Mental Stress Recognition on the fly using Neuroplasticity Spiking Neural Networks

Mahima Milinda Alwis Weerasinghe (mahima.weerasinghe@aut.ac.nz)  
Auckland University of Technology

Grace Wang  
University of Southern Queensland

Jacqueline Whalley  
Auckland University of Technology

Mark Crook-Rumsey  
King’s College London

Article

Keywords:

Posted Date: August 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1841009/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Mental Stress Recognition on the fly using Neuroplasticity Spiking Neural Networks

Author Information

Affiliations

Department of Computer Science and Software Engineering, Auckland University of Technology, Auckland, New Zealand

Department of Computer and IT, Sri Lanka Technological Campus, Padukka, Sri Lanka

Mahima Milinda Alwis Weerasinghe*1,2

School of Psychology and Wellbeing, University of Southern Queensland, Australia

Grace Wang

Department of Computer Science and Software Engineering, Auckland University of Technology, Auckland, New Zealand

Jacqueline Whalley

Department of Basic and Clinical Neuroscience, King’s College London, London, United Kingdom

UK Dementia Research Institute, Centre for Care Research and Technology, Imperial College London, United Kingdom

Mark Crook-Rumsey3,4

Contributions

M.W. developed the learning algorithms, designed the experiments, analysed the data, and wrote the manuscript. G.W. designed the data collection and M.C. acquired the data. J.W. and G.W. supervised the research. All authors reviewed the manuscript and provided feedback.

Corresponding author

Correspondence to: M.W. (mahima.weerasinghe@aut.ac.nz)
Abstract

Mental stress is found to be strongly connected with human cognition and wellbeing. As the complexities of human life increase, the effects of mental stress have impacted human health and cognitive performance across the globe. This highlights the need for effective non-invasive stress detection methods. In this work we introduce a novel, artificial spiking neural network model called, Online Neuroplasticity Spiking Neural Network (O-NSNN) that utilizes a repertoire of learning concepts inspired by the brain, to classify mental stress using Electroencephalogram (EEG) data. These models are personalized and tested on EEG data recorded during sessions in which participants listen to different types of narratives designed to induce acute stress. Our O-NSNN models learn on the fly producing an average accuracy of 90.76% (σ = 2.09) when classifying EEG signals of brain states associated with these audio comments. The brain-inspired nature of the individual models makes them robust and efficient and has the potential to be integrated into wearable technology. Furthermore, this article presents an exploratory analysis of trained O-NSNNs to discover links between perceived and acute mental stress. The O-NSNN algorithm proved to be better for personalized stress recognition in terms of accuracy, efficiency, and model interpretability.

Introduction

Implications of mental stress. We often encounter stress in daily life with variations of intensity and prolongation. Stress is understood as the response of the human body to mental and/or physical stimuli that involves the nervous system and hypothalamus-pituitary-adrenocortical axis\(^1\). According to the literature, stress is often classified as acute, episodic, or chronic\(^2\). Many contemporary studies have found stress to have a major impact on human health and cognitive performance. In some cases, stress has been shown to have a direct connection to depression, anxiety, stroke, cardiovascular disease, cancer, speech, and cognition impairment\(^3\)\(^-\)\(^5\). The negative effects of stress on human cognition are associated with dysfunctional changes in the prefrontal cortex and amygdala activation\(^5\)\(^,\)\(^6\) whereas the physical health
effects of stress are related to detrimental changes in immunity and physical homeostasis. As the complexities of human life increase, the effects of stress have begun to burden nations and, the globe at large, which highlights the requirement for more research in this area. Early detection of harmful stress can be crucial as a part of effective stress management to promote greater wellbeing.

**Stress and electroencephalogram.** Rapid development in sensor technologies and machine learning (ML) techniques have enabled research communities to begin to develop automated stress detection systems. These systems use invasive and/or non-invasive data acquisition methods. Stress recognition using invasive methods can be highly time-consuming and often require experts for data acquisition and processing; this is not ideal for an automated system. The most common non-invasive methods include Electroencephalogram (EEG), heart rate variability, galvanic skin response, blood volume pulse, and electromyography for data acquisition. Of these non-invasive methods, EEG is used most extensively for stress recognition due to its: information richness, cost-effectiveness, and high temporal resolution.

**Stress recognition on the fly.** Current methods for stress recognition use traditional ML techniques such as Linear Discriminant Analysis, Naive-Bayes, Support Vector Machine, K-Nearest Neighbor, and Multi-layer Perceptron. However, these methods are not capable of evolving and adapting to new information after training, preventing them from being used in an online setup. Online learning typically uses real-world data that changes with time thus the model is adaptive and learns as new data is fed into it over time. In contrast, most stress detection approaches presented in the literature use static data to train and test the model. They also typically employ interventions, to manipulate the data used to train and test the models, such as feature engineering methods. It is difficult to compare the performance of known stress detection models because the feature engineering and extraction approaches differ from one study to another. This lack of standards also means that the generalizability of the methods presented is questionable. Moreover, these traditional methods require a high volume of labelled data for model training. Today, the emergence of wearable technologies has revealed the potential for personalized health applications, designed to detect stress. Such applications must meet certain conditions to be
practical. Use of online learning allowing the model to adapt to change, capability to operate under low power and the need for low-resource utilization are among them. This work focuses on finding solutions for the challenges posed by these conditions.

Data drifts and online learning. One of the challenges in online learning is handling what is known as the drift phenomena successfully. Drifts can be observed in spatiotemporal data such as EEG and they can be defined in terms of input(s) and concept(s). The input(s) drift refers to the change of input data distribution over time without affecting the posterior probabilities of classes; concept drift refers to the change of posterior probabilities of the classes over time without any changes in the input distribution. The drift phenomena require ML techniques to be able to acquire new knowledge without forgetting the prior knowledge (i.e., avoiding catastrophic forgetting) and even to update prior knowledge based on that new or recently gained knowledge. Adding to the challenge are the restrictions posed by online learning which demands the algorithm to use only a limited amount of pre-allocated memory, process a sample only once, use a consistent amount of time for processing, produce a valid model at each processing step, and perform in par with batch mode learning.

Spiking Neural Networks (SNNs). SNNs are a class of artificial neural networks (ANNs) that are considered to be biologically plausible. They have proven to be highly efficient in terms of time and memory requirements for data processing compared to commonly used sigmoidal counterparts. The temporal dimension used in data processing is a major factor that contributes to their increased efficiency when compared with traditional ANNs, which makes SNNs an ideal candidate for online learning. Moreover, the unsupervised learning mechanisms in SNNs have demonstrated capability in fast and data-efficient learning. These attributes have led to the development of several online learning algorithms using SNNs with both supervised and unsupervised learning. Of these methods, only a few algorithms use structural adaptation (i.e., evolving and pruning neurons and connections). Structural adaptation is crucial for learning new knowledge and forgetting irrelevant information in an online
However, some of these structurally adaptive methods are built for batch mode learning only or do not fully exploit the temporal dynamics through learning.

The Online Neuroplasticity Spiking Neural Network (O-NSNN). The O-NSNN introduced in this work uses mathematical abstractions of selected plasticity techniques found in brain functions to fully exploit spatiotemporal patterns present in the data. This does not mean that the model mimics the entire neurobiological process of the brain, but rather it uses selected concepts of signal encoding, propagating, processing, and learning found in the brain. This algorithm differs from the previous ASNNs due to the inclusion of a full repertoire of plasticity techniques for temporal learning. These techniques are Spike Time Dependent Plasticity (STDP), Intrinsic Plasticity (IP), Neuron Evolving (neuron addition) and Neuron Pruning (neuron elimination). We hypothesize that this algorithm will produce stable and faster pattern separation capability in the online classification of stress-related EEG, by considering and handling the challenges associated with online learning.

The proposed O-NSNN consists of three layers of Leaky-Integrate and Fire neurons (LIF) (see Figure 1: (a) The proposed O-NSNN architecture for stress recognition. EEG originating from FP1, FP2, T7 and T8 channels are encoded into spikes (using the AER algorithm) and propagated through a three-layered SNN architecture. An STDP rule is used for temporal learning between the input layer and the hidden layer. Hidden layer neurons use IP to adapt excitability based on the incoming data. The output layer learns using RO and SDSP rules. Each hidden layer neuron prunes itself according to soft-pruning rule and, the output layer evolves. (b) Stress class input samples of P1 with different spike rate distribution (Input drift) (c) Two separate classes of P1 (Critical and Positive) with the same input spiking distributions (Concept drift).); a mathematical abstraction of a biological neuron that has demonstrated a greater balance between biological plausibility and computational tractability. Before processing, the EEG signals are converted to their spiking equivalent using Address Event Representation (AER); a spike encoding algorithm used in artificial retina. Thereafter, the first layer of neurons propagates spikes to the second layer via excitatory (blue) and inhibitory (black) synapses. During this propagation, the synaptic
weights are updated using the STDP rule\textsuperscript{38}. In addition, all the neurons adjust their excitability using an IP rule\textsuperscript{45}. This combination of unsupervised STDP and IP prevents the network from getting caught up in a potentiation loop\textsuperscript{46} ensuring homeostasis\textsuperscript{47} and helping neurons extract independent spiking features from the input\textsuperscript{48}. Moreover, the second layer of neurons undergoes a self-pruning process induced by error monitoring to avoid misclassifications caused by low spiking neurons\textsuperscript{45}. The synapses from the second layer to the third layer are excitatory and, follow a similar weight updating strategy discussed in dynamically evolving SNN (deSNN)\textsuperscript{49} evolving a new neuron in the presence of new knowledge. However, unlike in deSNN, output neurons are not merged based on weight vector similarities (i.e., calculated using Euclidean distance of the input weight vector of a given neuron). In the presence of data drift, neurons of similar Euclidean distances may be representing different classes. Therefore, we do not merge neurons, rather we eliminate or preserve neurons created based on the classification errors made during the data processing (Please refer to the Methods section for an in-depth explanation). This combined process of neuron addition in the third layer and, neuron pruning in the second layer are unique implementations that have not been discussed together in the published literature, to the best of our knowledge.

**Acute stress and data collection.** The dataset used in this study consists of EEG recordings from 22 healthy participants (twelve males - average age = 27.92 years, standard deviation (\(\sigma\)) = 3.09 and ten females - the average age of 25.9 years, \(\sigma = 8.20\)) across three different conditions. On each condition, the participants were asked to listen to one type of comment, either critical, neutral, or positive. Such critical comments stimulate the part of the human auditory system of which the primary objective is to alert and warn\textsuperscript{50}. Moreover, audio criticism has also been shown to increase stress levels\textsuperscript{51}. Therefore, audio or audio-visual stimuli has been used for stress studies\textsuperscript{52–54}. Based on these previous studies, we presumed that the critical comments would induce acute stress in the participant. In addition to EEG data, the perceived stress of each participant was recorded using the PSS-14 scale\textsuperscript{55}. Each EEG recording lasted for two minutes, and the recordings were segmented into five-second splits to feed the O-NSNN.
Consequently, a single sample of EEG data consisted of 1280 time points and four channels. From each participant, 72 such samples with 24 samples for each class of stressed, neutral, and positive were processed. Complete details of the dataset are given in the methods section.

**EEG channels and evaluation.** For the experiments of this study, we extracted signals from the FP1, FP2, T7, and T8 channels in an attempt to capture the activation dynamics found between Prefrontal Cortex (PFC) and Amygdala during stress\(^6\). The classification accuracy of EEG in identifying samples of stressed, neutral, and positive, using O-NSNN was compared against 70/30 split batch learning and online learning without structural plasticity (SP). For this, we used individualized O-NSNN models, since the effects of stress are found to be depending on an individual’s neurobiological predisposition\(^2\). Moreover, we used the prequential accuracy metric to evaluate the performance of online learning\(^56\). Secondly, these individualized models were subjected to an exploratory analysis that was undertaken to test the interpretability of the model and see if relationships could be discovered between acute and participant perceived stress.

This exploratory analysis involved comparing the personalized network activations to individually reported perceived mental stress levels. We categorized participants into one of three classes based on their PSS-14 scores (See: Participant categorization according to PSS-14 Score Table I). The connection weights of personalized models and Euclidean Distances (ED) of third-layer neurons were analyzed to find patterns within and between the perceived mental stress groups.

In this work, we present a spatiotemporal data processing method for mental stress recognition and, elucidate the possibility of investigating brain activity at an individual level. Therefore, the contribution of this study benefits both computer science and psychology/neuroscience research communities. The contributions of the study are as follows:

1. O-NSNN algorithm equipped with a biologically plausible repertoire of plasticity techniques for online mental stress recognition
2. Insights into how perceived stress relates to incidences of acute stress
Table 1: Participant categorization according to PSS-14 Score

<table>
<thead>
<tr>
<th>Label</th>
<th>PSS-14 score</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Stress (HS)</td>
<td>PSS &gt; 30</td>
<td>6</td>
</tr>
<tr>
<td>Medium Stress (MS)</td>
<td>20 &lt; PSS ≤ 29</td>
<td>11</td>
</tr>
<tr>
<td>Low Stress (LS)</td>
<td>0 &lt; PSS ≤ 19</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 1: (a) The proposed O-NSNN architecture for stress recognition. EEG originating from FP1, FP2, T7 and T8 channels are encoded into spikes (using the AER algorithm) and propagated through a three-layered SNN architecture. An STDP rule is used for temporal learning between the input layer and the hidden layer. Hidden layer neurons use IP to adapt excitability based on the incoming data. The output layer learns using RO and SDSP rules. Each hidden layer neuron prunes itself according to soft-pruning rule and, the output layer evolves. (b) Stress class input samples of P1 with different spike rate distribution (Input drift) (c) Two separate classes of P1 (Critical and Positive) with the same input spiking distributions (Concept drift).
Results

We tested the classification accuracy of the O-NSNN model and compared the performance with the same learning framework without structural plasticity (SP) techniques (denoted as O-RSNN).

Furthermore, batch mode learning without SP (B-RSNN) (i.e., 70% of the samples for training and 30% for testing) was included in the performance comparison. The task involved measuring the correctness of the classification of EEG data into one of three possible classes: stress, neutral, and positive conditions. Since the synaptic weights of the first layer to the second are initiated randomly following Gaussian distribution, each experiment was conducted 30 times allowing the accuracy to be reported statistically.

The performance is discussed in terms of average accuracy (see Table 2) and efficiency. Furthermore, we explored patterns in network dynamics for knowledge extraction.

Table 2: Accuracy comparison between online (with (O-NSNN) and without SP (O-RSNN)) and batch mode (B-RSNN) learning

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>O-NSNN</th>
<th>O-RSNN</th>
<th>B-RSNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.94 ± 0.02</td>
<td>0.66 ± 0.09</td>
<td>0.84 ± 0.07</td>
</tr>
<tr>
<td>P2</td>
<td>0.93 ± 0.03</td>
<td>0.57 ± 0.07</td>
<td>0.72 ± 0.08</td>
</tr>
<tr>
<td>P3</td>
<td>0.91 ± 0.06</td>
<td>0.60 ± 0.06</td>
<td>0.69 ± 0.09</td>
</tr>
<tr>
<td>P4</td>
<td>0.90 ± 0.08</td>
<td>0.77 ± 0.11</td>
<td>0.93 ± 0.05</td>
</tr>
<tr>
<td>P5</td>
<td>0.91 ± 0.06</td>
<td>0.66 ± 0.11</td>
<td>0.83 ± 0.09</td>
</tr>
<tr>
<td>P6</td>
<td>0.92 ± 0.03</td>
<td>0.79 ± 0.08</td>
<td>0.88 ± 0.08</td>
</tr>
<tr>
<td>P7</td>
<td>0.90 ± 0.06</td>
<td>0.59 ± 0.08</td>
<td>0.74 ± 0.09</td>
</tr>
<tr>
<td>P8</td>
<td>0.94 ± 0.02</td>
<td>0.68 ± 0.11</td>
<td>0.88 ± 0.09</td>
</tr>
<tr>
<td>P9</td>
<td>0.94 ± 0.02</td>
<td>0.76 ± 0.08</td>
<td>0.75 ± 0.10</td>
</tr>
<tr>
<td>P10</td>
<td>0.91 ± 0.06</td>
<td>0.42 ± 0.07</td>
<td>0.49 ± 0.11</td>
</tr>
<tr>
<td>P11</td>
<td>0.91 ± 0.07</td>
<td>0.75 ± 0.08</td>
<td>0.86 ± 0.08</td>
</tr>
<tr>
<td>P12</td>
<td>0.91 ± 0.04</td>
<td>0.68 ± 0.08</td>
<td>0.86 ± 0.08</td>
</tr>
<tr>
<td>P13</td>
<td>0.92 ± 0.08</td>
<td>0.82 ± 0.09</td>
<td>0.92 ± 0.06</td>
</tr>
<tr>
<td>P14</td>
<td>0.89 ± 0.07</td>
<td>0.53 ± 0.10</td>
<td>0.62 ± 0.10</td>
</tr>
<tr>
<td>P15</td>
<td>0.92 ± 0.04</td>
<td>0.66 ± 0.09</td>
<td>0.77 ± 0.10</td>
</tr>
<tr>
<td>P16</td>
<td>0.91 ± 0.04</td>
<td>0.46 ± 0.10</td>
<td>0.60 ± 0.12</td>
</tr>
<tr>
<td>P17</td>
<td>0.86 ± 0.13</td>
<td>0.49 ± 0.10</td>
<td>0.71 ± 0.10</td>
</tr>
<tr>
<td>P18</td>
<td>0.85 ± 0.09</td>
<td>0.55 ± 0.09</td>
<td>0.65 ± 0.09</td>
</tr>
<tr>
<td>P19</td>
<td>0.90 ± 0.07</td>
<td>0.41 ± 0.08</td>
<td>0.61 ± 0.08</td>
</tr>
<tr>
<td>P20</td>
<td>0.91 ± 0.07</td>
<td>0.74 ± 0.10</td>
<td>0.81 ± 0.08</td>
</tr>
<tr>
<td>P21</td>
<td>0.90 ± 0.12</td>
<td>0.64 ± 0.09</td>
<td>0.75 ± 0.09</td>
</tr>
<tr>
<td>P22</td>
<td>0.92 ± 0.03</td>
<td>0.67 ± 0.09</td>
<td>0.82 ± 0.09</td>
</tr>
</tbody>
</table>
**Increased accuracy and robustness in O-NSNN.** The highest average accuracy for O-NSNN was 93.63% ± 2.37 for P1 and, the lowest was 85.29% ± 9.25 for P18. It can be seen in Table 2, that O-NSNN outperformed O-RSNN across all 22 participants and B-RSNN was outperformed except for one participant (P4). Figure 2, shows the variation of performance for personalized models for each participant obtained from 30 pseudo-random network initiations. Accordingly, for all 22 participants, the O-NSNN model had the lowest degree of variation in performance variation. This indicates higher robustness compared to other methods.

![Figure 2: Performance variation of individual models. Performance distribution obtained from 30 testing cycles. At each cycle the initial weights between the input to hidden layers are selected pseudo randomly according to gaussian distribution. S – Online learning with SP, N – Online learning without SP, B – Batch mode learning without SP](image-url)
Figure 3: (a) Number of output neurons evolved by O-NSNN during 30 testing cycles for each participant model (b) Prequential accuracy progression with the number of samples increasing (c) Sample spiking raster plot of the hidden layer for P1
The efficiency of O-NSNN. The efficiency factor of the O-NSNN can be presented in terms of the number of output neurons used and spikes generated in the hidden layer. When the number of output neurons used was investigated, the O-NSNN method used, on average 20.39 (\( \sigma = 3.84 \)) neurons (see Figure 3a) whereas O-RSNN used 72 (i.e., absence of structural plasticity created a neuron for each input sample) and, B-RSNN used 50 output neurons respectively (i.e., 70/30 split training used 50 input samples for training where a neuron was created for each input). The spike generation of O-NSSN was measured as a ratio between the number of spikes received at the hidden layer to the number of spikes generated by the hidden layer where the mean was recorded at 0.063 (\( \sigma = 0.009 \)). This spike encoding is epitomized in Figure 3c where the raster plot indicates the temporal sparseness of the spikes. When considering the trend of model accuracy over time, O-NSNN typically reached a prequential accuracy of 80% within 150 to 200 seconds of data processing commencement. An exception to this trend was noted in the case of P17 and P21 O-NSNN models (see Figure 3c).

O-NSNN knowledge extraction. We also analyzed the Euclidean distance (ED) of the output neuron weight vectors and input to the hidden layer synaptic weights (i.e., STDP weights), of each individualized NSNN model. The evolved output neurons of an individualized NSNN model represented a certain class (i.e., stress, neutral or positive). The NSNN used this weight vector of the output neurons to predict the

![Figure 4: Euclidean Distance between initial(Blue) and final(Red) output neurons. The initiation process use the first 15 samples to evolve 15 output neurons. (a) without pruning or evolving new neurons (O-RSNN) (b) with pruning and evolving new neurons (O-NSNN)](image-url)
class of the incoming signals. Therefore, each ED of a sample is a numerical representation of the individual's brain signal under a given stimulus. Similarly, the weights of input to the hidden layer in NSNN are updated in an unsupervised method using STDP and IP. Once all the data samples are passed through the network, the NSNN weights (i.e., input to the hidden layer) capture the spatiotemporal correlations of the input signals.

Compared numerical representations of brain signals. We compared the EDs between the HS, MS, and LS groups and found that the mean distance between neutral and critical stimuli of the HS group was 0.95 ($\sigma = 0.41$). In contrast, the LS group’s average distance between neutral and critical stimuli was much
shorter at 0.25 ($\sigma = 0.22$). The average distance between neutral and positive stimuli of the HS group was 0.87 ($\sigma = 0.86$) and lower than that of the LS group’s distance of 1.86 ($\sigma = 0.84$). According to these results, the HS group’s EEG for positive stimuli did not differ to any notable extent from the EEG generated during neutral stimuli; this was the same for negative stimuli (i.e., under stress). However, the LS group recorded a much larger difference in both cases (see Figure 5a).

*Input channel correlation.* When considering the activations between input channels (i.e., using the input to hidden layer synaptic weights) the majority of MS participants exhibited similar activation patterns (see Figure 5c) whereas the LS and HS groups exhibited irregular patterns of activation from one individual to another (see Figure 5b and Figure 5d). While investigating this further by examining the input synaptic weights of the hidden layer we found that the HS group had higher inhibition than the LS group in the FP1 and FP2 channels (see Figure 6). The same inhibitory patterns were observed for T8 but not T7. When examining the right and left-brain activations we discovered that the HS group showed higher inhibition in the right hemisphere (FP2 and T8) than in the left hemisphere (FP1 and T7). However, in the LS group, the average difference between right and left hemisphere activations was significantly smaller. Moreover, higher activation was observed between FP1 and T8 than FP2 and T7 in five out of six participants in the HS group. The opposite activation pattern was observed with four out of five of the participants in the LS group.

Figure 6: Cumulative weights of the synapses fanning out from respective inputs calculated according to perceived stress groups
**Discussion**
This study presents neuroplasticity spiking neural network in an online learning setup for classifying the neural activity of healthy participants when exposed to comments that were intended to trigger different levels of mental states (i.e., stress, neutral, positive) and explores the link between these classifications and self-reported stress levels (i.e., perceived mental stress scores). This O-NSNN method produced higher pattern recognition capability on the fly, with increased efficiency, interpretability, and biological plausibility.

**The performance of the O-NSNN.** The O-NSNN outperformed the other SNNs (O-RSNN and B-RSNN), in terms of average accuracy, as shown in Table 2. When comparing the two online learning methods, O-NSNN (90.76%, \(\sigma = 2.09\)) was found to perform significantly better than O-RSNN (63.08%, \(\sigma = 11.09\)) (Student’s \(t\)-test, \(\alpha=0.05, p=0.005\)). As per Figure 2, the O-NSNN model produced the least performance variation indicating higher robustness\(^5\). Delving into the reasons behind the success of the O-NSNN approach, we found that the EDs of output neurons (i.e., numerical representations of input samples) have better discriminative capability between the initial and final states of O-NSNN than in O-RSNN. This enhanced discriminative capability is presented in Figure 4 for P1. With neurons evolving and self-pruning being the only difference between O-NSNN and O-RSNN; we propose this SP technique as a successful method for handling new classes and/or new representations of already known classes. In other words, the O-NSSN approach is effective at handling concept drift.

**STDP and IP learning.** In a previous study, it was reported how hidden layer neuron pruning with STDP+IP leads to increased robustness and efficiency of SNNs in a batch learning setup for EEG classification\(^4\). In the same study, hidden layer neurons with low firing probability causing classification errors were noted. In this study, instead of completely pruning these low-firing probability neurons, we have adopted a self-pruning method that stops a neuron activation for a limited period. This is achieved by increasing the neuron threshold voltage to the highest value found in the population. The advantage of this method is three-fold. Firstly, the inactivity of the neuron caused by threshold alteration help in
reducing the number of dimensions used to represent an input sample at the output layer. Since classifications of the proposed O-NSNN are based on EDs calculated from output layer synaptic weights, part of the increase in performance can be attributed to the mitigation of the *curse of dimensionality*.\(^58\)

Secondly, the self-pruned neurons remain in the network to respond to salient features that may occur due to drifts or new data. This repurposing of neurons may account for the improvement of the performance of the network with time\(^41\). Thirdly, the efficiency of this pruning is superior to regular synaptic pruning which requires scanning of the entire weight matrix against a threshold\(^41,59\).

*The efficiency of O-NSNN.* The efficiency of the O-NSNN in terms of the number of neurons used and spikes generated reduced drastically with the use of STDP+IP learning and self-pruning. Unlike continuous streams of spiking, these techniques enabled sparser spiking activity resulting inactive states most of the time (see Figure 3c). When compared to STDP-only learning, STDP+IP was shown to have reduced the average spiking by 35 times (Student’s *t*-test, \(\alpha=0.05, p=0.008\)). This reduction of spikes minimizes the calculations involved from the hidden to the output layer. Moreover, the O-NSNN output layer utilized 3.52- and 2.45-times lesser neurons on average compared to O-RSNN and B-RSNN models respectively. In comparison to the early methods of evolving neurons where the spiking is not regulated\(^35,60\) and the output repository grows indefinitely\(^37\), this method is much more suitable for memory-restricted applications.

**Knowledge extraction.** From trained NSNN models, HS participants showed lower activation levels in prefrontal channels FP1 and FP2 compared to the LS group. This was observed during the synaptic weight analysis of individual models where the HS group had more inhibitory weights connected to FP1 and FP2 channels (see Figure 6). The impairment of prefrontal activity during stressful events and, in individuals with high perceived stress have been reported previously\(^6,61\). Moreover, the amygdala activation of the HS group was expected to be higher compared to the LS group\(^6\) which we tried to capture through T8 and T7 channel inputs. However, in our findings, only the T8-connected synapses showed higher activations for the HS group (compared to the LS group), but this was not the case with
T7-connected synapses (see Figure 6). This can be due to the inability of EEG to capture the sub-cortical activity of the brain in general62. In terms of the activations between the channels, the similarity was observed among the individuals of the MS group, but not in HS and LS groups (Figure 5b-d). In addition, the HS group had the smallest difference between EDs (numerical representations of spike patterns) produced during stress and positive stimuli compared to neutral states, whereas in the LS group, the observation was the opposite (Figure 5a). This suggests a lack of change in functional patterns of the brain to external stimuli in the HS group and, a higher change in functional patterns in the LS group. This observation leads to an interesting hypothesis about the relationship between acute and perceived stress. Namely, the individuals with high perceived stress (HS group) have less discrimination between positive and negative stimuli. In a previous study, long-term stress has been found to alter the perception of emotional stimuli63. Our findings suggest that individuals with high perceived stress have low sensitivity to day-to-day acute stressors.

**Biological Plausibility.** The biological plausibility of O-NSNN can be discussed in the aspects of data processing techniques employed and the spiking behavior observed. Firstly, the data processing techniques inspired by neuroscientific concepts include STDP for temporal synaptic learning38, IP for neuron spike regulation39, self-pruning (apoptosis) to selectively restrict activation of neurons64, and addition of new neurons (neurogenesis) for retention of new knowledge65. Secondly, the model introduced demonstrates avalanche-like spiking which is also found in neocortical circuits66. Arguably this makes O-NSNN much more biologically plausible than other online learning methods introduced which do not utilize the same repertoire of plasticity techniques or show spiking behavior close to what is found in biology20,34,35.

**Conclusion**

This work presents a novel neural network algorithm for mental stress classification using EEG data and online learning. The algorithm adapts to individuals and uses functional concepts of the biological brain to learn, on the fly, in a resource-efficient manner. The O-NSNN algorithm introduced displayed superior
performance in terms of accuracy, robustness, and resource efficiency over models that did not use structural plasticity.

Our method introduced goes beyond traditional black box ANN models to reveal insights into individual brain dynamics for better interpretation. Improving the capability of this algorithm to recognize a higher number of classes under resource restrictions could potentially contribute to applications of wearable technology for the detection and monitoring of mental stress.

Methods

Neuroplasticity spiking neural network. Here we present a description of the O-NSSN model, and the experimental framework designed to test the model. The NSNN is a fully connected, feed-forward spiking neural network consisting of LIF neurons. The input nodes can process both excitatory and inhibitory spikes. These nodes are connected to the hidden layer via excitatory and inhibitory synapses in which the weights are updated using an unsupervised STDP learning algorithm. The hidden layer neurons operate in an adaptive threshold scheme in an unsupervised manner using an IP learning rule. The hidden layer is connected to the output layer via excitatory synapses updated according to Spike Driven Synaptic Plasticity and, initiated using the Rank Order (RO) rule. The hidden layer neurons undergo a self-pruning mechanism. The third layer acts as the classifier and can evolve new neurons. All the hyperparameter values of the NSNN introduced are given in Table 3.

Spike encoding using address event representation. AER is a biologically inspired spike encoding mechanism used in artificial retina applications. Its simplicity, efficiency, and adaptiveness make it an attractive option for online applications. The temporal difference $\Delta t = t - t_{\text{prev}}$ (refer Eq. (1)), between two temporarily contiguous data points (denoted $x_t$ and $x_{(t-1)}$) and, a user defined threshold factor $\delta$ is used to calculate an adaptive spike threshold at each time step (refer to Eq. (2)). If the EEG voltage value
of the current time step is more than the threshold, an excitatory spike is emitted otherwise an inhibitory spike is emitted.

\[ \text{tempdiff}_t = x_t - x_{(t-1)} \]  

(1)

\[ \text{threshold} = \text{mean}(\text{tempdiff}_t) + (f \times \text{std}(\text{tempdiff}_t)) \]  

(2)

**Leaky Integrate and Fire Neuron.** The LIF neuron is commonly used in machine learning applications due to its computational tractability and the ability to produce basic spike behaviours\(^43\). Since this study involves an IP (adaptive voltage threshold) method, a wider variety of spiking behaviors can be produced than can be produced by a normal LIF\(^43\). The membrane potential change \( \frac{d v_t}{dt} \) of a LIF neuron can be modelled using a resistor-capacitor circuit and mathematically expressed using equation (3). In the equation, the time constant \( \tau_m \) is equal to the product of resistance \( R \) and capacitance \( C \). The membrane potential is given by \( v_t \) and, the input current at time \( t \) is given by \( I_t \). The resting voltage of the neuron is given by \( v_{rest} \).

\[ \tau_m \frac{d v_t}{dt} = v_{rest} - v_t + R I_t, \quad \tau_m = RC \]  

(3)

**Unsupervised learning.** In the O-NSSN, the unsupervised weight update strategy STDP\(^38\) is accompanied by an IP rule\(^45\) that adapts the threshold of hidden layer neurons individually. This combination of plasticity is a key factor in maintaining firing homeostasis and enhancing SNN performance in terms of classification accuracy and efficiency\(^45,47,69\).

\[ F(\Delta t) = A_+ \exp(-\Delta t/\tau_{pos}) \Delta t > 0 \]  

(4)
\[ F(\Delta t) = -A_- \exp(\frac{\Delta t}{\tau_{neg}}) \Delta t < 0 \]

(5)

\[ \Delta w_{ij} = \sum_a \sum_p F(t^m_j - t^n_i) \]

(6)

Equations (4) and (5) represent STDP according to Long-Term Potentiation (LTP) and Long-Term Depreciation (LTD) respectively. Both equations are functions of the time difference \( \Delta t \) between spikes. In equation (6) the pre-synaptic neuron is denoted by \( i \) and the post-synaptic by \( j \). If \( j \) fires before \( i \), \( \Delta t \) is positive leading to LTP. A reversed firing sequence leads to LTD. In equations (4) and (5), the positive and negative time constants are given by \( \tau_{pos} \) and \( \tau_{neg} \) respectively. These time constants are predetermined windows of time used for synaptic modifications. \( A_+ \) and \( A_- \) terms determine the maximum synaptic modification. The cumulative weight change \( \Delta W_{ij} \) is calculated using the spike timing of each pre-synaptic neuron from \( p \) to \( q \) and each post-synaptic neuron spiking from \( a \) to \( b \). The instantaneous spike time of each post-synaptic neuron is given by \( t^m_j \) and each pre-synaptic neuron by \( t^n_i \).

The IP rule operates simultaneously with STDP according to the two equations defined in (7). Here, the first expression of equation (7) is used to upregulate the neuron voltage thresholds and, the second to down-regulate.

\[ v_{thr}(t) = \begin{cases} v_{thr}(t - 1) + N\theta_{pos}v_{init}, & s(t - 1) = 1 \\ v_{thr}(t - 1) - N\theta_{neg}v_{init}, & otherwise \end{cases} \]

(7)

The threshold voltage of a neuron at time \( t \) is given by \( v_{thr}(t) \). If the neuron fired in the previous time step and satisfies the condition \( s(t - 1) = 1 \), then a fraction of the initial voltage \( v_{init} \) is added to the threshold voltage of the previous time step \( v_{thr}(t - 1) \). This fraction is calculated using the product of the positive learning rate \( \theta_{pos} \) and the number of neurons in the hidden layer \( N \). If a spike did not occur in
the previous time step, then the threshold voltage is lowered using the negative learning rate $\theta_{neg}$. The two learning rates are determined based on the highest neuron activation and lowest information entropy after each sample propagation.

**Structural plasticity.** The addition of new neurons in the output layer and self-pruning of the hidden layer are the two key SP techniques incorporated in the NSNN algorithm. There are no neurons in the output layer at first. During the initiation process, a predefined number of neurons are evolved. The number of samples used to evolve these initial neurons was 15 for the NSNN in this study. This set of neurons remains in the network and gets their weights updated at each sample pass. Since the NSNN operates under the test-then-train regime, if an error is made during the test phase a new neuron is evolved in the following training phase. Here, an error symbolizes the emergence of a new class or a representational change in an already known class caused by concept drift. Moreover, self-pruning also takes place in the hidden layer if an error is identified in the previous time step. This self-pruning is executed on neurons with low spiking probability since they can cause poor generalization.

$$ W_{jk}(\text{init}) = \alpha \cdot \text{mod}^{\text{order}(j,k)} $$  \hspace{1cm} (8)

$$ W_{jk}(t) = W_{jk}(\text{init}) + \sum_{t=1}^{n} d $$ \hspace{1cm} (9)

The synaptic weights from the hidden to output layer are initiated according to the RO rule given in equation 8. The initial weight between $j$ pre-synaptic neuron and $k$ post-synaptic neuron $W_{jk}(\text{init})$, is determined using a learning parameter $\alpha$ and an exponent of $\text{mod}$. The modulation factor $\text{mod}$ is determined based on the importance of the spike order. For the first spike to arrive at the synapse, $\text{order}(j,k)$ starts at 0 thereby allocating the highest possible weight and increases as the spikes arrive at
other neurons (i.e., decreases $W_{jk(\text{init})}$). Thereafter, a drift parameter $d$ is used to increase or decrease the initial weight to form a weight value at time $t$, $W_{jk}(t)$.

**Performance evaluation.** To evaluate the performance in online learning, we used the prequential accuracy metric\(^70\) with the test-then-train approach\(^21\). In test-then-train, a sample is tested first before training. This method minimizes the memory cost since samples need not be held in memory. By applying prequential memory with this approach, accuracy can be updated incrementally. The accuracies for online learning stated in the study are the final accuracy performance after 360 seconds or 72 samples.

\[
ACC_{\text{pre}}(t) = \begin{cases} 
  ACC_{\text{pre}}(t), & \text{if } t = t_{\text{init}} \\
  ACC_{\text{pre}}(t-1) + \frac{ACC_{\text{pre}}(t) - ACC_{\text{pre}}(t-1)}{t - t_{\text{init}} + 1}, & \text{else}
\end{cases}
\]

(10)

In equation (10), the classification accuracy of the NSNN at time $t$ is given by $ACC_{\text{pre}}(t)$. Here, $t_{\text{init}}$ represents the initial time point which is taken as the reference time point. For the batch learning experiments (i.e., B-RSNN), we used the standard accuracy metric which is defined as the ratio of the number of correct predictions over the total number of predictions\(^71\).

**Ethics approval and consent to participate.** All experiments were performed in accordance with the relevant guidelines and regulations. The Auckland University of Technology Ethics Committee (AUTEC) provided approval for the study on 2\(^{nd}\) October 2019 (Approval identity number: 19/231). All participants were provided with a detailed written consent form explained verbally, detailing the objectives, activities and consequences related to the study. Thereafter, a written consent was obtained from all participants individually.

**EEG Data.** The participant group consisted of 12 males with an average age of 27.92 ($\sigma = 3.09$) and 10 females with an average age of 25.9 ($\sigma = 8.20$). The EEG data were recorded over three sessions in a sound-attenuated room with a gap of at least one day between each session to prevent carry-over effects.
Each participant attended three sessions. At each session, the participant followed a sequence of steps: starting with completing the PSS-14 survey, recording two minutes of resting EEG, recording EEG while listening to an audio of either a critical, neutral or negative comment, followed by recording two minutes of resting EEG. The selection of comments (i.e., Critical, Neutral, or Negative) for a given session was selected pseudo-randomly using a computer algorithm. The perceived mental stress score for a given participant was calculated as an average across the three sessions.

EEG recording was performed with a SynAmps amplifier and a 62-channel QuickCap with electrodes configured in the international 10-20 system. Electrodes channels were: FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT8, T7, C5, C3, C1, CZ, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, P7, P5, P3, P1, PZ, P2, P4, P6, P8, PO7, PO5, PO3, POZ, PO4, PO6, PO8, CB1, O1, OZ, O2, CB2. Data was recorded at 1000Hz. For the experiments presented in this paper, only FP1, FP2, T7, and T8 electrodes were specifically selected, in an attempt to capture the spatiotemporal dynamics between PFC and amygdala during stress. EEG data preprocessing was performed in MATLAB 2019a (The Mathworks, Inc) using custom scripts that involved functions from EEGLAB plugin. Data were down-sampled offline to 256Hz. A high-pass finite impulse response (FIR) filter at 0.01Hz and a low-pass FIR filter at 50Hz were applied. Using the CleanLine function, line noise was removed before data were manually inspected for the removal of bad channels (flat or extremely noisy). The removed channels were interpolated before an independent component analysis was performed, to decompose the sample, using the runica function from the MATLAB ICA Toolbox for Psychophysiological Data Analysis. The independent components derived from ICA were inspected and muscular and ocular artifacts were removed from the data based on their activity spectra and scalp topographies.
Table 3: O-NSNN hyperparameters

<table>
<thead>
<tr>
<th>Participant identifier</th>
<th>Online learning with SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AER encoder</td>
<td>( f ) 0.7</td>
</tr>
<tr>
<td>LIF</td>
<td>( v_{\text{thresh}} ) 0.05</td>
</tr>
<tr>
<td></td>
<td>( v_{\text{rest}} ) 0</td>
</tr>
<tr>
<td></td>
<td>( R ) 1</td>
</tr>
<tr>
<td></td>
<td>( C ) 10</td>
</tr>
<tr>
<td>STDP</td>
<td>( A_+ ) 0.001</td>
</tr>
<tr>
<td></td>
<td>( A_- ) 0.001</td>
</tr>
<tr>
<td></td>
<td>( \tau_{\text{pos}} ) 10</td>
</tr>
<tr>
<td></td>
<td>( \tau_{\text{neg}} ) 10</td>
</tr>
<tr>
<td></td>
<td>( w_{\text{max}} ) 0.5</td>
</tr>
<tr>
<td></td>
<td>( w_{\text{min}} ) -0.5</td>
</tr>
<tr>
<td>IP</td>
<td>( \theta_{\text{pos}} ) 0.001</td>
</tr>
<tr>
<td></td>
<td>( \theta_{\text{neg}} ) 0.000001</td>
</tr>
<tr>
<td>Pruner</td>
<td>( sp_{\text{thresh}} ) 1</td>
</tr>
<tr>
<td>Classifier</td>
<td>( \alpha ) 1</td>
</tr>
<tr>
<td></td>
<td>( \text{mod} ) 0.8</td>
</tr>
<tr>
<td></td>
<td>( \text{drift} ) 0.001</td>
</tr>
</tbody>
</table>

**Data Availability**

Contact the authors for data access.
References


20. Lobo, J. L., Laña, I., Del Ser, J., Bilbao, M. N. & Kasabov, N. Evolving Spiking Neural Networks


30. Legenstein, R., Naeger, C. & Maass, W. What can a neuron learn with spike-timing-dependent


60. Wang, J., Belatreche, A., Maguire, L. P. & McGinnity, T. M. SpikeTemp: An Enhanced Rank-


**Acknowledgments**

**Competing interests**
The authors declare no competing interests.

**Supplementary Information**