versus vestibulo-ocular PSC occurred with a window size of 256 and an overlap of 128 between windows, as determined by Cohen's Kappa and PPV. Similar levels of accuracy were observed using these hyperparameters with data from both first- and third-minute slices. This meant that, for each participant, their 60 s of data (sampled at 250 Hz) fed into the CNN was separated into $D = 117$ windows for the LOOCV experiments.

Table 1. Hyperparameter tuning results from each temporal slice.

| | Window Size | Overlap | Accuracy | Kappa | Sensitivity | Specificity | PPV | NPV |
|--------------|-------------|---------|----------|-------|-------------|-------------|------|------|
| First minute | 16 | 8 | 0.90 | 0.76 | 0.84 | 0.92 | 0.82 | 0.93 |
| | 32 | 16 | 0.93 | 0.83 | 0.92 | 0.93 | 0.84 | 0.97 |
| | 64 | 32 | 0.95 | 0.88 | 0.93 | 0.96 | 0.9 | 0.97 |
| | 128 | 64 | 0.97 | 0.92 | 0.92 | 0.98 | 0.96 | 0.97 |
| | 256 | 128 | 0.98 | 0.96 | 0.98 | 0.99 | 0.97 | 0.99 |
| Third minute | 16 | 8 | 0.89 | 0.72 | 0.7 | 0.97 | 0.91 | 0.88 |
| | 32 | 16 | 0.93 | 0.83 | 0.88 | 0.95 | 0.88 | 0.95 |
| | 64 | 32 | 0.96 | 0.90 | 0.92 | 0.98 | 0.94 | 0.96 |
| | 128 | 64 | 0.95 | 0.88 | 0.87 | 0.99 | 0.96 | 0.94 |
| | 256 | 128 | 0.97 | 0.94 | 0.96 | 0.98 | 0.95 | 0.98 |
| Final minute | 16 | 8 | 0.88 | 0.69 | 0.7 | 0.95 | 0.86 | 0.88 |
| | 32 | 16 | 0.93 | 0.82 | 0.8 | 0.98 | 0.94 | 0.92 |
| | 64 | 32 | 0.96 | 0.91 | 0.95 | 0.97 | 0.92 | 0.98 |
| | 128 | 64 | 0.96 | 0.90 | 0.91 | 0.97 | 0.94 | 0.96 |
| | 256 | 128 | 0.94 | 0.86 | 0.92 | 0.95 | 0.88 | 0.97 |

Note: Shaded row indicates window size and overlap with best performance for each temporal slice.

3.3. LOOCV Experiment

Using a k-folds approach to implement LOOCV, the CNN correctly classified three of five (60%) vestibulo-ocular cases and 10 of 12 (83%) physiological cases, using ECG and accelerometry data from the first minute of each participants BCTT. Similar performance was observed when using ECG and accelerometry data from the third minute of the treadmill protocol, with three of five (60%) vestibulo-ocular and 11 of 12 (92%) physiological cases correctly classified. The final minute performed worse than chance in classifying 0 of

five vestibulo-ocular cases and eight of 12 (67%) physiological cases. Comparable classification accuracy was seen for each temporal slice when ECG data from these participants were omitted and CNNs were trained using only accelerometry data. See Figure 5 for receiver operating characteristic curves and related accuracy metrics for each condition, and Table 2 for individual results.

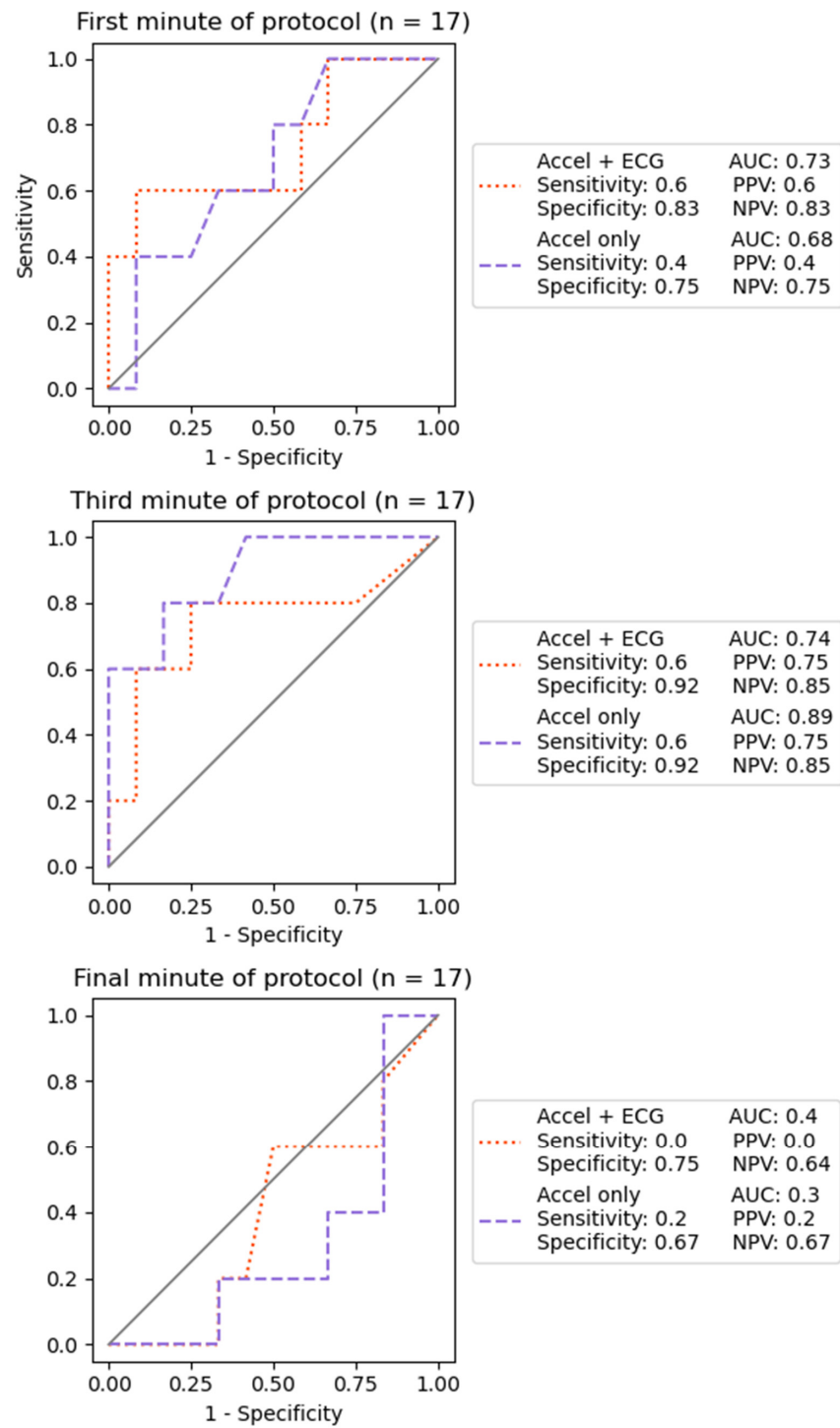


Figure 5. LOOCV classification performance of the CNN using accelerometry data with and without ECG data from each temporal slice.

Table 2. Predicted class probabilities for each participant and temporal slice using LOOCV.

| Accelerometry + ECG | | | | | | | | | | | | | | | | | | | |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----------------|-------------------|
| Participant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | # of VO Correct | # of Phys correct |
| PSC | V | V | P | V | P | P | V | P | P | P | P | V | P | P | P | P | P | | |
| First minute | 0.01 | 0.01 | 0.29 | 0.84 | 0.00 | 0.09 | 0.65 | 0.39 | 0.62 | 0.00 | 0.49 | 0.93 | 0.04 | 0.22 | 0.54 | 0.00 | 0.00 | 3/5 | 10/12 |
| Third minute | 0.00 | 0.56 | 0.03 | 0.85 | 0.00 | 0.10 | 0.30 | 0.73 | 0.17 | 0.00 | 0.35 | 0.63 | 0.23 | 0.39 | 0.11 | 0.02 | 0.00 | 3/5 | 11/12 |
| Final minute | 0.02 | 0.00 | 0.19 | 0.19 | 0.00 | 0.26 | 0.19 | 0.80 | 0.15 | 0.16 | 0.83 | 0.46 | 0.16 | 0.51 | 0.50 | 0.18 | 0.00 | 0/5 | 8/12 |
| Accelerometry Only | | | | | | | | | | | | | | | | | | | |
| Participant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | # of VO Correct | # of Phys correct |
| PSC | V | V | P | V | P | P | V | P | P | P | P | V | P | P | P | P | P | | |
| First minute | 0.01 | 0.08 | 0.30 | 0.89 | 0.00 | 0.06 | 0.36 | 0.36 | 0.22 | 0.00 | 0.91 | 0.70 | 0.01 | 0.52 | 0.54 | 0.00 | 0.00 | 2/5 | 9/12 |
| Third minute | 0.05 | 0.95 | 0.03 | 0.79 | 0.00 | 0.18 | 0.38 | 0.44 | 0.04 | 0.01 | 0.61 | 0.85 | 0.01 | 0.33 | 0.05 | 0.02 | 0.00 | 3/5 | 11/12 |
| Final minute | 0.07 | 0.01 | 0.21 | 0.07 | 0.00 | 0.10 | 0.11 | 0.88 | 0.34 | 0.08 | 0.88 | 0.51 | 0.41 | 0.28 | 0.54 | 0.87 | 0.00 | 1/5 | 8/12 |

Note: Values presented are predicted probability of left out participant belonging to vestibulo-ocular class. Predicted probabilities >0.5 indicated that vestibulo-ocular was more likely and <0.5 physiological was more likely. Shaded cells indicate correct predictions.

3.4. Post-Hoc Results

Post-hoc analysis of the magnitude of group differences between vestibulo-ocular and physiological PSC for the maximum PSD and MSE complexity index are presented in Figures 6 and 7, respectively. Medium-sized effects in MSE complexity index were observed between groups for vertical, lateral, and sagittal acceleration from the first minute, vertical and sagittal acceleration from the third minute, and sagittal acceleration from the final minute (Figure 6). While confidence intervals for differences in complexity index of these features each cross zero, they hint at potentially strong effects that warrant further investigation. Only trivial to small effects were observed between groups for maximum PSD from accelerometry data, except for sagittal acceleration from the final minute (Figure 7). Of note, only trivial to small effects were present between groups for ECG data for both time series features within all temporal slices. These small differences in PSD and MSE features of ECG data may explain why similar classification results were observed in the LOOCV experiment when only accelerometry data were provided to the CNN.

Maximum power spectral density

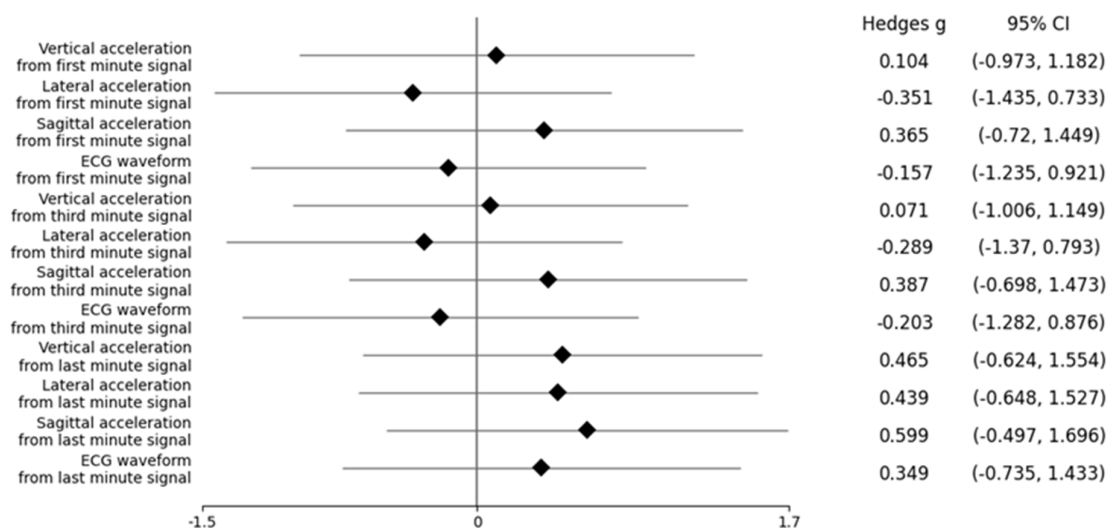


Figure 6. Hedge’s g effect sizes for group differences in maximum power spectral density between vestibulo-ocular PSC and physiological PSC.

Finally, Figure 8 depicts group differences in MSE complexity index for accelerometry data between vestibulo-ocular and physiological PSC using data from the third and final minute to elucidate why data from the former outperformed the latter. Similar complexity index profiles were observed between data from the third and final minute for lateral acceleration, sagittal acceleration, and ECG features. A notable group level reduction in

vertical acceleration complexity was detected in vestibulo-ocular PSC versus physiological PSC using data from the third minute, whereas vertical acceleration complexity was nearly completely overlapping between groups in data collected during the final minute. These differences in MSE complexity index may have contributed to why data from the third minute heavily outperformed data from the final minute in terms of classification performance.

Multiscale entropy complexity index

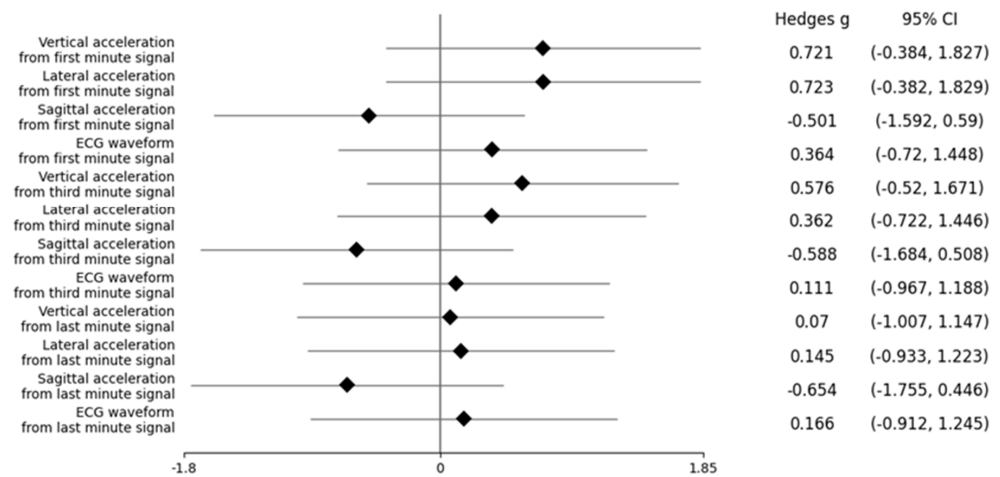
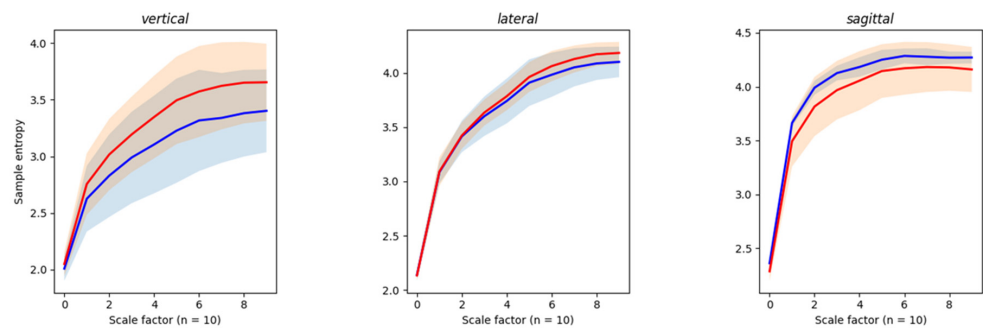
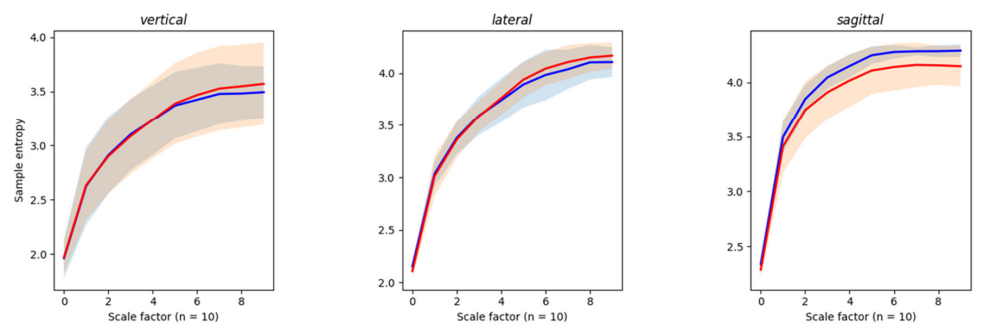


Figure 7. Hedge’s g effect sizes for group differences in multiscale entropy complexity index between vestibulo-ocular PSC and physiological PSC.

Accelerometer data from third minute of BCTT protocol



Accelerometer data from last minute of BCTT protocol



— Mean across vestibulo-ocular patients
 — Mean across physiological patients

Figure 8. Group differences in multiscale entropy complexity index profiles using data from the third and last minute of BCTT protocol. Note: The shaded areas represent the range of individual differences for participants within each PSC group.

4. Discussion

To our knowledge, this is the first study to explore whether a deep learning approach could accurately classify PSCs in mTBI patients using sensor data collected during BCTTs. Utilising a LOOCV approach, moderate levels of physiological versus vestibulo-ocular PSC classification accuracy were observed when a CNN was trained using accelerometry and ECG data acquired during a BCTT. The third minute demonstrated the best classification performance based on area under the ROC curve and total correct scores 14/17 for accelerometry + ECG and 14/17 for accelerometry only. In contrast, the final minute performed worse than chance. Poor performance in the final minute was likely because BCTT duration varied for participants based on symptom exacerbation/fatigue, thus demands experienced by each participant within this temporal slice were non-standardised. The third minute performed slightly better at classification than the first minute, which may be explained by greater physical load generating more symptomatic behaviour. It is worth noting that, while we observed an overall accuracy of 82% and a physiological PSC accuracy of 92% (11/12), the classification of vestibulo-ocular PSC was slightly better than chance at 60% (3/5). However, even with our small sample, our approach demonstrated a moderate PPV (75%) for identifying vestibulo-ocular PSC coupled with strong specificity (92%) and NPV (85%) to rule out vestibulo-ocular PSC. Overall, findings suggested that with further research, adding wearable sensors during clinical tests such as the BCTT, combined with deep learning models, may have clinical utility in assisting clinicians when classifying PSCs in mTBI patients.

Inflated machine learning accuracy is commonly observed with small samples [58], and this may have influenced our results, given our study's low sample size, imbalanced classes, and cross-validation approach. Our results are proof of concept that further work to determine the generalisability of this approach appears justifiable. Three key considerations should be accounted for in future investigations.

First, when optimal coupling between the BioHarness™ and the skin was achieved (as determined by clean ECG signals), a CNN trained with only accelerometry data demonstrated similar levels of classification accuracy as a CNN trained using both ECG and accelerometry data. This observation was corroborated by post-hoc analysis of PSD and MSE time series features, suggesting limited contribution of ECG data to classification. Some studies have shown reliable and valid measurement of static balance in neurologically compromised participants using accelerometers and gyroscopes within smartphones [59,60]. If future research extended our findings that accelerometry is sensitive to both physiological and vestibular-ocular groups, it could make clinical implementation of a wearable sensor during exertion testing easier and more affordable. A smartphone application could be developed to predict the most probable PSC based on data recorded using the internal accelerometer.

Second, due to the novelty of this study, we opted to focus on mTBI patients with the two PSCs (physiological and vestibulo-ocular) that were most prevalent in our previous research [45]. Theoretically, distinct differences in the degree of autonomic uncoupling and impaired sensorimotor integration could be present between these two clinical subgroups, and such differences would allow accurate classification of these groups as suggested by the study results. However, in clinical practice, an important subgroup is that of patients whose symptoms and clinical examination suggest that both physiological and vestibulo-ocular PSCs are present. Patients with mixed PSC require multi-modal treatment plans to restore autonomic function and sensorimotor integration. While current study results are encouraging, they do not account for this important subgroup. To have true clinical utility, an optimal model would be capable of separately classifying physiological, vestibulo-ocular, and mixed PSC with high accuracy. If such a model could be developed using this third class's wearable sensor data and expert labelling, it would provide a more objective tool for clinicians to utilise when establishing individualised treatment plans for their mTBI patients.

Third, better classification was observed using a combination of accelerometry + ECG data, as well as accelerometry data alone, from the third minute versus the first minute. Post-hoc analysis demonstrated similar PSD and MSE profiles for wearable sensor data from both temporal slices of data, indicating that other features and/or complex interactions between features may explain greater classification accuracy from the third minute data. Future work using more conventional statistical techniques and manual feature engineering could allow inferences that may lead to useful clinical applications. For example, identification of key features could be used to assess the effectiveness of interventions targeting underlying impairments related to each PSC by re-administering a BCTT at follow-up appointments to assess changes between successive timepoints. There are opportunities for parallel development of “black box” and “traditional”/explainable approaches to clinical decision support for mTBI. The former potentially provides decision support tools to augment clinical management of mTBI in short- to mid-terms, particularly when access to specialist clinicians may be restricted. However, in the long-term, the latter may unlock our ability to monitor recovery in real-world clinical environments. Overall, the study findings showcase that combining wearable sensors and machine learning algorithms under ecologically valid conditions has potential to advance clinical management of mTBI.

5. Limitations

This pilot study's sample size was a limitation, as only 17 participants had good ECG and accelerometry data. The most plausible explanation for the signal quality issue was suboptimal coupling between the BioHarness™ and the skin. While deployment of a BioHarness™ is very similar to that of a commercial heart rate monitor, it appears that additional time needs to be dedicated to ensure optimal coupling and resulting data quality. Another limitation was that we could not evaluate whether age or sex impacted our results, due to low sample size. These factors have been associated with differences in recovery outcomes following mTBI [61].

The decision to use the k-folds method of implementing LOOCV was due to limited sample size. A more common approach would have been to randomly assign 66% of participants (and their corresponding time series data) for training the model and to test model performance on the remaining participants. Utilising a LOOCV approach was an appropriate compromise, since our aim was to assess potential clinical utility of how a model trained on current data might perform if a new patient completed a BCTT wearing a BioHarness™, while making the most of the limited data available. There was an imbalance in the number of participants, with physiological ($n = 12$) and vestibular ($n = 5$), PSC in this pilot study. While this class imbalance is suboptimal for analysis, as it may contribute to overly optimistic classification accuracy, it is clinically representative of our previous clinical findings. Vestibulo-ocular PSC was considered as the positive class in this study, since it is less common and a predictor of worse recovery outcomes. However, physiological PSC was only labeled as the negative class since most participants belonged to this class. Even if results were biased towards identifying physiological PSC, this could still be viewed as clinically meaningful, as it could assist in management decisions. To account for this class imbalance, results depicted in Figure 5 could have been better visualised using a Precision/Recall curve rather than a ROC curve [62]. Because of the study aim, we chose to present ROC curves, since both classes are clinically meaningful and ROC curves are more familiar to clinicians [63]. To address this compromise, we also reported PPV and NPV, as PPV is associated with Precision/Recall [62], and these terms are commonly used in medicine [64]. Finally, these preliminary findings were from a sample of athletes who sustained mTBI, therefore the potential for these findings to be applicable to individuals who sustain mTBI due to non-sport-related causes is unknown.

6. Future Directions

Recruitment of a greater sample would overcome these limitations outlined above. Efforts to do so appear justified based on our pilot findings. A larger study should prioritize

recruiting males and females, ranging in age from children to adults who have sustained mTBI due to sport and non-sport-related causes. Additionally, patients presenting with mixed PSC should also be included. Such a sample would enable thorough inquiry into the generalizability of our pilot findings.

7. Conclusions

This is the first study to explore whether a deep learning approach could accurately classify PSCs in mTBI patients using sensor data collected during exertion testing. Our results provide proof of concept that incorporation of wearable sensors during BCTT and machine learning techniques have potential to assist in decision making for clinicians working with mTBI patients. More generally, this work outlines how deep learning can be leveraged on small datasets to explore whether features within available data *can* classify clinically relevant groups, before investing resources to acquire more data to understand *how* the model differentiates groups.

Author Contributions: Conceptualization, J.P.M., M.P., P.A.H., A.T., S.K. and B.R.; Methodology, J.P.M., M.P., P.A.H., A.T., S.K. and B.R.; Software, B.R.; Formal analysis, J.P.M., M.P. and B.R.; Data curation, J.P.M. and S.K.; Writing—original draft, J.P.M.; Writing—review & editing, M.P., P.A.H., A.T., S.K. and B.R.; Visualization, J.P.M., M.P. and B.R.; Supervision, P.A.H., A.T. and B.R.; Project administration, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received internal funding from the Auckland University of Technology Vice Chancellor’s PhD Scholarship.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Auckland University of Technology Ethics Committee (AUTEK 18/374) and Health and Disability Ethics Committee (HDEC 18/NTA/108).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The participants of this study did not provide consent for their data to be shared publicly.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Romeu-Mejia, R.; Giza, C.C.; Goldman, J.T. Concussion Pathophysiology and Injury Biomechanics. *Curr. Rev. Musculoskelet. Med.* **2019**, *12*, 105–116. [[CrossRef](#)] [[PubMed](#)]
2. Feigin, V.L.; Theadom, A.; Barker-Collo, S.; Starkey, N.J.; McPherson, K.; Kahan, M.; Dowell, A.; Brown, P.; Parag, V.; Kydd, R.; et al. Incidence of traumatic brain injury in New Zealand: A population-based study. *Lancet Neurol.* **2013**, *12*, 53–64. [[CrossRef](#)] [[PubMed](#)]
3. Douglas, D.B.; Ro, T.; Toffoli, T.; Krawchuk, B.; Muldermans, J.; Gullo, J.; Dulberger, A.; Anderson, A.E.; Douglas, P.K.; Wintermark, M. Neuroimaging of Traumatic Brain Injury. *Med. Sci.* **2018**, *7*, 2. [[CrossRef](#)] [[PubMed](#)]
4. Jain, S.; Iverson, L. *Glasgow Coma Scale*; StatPearls Publishing: Treasure Island, FL, USA, 2020.
5. Sussman, E.S.; Pendharkar, A.V.; Ho, A.L.; Ghajar, J. Mild traumatic brain injury and concussion: Terminology and classification. *Handb. Clin. Neurol.* **2018**, *158*, 21–24.
6. Theadom, A.; Starkey, N.J.; Dowell, T.; Hume, P.A.; Kahan, M.; McPherson, K.; Feigin, V.; Group, B.R. Sports-related brain injury in the general population: An epidemiological study. *J. Sci. Med. Sport* **2014**, *17*, 591–596. [[CrossRef](#)]
7. McCrory, P.; Meeuwisse, W.H.; Aubry, M.; Cantu, B.; Dvorak, J.; Echemendia, R.J.; Engebretsen, L.; Johnston, K.; Kutcher, J.S.; Raftery, M.; et al. Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *Br. J. Sport. Med.* **2013**, *47*, 250–258. [[CrossRef](#)]
8. McCrory, P.; Meeuwisse, W.; Dvorak, J.; Aubry, M.; Bailes, J.; Broglio, S.; Cantu, R.C.; Cassidy, D.; Echemendia, R.J.; Castellani, R.J.; et al. Consensus statement on concussion in sport—the 5(th) international conference on concussion in sport held in Berlin, October 2016. *Br. J. Sport. Med.* **2017**, *51*, 838–847.
9. Kara, S.; Crosswell, H.; Forch, K.; Cavadino, A.; McGeown, J.; Fulcher, M. Less than Half of Patients Recover within 2 Weeks of Injury after a Sports-Related Mild Traumatic Brain Injury: A 2-Year Prospective Study. *Clin. J. Sport. Med.* **2020**, *30*, 96–101. [[CrossRef](#)]
10. Theadom, A.; Parag, V.; Dowell, T.; McPherson, K.; Starkey, N.; Barker-Collo, S.; Jones, K.; Ameratunga, S.; Feigin, V.L.; BIONIC Research Group. Persistent problems 1 year after mild traumatic brain injury: A longitudinal population study in New Zealand. *Br. J. Gen. Pract.* **2016**, *66*, e16–e23. [[CrossRef](#)]

11. Voormolen, D.C.; Zeldovich, M.; Haagsma, J.A.; Polinder, S.; Friedrich, S.; Maas, A.I.R.; Wilson, L.; Steyerberg, E.W.; Covic, A.; Andelic, N.; et al. Outcomes after Complicated and Uncomplicated Mild Traumatic Brain Injury at Three-and Six-Months Post-Injury: Results from the CENTER-TBI Study. *J. Clin. Med.* **2020**, *9*, 1525. [[CrossRef](#)]
12. Nelson, L.D.; Temkin, N.R.; Dikmen, S.; Barber, J.; Giacino, J.T.; Yuh, E.; Levin, H.S.; McCrea, M.A.; Stein, M.B.; Mukherjee, P.; et al. Recovery After Mild Traumatic Brain Injury in Patients Presenting to US Level I Trauma Centers: A Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) Study. *JAMA Neurol.* **2019**, *76*, 1049–1059. [[CrossRef](#)] [[PubMed](#)]
13. Ellis, M.J.; Leddy, J.J.; Willer, B. Physiological, vestibulo-ocular and cervicogenic post-concussion disorders: An evidence-based classification system with directions for treatment. *Brain Inj.* **2015**, *29*, 238–248. [[CrossRef](#)] [[PubMed](#)]
14. Leddy, J.J. Buffalo Concussion Test—Instruction Manual. Available online: <https://ubortho.com/wp-content/uploads/2020/10/Buffer-Concussion-Treadmill-Test-Manual.pdf> (accessed on 10 May 2020).
15. Leddy, J.J.; Baker, J.G.; Merchant, A.; Picano, J.; Gaile, D.; Matuszak, J.; Willer, B. Brain or strain? Symptoms alone do not distinguish physiologic concussion from cervical/vestibular injury. *Clin. J. Sport Med.* **2015**, *25*, 237–242. [[CrossRef](#)]
16. Leddy, J.J.; Baker, J.G.; Willer, B. Active rehabilitation of concussion and post-concussion syndrome. *Phys. Med. Rehabil. Clin. N. Am.* **2016**, *27*, 437–454. [[CrossRef](#)] [[PubMed](#)]
17. Leddy, J.; Baker, J.G.; Haider, M.N.; Hinds, A.; Willer, B. A physiological approach to prolonged recovery from sport-related concussion. *J. Athl. Train.* **2017**, *52*, 299–308. [[CrossRef](#)]
18. Ellis, M.J.; Leddy, J.; Willer, B. Multi-disciplinary management of athletes with post-concussion syndrome: An evolving pathophysiological approach. *Front. Neurol.* **2016**, *7*, 136. [[CrossRef](#)]
19. Leddy, J.J.; Baker, J.G.; Kozlowski, K.; Bisson, L.; Willer, B. Reliability of a graded exercise test for assessing recovery from concussion. *Clin. J. Sport Med.* **2011**, *21*, 89–94. [[CrossRef](#)]
20. Leddy, J.J.; Hinds, A.L.; Miecznikowski, J.; Darling, S.; Matuszak, J.; Baker, J.G.; Picano, J.; Willer, B. Safety and Prognostic Utility of Provocative Exercise Testing in Acutely Concussed Adolescents: A Randomized Trial. *Clin. J. Sport Med. Off. J. Can. Acad. Sport. Med.* **2018**, *28*, 13–20. [[CrossRef](#)]
21. Leddy, J.J.; Kozlowski, K.; Donnelly, J.P.; Pendergast, D.R.; Epstein, L.H.; Willer, B. A preliminary study of subsymptom threshold exercise training for refractory post-concussion syndrome. *Clin. J. Sport. Med.* **2010**, *20*, 21–27. [[CrossRef](#)]
22. Kenzie, E.S.; Parks, E.L.; Bigler, E.D.; Wright, D.W.; Lim, M.M.; Chesnutt, J.C.; Hawryluk, G.W.J.; Gordon, W.; Wakeland, W. The dynamics of concussion: Mapping pathophysiology, persistence, and recovery with causal-loop diagramming. *Front. Neurol.* **2018**, *9*, 203. [[CrossRef](#)]
23. Gall, B.; Parkhouse, W.; Goodman, D. Heart rate variability of recently concussed athletes at rest and exercise. *Med. Sci. Sport. Exerc.* **2004**, *36*, 1269–1274. [[CrossRef](#)] [[PubMed](#)]
24. Bishop, S.A.; Dech, R.; Baker, T.; Butz, M.; Aravinthan, K.; Neary, J.P. Parasympathetic baroreflexes and heart rate variability during acute stage of sport concussion recovery. *Brain Inj.* **2017**, *31*, 247–259. [[CrossRef](#)] [[PubMed](#)]
25. Akin, F.W.; Murnane, O.D.; Hall, C.D.; Riska, K.M. Vestibular consequences of mild traumatic brain injury and blast exposure: A review. *Brain Inj.* **2017**, *31*, 1188–1194. [[CrossRef](#)] [[PubMed](#)]
26. Franke, L.M.; Walker, W.C.; Cifu, D.X.; Ochs, A.L.; Lew, H.L. Sensorintegrative dysfunction underlying vestibular disorders after traumatic brain injury: A review. *J. Rehabil. Res. Dev.* **2012**, *49*, 985–993. [[CrossRef](#)] [[PubMed](#)]
27. Doherty, C.; Zhao, L.; Ryan, J.; Komaba, Y.; Inomata, A.; Caulfield, B. Quantification of postural control deficits in patients with recent concussion: An inertial-sensor based approach. *Clin. Biomech.* **2017**, *42*, 79–84. [[CrossRef](#)]
28. Doherty, C.; Zhao, L.; Ryan, J.; Komaba, Y.; Inomata, A.; Caulfield, B. Concussion is associated with altered preparatory postural adjustments during gait initiation. *Hum. Mov. Sci.* **2017**, *52*, 160–169. [[CrossRef](#)] [[PubMed](#)]
29. Slobounov, S.; Sebastianelli, W.; Hallett, M. Residual brain dysfunction observed one year post-mild traumatic brain injury: Combined EEG and balance study. *Clin. Neurophysiol.* **2012**, *123*, 1755–1761. [[CrossRef](#)]
30. McCrea, M.; Hammeke, T.; Olsen, G.; Leo, P.; Guskiewicz, K. Unreported concussion in high school football players: Implications for prevention. *Clin. J. Sport Med.* **2004**, *14*, 13–17. [[CrossRef](#)]
31. Delaney, J.S.; Lacroix, V.J.; Leclerc, S.; Johnston, K.M. Concussions among university football and soccer players. *Clin. J. Sport. Med.* **2002**, *12*, 331–338. [[CrossRef](#)]
32. Meier, T.B.; Brummel, B.J.; Singh, R.; Nerio, C.J.; Polanski, D.W.; Bellgowan, P.S. The underreporting of self-reported symptoms following sports-related concussion. *J. Sci. Med. Sport.* **2015**, *18*, 507–511. [[CrossRef](#)]
33. Asken, B.M.; McCrea, M.A.; Clugston, J.R.; Snyder, A.R.; Houck, Z.M.; Bauer, R.M. “Playing Through It”: Delayed Reporting and Removal From Athletic Activity After Concussion Predicts Prolonged Recovery. *J. Athl. Train.* **2016**, *51*, 329–335. [[CrossRef](#)] [[PubMed](#)]
34. Galbraith, H.; Quesnele, J.; Kenrick-Rochon, S.; Grenier, S.; Baldisera, T. What are the knowledge, attitudes and beliefs regarding concussion of primary care physicians and family resident physicians in rural communities? *J. Concussion* **2020**, *4*, 2059700220974548. [[CrossRef](#)]
35. Yorke, A.M.; Littleton, S.; Alsalaheen, B.A. Concussion Attitudes and Beliefs, Knowledge, and Clinical Practice: Survey of Physical Therapists. *Phys. Ther.* **2016**, *96*, 1018–1028. [[CrossRef](#)] [[PubMed](#)]
36. Reid, D.; Stuart, C.; Fulcher, M.; Hume, A.P.; Theadom, A.; Whatman, C.; Walters, S. *Knowledge and Attitudes (KA) Surveys on Concussion in Sports: Doctor September 2017 Survey. Report #4 to Accident Compensation Corporation; SPRINZ, Auckland University of Technology: Auckland, New Zealand, 2018; p. 15.*

37. Russell, B.; McDaid, A.; Toscano, W.; Hume, P. Moving the Lab into the Mountains: A Pilot Study of Human Activity Recognition in Unstructured Environments. *Sensors* **2021**, *21*, 654. [[CrossRef](#)] [[PubMed](#)]
38. Sidey-Gibbons, J.A.M.; Sidey-Gibbons, C.J. Machine learning in medicine: A practical introduction. *BMC Med. Res. Methodol.* **2019**, *19*, 64. [[CrossRef](#)] [[PubMed](#)]
39. Ignatov, A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [[CrossRef](#)]
40. Clouthier, A.L.; Ross, G.B.; Graham, R.B. Sensor Data Required for Automatic Recognition of Athletic Tasks Using Deep Neural Networks. *Front. Bioeng. Biotechnol.* **2020**, *7*, 473. [[CrossRef](#)]
41. Attia, Z.I.; Kapa, S.; Lopez-Jimenez, F.; McKie, P.M.; Ladewig, D.J.; Satam, G.; Pellikka, P.A.; Enriquez-Sarano, M.; Noseworthy, P.A.; Munger, T.M.; et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat. Med.* **2019**, *25*, 70–74. [[CrossRef](#)]
42. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nat. Med.* **2019**, *25*, 24–29. [[CrossRef](#)]
43. Haider, M.N.; Leddy, J.J.; Wilber, C.G.; Viera, K.B.; Bezherano, I.; Wilkins, K.J.; Miecznikowski, J.C.; Willer, B.S. The Predictive Capacity of the Buffalo Concussion Treadmill Test After Sport-Related Concussion in Adolescents. *Front. Neurol.* **2019**, *10*, 339. [[CrossRef](#)]
44. Borg, G. *Borg's Perceived Exertion and Pain Scales*; Human Kinetics: Champaign, IL, USA, 1998.
45. McGeown, J.P.; Kara, S.; Fulcher, M.; Crosswell, H.; Borotkanics, R.; Hume, P.A.; Quarrie, K.L.; Theadom, A. Predicting Sport-related mTBI Symptom Resolution Trajectory Using Initial Clinical Assessment Findings: A Retrospective Cohort Study. *Sport. Med.* **2019**, *50*, 1191–1202. [[CrossRef](#)] [[PubMed](#)]
46. Johnstone, J.A.; Ford, P.A.; Hughes, G.; Watson, T.; Garrett, A.T. Bioharness multivariable monitoring device: Part. II: Reliability. *J. Sport. Sci. Med.* **2012**, *11*, 409–417.
47. Johnstone, J.A.; Ford, P.A.; Hughes, G.; Watson, T.; Garrett, A.T. Bioharness multivariable monitoring device: Part. I: Validity. *J. Sport. Sci. Med.* **2012**, *11*, 400–408.
48. Nazari, G.; Bobos, P.; MacDermid, J.C.; Sinden, K.E.; Richardson, J.; Tang, A. Psychometric properties of the Zephyr bioharness device: A systematic review. *BMC Sport. Sci. Med. Rehabil.* **2018**, *10*, 6. [[CrossRef](#)]
49. Cust, E.E.; Sweeting, A.J.; Ball, K.; Robertson, S. Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance. *J. Sport. Sci.* **2019**, *37*, 568–600. [[CrossRef](#)]
50. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [[CrossRef](#)]
51. Pérez-Zapata, A.F.; Cardona-Escobar, A.F.; Jaramillo-Garzón, J.A.; Díaz, G.M. Deep Convolutional Neural Networks and Power Spectral Density Features for Motor Imagery Classification of EEG Signals. In *Augmented Cognition: Intelligent Technologies*; Springer International Publishing: Cham, Switzerland, 2018; pp. 158–169.
52. Park, J.-S.; Seo, J.; Cha, H.; Song, H.-J.; Lee, S.-H.; Jang, K.E.; Lee, H.J.; Park, J.; Lee, H.-W.; Chang, Y. Altered power spectral density in the resting-state sensorimotor network in patients with myotonic dystrophy type 1. *Sci. Rep.* **2018**, *8*, 987. [[CrossRef](#)]
53. Luft, F.; Sharifi, S.; Mugge, W.; Schouten, A.C.; Bour, L.J.; van Rootselaar, A.F.; Veltink, P.H.; Heida, T. A Power Spectral Density-Based Method to Detect Tremor and Tremor Intermittency in Movement Disorders. *Sensors* **2019**, *19*, 4301. [[CrossRef](#)]
54. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [[CrossRef](#)]
55. Costa, M.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Phys. A Stat. Mech. Its Appl.* **2003**, *330*, 53–60. [[CrossRef](#)]
56. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2005**, *71*, 021906. [[CrossRef](#)] [[PubMed](#)]
57. Wu, C.-H.; Lee, C.-H.; Jiang, B.C.; Sun, T.-L. Multiscale Entropy Analysis of Postural Stability for Estimating Fall Risk via Domain Knowledge of Timed-Up-And-Go Accelerometer Data for Elderly People Living in a Community. *Entropy* **2019**, *21*, 1076. [[CrossRef](#)]
58. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [[CrossRef](#)] [[PubMed](#)]
59. Kis, M. Reliability of a New Test of Balance Function in Healthy and Concussion Populations. *J. Funct. Morphol. Kinesiol.* **2020**, *5*, 13. [[CrossRef](#)] [[PubMed](#)]
60. Hou, Y.-R.; Chiu, Y.-L.; Chiang, S.-L.; Chen, H.-Y.; Sung, W.-H. Development of a Smartphone-Based Balance Assessment System for Subjects with Stroke. *Sensors* **2019**, *20*, 88. [[CrossRef](#)]
61. Iverson, G.L.; Gardner, A.J.; Terry, D.P.; Ponsford, J.L.; Sills, A.K.; Broshek, D.K.; Solomon, G.S. Predictors of clinical recovery from concussion: A systematic review. *Br. J. Sport. Med.* **2017**, *51*, 941–948. [[CrossRef](#)]
62. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)]

63. Hajian-Tilaki, K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Casp. J. Intern. Med.* **2013**, *4*, 627–635.
64. Trevethan, R. Sensitivity, Specificity, and Predictive Values: Foundations, Plabilities, and Pitfalls in Research and Practice. *Front. Public. Health* **2017**, *5*, 307. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.