## SPORT AND EXERCISE SCIENCES: FULLY FUNDED SWANSEA UNIVERSITY AND SWIM WALES PHD SCHOLARSHIP: THE APPLICATION OF ADVANCED DATA ANALYTICS TO MONITOR TRAINING LOAD AND PREDICT SUCCESS

Applicants for research degrees must attach a research proposal for your proposed project - maximum of 1500 words including references. If the proposal has already been determined (e.g. you are applying in response to an advertised studentship), please attach a copy.

It is recommended that the proposal includes some or all of the following:

* A title (or area) for the proposed research project
* Aims and objectives of the research
* An outline of the proposed methodology, including information on the research sample and methods of data collection
* A list of questions to be addressed by the research
* Significance of research
* A bibliography and brief summary of research that has already been undertaken in this field

The aim of this PhD project is to investigate the application of advanced data analytics to training load data to help improve the overall performance in international swimmers. Adequate training loads promote favourable physical and physiological adaptations, reduce the likelihood of illness and injury, and, therefore, increase the possibility of success. Collecting internal and external training loads has become a critical issue in elite sport practice and research. In this regard, monitoring athletes global training load is essential for undertaking whether athlete is positively adaptation to their training program. Currently, training and competition data is gathered routinely using an array of cameras and on-body sensors to give performance staff an insight into the overall performance process, however these data streams are analysed and reported in isolation limiting the overall value and insight. The candidate will review current data streams, and apply specific data analytics to this, to retrospectively establish their efficacy in performance prediction.

The candidate will be embedded in a high-performance sport environment and will be responsible for collecting multi-channel data from a range of devices that assess performance and monitor training load and ultimately will oversee data quality, provenance, and curation within a standardised database.

Data analytics: Random Forest classifier

Nevertheless, designing and implementing a framework that effectively supports and promotes sports performance is a challenging task for two reasons. First, a real-time acquisition and communication architecture must be developed. Second, a robust and efficient intelligent system that can support coaches’ decisions in real time must be conceived.

Additionally, although research on wearable sensor development is present in the literature, as detailed below, few intelligent swimming analytics approaches can be found.

Moreover, the combination of computationally demanding intelligent methods with the need for real time assessments, eventually on mobile devices, also leads to the development of flexible solutions that adapt to dynamic scenarios in terms of computational needs and features to be made available, e.g., performance classification or forecast of upcoming results. Intelligent methods such as Naïve Bayes, K-Nearest Neighbors, Decision Trees, Random Forests, or Support Vector Machines may be combined in ensembles, effectively responding to these flexible and heterogeneous scenarios [[14](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8348079/#B14-sensors-21-05162)].

In this paper, a swimmer performance intelligent data analytics system is proposed, which includes (i) pre-processing of raw signals; (ii) feature representation of wearable inertial sensors and biosensors; (iii) online recognition of the swimming style and turns; and (iv) post-analysis of the performance for coaching decision support, including average speed and stroke counting.

The main goal in this situation is the improvement of their performance in order to achieve better results. Proper movement inside water will definitely contribute for higher speed. Beyond vital signs, other signals are crucial for optimizing the athlete effort and thus reduce the human fatigue when subjected to intensive training or competition charges

Nevertheless, wearable devices are both less expensive and easier to set up in different scenarios and locations and allow for the acquisition of physiological data such as heart rate, breath rate, body temperature, motion, and position information, among others.

The wearable device (Tx) main component is an MCU that manages the information of the following sensors:

* three-axis accelerometer,
* gyroscope,
* magnetometer,
* heart rate sensor, and
* pulse oximeter sensor (SPO2).

he MCU used for this system is a board based on an ESP32 microprocessor that interconnects a LoRa transceiver, a Li-Ion battery charger, and an SD card slot, all with a small physical footprint. Two 1300 mAh Li-ion batteries are used to power the system. In order to reduce the case dimensions, the microprocessor board, sensors, LoRa communication system, 5.3 kHz receiver, and batteries are connected using a second, custom-designed, printed circuit board (PCB).

The athlete’s inertial motion data acquired during the training is processed by an Attitude and Heading Reference System (AHRS). An ICM-20948 Motion Tracking chip from TDK InvenSense was used due to the built-in gyroscope, accelerometer, and magnetometer.

Physiological information such as heart rate, breathing effort, or oxygen saturation is very important to understand if an athlete is training at their physical limits, allowing the coach to adapt and improve the training session parameters, which in turn improves the athlete’s performance. Therefore, in this system, heart rate and oxygen saturation are also measured by means of the two following sensors

Regarding performance monitoring systems, in [[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8348079/#B1-sensors-21-05162)], inertial measurement units (IMU) were used in a macro–micro analysis approach comprehensive enough to cover a full training session, regardless of the swimming technique used at any particular time. Several swimmers were monitored, and the approach detected swimming bouts, laps, and swimming technique at the macro level. A statistic correlation between the sensor values and the output results was proposed.

In [[3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8348079/#B3-sensors-21-05162)], a pervasive monitoring solution for physiological and biomechanical signals was proposed. A wearable IMU was used for data acquisition and later processing. The main contribution is related to the commercial availability of such system for real time analysis by coaches and teams.

* three-axis accelerometer,
* gyroscope,
* magnetometer,
* heart rate sensor, and
* pulse oximeter sensor (SPO2).

As a performance monitoring system, inertial measurement units (IMU) should be used in a micro-macro analysis approach to cover a full training session, sealed inside a waterproof case. Several swimmers will be monitored and the approach will detect swimming bouts, laps, technique at the macro level with a statistic correlation between the sensor values and the output results will be proposed. To acquire the relevant data in real time, the system needs to be divided in three modules; wearable module that acts as the transmitter (Tx), base station that acts as the receiver (Rx) and a computational application that includes the analytics engine and the graphical user interface. A 433-MHz half-duplex radio-frequency (RF) communication link will be implemented between the Tx and the Rx.

The main component of the Tx, the MCU will manage the information of the following sensors: three-axis accelerometer, gyroscope, magnetometer, heart rate sensor and pulse oximeter sensor. A memory card will also be including in the wearble device to backup all sensor data, along with a haptic stimulation transducer (allows interaction with the swimmer by means of commands received from the base station). An asynchronous serial protocol (USART) will be used to deliver data packets from the Rx receiver to the swimming analytics framework. Additionally, the base module station will connect the computer running the swimming analytics framework.

A waterproof chest band (Polar T31 coded band) will detect the heart beats and transmit them through a dedicated 5.3 kHz radio signal. Non-invasive sensors will also be placed on the skin surface (fingertips, forehead or earlobe) of the swimmers to determine the SPO2. The system used will be the MAX30102 sensor from Maxim Integrated to acquire the data, which are then used to calculate SPO2 The acquisition sampling rate was kept constant at four samples per second, and all data were sent to the Rx module every second. For each athlete, the following data were recorded and stored:

* **Heart rate**: in beats per minute;
* SPO2: in percentage;
* **Acceleration (3 axis)**: between −4 g and 4 g, with a resolution of 0.1 g;
* **Rotation (3 axis)**: between −500 deg/s and 500 deg/s, with a resolution of 1 deg/s; and
* **Heading (3 axis)**: between 0 and 7, representing N, NE, E, SE, S, SW, W, and NW, respectively.
* The acquired data was then used to build a stroke classification dataset in order to test and validate the swimming analytics framework, presented in the next section. In the experiments, 10 federated athletes competing at the national level, of both genders, with ages between 15 an 17 years old, were recorded, but data from only four athletes were used in the stroke classification dataset, as not all athletes performed all swimming styles, which are considered the most representative for stroke classification.
* The dataset was then manually annotated, and six states were considered: Stopped (0), Butterfly (1), Backstroke (2), Breaststroke (3), Freestyle (4), and Turn (5). All strokes are equally represented in the dataset, as all considered athletes swam the same distances in each swimming style.

Along with the pre-processing filtering applied to the raw signals acquired by the accelerometer and the gyroscope, a mandatory normalization is performed, limiting the output of the accelerometer to the −4 g to 4 g range, and the output of the gyroscope to −500 deg/s to 500 deg/s. Both values are given as double type values. In order to efficiently store the accelerometer data, such data are multiplied by a factor of 10 and saved in 8-bit format. A similar process is applied to the gyroscope output, with the decimal part being discarded and only the integer part saved in 16-bit format. A representation of the acquired data after pre-processing is presented in

A window size of one second will be used to achieve the best classification score regarding the F1 measure.

The use of stacking with Naïve Bayes as a meta-classifier has the best performance.

Considering only the 1 s feature representation, the one that presents the best results regarding stroke classification, Random Forest seems to be the best classifier when using ensembles, which might be related to the robustness of the Random Forest algorithm itself.

An outline of the proposed methodology, including information on the research sample and methods of data collection

At least 10 international competitive swimmers need to partake in this study in approximately 95 weeks. The training process should be divided into different preparation macrocycles, including final competitions. The macrocycles should consist of 6-14 weeks of training preparation and 1-3 weeks of competition. The data should document training loads in three zones of swim training intensity and two categories of dryland training. The first category of training should focus on the compensation and maintenance of aerobic endurance training at and slightly above the aerobic threshold (2–3 mmol/L blood lactate), the second category should focus on the developmental and overload of the aerobic endurance training and slightly above the anaerobic threshold.