Workload categorization for hazardous industries: the semantic modelling of multi-modal physiological data

Jemma L König[a,*], Annika Hinze[a], Judy Bowen[a]

[a]School of Computing and Mathematical Sciences, University of Waikato

Abstract

The forestry industry is one of the most hazardous industries in New Zealand, and the physical and cognitive fatigue of forestry workers has been shown to contribute to this. Physical and cognitive fatigue can be exacerbated by prolonged physical and cognitive workload. As such, we propose that the identification and mitigation of fatigue factors could reduce the risk of incident and injury in hazardous work environments. This paper introduces a semantic model for workload categorization. The model takes as input, a set of multi-modal physiological measurements, and uses parallel processing, complex event processing, and rule-based modelling to categorize a series of workloads (resting, cognitive workload, and physical workload). The model has undergone a set of evaluations, including categorization accuracy, and performance. The model has been tested under three scenarios: when a participant is resting and refraining from any physically or mentally demanding tasks; when a participant is undertaking a cognitively intensive task; and when a participant is walking, jogging, and running. The study has been conducted with participants between the ages of 22 and 39 and has shown an average accuracy of 89% for resting workload, 76% for cognitive workload, and 97% for physical workload. Finally, in this paper we discuss the application and extension of this model to predict fatigue in hazardous industries. The work described in this paper contributes to a larger research project centered on investigating technology uses in hazardous work environments.

Keywords: semantic modelling, complex event processing, parallel processing, multi-modal data, physiological data

1. Introduction

The forestry industry is one of the most hazardous industries in New Zealand, and the physical and cognitive fatigue of forestry workers has been shown to contribute to this [1, 2, 3, 4]. During the analysis of eight years worth of forestry incident data, we found that more than half (54%) of the reported incidents showed indications of worker fatigue [1]. Bentley et al. [2, 3] suggest that the likelihood of fatigue in the forestry industry is increased by the combination of early day starts and long work hours with limited breaks. Lilley et al. [4] conducted a survey that supports this. They found that 78% of workers reported experiencing fatigue, concluding the fatigue is a common problem in the New Zealand forestry industry.

Physical and cognitive fatigue can be exacerbated by prolonged physical and cognitive workload. Excessive physical exertion causes physical tiredness, which decreases physical performance [5]. Prolonged cognitive strain causes mental exhaustion, which affects mental functioning and cognitive functions [6]. Both are cumulative. However, both also lessen after a period of rest.

We propose that the identification and mitigation of fatigue factors could reduce the risk of incident and injury in hazardous work environments. To this end, this paper introduces a semantic model for workload categorization in hazardous industries. The model takes as input, a set of multi-modal physiological measurements such as electrocardiogram data, electrodermal activity, and accelerometer data. This data can be recorded, either with the use of commercial wearable sensors, such as the Polar H10 Heart Rate Sensor[1] and the Mindfield eSense Skin Response Sensor.[2], or with custom build wearable technology, such as the <Anonymised >Smart Shirt [7].

The model described in this paper uses parallel processing, complex event processing, and rule-based modelling to categorize a series of workloads. Workloads are categorized using a set of high-level events that have been extracted from the physiological data. For example, we can assume that a worker may be experiencing high physical workload if their heart rate is high, and they are moving vigorously. In contrast, we can assume that a worker may be experiencing high cognitive workload if their heart rate is low, they are not moving around, but their electrodermal activity responses are high.

This paper is structured as follows. First, we give a brief background on physical and cognitive fatigue, physiological measurements, and semantic modelling. Next, we discuss work in related fields – using physiological measures to categorize physical and mental workload. After this, we introduce our semantic model, including the parallel processing of the multi-modal physiological data, complex event processing to extract

*Corresponding author

Email address: jkonig@waikato.ac.nz (Jemma L König)

[1]https://www.polar.com/nz-en/sensors/h10-heart-rate-sensor/

[2]https://bio-medical.com/mindfield-esense-skin-response-gsr-sensor-for-iphone-andriod.html

low and high-level events, and rule-based modelling to categorize workloads. Next, we evaluate the model, both in terms of the categorization accuracy, and the real-time performance of the model. Finally, we discuss our findings and propose the application and extension of the model for use with fatigue mitigation in hazardous industries.

2. Background

As discussed in Section 1, here we give a brief background on physical and cognitive fatigue, physiological measurements, and semantic modelling.

2.1. Physical and cognitive fatigue

Fatigue is a physiological state characterised by a reduction in performance capacity due to physical or cognitive depletion. Physical fatigue, bought on by extreme physical exertion, can cause physical tiredness, and has an impact on physical performance [5]. Cognitive fatigue, bought on by long-term cognitive strain, causes mental exhaustion, which affects mental and cognitive functions [6]. Both are cumulative, and increase the more a task is performed. However, both also lessen after a period of rest.

In the forestry industry, both physical and cognitive fatigue play a role in the occurrence of incidents. Physical fatigue has been identified in incidents such as “slip, trip, and fall”, while cognitive fatigue is mentioned as a contributing factor in incident causes such as “a lack of concentration” or “errors in judgement” [8, 1, 9, 10]. Driscoll et al. [11] report “errors of judgement” as one of the most common causes of death among forestry workers, while peaks in incident occurrences around mid-morning and mid-afternoon have been identified by Bentley et al. [8, 10], which they attribute to fatigue.

2.2. Physiological measurements

The work described in this paper involves the use of three physiological measurements: electrocardiogram data, electrodermal activity, and accelerometer data.

An Electrocardiogram (ECG) measures the rate at which the heart’s electrical activity varies over time. Electrocardiogram data forms a P-QRS-T wave, which represents cardiac activity [12]. Several modalities can be extracted from this wave, including heart rate and heart rate variability. In this paper, we will focus on heart rate (or beats per minute).

Electrodermal Activity (EDA), also referred to as Galvanic Skin Response (GSR), measures variation in the electrical conductivity of the skin. An EDA signal comprises a background, tonic signal called Skin Conductance Level (SCL) and a rapid phasic component called Skin Conductance Responses (SCRs). EDA can be used as an indication of change in emotional and cognitive states, and has been linked to emotional and cognitive processing [13].

Finally, accelerometer data, while not technically a physiological data point, measures the acceleration of an object, in this case, the acceleration of a body. An accelerometer records acceleration on three axes (x, y, and z), which can be extrapolated out to infer activity level, steps taken, and so on [14].

2.3. Semantic modelling

Semantic Modelling is used to conceptualise data as it pertains to the real world, including depictions of the relationships between data points. Semantic models use semantic annotations and semantic reasoning to resolve complex problems [15]. Semantic modelling is often the initial stage in design and should take into account how the model will be used, while also being transparent and meaningful to the end user. Semantic models are often used on the web, one of the most well known of which is the Friend Of A Friend model (FOAF) [16]. FOAF is a simple semantic model that depicts a person and the things they make and do. It includes vocabulary for basic kinds of entities such as Person, Organisation and Group, and the relationships between them [17]. Semantic models can be expressed both visually and textually. FOAF, for example, was developed using machine-readable RDF syntax and has since been expanded into several more application-specific visualisation [18, 19, 20].

This paper outlines the use of semantic modelling properties to design a system for workload categorization. We describe how a semantic model can be used to conceptualise multi-modal physiological data as it pertains to the real world, i.e. how it pertains to workload categorization.

3. Related work

To our knowledge, semantic modelling has not yet been developed for physical and cognitive workload categorization. However, researchers have investigated other approaches for identifying workload. They have also investigated semantic modelling in other domains. This section first illustrates the types of physiological data that have been used for workload categorization. We then outline some of the alternative approaches to categorizing physical and cognitive workload (aside from semantic modelling). Finally, we touches on semantic modelling as it has been used in other areas pertaining to multimodal data.

3.1. Physiological data and workload categorization

The use of physiological data for workload categorization is not a new concept. Researchers have been investigating the use of physiological data for categorizing workload in a number of situations. Albuquerque et al. [21], for example, conducted a study that used electroencephalography (EEG), electrocardiography (ECG), breathing rate, skin temperature, electrodermal activity (EDA), and blood volume pulse (BVP) to investigate cognitive workload. The study saw 48 participants perform the NASA Multi-Attribute Task Battery (MATB-II), a cognitively intensive task that simulates simultaneous monitoring, resource management, and tracking tasks. The protocol for this study included 3 minutes of baseline monitoring, 10 minutes of the Multi-Attribute Task Battery, and 5 minutes for a subjective evaluation. This was repeated six consecutive times (six sessions) for each participant. In this study, cognitive workload was determined using the binary low vs. high MATB-II difficulty levels as ground truth. Each physiological measure was evaluated separately: EEG, temperature, EDA, BVP, ECG, and breathing rate. Electroencephalography (EEG) and breathing rate returned the most promising results.

In another example, Greco et al. [22], conducted a study that used electrodermal activity (EDA) and heart rate variability to detect muscle fatigue. The study saw 32 male participants perform two isometric force tasks. The first involved a maximum voluntary isometric contraction, where participants had a weight attached to their wrist and were asked to extend their arm out perpendicular to their body, holding the position for 5 seconds. The second involved a submaximal fatiguing contraction, where participants again had a weight attached to their wrist, but were this time asked to flex their elbow at 90 degree so that their upper arm lay along their body and their lower arm extended out perpendicular to their body. In this study, surface EMG activity was used as a baseline to distinguish between fatigued and non-fatigued participants. Any participants that showed a decrease over time in median EMG frequency were considered fatigued. These categorizations were then compared against the ECG and EDA data that was collected from each participant. First, heart rate variability was calculated using the ECG signal, of which several time and frequency domain features were extracted (e.g. LF, HF, LF/HF, etc.). The EDA signal was also decomposed into its phasic and tonic components and several features were extracted from each (e.g. STDphasic, STDtonic, MeanTonic, etc.). The results of this study indicated that heart rate variability and electrodermal activity can be used to determine physical muscle fatigue.

3.2. Current workload categorization techniques

Semantic modelling has not yet been investigated for use with workload categorization. Instead, workload tends to be categorized using machine learning algorithms. Albuquerque et al. [21], for example, categorized cognitive workload using the Random Forest machine learning classifier. The classifier was trained using three different techniques: mixed-subjects, intra-subject, and leave-one-subject-out. Mixed-subjects involved pooling together the data from all subjects and using five-fold cross validation. Intra-subject involved using five-fold cross validation on the data of one subject at a time. Leave-one-subject-out involved training the classifier on the data from all subjects except one, then using the last subject for classification. Similarly, Greco et al. [22] used a Support Vector Machine (SVM) machine learning classifier to categorize physical workload. In this case, the leave-one-subject-out method was used to train and test the data. In this study, Greco et al. [22] compared the values across fatigued and non-fatigued participants and the results of the study demonstrated predictions of fatigue/non-fatigue with a balanced accuracy of 83.33%.

While machine learning classifiers have been shown to produce strong predictions for workload categorization, machine learning does have some disadvantages. The most prominent of which is the need for labelled training data. Given that physiological measurement is highly personalised [23], we propose that the physiological data recorded from one person may not be a good indicator of the type of physiological data that is recorded from another. This would suggest that, when using a machine learning classifier, one would need to obtain a set of labelled data from each participant individually (i.e. the intra-subject technique). Furthermore, peoples baseline physiological data can change over time (maximum heart rate, for example, is dependant on age). This introduces challenges when using machine learning. Instead, we propose the use of semantic modelling, parallel processing, complex event processing, and a rule based modelling, in conjunction with baseline readings, to categorize cognitive and physical workload. This will allow the categorization of workload for individualised participants in real time using streaming physiological data.

3.3. Semantic modelling of multi-modal data

While semantic modelling has not yet been used for workload categorization, it has been used with multi-modal data in other domains. Reda et al. [24] for example, have investigated the use of semantic modelling for multi-modal healthcare data. In their work, they transform low-level IoT data into an enriched information model. This information model is then transformed into a specialised domain model which enables semantic reasoning to be applied. One such example that they give centres on the causalities around high blood pressure. They reason that high blood pressure increases the risk of heart problems, stroke, and other conditions, and that lifestyle factors (e.g. smoking) can increase this risk. In their work, semantic modelling has allowed them to apply semantic reasoning to classify these types of situations, and their inferences have been revised by an expert who confirmed their outcomes.

In another example, Bischof et al. [25] investigated the use of semantic modelling for multi-modal smart city data. They discuss the use of a linked-data model for streaming smart city data. This includes the semantic annotation of city map data, transport data, and entertainment data. While this work is still underway, they discuss the challenges around smart city data and how a semantic model may be well suited to solve these problems. One such problem is that of heterogeneity. The use of different IoT devices and different data providers, each of which share data in different formats and at different levels of processing, results in heterogeneous data with challenges in several dimensions. This is a challenge that we also face with wearable technology for physiological data, and that we propose can be overcome with the use of semantic modelling.

4. Describing the model

This paper introduces a semantic model for workload categorization. As described in Section 2.3, a semantic model conceptualises data as it pertains to the real world. Figure 1 illustrates our model and how it conceptualises workload categorization.

For this model, we have used a custom ontology, named here as Custom IoT Fitness Ontology (CIFO), that centres on FOAF and IFO. As touched on in Section 2.3, FOAF (Friend Of A Friend) is a simple semantic model that depicts a person and the things they make and do [17]. IFO (IoT Fitness Ontology) provides formal representations within the IoT fitness domain,

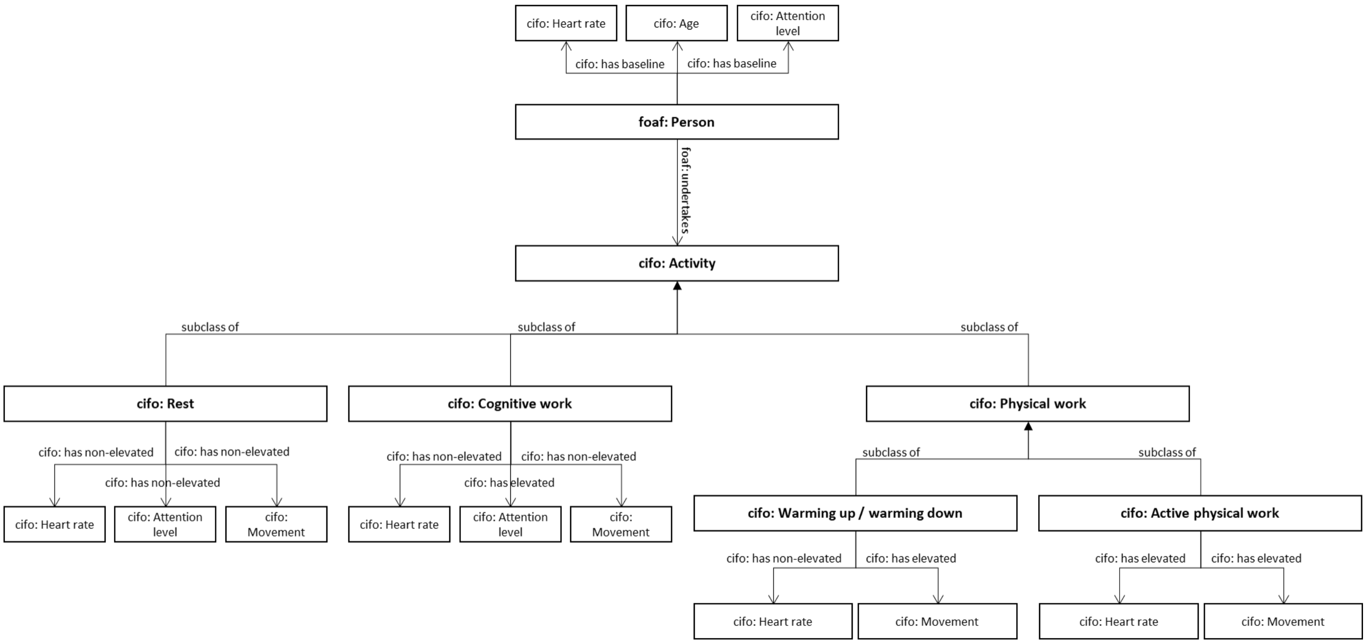


Figure 1: Semantic Model: modelling resting, cognitive and physical workload

including activity types such as physical activity, and vitals such as heart rate [24, 26].

As shown in the figure, our model contains two main classes – Person and Activity. Person represents the demographic and baseline information for an individual. The attributes of the Person class include age, heart rate (baseline), and attention level (baseline). The Activity class has three subclasses: Rest, Cognitive Workload, and Physical Workload. Rest and Cognitive Workload both have three attributes: heart rate, attention level, and movement. The Physical Workload class has two subclasses: Warming up/warming down, and Active Physical Workload, both of which have two attributes: heart rate and movement.

As shown in the figure, *Resting* is categorised by a resting heart rate, a low level of cognitive attention, and a low level of movement. *Cognitive workload* is categorised by a resting heart rate, an elevated level of attention, and a low level of movement. *Physical workload* can be compartmentalised into two states: warming up and warming down, which occur before and after exercise, and active physical work. *Warming up and warming down* are categorised by a resting heart rate and an elevated level of movement. *Active physical work* is categorised by an elevated heart rate and an elevated level of movement. Finally, the categorization of workload is highly individualised. What may be considered an elevated heart rate for one person may not be for another. Likewise, an elevated level of attention for one person may not be the same for another. As shown in the model, each individual person maintains a unique resting heart rate, resting attention level, and age, which can be used to individualize the model for each person.

While we have used semantic modelling to conceptualise and visualise our approach, the implementation of our model involves three additional techniques: parallel processing, complex event processing, and rule-based modelling. Wearable devices tend to produce large quantities of heterogeneous data. Different devices and different providers produce data in a series of different formats and in different states of pre-processing. Semantic modelling is well suited to handling this heterogeneous data in so much as it provides us with a template for transforming the raw data into homogeneous semantic modelling vocabulary. As such, the model that is described in this paper focuses mainly on this data transformation. The model takes as input, a set of multi-modal physiological measurements, and uses parallel processing, complex event processing, and rule-based modelling to transform the data and categorize a series of workloads (resting workload, physical workload, and cognitive workload). This section outlines the process involved in developing the model.

Figure 2 illustrates the architecture of the software that was developed to house the model. As shown in the figure, the system comprises three components: the stream processor, the complex event processor, and the rule-based model. First, ECG, EDA, and accelerometer data is taken as input into the stream processor as three separate data streams. These data streams are then aggregated into a sliding window, which passes incremental blocks of data down into the complex event processing component. The complex event processor splits the data stream back into ECG, EDA, and accelerometer data points, and extracts low-level events that are used to generate high-level events. The high-level events are then aggregated and passed down to the rule-based model. Finally, the rule-based model uses the high-level events to assign a workload categorization to the data.

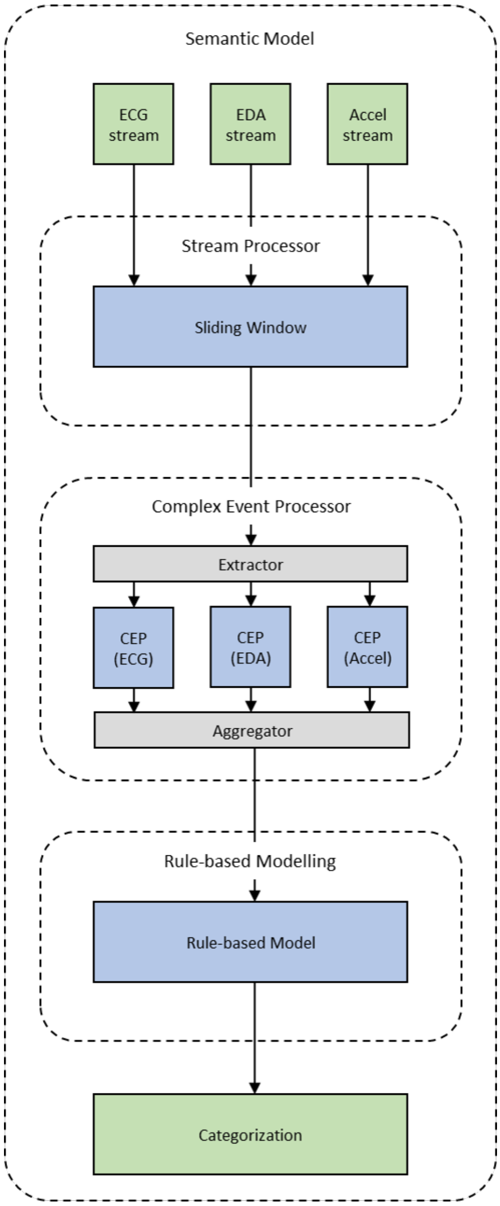


Figure 2: Software Architecture

4.1. Parallel processing

The first component involved in the modelling system is the "Stream Processor". The stream processor handles input data, and generates a sliding window of incremental data blocks that are passed on to the second component.

As already mentioned, ECG, EDA, and accelerometer data are taken in as three separate data streams. This is because each data type uses a different recording rate. ECG and EDA require fine-grained recording rates – ideally between 100-250 Hz for ECG data [27], and between 500-2000 Hz for EDA data [13]. In contrast, accelerometer data can be read at much lower rates (i.e. 5 Hz).[3]

[3]Note: Hertz measure the number of cycles per second. For example, a reading rate of 500 Hz means that there will be 500 readings per second, or one reading every 2 milliseconds

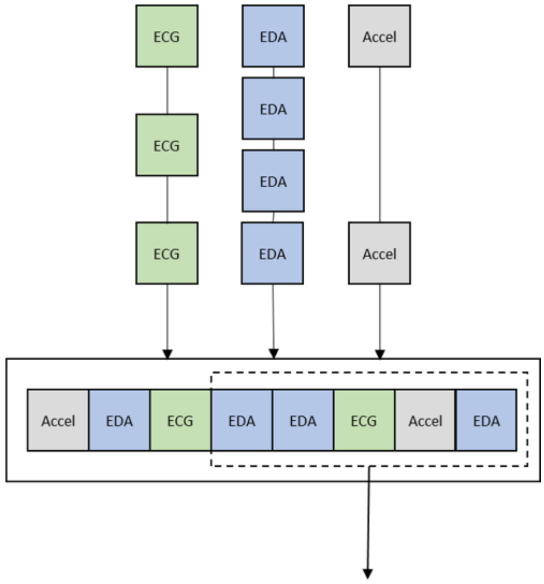


Figure 3: Component 1: Stream Processor

As illustrated in Figure 3, the Stream Processor reads in each data type using a separate thread, allowing each thread to be read in at a different rate. Each thread then pushes data onto a combined stack that allows the data to be aggregated into a single data stream. The aggregated data stream is then processed using a sliding window, with a window size of 60 seconds, and a slide of 5 seconds. Each 60 second block of data is then passed down into the Complex Event Processor.

4.2. Complex Event Processing

The second component involved in the modelling system is the "Complex Event Processor". The complex event processor takes an aggregated data stream as input and extracts low and high-level events.

Complex Event Processing involves identifying and/or extracting events within a data stream in real time. This can often be compartmentalised into low-level events, which are then combined together to form complex or high-level events [28]. This technique is applied here, first extracting low-level events from each data type, then combining these together to create high-level events.

Table 1 shows the low-level events that have been extracted from each data type. First, the intervals of time between heart beats (heart rate variability or HRV) are extracted from the ECG data. This is achieved using Scipy signal and peak detection.[4] These HRV values are then used to calculate the mean HRV, which is used to calculate the mean beats per minute (BPM or heart rate). Next, the EDA data is split into it's tonic and phasic components using Neurokit2.[5] The number of Skin Conductance Responses (SCRs), and recovery time after each, are then

[4]`https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html`

[5]`https://neurokit2.readthedocs.io/en/latest/examples/eda.html`

Table 1: Component 2: Complex Event Processor low-level events

ECG	hrvs	A list of heart rate variability values for the sliding window
	hrv_mean	The mean heart rate variability
	bpm_mean	The mean beats per minute. I.e. heart rate. Calculated by 60,000/hrv_mean
EDA	scr_peak_count	The number of Skin Conductance Responses (SCRs) in the sliding window
	scr_mean_rt	The average phasic recovery time after each SCR
Accel	vm	The vector magnitude of the 3 axis acceleration values

Table 2: Component 2: Complex Event Processor high-level events

ECG %HRmax	low	Less than 40% max HR
	medium	Less than 60% max HR
	high	Less than 80% max HR
	very high	Greater than 80% max HR
EDA %change	Low	Less than double the baseline (%change <1)
	Medium	Less than three times the baseline (%change <2)
	High	Less than four times the baseline (%change <3)
	Very High	Greater than four times the baseline (%change >3)
Accel vmRate	Low	Less than 2691
	Medium	Less than 6167
	High	Less than 9643
	Very High	Greater than 9643

extracted from the phasic signal component. Finally, the 3-axis accelerometer data is aggregated into a single vector-magnitude measurement, which represents the magnitude of acceleration over all three axes. This is achieved using SensorMotion.[6]

Once the low-level events have been extracted for each data type, they are combined together to form high-level events. Table 2 shows the high-level events that have been generated for each data type. First, the ECG low-level events are used to calculate the *Percentage of Maximum Heart Rate* (%HRmax). Although there are several heart rate based modalities that we could use, according to Karvonen and Vuorimaa [29], the %HRmax is the most suited indicator of exercise intensity. This is because the effect of age and individualised heart rate factors on the %HRmax is minimal [29]. %HRmax is calculated as follows.

$$\%HRmax = \frac{HRwork - HRrest}{HRmax - HRrest}$$

where $HRwork = HRmax - HRrest$, $HRmax = 205.8 - (0.685xAGE)$ [29], and HRrest is the worker's baseline heart rate. We have used the %HRmax to categorize the current exercise intensity. A worker's exercise intensity is *Low* when their current heart rate is less than 40% of their %HRmax. It is *Medium* when their current heart rate is less than 60% of their %HRmax. It is *High* when their current heart rate is less than 80% of their %HRmax, and it is *Very High* when their current heart rate is greater than 80% of their %HRmax.

Next, the EDA low-level events are used to calculate the *SCR Percentage Change from Baseline* (%change). As already discussed in Section 2.2, EDA can be used as an indication of change in emotional and cognitive states, and has been linked to emotional and cognitive processing [13]. Skin Conductance Responses (SCRs), which are extracted from the phasic component of an EDA signal, can occur naturally in the absence of stimuli. However, they also occur in response to stimuli or events (Event-Related SCRS, or ER-SCRs) [13]. As such, the high-level event that is used for EDA is the number of SCR peaks, more specifically, the *SCR Percentage Change from Baseline* (%change). %change is calculated as follows.

$$\%change = \frac{scr\_peak\_count - scr\_peak\_count\_base}{scr\_peak\_count\_base}$$

where *scr_peak_count* is the number of SCR peaks in the current 60 second sliding window, and *scr_peak_count_base* is the worker's baseline SCR peak count. We have used the %change to categorize the current EDA intensity. A worker's EDA intensity is *Low* when their current SCR peak count is less than double that of their baseline (%change <1). It is *Medium* when their current SCR peak count is less than three times that of their baseline (%change <2). It is *High* when their current SCR peak count is less than four times that of their baseline (%change <3), and it is *Very High* when their current SCR peak count is greater than four times that of their baseline (%change >3).

Finally, the Accelerometer low-level events are used to calculate the activity level as *Vector Magnitude Rate* (vmRate). As already discussed in section 2.2, an accelerometer records acceleration on three axes (x, y, and z), which can be extrapolated out to infer activity level, steps taken, and so on [14]. We have used the accelerometer low-level events to infer activity level using cut-points defined by Sasaki et al. [14]. A cut-point refers to the boundary between activity levels. Researchers have defined several cut-points, dependant on different participant demographics. Butte et al. [30] for example, defined cut-points for physical activity in preschoolers; Keadle et al. [31] defined cut-points for physical activity in woman; and Sasaki et al. [14] defined cut-points for physical activity in adult men and woman. It is the latter that we use in this work. We have used the vmRate to categorize the current accelerometer activity level using these cut-points. A worker's activity level is *Low* when their current vmRate is less than the cut-point of 2691. It is *Medium* when their vmRate is less than the cut-point of 6167. It is *High* when their vmRate is less than the cut-point of 9643, and it is *Very High* when their vmRate is greater than the cut-point of 9643.

[6] `https://sensormotion.readthedocs.io/en/latest/_modules/sensormotion/pa.html`

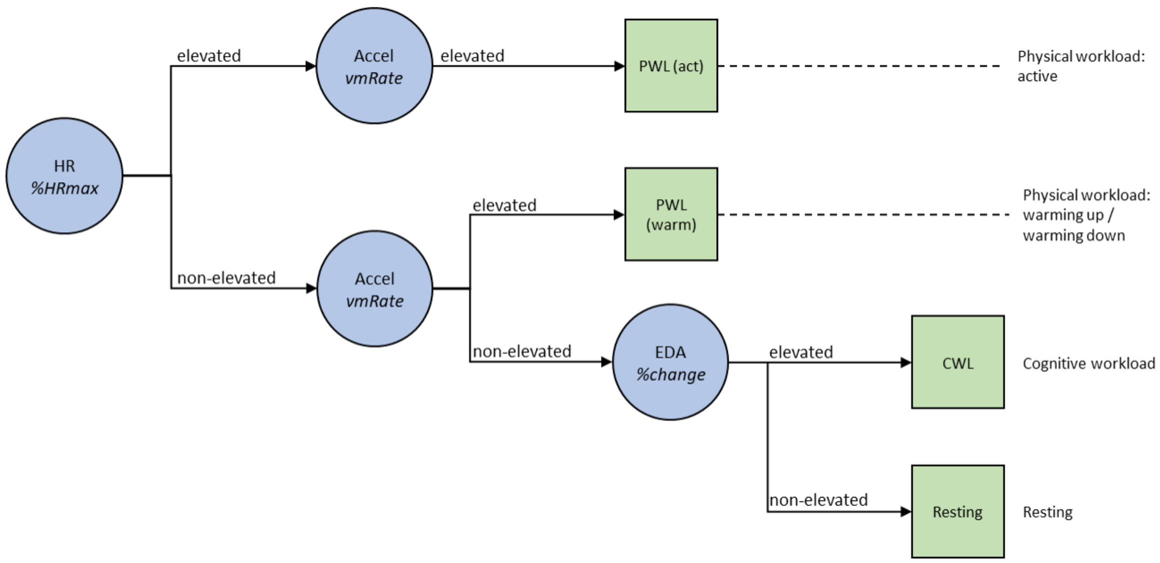


Figure 4: Rule-based model

Once the high-level events have been calculated for each data-type, these events are aggregated and passed down into the Rule-based Modelling component for the final categorization of workload.

4.3. Rule-based Modelling

The third and final component of the system is the "Rule-based Model".

As previously discussed, Semantic Modelling has been used to conceptualise and visualise our approach, and to handle the challenges of working with heterogeneous data. As such, a set of semantic-reasoning-based rules have been created based on the structure of the semantic model outlined in Figure 1. As shown earlier in the figure, and discussed in Section 4, *Resting* is categorised by a resting heart rate, a low level of cognitive attention, and a low level of movement. *Cognitive workload* is categorised by a resting heart rate, an elevated level of attention, and a low level of movement. *Physical workload* can be compartmentalised into two states: warming up and warming down, which occur before and after exercise, and active physical work. *Warming up and warming down* are categorised by a resting heart rate and an elevated level of movement. *Active physical work* is categorised by an elevated heart rate and an elevated level of movement.

As discussed in Section 4.2 and shown in Table 2, three high-level events were calculated by the Compex Event Processor: HR %hrmax, Accel vmRate, and EDA %change. Each of these high-level events contains a value that can either be considered as elevated (medium/high/very high) or non-elevated (low) (refer back to Table 2).

Figure 4 illustrates the baseline rule-based model.[7] As shown in the figure, the model categorizes the complex event processing high-level events into one of four workload categories based on whether each event's value is elevated (medium/high/very high) or non-elevated (low). The semantic reasoning behind these rules is as follows. If a worker's heart rate and activity level are both elevated, then they are most likely undertaking some form of physical activity (a categorization of *PWL (act)* or *Physical Workload: active*). If the worker's heart rate is low but their activity level is elevated, they may be warming up at the start of their physical activity stage, or warming down after physical activity (a categorization of *PWL (warm)* or *Physical workload: warming up / warming down*). If their heart rate and activity level are both low, they could be resting, or they could be undertaking a mentally taxing activity. In forestry, this can be seen in machine operators or cable-hauler operators, where workers are sitting stationary in a machine but are concentrating on operating the machine to complete their task. In this case, if their attention levels (EDA levels) are elevated, we may assume they are undertaking a mentally taxing task (a categorization of *CWL* or *Cognitive Workload*). If their attention levels are low, they may be on a break or resting (a categorization of *Rest* or *Resting*). The rules associated with this model can be seen in the pseudocode below. The complex event processing high-level events and their associated values are illustrated in brackets.

```
IF heart rate (ECG %HRmax) == non-elevated (low):
    IF movement (Accel vmRate) == non-elevated (low):
        IF attention (EDA %change) == non-elevated (low):
            Categorize as Resting
        ELSE:
            Categorize as Cognitive Workload
    ELSE:
        Categorize as Physical Workload: warming up /
             ↪ warming down
ELSE
    IF movement (Accel vmRate) == elevated (medium/high/
         ↪ very high):
        Categorize as Physical Workload: active
```

The rule-based model generates these categories for each 60 second sliding window of data, updating the categorization

[7] This model can be extended to include additional categories or take advantage of additional high-level events. This is discussed further in Section 7

every 5 seconds to align with the most recent data points (window size of 60 seconds, and slide of 5 seconds). These categorizations can then be used to make recommendations to aid in mitigating fatigue. For example, if a worker is experiencing *Cognitive Workload* for an extended period of time, being made aware of it may aid them in taking more breaks or resting intermittently to help prevent cognitive fatigue. This is discussed more in Section 7.

5. Evaluating the accuracy of the Model

The following two sections outline studies that have been conducted to evaluate the model. The first evaluation involved running a study to evaluate the accuracy of our model, collecting data to test the *Resting*, *Cognitive Workload*, and *Physical workload* categories.

5.1. Methodology

The first study was conducted with ten participants. Participants included both males and females and ranged in age from 22 to 39. Participants were recruited from a university.

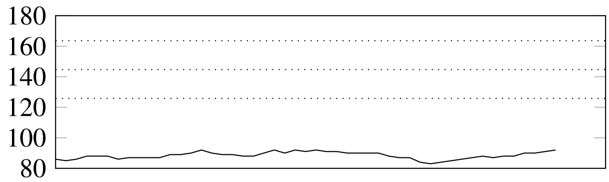
The participants undertook three tasks to produce three data sets. The first, referred to as the *resting task* involved participants sitting stationary and relaxed for 5 minutes. During this task, participants were asked not to engage or conduct any mentally or physically demanding work. Instead, they were asked to rest as they would if they were on a break at work. The second task, referred to as the *cognitive workload task* involved participants undertaking a cognitively intensive exercise for 10 minutes. During this task, participants were asked to perform the NASA Multi-Attribute Task Battery (using OpenMATB), a cognitively intensive task that simulates simultaneous monitoring, resource management, and tracking tasks [32]. The third task, referred to as the *physical workload task* involved participants exercising for 10 minutes. During this task, participants were asked to walk for 3.5 minutes, jog for 1 minute, run for 1 minute, jog for 1 minute, and walk for 3.5 minutes.

To record the physiological data, participants were asked to wear the Polar H10 Heart Rate Sensor and the Mindfield eSense Skin Response Sensor. The Polar sensor records ECG data at 130 Hz and accelerometer data at 25 Hz. The Mindfield eSense Skin Response Sensor records raw EDA data internally, but provides aggregated SCR peak counts at 5 Hz. The start time of the three data sets were synchronised before being read into the modelling system.

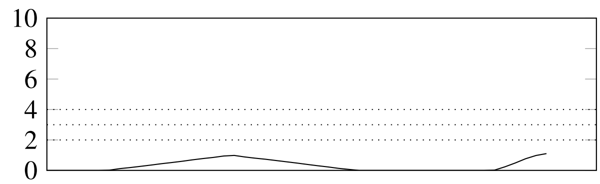
Finally, it should be noted that the tasks were undertaken within a neutral environment, rather then on-site in a hazardous industry. We recognise that this environment is different to what would be expected on-site in the forestry industry. However, as this is a preliminary study, and measuring physiological data in a hazardous work environment introduces major challenges [33], it was decided that this initial study would be conducted off-site in a more neutral environment (discussed further in Section 7). More comprehensive on-site studies are planned in the future to further validate the model.

Table 3: Baseline readings: age, baseline heart rate, baseline SCR peak count

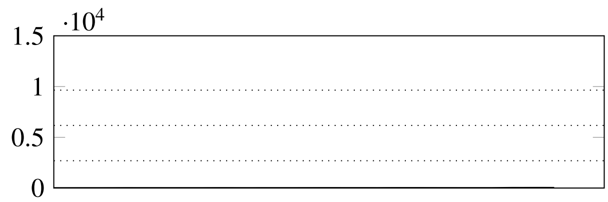
Participant	Age	Heart rate	SCR peak
P1	34	88	1.00
P2	25	71	2.02
P3	39	87	1.50
P4	37	87	0.00
P5	24	83	0.10
P6	31	88	0.65
P7	24	87	0.01
P8	31	69	0.28
P9	22	74	1.30
P10	22	82	1.34



(a) ECG events (beats per minute)



(b) EDA events (SCR peaks)



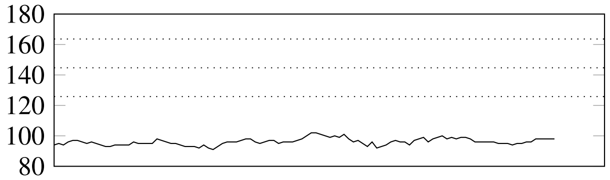
(c) Accelerometer events (vector magnitude)

Figure 5: Resting study results

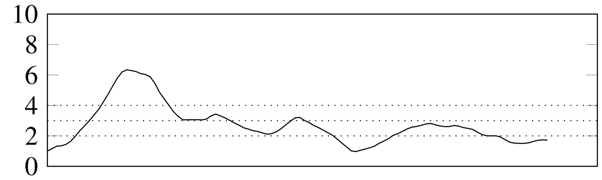
5.2. Results

As discussed in Section 4, the model uses baseline measurements to calculate some of the low-level events. For ECG, a baseline resting heart rate, and the age of the participant, are used to calculate the *Percentage of Maximum Heart Rate* (%HRmax). For EDA, the participant's baseline SCR peak count is used to calculate the *SCR Percentage Change from Baseline* (%change). These baseline measurements can be seen in Table 3.

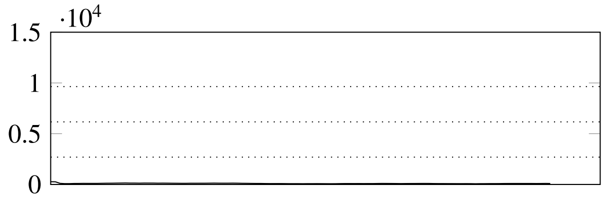
Participant's baselines were entered into the system when processing each dataset (resting, cognitive, and physical). As an example, Figure 5, Figure 6, and Figure 7 show the physiological measurements for Participant 1. Figure 5 shows the results from the *resting* dataset. Figure 5a illustrates the participant's heart rate throughout the study. The horizontal lines represent



(a) ECG events (beats per minute)

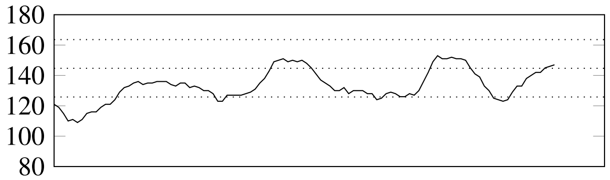


(b) EDA events (SCR peaks)

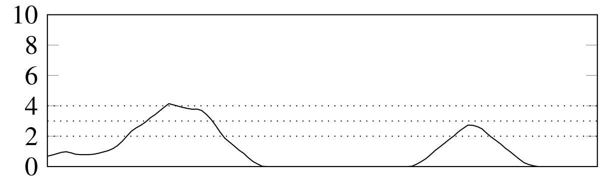


(c) Accelerometer events (vector magnitude)

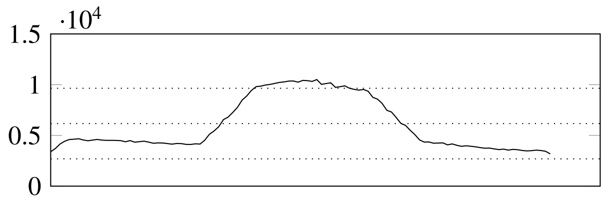
Figure 6: Cognitive study results



(a) ECG events (beats per minute)



(b) EDA events (SCR peaks)



(c) Accelerometer events (vector magnitude)

Figure 7: Physical study results

the high-level event thresholds for low, medium, high, and very high heart rates. For example, a heart rate below the bottom horizontal line represents the *low* high-level event for ECG, while a heart rate above the top horizontal line represents the *very high* high-level event. Figure 5b illustrates the participant's SCR peak count throughout the study. Again, the horizontal lines represent the high-level event thresholds for low, medium, high, and very high. An SCR peak count below the bottom horizontal line represents the *low* high-level event for EDA, and so on. Finally, Figure 5c illustrates the participant's acceleration throughout the study. Thresholds are illustrated here, again, by horizontal lines representing low, medium, high, and very high. Please note, given that the participant was not moving during this study, their accelerometer readings were very low (values between 30 and 150). As such, the readings are not visible on the graph.

As discussed in Section 4, if the participant's heart rate, scr peaks, and acceleration are low, the model should categorize them as *Resting (Rest)*. This can be seen in Figure 5, where the ECG, EDA, and accelerometer readings all fall below the bottom threshold (i.e. produce *low* high-level events) for the duration of the study. Again, as discussed in Section 4, if the participant's heart rate and acceleration are both low, they could be resting, or they could be undertaking a mentally taxing activity. In this case, if their EDA levels are high, we may assume they are undertaking a mentally taxing task (a categorization of CWL or *Cognitive Workload*). This can be seen in Figure 6, where the ECG and accelerometer readings both stay below the bottom threshold for the duration of the study, but the EDA readings move above the bottom threshold. At the most cognitively intensive point, the EDA readings (SCR peak count) climb above the top-most threshold, producing a *very-high* high-level event. The EDA readings drop below the bottom threshold at two points, suggesting that there may have been a lull in cognitive intensity mid-way through the study. Finally, as discussed in Section 4, if the participant's heart rate and activity level are both elevated, then they are most likely undertaking some form of physical activity (a categorization of PW or *Physical Workload*). If their activity level is elevated but their heart rate is low, they may be warming up at the start of their physical activity stage, or warming down afterwards (a categorization of WUWD or *Warming Up / Warming Down*). This can be seen in Figure 7, where the accelerometer readings stay above the bottom threshold (i.e. produce at least a *medium* high-level event) for the duration of the study, and the ECG readings stay above the bottom threshold for majority of the study.

Each dataset from each participant has been passed through the model to categorize workloads. Like Albuquerque et al. [21], we have used the activity type as our benchmark. For example, during the resting study, it is expected that the model will categorize data points as *resting*. During the cognitive study, it is expected that the model will categorize data points as *cognitive workload (CWL)*. During the physical study, it is expected that the model will categorize data points as either *physical workload: warming up / warming down (PWL (warm))* or *physical workload: active physical work (PWL (act))*. **??** shows the categorization results for each dataset for each participant. As can be seen in the table, the model categorised participants' resting datasets with an average accuracy of 89%. It categorised participants' cognitive datasets with an average ac-

Table 4: Workload categorization accuracy for each dataset, for each participant

Participant	Resting	Cognitive	Physical
P1	100%	72%	100%
P2	85%	93%	100%
P3	86%	100%	76%
P4	100%	24%	-
P5	86%	94%	100%
P6	67%	83%	100%
P7	100%	64%	100%
P8	100%	100%	100%
P9	88%	71%	100%
P10	76%	54%	100%
Average	**89%**	**76%**	**97%**

curacy of 76%, and it categorized participants' physical datasets with an average accuracy of 97%.

Table 5: Workload categorization results, confusion matrix

	Resting prediction	Cognitive prediction	Physical prediction
Resting Class	434	55	0
Cognitive Class	265	813	0
Physical Class	26	0	943

Table 6: Workload categorization results, precision scores

Resting	0.60
Cognitive	0.94
Physical	1.00
Combined	0.86

Table 7: Workload categorization results, recall scores

Resting	0.89
Cognitive	0.75
Physical	0.97
Combined	0.86

Table 8: Workload categorization results, f-measure scores

Resting	0.71
Cognitive	0.84
Physical	0.99
Combined	0.86

Table 5 shows the confusion matrix for workload categorization.

We can see that the resting datasets contain 434 correctly classified resting instances, and 55 incorrectly classified cognitive workload instances. Similarly, the cognitive datasets contain 813 correctly classified cognitive instances and 265 incorrectly classified resting instances. Finally, the physical datasets contain 943 correctly classified physical workload (312 PWL (warm) and 631 PWL (act)) instances and 26 incorrectly classified resting instances.

Table 6, Table 7, and Table 8 show the Precision, Recall, and F1 scores respectively. Precision is calculated as *TruePositives / (TruePositives + FalsePositives)* and provides an indication of the models ability to identify only the correct positive predictions. As can be seen in Table 6, both Cognitive Workload and Physical Workload have high Precision scores (0.94 and 1.0). In contrast, Resting has a much lower score of 0.6. This lower precision score can be attributed to the 265 False Positives, where Cognitive Workload was predicted as Resting.

Recall is calculated as *TruePositives / (TruePositives + FalseNegatives)* and provides an indication of missed positive prediction. Table 7 illustrates the Recall scores for Resting, Cognitive Workload, and Physical Workload. As can be seen in the table, both Resting and Physical Workload have high Recall scores (0.89 and 0.97). In contrast, Cognitive Workload has a slightly lower score of 0.75. This lower recall score can be attributed, again, to the 265 instances where Cognitive Workload was predicted as Resting.

F1 is calculated as *(2 * Precision * Recall) / (Precision + Recall)* and combines Precision and Recall into a single measure. Table 7 illustrates the F1 scores for Resting, Cognitive Workload, and Physical Workload. As can be seen in the table, Physical Workload has the highest F1 score (0.99), Cognitive Workload has the second highest (0.84) and Resting has the lowest (0.71). Again, the difference in F1 scores can be attributed, mostly, to the 265 instances where Cognitive Workload was predicted as Resting.

6. Evaluating the performance of the Model

The second evaluation of the model involved a study to test the performance of the model. The model must run in real-time, providing real-time categorizations without any lag. The second evaluation involved logging the completion time through different stages of the model to ensure that the model is capable of running without lag within the 5 second slide of the data window.[8]

6.1. Methodology

As illustrated earlier in Figure 2, the model described in this paper comprises three components: the parallel processing component that processes the data through a sliding window; the complex event processing component that extracts low and high-level events; and the rule-based modelling component

[8]Here, it should be noted that the complex event processing and semantic modelling components are run as threads, so that if the computation time exceeds 5 seconds, it will not affect the next sliding window. However, it is still preferable that the computation time falls well below 5 seconds so that the categorizations can be given in real time.

Table 9: Performance evaluation: Complex Event Processing

Dataset	Min (ms)	Max (ms)	Mean (ms)
Resting	57.58	195.31	92.12
Physical	57.12	215.28	90.86
Cognitive	65.37	261.10	103.95
All	57.12	261.10	95.64

Table 10: Performance evaluation: Semantic Modelling

Dataset	Min (ms)	Max (ms)	Mean (ms)
Resting	0.89	10.96	3.67
Physical	0.99	7.97	2.93
Cognitive	0.99	9.94	3.33
All	0.89	10.96	3.31

which produces the categorizations. The performance evaluation measures the computation time of the complex event processing and rule-based modelling components. In order to run in real-time, without lag, the combined computation time of these two components must remain under 5000 milliseconds (i.e. the width of the window slide).

Log messages have been added to the complex event processing and rule-based modelling components. These messages print the “current time” in milliseconds, at the beginning of the complex event processing, the end of the complex event processing, the beginning of the rule-based modelling, and the end of the rule-based modelling. This allows us to produce the computational time for: (a) complex event processing, (b) rule-based modelling, and (c) the combined computation time for both.

6.2. Results

Section 5 outlined the collection of three types of physiological data sets: resting, physical, and cognitive. These data sets have been run through the model (with logging turned on) to generate a log file that contains the computation time for the complex event processing and semantic modelling components. These datasets have been run through the model using a Dell Latitude 5300 laptop with an Intel Processor (Intel(R) Core(TM) i5-8365U CPU @ 1.60GHz 1.90 GHz) and 16 GB of RAM.

Table 9 illustrates the performance (in ms) for the Complex Event Processing component. As shown in the table, the Complex Event Processing component took an average of 95.64 ms to complete. This includes the calculation of the low-level and high-level events for the ECG, EDA, and Accelerometer data. Table 10 illustrates the performance (in ms) for the Rule-based Modelling component. As shown in the table, the Rule-based Modelling component completed significantly faster then the Complex Event Processing component. It took an average of 3.31 ms to complete. This is unsurprising, as the processing for the Rule-based Model is much more trivial then that of the Complex Event Processing component. The Rule-based Model simply takes the high-level events and categorizes them based on their values. Table 11 shows the combined processing time

Table 11: Performance evaluation: combined Complex Event Processing and Semantic Modelling

Dataset	Min (ms)	Max (ms)	Mean (ms)
Resting	60.24	198.74	95.79
Physical	58.50	223.25	93.79
Cognitive	67.37	269.08	107.28
All	58.50	269.08	98.95

Table 12: Performance evaluation: combined Complex Event Processing and Semantic Modelling

Process	Min (ms)	Max (ms)	Mean (ms)
CEP	540.50	761.77	630.58
SM	1.54	4.77	3.29
Combined	543.27	765.71	633.87

for both the Complex Event Processing, and Rule-based Modelling components. As shown in the table, when combined, the two components take an average of 98.95 ms to complete. As discussed earlier, the purpose of this study was to evaluate whether the model could run in real-time. In order to run in real-time, without lag, the combined computation time of the two components must remain under 5000 milliseconds (i.e. the width of the window slide). These results illustrate that the computation inside the model is completing well below this threshold, allowing for accurate, real-time categorization of workloads.

Finally, it should be noted that the level of computation that is undertaken to process the datasets from our study does not include the full computation of EDA low-level events. As discussed in Section 5, the Mindfield eSense Skin Response Sensor records raw EDA data internally, and then provides aggregated SCR peak counts at 5 Hz. This results in the Complex Event Processing component doing less work then may be needed otherwise. As discussed in Section 4, our model is capable of taking raw EDA data, splitting it into it’s tonic and phasic components, and calculating the SCR peaks natively. While the Mindfield eSense Skin Response Sensor does this internally, if we provide our model with raw EDA data, the processing time may increase. As such, Table 12 illustrates the processing time when our model is provided with raw EDA data. As can be seen in the table, this increases the processing time significantly (previously 98.95 ms, now 633.87 ms). However, even when considering the longest processing time (765.71), this still falls well below the threshold of 5000 ms, allowing the continued accurate, real-time categorization of workloads.

7. Discussion

In this section we first discuss the development of our model. We then discuss some of the limitations associated with it, and touch on future work.

7.1. Categorizing workload

This paper introduces a model for workload categorization. The model has been designed with a focus on semantic modelling, where we conceptualise workload categorization as it pertains to the real world. The model takes, as input, a set of multi-modal physiological measurements (ECG, EDA, and accelerometer data) and uses parallel processing, complex event processing, and rule-based modelling to transform the data and produce a workload categorization (resting workload, cognitive workload, or physical workload) for that data in real time (as discussed in Section 4).

As discussed in Section 3.2, workload categorization tends to be conducted using machine learning classifiers. While this has been shown to produce strong categorizations, it does introduce some challenges and limitations – specifically the need for larger quantities of data for both training and testing. Instead, our semantic model requires only some baseline measurements, which it can then incorporate into the real-time categorization. Furthermore, as discussed in Section 3.2, physiological measurement is highly personalised [23]. This means that the physiological data recorded from one person may not be a good indicator of the type of physiological data that is recorded from another. Furthermore, peoples baseline physiological data can change over time (maximum heart rate, for example, is dependant on age). By gathering simple baseline measurements and applying these measurements to our calculations for individual participant's, we are able to create a personalised model that is tailored to the physiological readings of individual people. Semantic modelling, parallel processing, complex event processing, and a rule based modelling, in conjunction with baseline readings, has allowed us to develop a model to categorize cognitive and physical workload. This allows for the categorization of workload for individualised participants in real time using streaming physiological data.

Finally, our model has been evaluated in two ways. First, we have evaluated the accuracy of the model by collecting a series of data sets that test different categorizations (as discussed in Section 5). Three datasets were collected for ten participants: one to evaluate resting, one to evaluate cognitive workload, and one to evaluate physical workload. The model produced an average accuracy of 89% for resting workload, 76% for cognitive workload, and 97% for physical workload. The second evaluation centred around the performance of the model (as discussed in Section 6). In order for the model to produce categorizations in real-time, the computational processing must complete within a specific time-frame. As discussed in Section 4, the parallel processing component processes data through a sliding window, with a window size of 60 seconds and a slide of 5 seconds. As such, the complex event processing and rule-based modelling components must complete inside of 5 seconds (5000 ms) so as to produce categorizations in real time. Depending on the level of processing that was undertaken during the Complex Event Processing component, the model completed in an average of 98.95 ms (with limited EDA processing) to 633.87 ms (with full EDA processing). Furthermore, the longest processing time was 765.71 ms. These results illustrate that the computation inside the model is completing well within the threshold of 5000 ms, allowing for accurate, real-time categorization of workloads.

7.2. Limitations

As with any system, our modelling system has some limitations and therefore, room for future advancements. First, the model has been developed as a Python application. This allows for it to be run on a laptop, PC, or a micro-computer such as a Raspberry Pi. However, the model currently reads physiological data points from file, rather than directly from wearable IoT sensors. Each file is read in using a separate thread, and we have incorporated a timed read-in based on the recording rate of each data set. This allows us to simulate real time recording while experimenting with a variety of wearable devices. As mentioned in Section 1, the model can currently be applied to data sets collected from commercial wearable sensors, such as the Polar H10 Heart Rate Sensor and the Mindfield eSense Skin Response Sensor, or with custom build wearable technology, such as the <Anonymised >Smart Shirt [7]. Future work will include interfacing wearable sensors with the model, allowing us to read directly from the wearable sensors themselves (discussed further in Section 7.3).

The other limitation of our model centres on the simplicity of the model itself. While this can be seen as an advantage – a simple model can give a simple solution – it can also be seen as a limitation. Contextual and/or environmental information, for example, is not currently represented within this model. As briefly discussed in Section 2.1, both ourselves and other researchers have identified mid-morning and mid-afternoon as times where fatigue may be more likely to occur within the forestry industry [8, 10]. As such, incorporating time-of-day into the model may prove beneficial, and may provide us with more meaningful insights into fatigue in the forestry industry. Similarly, temperature could play a role in the risk of incident, accident, or fatigue. While high heart rate, low acceleration, and high EDA signals can indicate anxiety or stress, the addition of high temperature could indicate the potential risks of heat stroke in forestry workers [34] (as discussed in [7]). Again, incorporating temperature into the model may likewise prove beneficial and meaningful. This again is discussed further in Section 7.3.

7.3. Future work

As discussed in the previous section, the model currently reads physiological data points from file, rather than directly from wearable IoT sensors. Although each file is read in using a separate thread, and we have incorporated a timed read-in based on the recording rate of each data set, this does not allow for physiological data to be transmitted directly from wearable sensors into our model. As such, we are currently investigating approaches for interfacing wearable sensors with the semantic model. This would allow us to read directly from the wearable sensors. The devices that were used in the participant-based study transmit data over BlueTooth (Polar H10 HR Sensor) and via the headphone jack (Mindfield eSense Skin Response Sensor). In contrast, the <Anonymised >Smart Shirt transmits data via analog and digital connections [7]. We are in the process of developing a set of software interfaces: one that connects with BlueTooth and headphone input, and another that connects via analogue and digital connections. This would allow the model to interact with a pre-defined set of sensors, including the Polar H10 HR Sensor, the Mindfield eSense Skin Response Sensor, and the <Anonymised >Smart Shirt, thereby supporting the real-time monitoring of physiological data.

Furthermore, as mentioned in the previous section, our model has been developed as a Python application. This allows for it to be run on a laptop, PC, or a micro-computer such as a Raspberry Pi. Additional future development includes the portability of this model, either onto a customised wearable device, a small portable micro computer such as the Raspberry Pi, or a mobile device. This would allow us to implement truly real-time categorization, developing vocabulary for our semantic model that can be used to provide feedback and/or notifications to the user when they are experiencing differing workloads.

Next, as discussed in Section 7.2, the model does not incorporate any contextual (time-of-day) or environmental (temperature) data. Yet both time-of-day and temperate (along with other forms of contextual and environmental data) could contribute to incidents, accidents, and fatigue within the forestry industry and other hazardous industries. As such, future work for this project includes the investigation and incorporation of additional data types. This, in turn, will result in additional modelling and additional validation studies being conducted, both as preliminary neutral studies (similar to the study described in this paper) and larger on-site studies in the forestry industry and other hazardous worksites.

Finally, we propose that the identification and mitigation of fatigue factors could reduce the risk of incident and injury in hazardous work environments. While this paper describes our model to categorize different workloads (resting workload, cognitive workload, and physical workload), it does not investigate the use of this model to identify and mitigate fatigue-based incidents in the forestry industry. This is outside the scope of this paper. However, it is the next significant step in this research project. I.e. conducting larger, more longitudinal studies, onsite in the forestry industry; evaluating how physical workload, cognitive workload, stress, rest, and so on, impact the risk of fatigue related incidents and injury in hazardous work environments.

8. Conclusion

This paper introduces a model for workload categorization. The model takes, as input, a set of multi-modal physiological measurements (ECG, EDA, and accelerometer data) and uses parallel processing, complex event processing, and rule-based modelling to produce a workload categorization (resting workload, cognitive workload, physical workload) for that data in real time. This data can be recorded, either with the use of commercial wearable sensors, such as the Polar H10 Heart Rate Sensor and the Mindfield eSense Skin Response Sensor, or with custom build wearable technology, such as the <Anonymised >Smart Shirt.

The model has been evaluated in two ways. First, we have evaluated the accuracy of the model by collecting a series of data sets that test different categorizations. The study has been conducted with participants between the ages of 22 and 39 and includes a resting study, a cognitive study, and a physical study. This has resulted in an average categorization accuracy of 89% for resting workload, 76% for cognitive workload, and 97% for physical workload. Next, we evaluated the performance of the model. The model categorizes workloads using a sliding window, with a slide of 5 seconds. As such, the model must complete inside of 5 seconds (5000 ms) so as to produce categorizations in real time. The average processing time (for full EDA processing) was 633.87 ms, and the longest processing time was 765.71 ms. These results illustrate that the computation inside the model is completing well within the threshold (5000 ms), allowing for accurate, real-time categorization of workloads.

Finally, the work described in this paper contributes to a larger research project centered on investigating technology uses in hazardous work environments. Future work for this project includes the creation of software interfaces, allowing the model to communicate directly with wearable physiological sensors; extending the model to include contextual and environmental data, such as time-of-day and temperature; and using the model to investigate how physical workload and cognitive workload impact the risk of fatigue related incidents and injury in hazardous work environments.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This paper describes work that is part of a larger project on investigating technology uses in hazardous work environments, which is funded by the New Zealand Ministry of Business, Innovation Employment (UOWX1806).

Ethics Approval

Ethics was applied for and approved by the University of Waikato Human Research Ethics Committee on 19th October 2021 (HREC(Health)202151).

References

[1] Hinze A, König JL, Bowen J. Worker-fatigue contributing to workplace incidents in new zealand forestry. Journal of safety research 2021;79:304–20.

[2] Bentley TA, Parker RJ, Ashby L. Understanding felling safety in the New Zealand forest industry. Applied ergonomics 2005;36(2):165–75.

[3] Bentley TA, Parker R, Ashby L, Moore D, Tappin D. The role of the New Zealand forest industry injury surveillance system in a strategic ergonomics, safety and health research programme. Applied Ergonomics 2002;33(5):395–403.

[4] Lilley R, Feyer AM, Kirk P, Gander P. A survey of forest workers in New Zealand: Do hours of work, rest, and recovery play a role in accidents and injury? Journal of Safety Research 2002;33(1):53–71.

[5] Hawley JA. Fatigue revisited. J Sports Sci 1997;15:245–6.

[6] Marcora SM, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. Journal of applied physiology 2009;.

[7] Hinze A, Bowen J, König JL. Wearable technology for hazardous remote environments: Smart shirt and rugged iot network for forestry worker health. Smart Health 2022;23:100225.

[8] Bentley TA, Parker RJ, Ashby L. Understanding felling safety in the New Zealand forest industry. Applied ergonomics 2005;36(2):165–75.

[9] Bell JL. Changes in logging injury rates associated with use of feller-bunchers in West Virginia. Journal of safety research 2002;33(4):463–71.

[10] Bentley TA, Parker R, Ashby L, Moore D, Tappin D. The role of the New Zealand forest industry injury surveillance system in a strategic ergonomics, safety and health research programme. Applied Ergonomics 2002;33(5):395–403.

[11] Driscoll TR, Ansari G, Harrison JE, Frommer MS, Ruck EA. Traumatic work-related fatalities in forestry and sawmill workers in Australia. Journal of Safety Research 1995;26(4):221–33.

[12] Martis RJ, Acharya UR, Adeli H. Current methods in electrocardiogram characterization. Computers in biology and medicine 2014;48:133–49.

[13] Braithwaite JJ, Watson DG, Jones R, Rowe M. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. Psychophysiology 2013;49(1):1017–34.

[14] Sasaki JE, John D, Freedson PS. Validation and comparison of actigraph activity monitors. Journal of science and medicine in sport 2011;14(5):411–6.

[15] Tomlein M, Grønbæk K. Semantic model of variability and capabilities of iot applications for embedded software ecosystems. In: 2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA). IEEE; 2016, p. 247–52.

[16] Bermudez-Edo M, Elsaleh T, Barnaghi P, Taylor K. Iot-lite: a lightweight semantic model for the internet of things. In: 2016 INTL IEEE conferences on ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (uic/atc/scalcom/cbdcom/iop/smartworld). IEEE; 2016, p. 90–7.

[17] Graves M, Constabaris A, Brickley D. Foaf: Connecting people on the semantic web. Cataloging & classification quarterly 2007;43(3-4):191–202.

[18] Amith M, Fujimoto K, Mauldin R, Tao C. Friend of a friend with benefits ontology (foaf+): extending a social network ontology for public health. BMC Medical Informatics and Decision Making 2020;20(10):1–14.

[19] El Kassiri A, Belouadha FZ. A foaf ontology extension to meet online social networks presentation and analysis. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE; 2017, p. 3056–61.

[20] Olca E, Can O. Extending foaf and relationship ontologies with consent ontology. In: 2018 3rd International Conference on Computer Science and Engineering (UBMK). IEEE; 2018, p. 542–6.

[21] Albuquerque I, Tiwari A, Parent M, Cassani R, Gagnon JF, Lafond D, et al. Wauc: A multi-modal database for mental workload assessment under physical activity. Frontiers in Neuroscience 2020;14.

[22] Greco A, Valenza G, Bicchi A, Bianchi M, Scilingo EP. Assessment of muscle fatigue during isometric contraction using autonomic nervous system correlates. Biomedical Signal Processing and Control 2019;51:42–9.

[23] Bowen J, Hinze A, Griffiths C. Investigating real-time monitoring of fatigue indicators of new zealand forestry workers. Accident Analysis & Prevention 2019;126:122–41.

[24] Reda R, Piccinini F, Carbonaro A. Semantic modelling of smart healthcare data. In: Proceedings of SAI Intelligent Systems Conference. Springer; 2018, p. 399–411.

[25] Bischof S, Karapantelakis A, Nechifor CS, Sheth AP, Mileo A, Barnaghi P. Semantic modelling of smart city data 2014;.

[26] Reda R, Piccinini F, Martinelli G, Carbonaro A. Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach. Computing 2022;104(4):835–57.

[27] Kwon O, Jeong J, Kim HB, Kwon IH, Park SY, Kim JE, et al. Electrocardiogram sampling frequency range acceptable for heart rate variability analysis. Healthcare informatics research 2018;24(3):198–206.

[28] Kristiansen S, Plagemann T, Goebel V. Smooth and crispy: integrating continuous event proximity calculation and discrete event detection. In: Proceedings of the 10th ACM International Conference on Distributed and Event-based Systems. 2016, p. 153–60.

[29] Karvonen J, Vuorimaa T. Heart rate and exercise intensity during sports activities. Sports medicine 1988;5(5):303–11.

[30] Butte NF, Wong WW, Lee JS, Adolph AL, Puyau MR, Zakeri IF. Prediction of energy expenditure and physical activity in preschoolers. Medicine and science in sports and exercise 2014;46(6):1216.

[31] Keadle SK, Shiroma EJ, Freedson PS, Lee IM. Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. BMC Public Health 2014;14(1):1–8.

[32] Cegarra J, Valéry B, Avril E, Calmettes C, Navarro J. Openmatb: A multi-attribute task battery promoting task customization, software extensibility and experiment replicability. Behavior research methods 2020;52(5):1980–90.

[33] König JL, Bowen J, Hinze A. Perceived fatigue and performance: investigating the gap between laboratory-based and in-situ studies. Working Paper 2022;.

[34] Akhtar MJ, Al-Nozha M, Al-Harthi S, Nouh MS. Electrocardiographic abnormalities in patients with heat stroke. Chest 1993;104(2):411–4.