PRIMARY RESEARCH

A Data Analysis platform for high altitude bio monitoring using DEVOPS principles and time series machine learning

Felix Constantin Adochiei 1, Radu Ciucu 2, Ioana Raluca Adochiei 3, Florin Ciprian Argatu 4, Ciprian-Marius Larco 5, Miron Casian 6
1, 2, 4 Faculty of Electrical Engineering, University Politehnica of Bucharest, Bucharest, Romania
3, 5 Department of Aircraft Integrated Systems and Mechanics, Military Technical Academy “Ferdinand I” Bucharest, Bucharest, Romania
6 Faculty of Automatic Control and Computer Engineering, "Gheorghe Asachi" Technical University of Iasi, Iași, Romania

Keywords
Magnetic Resonance-Imaging (MRI)
Digital Video Recorder (DVR)
Devops
Biomonitoring
Long-short term memory

Abstract
In this paper we propose a software platform for correlating real-time physiological data using microservices and time series analysis with machine learning. We use a devops type architecture with multiple containers governing specific aspects of our platform. In order to run the analytics and correlation protocol several microservices had been designed to orchestrate and process multiple data streams in real time. The data ingested by our platform is provided in an intuitive data visualization interface that binds together digital elevation models and geospatial data with the blood oxygen saturation and heart rate of a subject. The platform provides an accurate representation of a subjects physical performance in high altitude/low oxygen environments.

© 2019 The Author(s). Published by TAF Publishing.

I. INTRODUCTION

The effects of high-altitude environments range from reduced oxygen to a lower atmospheric pressure. Such effects can inflict hypoxia and have a direct impact on the cardiovascular system [1, 2, 3, 4, 5, 6].

Electrocardiogram is one of the most common evaluation techniques and offers a wide range of information about a subject’s state. It offers insights about chemical imbalances and electrophysiological conditions. Among the most common conditions are arrhythmias and HRV [3, 4, 5, 7].

HRV deals with the variations of the time intervals between the heart beats. It usually precedes a myocardia infarct. Various techniques are used for heart rate analysis, such as Ballistocardiograph (BCG), Photoplethysmography (PPG) and electrocardiogram, which is considered to be the most reliable since it allows the assessment of the full wave-form [1, 8, 9]. DEVOPS is an IEEE standard and a software design philosophy recently popularized for making web platforms highly scalable, extremely easy to automate and capable of continuous integration and improvement. The devops field is directly tied to cloud computing and artificial intelligence as the need for scalable and fast solutions is on the rise. We developed our platform using the latest and most scalable solutions on the open-source scene [10, 11].

By using data fusion procedures, satellite data can be three dimensionally reconstructed and directly rendered in the
browser window using HTML5 and WebGL 2.0 methods. Currently, wearable technologies are profiting of a raised interest from both the consumers and companies. One of the most recent trends which was theorized in 2005-2006 is the idea of a Wireless Body Area Network (WBAN) system that allows multiple intelligent sensors to relate to various wireless technologies, such as Bluetooth and Bluetooth Low Energy (BLE), Wi-Fi or ZigBee [12, 13].

By connecting multiple sensors to a mobile terminal, parameters such as oxygen saturation, heart-rate, temperature, body position and ECG could be measured, stored or sent to a server [14, 15, 16, 17].

In this paper, we proposed a platform dedicated to recording ECG, HR and SpO2 using a WBAN application for Android devices as well as tracking using GPS and BARO altitude data from the mobile terminal for real-time positioning on a 3D model of a mountain.

In the Materials and Methods section, we will describe the data types used for the tracking application, the correlation between the altitude and GPS data, the topography files used in the 3D reconstruction of the mountain area and the methods we used for the representation of HR and ECG and SpO2 on the 3D map [18, 19].

The results section will showcase each platform component. We will describe the devops development cycle and emphasize on the scalability of our solution. Furthermore, we will describe each platform component, from the frontend interface that displays real time data to the backend services which store and communicate the data through the API. Lastly, we will present our scalable solution for LSTM pattern recognition service. We will conclude the paper by discussing the performance of the platform, the devops philosophy impact over the development cycle. Moreover, we will sketch planned features based on our experience with the platform [19, 20].

II. METHODS

A. Devops

Devops is an IEEE standard that deals with the organizational and developmental processes of software platforms and automation frameworks. It is an acronym that stands for Development Operations, and it threatens all aspects of itself as services and microservices. A virtualization pipeline is usually implemented to orchestrate, schedule and deploy software platforms.

Devops is well suited for data ingestion and data processing in real time by employing persistent memory databases and fast data schemas. Each unit can be replicated and then purged from the memory while continuous delivering real time support [1, 21, 22].

B. Digital Elevation Models

Digital elevation models are either two dimensional arrays or sets of irregular networks of points. NASA provides multiple sets of data, such as SRTM and ASTER, which are then translated into 3D displacement maps in order to achieve a realistic model of the Earth [23]. Shuttle Radar Topographic Mission (SRTM) DEM data sets are obtained by satellites with radar technologies, which makes the results accurate as the atmosphere conditions do not affect electromagnetic waves and their ability to bounce back to the receiver on the satellite. SRTM data as of 2006 is 90% accurate with errors of around 16 meters outside of the US and 9 meters for the North American data sets. Since 2016, DEM NASA data sets can be obtained freely from the United States Geological Survey online platform. The data sets consist of two-dimensional greyscale images, with luminosity representing the elevation levels (Figure 1).

C. Long-Short Term Memory (LSTM)

LSTM neural networks are a type of recurrent neural networks specialized in time series prediction and pattern classification. Time series data is characterized by time step units and the periodic nature of the signal or the dataset. LSTM allows for specific parameters to become dynamically available as neuronal weights during use.

D. Hardware Materials

The acquisition equipment is designed around a TI microcontroller MSP430, with a frequency of 16 MHz and a resolution of 16 bits. The data transmission was realized by using the Bluetooth system on a chip model the Anaren A2541. The module features low power consumption of around 14mA [4, 15, 24].
The ECG Signal acquisition was done using an analog ECG interface from TI ADS1293. The chip features 3 dedicated channels with a sampling rate of 25.6 ksp/s, and is connected to the microcontroller through an SPI interface.

The wearable device is strapped to the chest of the subject by using a belt. Moreover, the electrodes were fabricated from a flexible conductor to ensure an adequate contact while the subject is moving (Figure 3).

III. RESULTS

We chose to implement our platform using a DEVOPS approach. The brickwork consists of docker container which allow for fast replication, communication and routing. The main reason behind the use of Linux containers is scalability. We can allocate a container for each subject of a study session.

The two main paradigms of our platform are the backend architecture and the frontend interface. Our backend is the support pillar for the storage, processing and data input features while the frontend allows for a detailed overview of the geographical and biological data.

A. Backend Services and Microservices

Our backend architecture is comprised of two main services:

A data ingestion service for collecting and storing real time data and a data processing service for time series classification.

The core of our services are Linux containers built using Docker. The services contain all the software components needed to run a specific task. All services communicate using a JavaScript API that returns JSON formatted data.

B. Data Ingestion Service

We developed a Linux container with a main purpose of directly connecting over the internet using a WebSocket type connection to the mobile device to establish a continuous data stream. Our WebSocket server was built using Node.js with the Express web framework. This service has a database connector in order to send each sample coming from the subject.
Our database schema was implemented with a time series structure using the document-oriented database MongoDB and a caching service called Redis. We chose Redis as a fast caching service for our machine learning and pattern classification components. Redis can store value-keys using persistent memory or RAM for fast access and a distributed approach.

Our Web Socket connection receives time series objects consisting of JSON formatted text. This data corresponds to our database schema and is identified by using a timestamp coming from the mobile device.

C. Data Processing Service

We implemented our pattern recognition service using a LSTM model. Our pattern recognition component was built using python and containerized with a special version of Docker for GPU access. The data ingested by our first service is made available through an API that provides a specific number of samples. The time step of our model is adjustable using a frontend API.

D. Frontend Interface

Our frontend interface provides a realistic overview of the monitoring conditions and the GPS tracked route.

The digital elevation model is implemented using a merged height map of a specific region applied onto a plane. The displaced model represents the geographical shapes in three dimensions. We developed the data visualization using a WebGL framework called three js.

Our interface holds two key features:

- Geographical overview
- Biological Monitoring

The geographical overview consists of Geolocation data fused together using three dimensional markers with our geographical displacement models. The geolocation data is directly extracted from the Android terminal using system routines (Android System API) and is represented by the altitude, speed, latitude and longitude.

In order to obtain a 3D model, we generated a displacement map by applying the DEM data to a two-dimensional plane resulting in a tridimensional mesh, in which vertices are displaced based on the intensity of the values contained in the data file. The displacement of the mesh was done along the z axis and was treated as a linear information since it is not considered an RGB image, nor the contents representing the coloring of the terrain but the elevation levels. The resulted 3D model can be viewed in Figure 4.

Fig. 4. Resulted 3D mesh: A. View from East, B. View from North, C. Top view
E. **GPS and BARO Altitude Data**

The GPS data provides real-time positioning of the subject, by recording the GPS coordinates and the BARO altitude data a relative accurate position in a tridimensional space can be obtained.

F. **Spatial Placement of the 3D Model**

The DEM data is selected using a square area, which is defined by 4 points represent by their latitude and longitude values. To correlate the model we used the open street maps api to download an XML file from their servers, which then we used as a domain for the GPS coordinates. The routes are created by selecting the recording points in the open-source app Viking.

After scaling the GPS data and the 3D model to an appropriate domain we placed them into the WebGL Canvas and represented the path in a tridimensional space.

The biomonitoring component provides real time information regarding the physiological signs of the subjected, acquired directly from the monitoring devices. The interface is comprised of multiple charts and graphs.

The oxygen saturation is represented using a gradient of colors applied directly to the GPS route that is plotted on the three dimensional model.

G. **API**

Our api is the bond the ties the application together, it provides information about the subject, weather and altitude. We developed our API using JSON objects and implemented our backend logic using Node.js. The system routing was integrated using the NGINX server for static assets (geographical maps, HTML styles).

H. **LSTM Model**

Our LSTM model was developed using Tensor Flow in python. We implemented this model for fast detection and recognition of blood oxygen patterns and dynamic ECG analysis (Figure 5).

The models time step is manually adjusted along with other parameters such as the weights and memory allocation. The model is directly integrated within our platform, and allows for fast matching of a set number of patterns (Figure 6).

---

**Fig. 5. Logical diagram of our services architecture**

**Fig. 6. The Three dimensional positioning API**
IV. CONCLUSION

We presented a software platform developed using scalable technologies such as Linux containers and fast JavaScript APIs with JSON objects. Our LSTM model is implanted using Python and is directly accessible during the monitoring sessions.

We believe our devops philosophy has ensured our platform for the foreseeable future. Our architecture allows for continuous integration of many more containers and many more computational models. While the satellite data provided a realistic representation we believe aerial recordings could provide a more accurate data set for 3D topography.

The SRTM data and in some cases errors 20-30 meters and many types of artefacts resulted from the radar recordings. Moreover, the mobile GPS and Barometric recordings are not very accurate, our application does not calculate a mean value or renders the whole recorded track but only a few selected points due to performance issues. An approximation algorithm is required.

In the near future we will develop a new model that could profile and predict specific patterns of physical fitness.

ACKNOWLEDGEMENT

This paper was developed within the program National Project: GNAC 2018 ARUT no. 19/06.02.2018 from Politehnica University of Bucharest.

REFERENCES


