Seizure Classification Using Person-Specific Triggers

Jamie Pordoy[1] [0000−0002−3417−7862], Ying Zhang[2] [0000−0002−6669−1671], Nasser Matoorian[3], and Massoud Zolgharni[4] [0000-0003-0904-2904]

[1]jamiepordoy@hotmail.com; [2]ying.zhang@uwl.ac.uk; [3]nasser.matoorian@uwl.ac.uk; [4]massoud.zolgharni@uwl.ac.uk
School of Computing and Engineering, University of West London, London W5 5RF, UK

Abstract. With advancements in personalised medicine, healthcare delivery systems have moved away from the one-size-fits-all approach, towards tailored treatments that meet the needs of individuals and specific subgroups. As nearly one-third of those diagnosed with epilepsy are classed as refractory and are resistant to antiepileptic medication, there is need for a personalised method of detecting epileptic seizures. Epidemiological studies show that up to 91% of those diagnosed identify one or more triggers as the causation of their seizure onset. These triggers are person-specific and can affect those diagnosed in different ways dependent on each person's idiosyncratic tolerance and threshold levels. Whilst these triggers are known to induce seizure onset, only a few studies have even considered their use as a preventive component. Therefore, this pilot study investigates the use of person-specific triggers (PST) in diagnosed epileptics, and whether they can be used as an additional modality when detecting seizures. This study uses a precision medicine approach with artificial intelligence (AI), to train and test several classification algorithms that can predict an epileptic seizure using the PST of each participant. Experimental results show accuracy, sensitivity, and specificity scores of 94.73%, 96.90% and 93.33% for participant 1 and 96.87%, 96.96% and 96.77% for participant 2, respectively.

Keywords: Multi-modal · Seizure · Person-specific · Classification · Epilepsy

1 Introduction

Epilepsy is a prevalent neurological condition that affects an estimated 70 million people worldwide [1]. An overload of electrical activity between communicating neurons causes a temporal imbalance of neurological activity, culminating in the occurrence of an unprovoked seizure, often leaving an individual with a loss of anatomical motor functions and clarity of memory [2]. An estimated 30% of those diagnosed are classed as refractory and are resistant to anti-epileptic drugs (AEDs) [3]. Those who are resistant have no form of defence and are at a higher risk of triggering a convulsive seizure which can lead to an acute cardiac and respiratory dysfunction [4].

A sudden unexpected death in epilepsy (SUDEP) is the most frequent direct cause of epilepsy-related deaths, predominately affecting those who are resistant or have

poorly controlled chronic epilepsy. A study by Lambert et al. [5] identified 58% of SUDEP cases are nocturnal and occur once an individual has been asleep and experienced a generalised tonic-clonic (GTC) seizure. As the underlying cause of SUDEP remains unknown and without treatment at a therapeutic level, recent case studies have suggested that onset, and in turn SUDEP, could be triggered by several predisposed risk and trigger factors [6].

As observed by Hesdorffer *et al.* [7], the most significant risk factor is an increase in the frequency of GTC seizures, as this can lead to a cardiac and respiratory dysfunction. Patients with epilepsy (PWE) who experience ≥ 3 GTC seizures per year are 15 times more likely to have a fatal epilepsy-related event such as SUDEP. Other frequent risk factors include partial seizures, missing doses of AEDs and an intelligence quotient (IQ) < 70 [8].

In addition, it is important to ask whether there are specific triggers (precipitants) that increase the probability of onset and whether these could be used in conjunction with the aforementioned risk factors to improve on current methods of seizure detection [9]. Although seizures are sporadic and seemingly random, studies show there are person-specific triggers (PST) that increase the likelihood of onset [10]. As defined by Aird and Gordon [11], PST can be categorised as seizure inducing or seizure triggering events. Those classed as seizure inducing (lights, noises, and patterns), are caused by environmental or endogenous events and cause a transient lowering of the seizure threshold level [12]. Seizure triggering events (sleep deprivation, stress, and fatigue) are risk factors that vary based on each person's specific threshold and tolerance levels. A study by Ferlisi and Shorvon [13] interviewed 104 patients to identify the frequency of seizure precipitants (triggering factors).
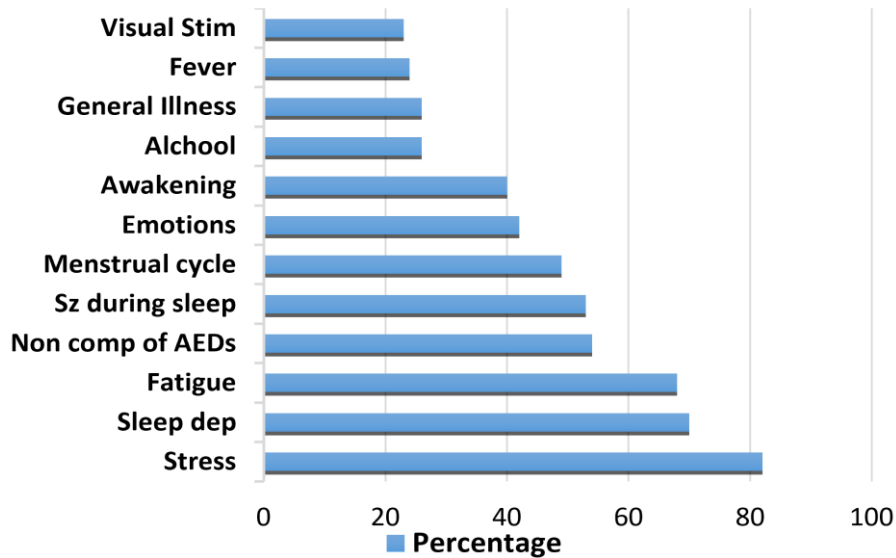
Fig. 1. Percentage of trigger factors observed in n-participants (n = 104)

Results show that seizure triggering events are more frequent, with an estimated 91% of participants experiencing one or more triggers prior to a seizure [13]. The distribution of these triggers is illustrated in Fig. 1, with stress, sleep deprivation, fatigue, and a non-compliance of AEDs as the four most common causes of seizure onset with percentage scores of 82%, 70%, 68% and 54% respectively.

These results reflect the findings of Nakken *et al.* [14], who also identified stress, sleep deprivation and fatigue as the most frequent triggers, with 592 (51%) PWE listing at least one trigger as the causation of their onset [14]. Furthermore, a study by Balamurugan *et al.* [15], analysed 405 PWE, observing that 86.9% experience at least one trigger prior to onset. These results show a noncompliance of AEDs as the most frequent trigger (40.98%), followed by stress (31.35%), sleep deprivation (19.75%) and fatigue (15.30%).

1.1   Person-specific Triggers

Precision medicine is a newly adapted paradigm of healthcare that allows medical treatments to be idiosyncratically tailored towards the needs of individuals or specific subgroups [16]. Whilst precision medicine in epilepsy is still relatively unexplored, a recent study by Porumb *et al.* [17], combined precision medicine and machine learning for hypoglycemic event detection from ECG wavelets. By applying a person-specific approach, a classification model was trained using data from a single participant, which was then tested on unseen data from the same participant. The results demonstrate the potential application of person-specific classification, with models attaining an accuracy measure of 84.8%, 88.5%, 89.9% and 78.3% for participants 1-4 respectively.

Similarly, a study by Ince *et al.* [18], has explored person-specific classification of cardiac cycles to detect ventricular ectopic beats (VEBs) and supra-VEBs (SVEBs). Using fully connected feed-forward neural networks that were optimally designed for each person's idiosyncrasies, this study's person-specific classification models have surpassed many state-of-the-art algorithms with accuracy and sensitivity scores of 98.3% – 84.6% and 97.4% – 63.5% for VEBs and SVEBS respectively.

Given the successful detection of hypoglycemic anomalies and the use of person-specific classification in other medical fields, we believe that this method of detection can be used to classify the PST observed in diagnosed epileptics prior to seizure onset. If we can observe a noticeable correlation between these triggers, and a seizures onset, this could present a new untested modality that could work in conjunction with other methods of detection.

2   Methodology

This paper presents a preliminary pilot study that investigates the practical application of PST when classifying epileptic seizures. We believe this investigation is supported by the notion that PST are preceding events that are responsible for initiating or precipitating a seizure [14]. Due to the impact of the current global crisis (Covid-19), we cannot conduct a full-scale clinical trial, which in turn has reduced the size of our

dataset. However, as this is a pilot study that focuses on person-specific classification, we decided to proceed using the participants we had available.

Classification models were developed using the Python programming language and coded in Jupyter notebook. The Python libraries used for the development include TensorFlow, Keras and Scikit-learn.

This pilot study will be used as a preliminary component to test the validity of PST and deem whether further research with a full-scale clinical trial is feasible. As far as we know, this is a novel concept and there is no existing research that has used these specific triggers for person-specific classification of epileptic seizures.

2.1   Data Acquisition

For this study we collected data from two participants with epilepsy, one female (participant 1) in her late twenties and one male (participant 2) in his mid-thirties. Of the 300 days participant 1 was observed, we documented 17 positive instances (seizures), whilst participant 2 was observed for a shorter duration of 248 days, of which 22 positive instances were recorded.

Based on Ferlisi and Shorvon's findings [13], we decided to record the trigger factors of sleep deprivation, stress, and fatigue.

Whilst it has not been observed as a commonly occurring trigger, we have decided to record exercise as an additional PST, as we can see a clear correlation with stress reduction, which in turn could help to prevent or reduce the frequency of seizure onset [19].

2.2   Pilot Study workflow

Fig. 2 shows a high-level diagrammatic representation of this study's methodology, which consists of four phases, data collection, data preparation, machine learning and finally a comparative analysis. Both participants were issued with an IoT-enabled smart watch (Fitbit Versa) and smart phone (Samsung SM-N950P) for the duration of the study. We deployed an Android based CRUD application with pre-installed quantitative questionnaires on each smart phone, which was then used to record daily measures of stress and fatigue. We used a Fitbit Versa to calculate the metabolic equivalents (METs) of participants, as it provided an accurate measure for the number of minutes that active movement occurred [20]. We also used the embedded accelerometer and heart rate sensor to measure fluctuations in heart rate variability (HRV) and body movement to estimate the quality and duration of a participants sleep cycle [21].

The following sections briefly describe the components that were used to measure the PST and the machine learning models used for person-specific classification.
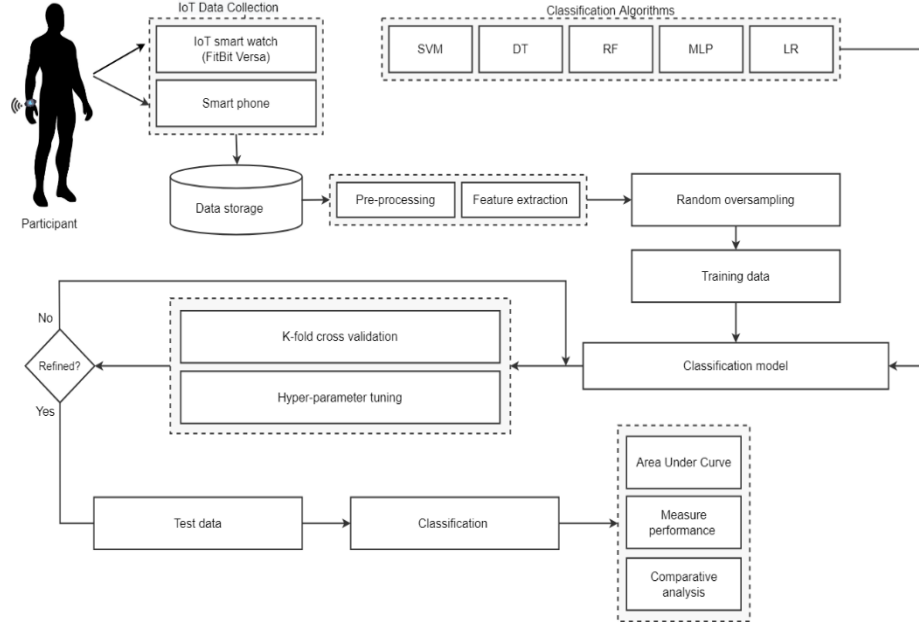
Fig. 2. Pilot study workflow; SVM = support vector machine; DT = decision tree; RF = random forest; MLP = multi-layer perceptron; LR = logistical regression; IoT = Internet of Things.

2.3 Resampling the Dataset

Accurately classifying rare events can be problematic as the frequency of occurrence can leave a dataset highly imbalanced. For this study, 548 instances were recorded, 509 negative instances where no seizure was documented and 39 positive instances. To account for the varying disparity between positive and negative instances, we used random over-sampling. This sampling technique balanced our datasets distribution as duplicate instances from the minority class were added to our training dataset [22].

2.4 Perceived Stress Scale

Perceived Stress Scale (PSS) is a widely used stress assessment tool that measures a subject's stress levels [23]. Although the PSS is generally calculated at the end of each month, for this study, we used a modified variant that measures stress daily. As shown in Fig.3, the PSS is comprised of 10 questions that require a numerical response from 0 to 4, where $0 := $ no stress and $4 := $ maximum stress. To calculate the PSS score, the value responses for questions 4, 5, 7 and 8 are reversed so that question $0 \leftrightarrow 5$, $1 \leftrightarrow 4$, $2 \leftrightarrow 3$, $3 \leftrightarrow 2$, $4 \leftrightarrow 1$ [24].

The values for each question are then combined and divided by 10. The PSS scale also accounts for perception, as two participants who encounter the same set of events can accumulate different scores.

0 = Never, 1 = Almost Never, 2 = Sometimes, 3 = Fairly Often, 4 = Very Often					
1	In the last day, how often have you been upset because of something that happened unexpectedly?	0 1 2 3 4
2	In the last day, how often have you felt that you were unable to control the important things in your life?	0 1 2 3 4
3	In the last day, how often have you felt nervous and "stressed"?	0 1 2 3 4
4	In the last day, how often have you felt confident about your ability to handle your personal problem?	0 1 2 3 4
5	In the last day, how often have you felt that things were going your way?	0 1 2 3 4
6	In the last day, how often have you found that you could not cope with all the things that you had to do?	0 1 2 3 4
7	In the last day, how often have you been able to control irritations in your life?	0 1 2 3 4
8	In the last day, how often have you felt that you were on top of things?	0 1 2 3 4
9	In the last day, how often have you been angered because of things that were outside of your control?	0 1 2 3 4
10	In the last day, how often have you felt difficulties were piling up so high that you could not overcome them?	0 1 2 3 4

Fig. 3. A modified variation of the PSS

2.5 Rating of Fatigue

The Rating of Fatigue (RoF) is a measurement tool that tracks the intensity of fatigue. The RoF uses a linear scale constructed of 11 numerical intervals (0-10) to assess the level of fatigue in patients, with 0 representing no fatigue and 10 representing total fatigue and exhaustion [25]. The RoF scale was deployed on the participants' smart phones, enabling a quantification of their fatigue levels at the start of each day.

2.6 Classification Algorithms and Techniques

This section provides a summary of the supervised learning classification models and techniques used for this study's comparative analysis. Classification is an instance of supervised learning, where a classification model (classifier) observes a set of input features and makes a prediction on unseen data that shares the same features. Classification models predict binomial outcomes such as yes/no, true/false and positive/negative to categorise new observations.

2.6.1   *K-Fold Cross Validation*

K-fold cross validation segments a dataset into k-subsets of approximately equal size, and in turn each classification model is then trained using k-1 subsets where the remaining subset is used to validate the classifiers performance on unseen data [26]. This process is then repeated for k-iterations, and each iteration uses a different subset for validation, whilst the previous validation subset becomes a training subset of k-1. For this study, we used a k-fold cross validation technique where k=10, to evaluate each classifiers performance on unseen data.

2.6.2   *Naive Bayes*

Stemming from Bayes theorem of probability, naive Bayes (NB) classifiers are a family of probabilistic classification algorithms which assume that each predictor is an independent entity that equally contributes to the outcome of the target class [27]. NB classifiers can calculate the posterior probability of event A, given the occurrence of event B [28]. This can be expressed mathematically as

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

(1)

Where $P(c|x)$ represents the posterior probability, $P(c)$ the prior probability of class c, $P(x)$ the probability of each predictor and $P(x|c)$ the probability of a predictor given the occurrence of class c [29]. As NB assumes that features are independent, only the variances of each training label need to be calculated, instead of the entirety of the covariance matrix. This enables a NB classifier to use a small quantity of training data when predicting the mean and variance for each predictor, which is ideal for this study's person-specific approach.

2.6.3   *Support Vector Machines*

Support vector machines (SVM) are supervised learning models used to increase the predictive accuracy for classification and regression analysis. SVM classifies data by finding the optimum position for a hyperplane in n-dimensional space, where n represents the number of features, and each feature has a specific co-ordinate value. The features closest to the hyperplane act as support vectors and are used to determine the orientation of the hyperplane. This enables the hyperplane to separate n-classes of training data by the maximum distance, leading to maximal generalisation and improved performance [30].

2.6.4   *Logistic Regression*

Logistic Regression (LR) is a statistical, supervised learning model used to calculate binary and binomial response data. LR computes the probabilistic relationship between a dependent, dichotomous variable (0, 1) and one or more independent predictors (variables).

$$z = b_0 + b_1 \cdot x_1 + \cdots b_n \cdot x_n \tag{2}$$

A sigmoid function transforms a linear equation into a logistic equation that converts any input values between negative and positive infinity to a value between 0 and 1 [31]. The equation for LR is derived from a generalised linear equation of independent predictors where e represents the natural logarithm base, b and x are the classifiers parameters, and P is the probability of 1 [32].

$$P = \frac{1}{1+e^{-(b_0+b_1 \cdot x_1 + \cdots b_n \cdot x_n)}} \tag{3}$$

2.6.5   *Decision Tree with Gini Index*

Decision Trees (DT) are hierarchical decision analysis structures that use a series of interconnected nodes to classify a decision and its consequences. A generalised tree structure consists of single root node, a series of decision nodes and leaf nodes. Deci-decision nodes represent the consequences of an action and have two or more branches that connect to the leaf nodes, whereas the leaf nodes represent the final classification decision of that action. For this study we used the Gini Index to measure impurity, as it enables the most relevant decision nodes to be closer to the root node. As the tree structure traverses downwards, the level of uncertainty surrounding each decision decreases, ensuring a more accurate method of classification [33]. The Gini index is calculated using the following formula where the sum of the squared probabilities for each class is deducted from 1 [34].

$$Gini = 1 - \sum_{i=1}^{c}(p_i)^2 \tag{4}$$

2.6.6   *Random Forest.*

Random Forest (RF) is an ensemble classifier that trains multiple decision trees in parallel for increased measures of performance [35]. RF classifiers use a combination of Breiman's "bagging" and random feature selection to form a process called majority voting where each tree classifies input data to identify the most frequently occur-ring class (prediction). This method of classification exhibits good generalisation, and often outperforms other classification models when measuring accuracy [36]. RF classifiers can be expressed mathematically as

$$\hat{f} = \frac{1}{B} = \sum_{b=1}^{B} \hat{f}_b(x') \tag{5}$$

Where $\hat{f}$ represents the final prediction, B the number of trees used, b the current tree and x the training sample used to teach the classifier.
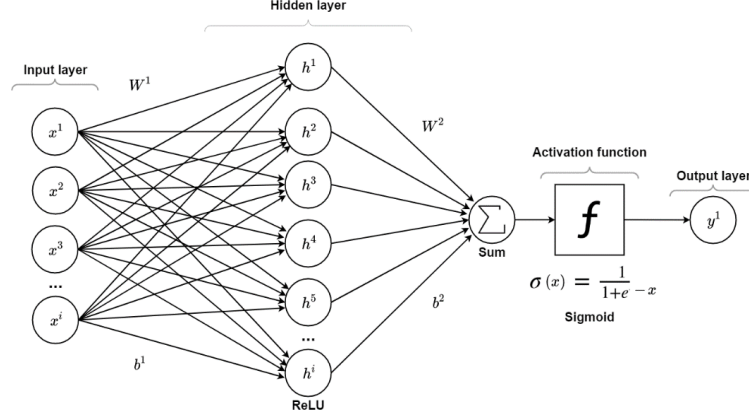
Fig. 4. Multi-layer perceptron architecture; x_i (Input layer) = 6 nodes; h_i (Hidden layer) = 12 nodes; y_i (Output layer) = 1 node; b_i = bias; w_i = weights; ReLU = Rectified linear unit;
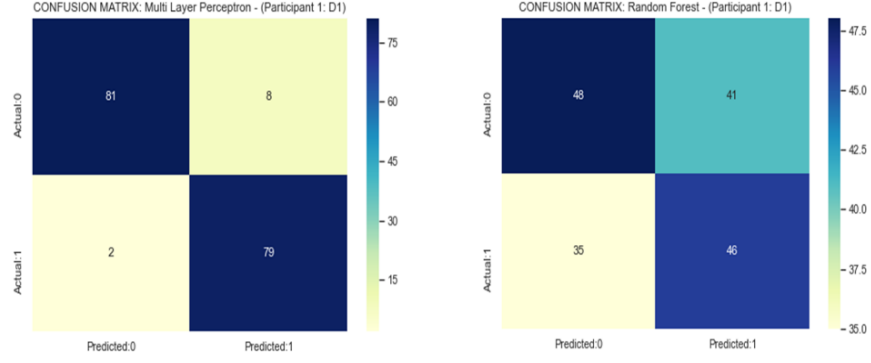
2.6.7   *Multi-layer Perceptron*

Artificial neural networks (ANN) are information processing paradigms that share performance characteristics with the human biological nervous system [37]. Multi-layer perceptron (MLP) is a type of feed-forward neural network that analyses the relationship between a series of independent input variables and a set of dependent output variables. MLP is a modified variation of the standard two layered perceptron that uses three or more layers of neurons with nonlinear activation functions to process complex computations. The MLP used in this study is shown in Fig. 4 and can be expressed mathematically as

$$y_i = \int \left(\sum_{i=1}^{n} w_i \cdot w_{ih} + \theta_h \right) \tag{6}$$

Where y_i stands for the hidden layer, x_i represents node i of the input layer, w_i is the connection weight between node i of the input layer and node h of the hidden layer. The number of input layer nodes is expressed as n, the bias values of node h is θ_h and the network's Sigmoid function is expressed as $\int (u)$ [38].

3   Experiments

As defined by literature [39], the following four metrics are the base measures used to assess the performance of binomial classification models. These base measures are calculated from a series of experiments using a set of positive and negative instances, where TP = true positive, FP = false positive, FN = false negative, TN= true negative. These base measures are then plotted using a confusion matrix, which is a visual representation of a classification model's performance. The rows of the confusion matrix represent the predictions made, whilst the columns represent the actual outcome.

Fig. 5. ROC curve and AUC score for D1 (left) & D3 (Right) using a Multi-layer perceptron

3.1 Classification Experiments

The following performance measures (accuracy, sensitivity, specificity, and positive predicted value) were used to assess varying aspects of each model's performance. Each performance measure is calculated using the output values from the confusion matrix. Accuracy (Acc) measures the number of correctly classified instances by the total number of instances; sensitivity (Sen) measures the number of true positive instances by all positive instances; specificity (Spe) measures the number of true negative instances by the number of negative instances and the positive predicted value (Ppv) is the rate of positive instances that are positive. The mathematical formulas for each base measure are expressed from algorithms 7-10.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{8}$$

$$Specificity = \frac{TN}{TN+FP} \tag{9}$$

$$Positive\ predicted\ value = \frac{TP}{TP+FP} \tag{10}$$

3.2 AUC ROC Experiments

To account for the imbalances of our dataset and the use of over-sampling, we constructed a multi-point receiver operating characteristic (ROC) curve of probability. A ROC curve plots a classifier's TP rate against its FP rate at multiple decision thresholds. Once plotted, we can then measure the area under AUC curve to assess how efficient each classifier was at distinguishing between positive and negative instances. To calculate AUC, we used the following formula [40].

$$\hat{A} = \frac{S_0 - n_0(n_0+1)/2}{n_0 \cdot n_1} \tag{11}$$

Where n_0 represents the number of positive instances, n_1 the number of negative instances, and $S_0 = \sum r_i$ where r_i is the rank of ith instance [41].

4 Results

The following section illustrates the experimental results recorded for this pilot study. Table 1 summarises the comparative analysis of our classification models for participant 1, with the MLP outperforming the other classification models with an accuracy measure of 94.73%, sensitivity of 96.29% and an AUC measure of 0.970.

Model	Acc (%)	Sen (%)	Spe (%)	AUC	Ppv (%)
MLP	94.73	96.29	93.33	0.970	96.55
LR	94.70	95.00	94.44	0.947	95.55
NB	94.11	88.76	92.22	0.944	89.01
SVM	93.85	88.75	97.77	0.933	90.72
DT	92.94	91.25	94.44	0.931	92.39
RF	89.44	88.50	90.36	0.942	88.23

Table 1. Results of each classification model for participant 1

To further investigate the frequency of PST and how they can be used as an additional sensing modality, we conducted a comparative distribution plot. By plotting the distribution of these triggers, we were able to identify a participant's threshold and tolerance levels pertaining to the occurrence of their seizure onset.
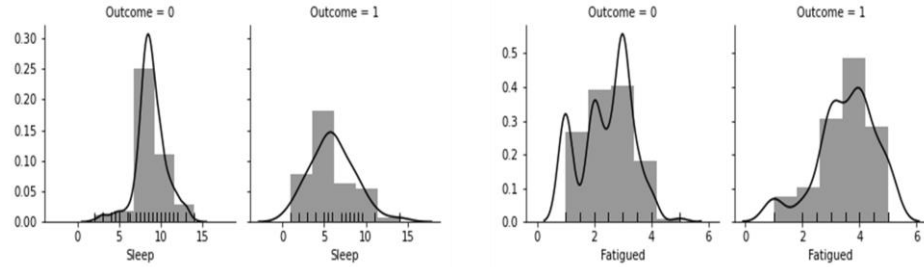
Fig. 6. Distribution plot visualising sleep and fatigue at the time of seizure onset (participant 1)
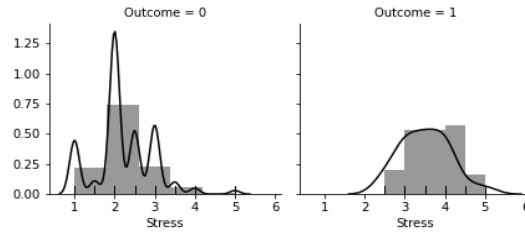
Fig. 7. Distribution plot visualising stress and seizure onset (participant 1)

As shown in Fig. 6 and 7, the positive instances for each trigger (sleep, stress, fatigue) have a distribution that is skewed right (positively skewed). This is an important observation regarding the potential use of PST when classifying GTC seizures, as we can deduce that these triggers are likely to occur prior to and during seizure onset. Our findings for participant 1 show that a GTC seizure was likely to occur in parallel with one or more of the documented triggers, thus validating our initial hypothesis. The results for participant 1 show that of the 49 positive instances recorded, 90% had a PSS ≥ 3.5, whilst 86% had a RoF ≥ 3.

Further analysis showed some interesting findings, as we found participant 1 was susceptible to triggering a GTC seizure if insufficient sleep was accumulated the day prior. We observed that on average, if less than 8 hours of sleep was recorded, the participant's likelihood of triggering a seizure increased, as observed in 81% of positive instances. Furthermore, 316 negative instances were documented during the observation period, of which 264 occurred when more than 8 hours of sleep was attained, with only 13.84% of negative instances occurring between $7.5 \leq 2$ hours of sleep.

Table 2 illustrates the classification results for participant 2, with the MLP once again outperforming the other classification models regarding accuracy, sensitivity, AUC, and positive predicted value.

Model	Acc (%)	Sen (%)	Spe (%)	AUC	Ppv (%)
MLP	96.87	96.96	96.77	0.987	96.70
SVM	91.20	95.45	87.23	0.913	95.34
DT	91.17	96.87	86.11	0.915	96.87
LR	90.44	90.90	90.00	0.905	91.30
RF	86.75	82.14	91.25	0.961	87.95
NB	85.84	82.45	89.28	0.859	83.33

**Table 2.** Results of each classification model for participant 2

From the results shown in Fig. 8 and 9, we can see that participant 2 is less susceptible to sleep fluctuations than participant 1, with positive instances documented at multiple threshold points. Results of our analysis show that participant 2 has a strong reactionary metric with fatigue, with 46% of positive instances recording a RoF =>3. Furthermore, we observed that stress has a similar distribution to fatigue, positively skewing right with a sudden incline towards the end of the PSS, with 64% of positive instances => 4. These findings indicate that participant 2's seizure onset is most likely triggered by stress and fatigue rather than sleep deprivation. This indicates that participant 2 is stills susceptible to an epileptic seizure even if enough sleep has been attained, as it is the varying fluctuations in stress and fatigue that are the predominant trigger factors.
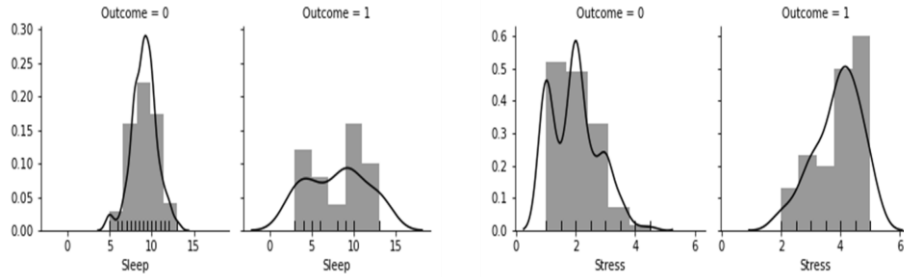
Fig. 8. Distribution plot visualising the occurrence of stress and sleep at the time of seizure onset (participant 2)
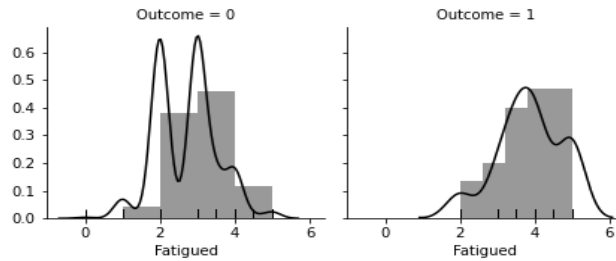
Fig. 9. Distribution plot visualizing fatigue and seizure onset (participant 2)
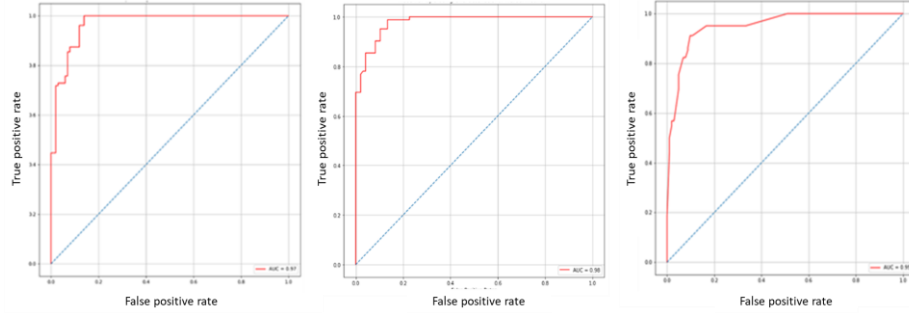
Fig. 10. ROC curve and AUC score for D1 (left) & D3 (Right) using a Multi-layer perceptron

The following results show how each classifier performed when the datasets for both participants were combined. Once again, the MLP outperformed the remaining classifiers, with accuracy scores of 94.11%, sensitivity of 92.15% and an AUC measure of 0.952.

Model	Acc (%)	Sen (%)	Spe (%)	AUC	Ppv (%)
MLP	94.11	92.15	96.07	0.952	92.45
LR	92.48	92.10	92.85	0.925	92.25
NB	92.15	90.12	94.44	0.923	89.47
SVM	91.66	91.17	92.15	0.917	91.26
DT	91.17	88.88	93.75	0.913	88.23
RF	89.54	90.80	88.23	0.958	90.60

Table 3. Classification results for dataset D3 (participant 1 + participant 2)

As shown in Fig.10, the final set of experiments were conducted using AUC ROC, and assessed the AUC scores of each classifier for datasets D1 (participant 1), D2 (participant 2) and D3 (D1 + D2). For D3, we combined the data from both participants to see if it would improve the performance of our person-specific classifiers. Across all 3 datasets, the MLP model produced the highest AUC scores, with measures of 0.97, 0.98 and 0.95 for datasets D1, D2 and D3, respectively.

5 Discussion

This section summarises the findings and contributions made from this pilot study. Experimental results indicate that PST can successfully train a classification algorithm that uses a single person's data, and then successfully classifies an epileptic seizure using the same person's unseen data.

Results show that the MLP was the classification model best suited for this type of task, as it outperformed the other models in almost every performance experiment. Our findings support the notion that onset is influenced by a person's idiosyncratic triggers as shown in in Fig. 11 and Fig. 12. When comparing the two plots, we can see the varying differentials in the threshold levels between participant 1 and 2.

Although there have only been a few studies that assess the correlation between sleep and those diagnosed with epilepsy, our findings indicate seizure onset was more likely to occur when participant 1 had less than 8 hours of sleep, whilst participant 2 would become susceptible if ≤ 6 hours of sleep but if more than 9.5 hours of sleep was experienced the day prior.

A noticeable correlation between stress, fatigue and the frequency of onset was observed, with participant 1 having a stress score of 3.5 or above in 86% of recorded seizures. Although participant 2 was less affected by stress, a higher fatigue score was observed throughout, with 77% of the 22 seizures recorded having a RoF ≥ 4. Furthermore, participant 2 had a PSS ≥ 3.5 or above in 64% of total observations when compared to participant 1's 48%.

5.1 Limitations

The main concern about the findings of this study was the sample size. Due to Covid-19, our sample size and participant availability was greatly affected, leaving this pilot study with 2 available participants. Of the 300 days participant 1 was observed, we documented 17 positive instances (seizures), whilst participant 2 was observed for a shorter duration of 248 days. Whilst we believe that the data collection process was conducted over a sufficient timeframe, a larger sample size would further validate our initial hypotheses and the practical application of PST.
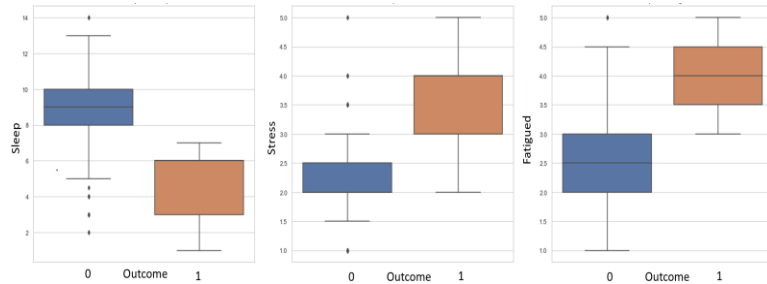
Fig. 11. Participant 1 PST; 0 = No seizure; 1 = Seizure.
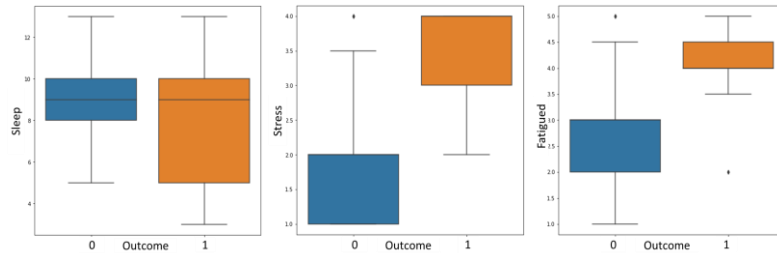
Fig. 12. Participant 2 PST; 0 = No seizure; 1 = Seizure

5.2 Future Research

We believe that PST should be considered for future research as an additional sensing modality, working in conjunction with other forms of multi-modal detection. Current multi-sensor modalities predominantly focus on the use of biometric sensors such as electrocardiogram (ECG) to formulate predictions, as this allows for biometric fluctuations to be measured in real time. Using PST in conjunction with standard sensing modalities could account for the varying diversities seen in different types of epilepsy and reduce the frequency of false alarms. Based on the practical application of PST and the ease in which they can be recorded, we propose that future studies should use a sample size of no less than 25 participants with refractory epilepsy, spanning a similar 300-day timeframe.

6   Conclusion

This pilot study has undertaken a preliminary investigation into whether PST from the same participant can be used to train and test a classification model. Results show that participants are susceptible to triggers in different ways, with varying tolerance levels observed throughout. This indicates that a person-specific approach may be best suited for this type of detection, as machine learning models can be tailored to each patient's idiosyncratic characteristics and fluctuations. To our knowledge, this is the first pilot study that has proposed the use of PST for epilepsy detection, and the results presented here warrant further investigation in the form of a full-scale clinical trial.

To conclude, we believe that PST can be used as an additional sensing modality when classifying epileptic seizures, and these triggers could assist deep learning algorithms when classifying biometric sensing data in real-time, adding an additional layer of validation that could assist in reducing the false detection rate and improve the overall performance when detecting epileptic seizures.

7   References

1. Thijs, R. D. et al.: Epilepsy in adults. The Lancet, 393(10172), 689-701 (2019).
2. Iasemidis, L. D. et al.: Adaptive epileptic seizure prediction system. IEEE Transactions on Biomedical Engineering, 50(5), 616-627 (2003).
3. Laxer, K. D. et al.: The consequences of refractory epilepsy and its treatment. Epilepsy & Behavior, 37, 59–70 (2014).
4. Barot, N., Nei, M.: Autonomic aspects of sudden unexpected death in epilepsy (SUD-EP). Clinical Autonomic Research, 29(2), 151–160 (2019).
5. Lamberts, R. J. et al.: Sudden unexpected death in epilepsy: People with nocturnal seizures may be at highest risk. Epilepsia, 53(2), 253-257 (2012).
6. DeGiorgio, C. M. et al.: Ranking the Leading Risk Factors for Sudden Unexpected Death in Epilepsy. Frontiers in Neurology, 8(473) (2017).
7. Hesdorffer, D. C. et al.: Combined analysis of risk factors for SUDEP. Epilepsia, 52(6), 1150-1159 (2011).

8. Van de Vel, A. et al.: Non-EEG seizure detection systems and potential SUDEP prevention: State of the art: Review and update. Seizure, 41, 141–153 (2016).

9. Pack, A. M.: SUDEP: What Are the Risk Factors? Do Seizures or Antiepileptic Drugs Contribute to an Increased Risk? Epilepsy Currents, 12(4), 131–132 (2012).

10. Manford, M. et al.: An analysis of clinical seizure patterns and their localizing value in frontal and temporal lobe epilepsies. Brai n, 119(1), 17–40 (1996).

11. Aird, R. B., Gordon, N. S.: Some excitatory and inhibitory factors involved in the epileptic state. Brain and Development, 15(4), 299–304 (1993).

12. Smith, S. J. M.: EEG in the diagnosis, classification, and management of patients with epilepsy. Journal of Neurology, Neurosurgery & Psychiatry, 76(2), ii2–ii7 (2005).

13. Ferlisi, M., Shorvon, S.: Seizure precipitants (triggering factors) in patients with epilepsy. Epilepsy & Behavior, 33, 101–105 (2014).

14. Nakken, K. O. et al.: Which seizure-precipitating factors do patients with epilepsy most frequently report? Epilepsy & Behavior, 6(1), 85-90 (2005).

15. Balamurugan, E. et al.: Perceived trigger factors of seizures in persons with epilepsy. Seizure, 22(9), 743–747 (2013).

16. Ginsburg, G. S., Phillips, K. A.: Precision Medicine: From Science To Value. Health affairs, 37(5), 694–701 (2018).

17. Porumb, M. et al.: Precision Medicine and Artificial Intelligence: A Pilot Study on Deep Learning for Hypoglycemic Events Detection based on ECG. Scientific Reports, 10(1), 170 (2020).

18. Ince, T. et al.: A Generic and Robust System for Automated Person-specific Classification of ECG Signals. IEEE Transactions on Biomedical Engineering, 56(5), 1415 – 1426 (2009).

19. Pimentel, J. et al.: Epilepsy and physical exercise. Seizure, 25, 87-94 (2015).

20. Fitbit.: What are active minutes? https://help.fitbit.com/articles/en_US/Help_article/1379, last accessed 2020/05/21.

21. Fitbit.: What should I know about sleep stages? https://help.fitbit.com/articles/en_US/ Help article/2163, last accessed 2020/05/21.

22. Xu, X. et al.: Over-sampling algorithm for imbalanced data classification. Journal of Systems Engineering and Electronics, 30(6), 1182-1191 (2019).

23. Cohen, S., Janicki-Deverts, D.: Who's Stressed? Distributions of Psychological Stress in the United States in Probability Samples from 1983, 2006, and 2009. Journal of Applied Social Psychology, 42(6), 1320-1334 (2012).

24. Cohen, S. et al.: A Global Measure of Perceived Stress. Journal of Health and Social Behavior, 24(4), 385-396 (1983).

25. Micklewright, D. et al.: Development and Validity of the Rating-of-Fatigue Scale. Sports Medicine, 47(1), 2375-2393 (2017).

26. Refaeilzadeh, P. et al.: Cross-Validation. In: Liu, L., Özsu, M. T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA.

27. Yusa, M., Utami, E.: Classifiers evaluation: Comparison of performance classifiers based on tuples amount. In: 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pp. 1-7. IEEE, Yogyakarta, Indonesia (2017).

28. Zhang, H., Li, D.: Naïve Bayes Text Classifier. In: IEEE International Conference on Granular Computing, pp. 708-708. IEEE, Fremont, CA, USA (2007).

29. Jing, N. et al.: Information credibility evaluation in online pRoFessional social network using tree augmented naïve Bayes classifier. Electronic Commerce Research (2019).

30. MavRoForakis, M. E.: Theodoridis, S.: Support Vector Machine (SVM) classification through geometry. In: 13th European Signal Processing Conference, pp. 1-4. IEEE, Antalya, Turkey (2005).

31. Lammertyn, J. et al.: Logistic regression analysis of factors influencing core breakdown in 'Conference' pears. Postharvest Biology and Technology, 20(1), 25-37 (2000).

32. Sun, Y. et al.: Application of Logistic Regression with Fixed Memory Step Gradient Descent Method in Multi-Class Classification Problem. In: 6th International Conference on Systems and Informatics (ICSAI), pp. 516-521. IEEE, Shanghai, China (2019).

33. Tahsildar, S.: Gini Index for Decision Trees. https://blog.quantinsti.com/gini-index/, last accessed 2020/07/09.

34. Decision Tree Flavors: Gini Index and Information Gain. https://www.learnbymarketing.com/481/decision-tree-flavors-gini-info-gain/, last accessed 2020/05/05.

35. Misra, S., Li, H., He, J.: Machine Learning for Subsurface Characterization. 1st edn. Elsevier, Cambridge, MA, USA (2020).

36. Sinnott, R. O., Duan, H., Sun, Y.: Chapter 15 - A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the Weather. In: Buyya, R. Calheiros, R. N., Dastjerdi, A. V. (eds) Big Data: Principles and Paradigms. 1st edn. Elsevier, Burlington, MA, USA (2016).

37. Balaji, S. A., Baskaran, K.: Design and Development of Artificial Neural Networking (ANN) System Using Sigmoid Activation Function to Predict Annual Rice Production in Tamilnadu. International Journal of Computer Science Engineering and Information Technology, 3(1), 13-31 (2013).

38. Matarat, K. et al.: Comparison of Classification Algorithms for Movie Reviews. In: 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications, and Information Technology (ECTI-CON), pp. 826-829. IEEE, Pattaya, Thailand (2019).

39. Canbek, G. et al.: Binary Classification Performance Measures/Metrics: A Comprehensive Visualized Roadmap to Gain New Insights. In: 2nd International Conference on Computer Science and Engineering (UBMK), pp. 821-826. IEEE, Antalya, Turkey (2017).

40. Hand, D. J., Till, R. J.: A Simple Generalisation of the Area under the ROC Curve for Multiple Class Classification Problems. Machine Learning, 45, 171-186 (2001).

41. Huang, J., Ling, C. X.: Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 17(3), 299-310 (2005).