CLOUD BASED PATIENT PRIORITIZATION AS A SERVICE IN THE PUBLIC HEALTH SECTOR OF THE DEVELOPING WORLD

ABSTRACT

Cyber-healthcare has recently emerged as a new field of medicine that builds on the advances made in sensor/actuator and RFID technologies. It is aimed at expanding Cyber-medicine beyond the sole consultation of virtual patients by Cyber doctors through the Internet. It provides new opportunities for enhancing health care in the developing world through low acquisition costs and flexible deployment, while improving accuracy by replacing manual operations by fully digitized processes. This paper proposes and evaluates the performance of a Cyber-healthcare system which is aimed at providing patient prioritization over the cloud as a public health service for the rural and urban communities of the developing world. We propose a deployment model for the proposed Cyber-healthcare system, and describe a patient prioritization process as part of its situation recognition component. The results obtained from a real experimental implementation reveal the field readiness of the off-the-shelf bio-sensor technology used by the system and the relative communication capabilities provided by the IEEE802.11 and IEEE802.15.4 protocols when deployed on the indoor and outdoor links of the implemented system. The relative efficiency of using supervised machine learning compared to unsupervised machine learning when performing patient prioritization, is also revealed through two popular algorithms: support vector machine and K-means clustering algorithms.

Index Terms— E-health, Cyber-healthcare, Internet-of-Things, Patient prioritization, Situation recognition

1 Introduction

The recent advances in sensor/actuator and radio frequency identification (RFID) technologies combined with a more health conscious world population have revolutionized the way health care is delivered in the developed world with positive impacts on patient care. However, the public health sector in many countries of the developing world has been lagging behind in this revolution. It is still plagued with many issues such as the lack of proper medical equipment, untrained staff, public hospital overcrowding, delayed response in emergency services, and unreliable and error-prone laboratory diagnostics resulting from manual clinical data capture processes.

1.1 Cyber-Helthcare

Cyber-medicine is a new field in medicine which builds around the field of medical care and the advances made in the Internet technologies to enable Cyber doctors/physicians consult and treat virtual patients via the Internet. Cyber-heelthcare extends Cyber-medicine to provide a broader perspective where the digitalization of all aspects of clinical work is used to better health care management. It includes aspects related to the technology, imaging, medications, surgery, rehabilitation, preventive measures, physical therapy, nursing homes, and medical supplies. The public health sector in both rural and urban settings of the developing world can leverage the Cyber-healthcare technology to improve health care management and service delivery. Leapfrogging from poorly prepared to adequately equipped communities, researchers of the developing world can also take advantage of the tools provided by these technologies, to advance research and thus reduce the scientific divide in the medical field. Some of the issues associated with Cyber-healthcare systems' deployments includes

Bio-sensor field readiness. While not aimed at replacing the medical practitioner, a Cyber-healthcare system is assumed to provide medical decision support by providing accurate and calibrated vital sign values. The field readiness of the bio-sensor devices used by the system is an important parameter upon which the accuracy of the system depends.

Sensor readings dissemination. In many deployment scenarios, the vital signs captured from patients are routed over a network to a processing place where situation recognition is achieved. The efficiency of the bio-sensor readings dissemination is another important parameter upon which service delivery depends.

Patient situation recognition. Besides using field-ready and calibrated bio-sensor devices, the Cyber-healthcare system is assumed to provide situation recognition and medical decision support to both patients and medical practitioners. Both objectives can be reached only through the use of intelligent software systems usually driven by machine learning algorithms. The design of such algorithms is another issue upon which successful situation recognition depends.

Many other important issues associated with digital health systems include security, privacy, inter-operability when deployed in a cloud-based infrastructure. These issues are beyond the scope of this paper.

1.2 Contribution and Outline

The main goal of this paper is to present and evaluate the performance of a Cyber-healthcare system that combines lightweight cloud computing and Internet-of-Things concepts to achieve patient prioritization also known in the medical field as the Triage system. This paper includes three contributions which are aimed to provide answers to the issues associated with Cyber-healthcare deployments. Firstly, we assess the field readiness of the sensor devices used by the proposed Cyber-helathcare system by benchmarking these sensors against the world health organization (WHO) patient scoring standard. Secondly, we evaluate the performance of the information dissemination protocols underlying the proposed system. Lastly, machine learning techniques are compared to select the most suitable algorithm for the proposed Cyber-healthcare system. Paper-based systems have been proposed to perform patient prioritization and a paper-based South African Triage Scale (SATS) has been recently incorporated into a mobile application called MTriage [1]. MTriage uses the existing South African Triage scoring systems designed based on observations and general knowledge by the medical personnel. A Cyber-healthcare model for patient prioritization based on supervised machine learning algorithms was recently proposed in [2] with the aim of providing an affordable, accurate, and efficient health care tool that can help in the health care planning, information exchange and bio-medical research. While differing from the paper-based MTriage system, our proposed prioritization system expands the work done in [2] to consider a hybrid communication model where both IEEE802.11 and IEEE802.15.4 protocols operating in the ISM frequency band are used on different communication links of the Cyber-healthcare system. Furthermore, this current work considers the relevance of using unsupervised learning as an alternative to the supervised learning model presented in [2] during the patient prioritization process.

The remainder of this paper is organized as follows. Section 2 presents the Cyber-healthcare framework and reveals the main components of the Triage system. The algorithmic solutions to the prioritization problem are presented and discussed in Section 3 while section 4 presents our conclusions.

2 The Cyber-healthcare system

A digital healthcare system is a platform that should empower people with no or limited medical training to capture and store clinical data into a digitized form, process, analyze this data and share it over the cloud as a service to the public health sector. It should also enable the capture of data in different other forms including crowd sensed data on mobile phones and on-body bio-sensed data. The cloud infrastructure will be equipped with intelligent data analysis algorithms capable of performing situation recognition in terms of patient and process prioritization and decision support using an expert system engine to help the concerned health professionals in the decision making. The medical health workers should periodically take the readings of all the patients that have not been attended to by the doctor. The system should also allow doctors to periodically monitor and access the patients data remotely from their smart devices; tablets and smart phones with no real time constraint. The information collected by the system should also be shared by health care planners for evaluation and planning and bio-medical researchers to achieve predictive patient analytics. When be deployed as an interconnected sets of medical databases, Cyber-healthcare systems provide an unprecedented opportunity to advance the discovery and treatment of new diseases and a better understanding of how the human body works [3, 4]. Such advances are boosted by the use of cloud computing technologies [5–8], [9, 10] and infrastructures as service to patients and the medical communities.

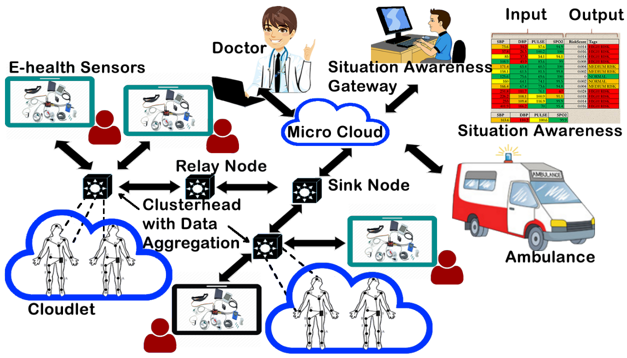

Fig. 1. The Cyber Health care System

2.1 Cyber-healthcare Deployment Scenario

The Cyber-healthcare deployment depicted by Figure 1 is built around the idea of a cloud based Internet-of-Things (Cloud-IoT) infrastructure where 1) body vital signs are captured by nurses in clinics using e-Helath kits and/or crowd sensed by body sensor networks (the cloudlets) 2) they are stored in the nodes of the sensor network and aggregated at cluster heads and 3) they are relayed to a sink node connected to a micro cloud where data analytics are performed to achieve situation management (patient prioritization, situation recognition and prediction) and the information is shared among a number of organizations or entities. The objective is to provide doctors, institutions, emergency workers, public health planners and researchers access to an integrated health care system for planning, utilization and advanced research. Some of the advantages of such a deployment when applied to the developing world context includes

- Time saving. When deployed in private or well-funded hospitals, the system can spare the nurse from the duty of going around taking readings since each patient could have a e-Health kit of medical bio-sensors and a smart device which should continuously take readings and update the medical records over-the-cloud.

- System accuracy. Using a medical bio-sensors e-Health kit provides a way of replacing the error-prone manual patient vital signs capture process by more accurate digitized procedures.

Table 1. WHO standardized TEWS calculators

	L3	L2	L1	Normal	H1	H2	H3
Systolic (mmHg)	50-59	60-79	80-99	100-130	131-160	161-200	201-300
Diastolic (mmHg)	40-44	45-49	50-59	60-85	86-90	91-110	111-140
SpO2 (%)	65-79	80-91	92-94	95-100			
Pulse (per minute)	40-44	45-49	50-59	60-100	101-120	121-180	181-250

- Cost saving. When deployed in a hospital or health care centre, a simple e-health kit could be used by many patients or shared by a community to reduce cost.
- Data access. Using a cloud-based IoT infrastructure allows easy storing and remote access to medical data or data analytics as illustrated by Figure 1 where the doctor or ambulance may access the situation awareness gateway via the Internet from anywhere and anytime.
- Data sharing. The cloud-based IoT infrastructure is a shared infrastructure between authorized units that can allow participatory consultation, medical diagnosis, health care support and many other services that could not be availed to citizens without its presence.
- Real-time updates. Using a cloud-based IoT infrastructure also enables real-time updates of patients' medical history (consultations, prescriptions, hospitalization) which are useful for future treatment validation.

As presented in figure 1, the Cyber-healthcare relies on a networked digital health infrastructure where a) the bio-sensor devices are equipped with different sensors aimed at capturing different body vital signs b) communication between nodes of the network is achieved indoor or outdoor depending on the localization of the vital signs capturing modules c) the routing of the sensor readings over these links is achieved by different protocols including WiFI and the 802.15.4 protocols and d) the micro-cloud infrastructure is equipped with a patient prioritization server and can be a component of a federated cloud infrastructure shared by several hospitals in rural settings of the developing world.

2.2 The Situation Recognition System

The aim of a situation recognition system based on the Triage model is to determine a quantitative measure of patients medical conditions and then give priorities to the most urgent cases. Some of the requirements to be met by the algorithm behind such a system include: interpretability, speed, simplicity, scalability and accuracy. The algorithm should be scalable because it should be portable enough to run on small devices, e.g. smart phones, tables, iPad, biomedical sensors and smart watches without any problems. It should also be efficient because it should be fast and accurate. A simple and easy to interpret solution is preferred. As illustrated by Figure 2, the situation recognition system proposed in this paper has four main components: a) a database b) a scoring system c) a mobile visualization application and d) a server application.

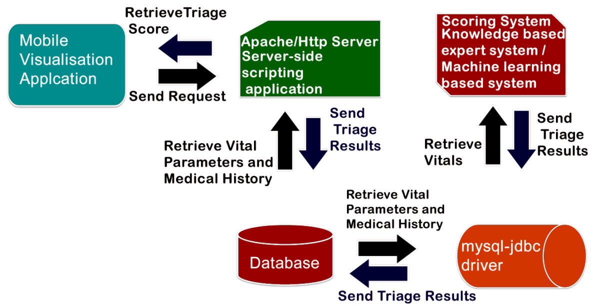

Fig. 2. High-level overview of the situation recognition.

Database: The database stores the medical record history of the patients, time stamped patient physiological parameter readings from the bio-medical sensors and also stores the Triage results (or scores) for every patient record.

Scoring System: A scoring system resulting from this WHO system depicted by Table 1 can use domain knowledge to analyze the patient data in a database; which it retrieves from the database. The scoring system will than use rule based data analysis in order to perform situation recognition. It assigns scores to every physiological parameter, and then calculates the scores using probability scoring for a certain number of readings, that is for some temporal state abstractions. The scoring system communicates with the database through a middleware MySQL-JDBC-Driver. A machine learning technique can be used to perform intelligent data analysis in order to perform situation recognition and patient treatment prioritization. When a knowledge based system is involved, it calculates the score using weights provided by the experts (physicians): no learning is involved. The final Triage scores for every record are then sent to the database. The Server-side scripting language application communicates with MySQL database and converts SQL data to JSON format. The visualization application is an android based application and it understands JSON files, therefore the mobile application retrieves the priority list in JSON format.

Mobile Visualization: This application interfaces with the server-side scripting application in the HTTP Server to retrieve the patient records from the database. The application presents the doctor with prioritized patient records and is also able to visualize patterns discovered by a machine learning technique based system in the form of graphical representations.

Serverside scripting application: The server-side scripting application interfaces with the Database and the Mobile Visualization application. The server-side scripting application interfaces with a) the database to acquire results obtained by the

Scoring System and 2) the Mobile Visualization application to provide the prioritized patient records and graphical representations.

2.3 The Triage Scoring System

Prioritization of patients is based on a Triage system that assigns scores to vital signs used as Triage parameters to quantify their severity level. Various Triage systems have been used throughout the world: the Australian Triage Scale, the Manchester Triage Scale, the Canadian Triage and Acuity Scale, and the Emergency Severity Index (ESI). All four scales have been validated for reliability and validity in adults. A standardized South African Triage was found in 2004 by the Cape Triage Group (CTG) assembled by the joint Division of Emergency Medicine at the Universities of Cape Town and Stellenbosch [11]. The South African Triage Scale was designed to work for persons of all ages by defining different Triage Early Warning Scores (TEWS) calculators for infants, children and adults. Nevertheless parameters used to Triage also differ from one Triage system to another. Table 1 shows the World Health Organization (WHO) standardized table of vital parameter risk zones used in our scoring system. It reveals values for three different risk zones and a normal zone. The risk zones L1 and H1 represent deviant or low-risk zones while L2 and H2 represent the mid-risk zones and L3 and H3 represent high-risk zones.

3 Machine Learning algorithms

Artificial intelligence has been used in medicine for many different specialized applications. For example, a reliable epileptic seizure detection model using an improved wavelet neural network was proposed in [12] while an acute ischaemic stroke prediction model from physiological time series patterns was proposed in [13] and believed to be useful in optimizing stroke recovery by manipulating physiological variables. Artificial intelligence prediction was also proposed in [14] to improve elective surgery scheduling. In this research, a general way to classify all patients into different categories irrespective of the different diseases is investigated by using machine learning algorithms. Two machine learning algorithms were selected to solve the patient prioritization problem. Their basic characteristics are described below and their performance compared in section 4.

3.1 Multivariate Linear Regression

The Matrix Algebra method has (MAM) been often used to solve problems like ours. However, the multivariate linear regression (MLR) by gradient descent was used in our work instead of the Matrix Algebra method because as opposed to the gradient descent method, MAM tends to run slow as the amount of data increases.

Algorithm description. As illustrated by Figure 3, this algorithm uses the knowledge based system to score the training

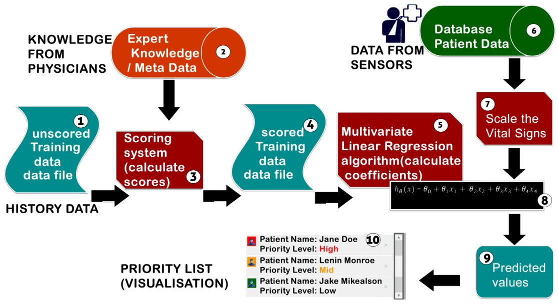

Fig. 3. Multivariate Linear Regression.

data before training. In other words, it is an improved expert system knowledge based algorithm which learns from the data, calculates the weights for each variable or generates a linear hypothesis which it uses to score the vital parameters.

3.2 K-means Clustering Algorithm

The K-means clustering algorithm considered in this paper partitions the n observations into k sets $(k < n)$ $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS), which is an error expressed by:

$$arg_s \min \sum_{i=1}^{k} \sum_{x_j \in S_i} \| x_j - \mu_i \|^2$$

where μ_i is the mean of points in S_i.

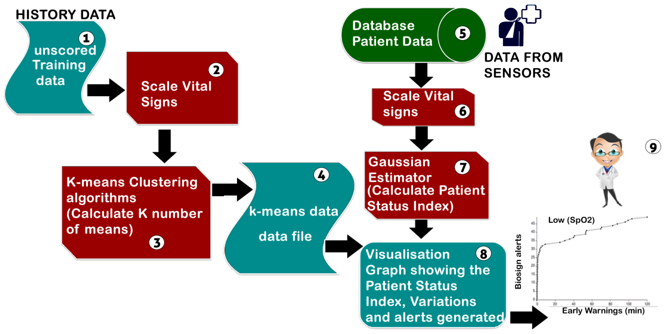

Fig. 4. K-means clustering

Algorithm description. As described by Figure 4, the K-means clustering algorithm considered in this paper partition the data into clusters and uses the Gaussian estimator (Parzen window estimator) algorithm to estimate a probability density function p(x) which is then used to calculate a patient status index. The patient status index (PSI) is expressed by the equation $PSI = log_e[1/p(x)]$.

4 System Development and Performance

The Cyber-healthcare system used in our work considered the off-the-shelf e-Health kit from Libelium as a low cost device that can be easily and quickly deployed in a rural environment. The development environment used in our experiments included: Ubuntu 13.10, Android SDK, Apache2 server (on Ubuntu), MySQL server, MySQL -Java-Bridge (MySQL-J-Connector). Different languages were used during our development. These include Java, PHP, SQL, JSON, and XML.

4.1 Sensor Field Readiness

We conducted a first set of experiments to evaluate the field readiness of the off-the-shelf e-health sensor technology with the objective of making sure that the sensor readings fall in acceptable ranges. To overcome the lengthy ethical clearance procedures aimed at protecting patient privacy through confidentiality, we used in this experiment two healthy users whose vital signs were monitored for four days. The results presented in tables 2 confirmed a normal healthy state for both individuals with 1) normal vital signs indication according to the WHO norms during the four days and 2) very little daily variations since the users did not fall sick during the experimentation. These values were calibrated against those obtained from medical equipment used by nurses in hospitals and benchmarked against the WHO values in table 1. They revealed similar values and performance patterns for non-sick individuals.

4.2 Information Dissemination

Different frequency bands have been recently recommended by the 802.15.6 standard to mitigate the interference in the ISM band resulting from co-location of bio-sensors with other devices. While this applies to urban settings of the developing world, the less crowded ISM band of the rural settings of the developing world does not need complex mitigation processes to overcome wireless interference. The focus of our work thus lies on the IEEE802.11 and IEEE802.15.4 communication standards, as they are most available and provide a low probability of interference. We conducted a second set of experiments to determine the signal strength at the receiver (RSSI) and throughput achieved by both protocols on both indoor and outdoor links of the communication network used by our Cyber-healthcare system. We used for this set of experiments the XBee Series 1 (S1) and Series 2 (S2) of the IEEE802.15.4 devices and lightweight versions of the IEEE802.11 devices of the XBee Series 6 (S6) referred to as XBee-WiFi. Assuming that multi-hop communication will be a less probable deployment option in many rural hospitals of the developing world, we considered only single hop communication with potential walls between sender and receivers for indoor communication links. For the indoor experiments, one and two doors separation were considered between sender and received of the bio-sensor information. As a worse case deployment scenario for a rural hospital, we considered for our experimentation an over-crowed building complex in Cape Town where many tenants use WiFi devices (laptops, tablets and phones) to access the Internet and communicate through social media. The experimental results are depicted by Figures 5 and 6 for the RSSI and throughput respectively. **1. The Received Signal Strength Indicator.** The signal strength at the receiver's side for the IEEE802.11 communication is constant over different distance ranges while the signal strength in the IEEE802.15.4 links decreases with distance. This is in line with the IEEE802.11 protocol which has been designed for longer communication ranges than the IEEE802.15.4 protocol. Furthermore, the IEEE802.11 reveal similar reults for both indoor and outdoor communication. This differs from the IEEE802.15.4 which shows a difference between between indoor and outdoor communication with a performance pattern where outdoor links reach longer communication ranges than indoor links and the the indoor RSSI strength reduces with the number of walls.

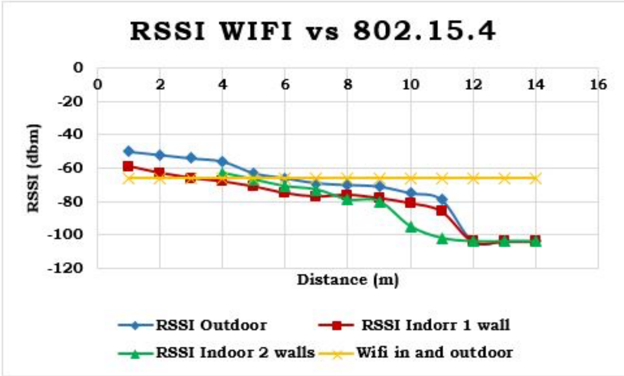

Fig. 5. WIFI vs IEEE802.15.4 RSSI.

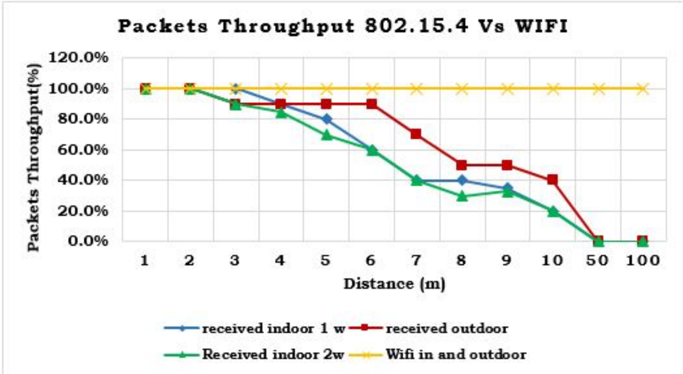

Fig. 6. WIFI vs IEEE802.15.4 Packets throughput.

1. The Wireless Communication throughput. To measure the throughput achieved over indoor and outdoor communication links, we transmitted a number of packets configured to contain the bio-sensor readings as payload and measured the ratio of the number of packet successfully received and acknowledged to the number of packets sent. Figure 6 reveals a performance patter similar to the received signal strength indicator (RSSI) where the throughput achieved indoor and outdoor are the same for the IEEE802.11 protocol. The IEEE802.15.4 reveals different performance patterns between indoor and outdoor communication and for the indoor communication with different number of walls. Similarly to the RSSI, a constant percentage of packets was received for the IEEE802.11 communication while the indoor IEEE802.15.4 links achieved a higher throughput for one wall compared to two walls. The IEEE802.11 protocol achieved higher through-

Table 2. Sensor Field Readiness: Subject One

Subject one	Day 1	Day 2	Day 3	Day 4
Systolic blood pressure	Max: 120mmHg Min: 111 mmHg Av: 115.5 mmHg Range:(111,120) mmHg	Max: 137mmHg Min: 117 mmHg Av: 127 mmHg Range:(117,137) mmHg	Max: 127mmHg Min: 121 mmHg Av: 124 mmHg Range:(121,127) mmHg	Max: 131mmHg Min: 127 mmHg Av: 129 mmHg Range:(127,131) mmHg
Diastolic blood pressure	Max: 93mmHg Min: 77mmHg Av : 85mmHg Range: (77 -93)mmHg	Max: 86mmHg Min: 76mmHg Av: 81mmHg Range: (76-86) mmHg	Max: 91mmHg Min: 83mmHg Av : 87mmHg Range:(83-91) mmHg	Max: 78mmHg Min: 69mmHg Av : 73.5mmHg Range: (69 -78) mmHg
Pulse	Max: 66 bpm Min: 61 bpm Av : 63.5 bpm Range: (61, 66) bpm	Max: 86 bpm Min: 61 bpm Av : 73.5 bpm Range: (61,86) bpm	Max: 71 bpm Min: 65 bpm Av : 68 bpm Range: (65, 71) bpm	Max: 70 bpm Min: 66 bpm Av : 78 bpm Range: (66,70) bpm
SPO2	Max: 99% Min: 98 % Average: 98.5% Range:(98-99)	Max: 90% Min: 88 % Average: 89% Range:(88-90)	Max: 95% Min: 89 % Average: 92% Range: (89-95)	Max: 99% Min: 93 % Average: 96% Range:(93-99)
Temp (^oC)	Max:37.00 Min:36.98 Average:36.49 Range: 36.98 37.00	Max:36.94 Min:36.64 Average:36.77 Range: 36.64 36.94	Max:36.98 Min:36.90 Average:36.94 Range: 36.90 36.98	Max:36.99 Min:36.81 Average:36.90 Range: 36.81 36.99
Subject two	**Day 1**	**Day 2**	**Day 3**	**Day 4**
Systolic blood pressure	Max: 134mmHg Min: 132 mmHg Av: 133.5 mmHg Range:(132,134) mmHg	Max: 131mmHg Min: 129 mmHg Av: 130 mmHg Range:(129,131) mmHg	Max: 134mmHg Min: 131 mmHg Av: 132.5 mmHg Range:(131,134) mmHg	Max: 130mmHg Min: 126 mmHg Av: 128 mmHg Range:(126,130) mmHg
Diastolic blood pressure	Max: 79mmHg Min: 78mmHg Av : 78.5mmHg Range: (78 -79)mmHg	Max: 82mmHg Min: 76mmHg Av: 79mmHg Range: (76-89) mmHg	Max: 81mmHg Min: 75mmHg Av : 78mmHg Range:(75-81) mmHg	Max: 83mmHg Min: 63mmHg Av : 73mmHg Range: (63 -83) mmHg
Pulse	Max: 88 bpm Min: 82 bpm Av : 85 bpm Range: (82, 88) bpm	Max: 79 bpm Min: 67 bpm Av : 73 bpm Range: (67,79) bpm	Max: 73 bpm Min: 67 bpm Av : 70 bpm Range: (67, 73) bpm	Max: 74 bpm Min: 66 bpm Av : 70 bpm Range: (66,74) bpm
SPO2	Max: 99% Min: 95 % Average: 97% Range:(95-99)	Max: 93% Min: 89 % Average: 91% Range:(89-93)	Max: 97% Min: 95 % Average: 96% Range: (95-97)	Max: 95% Min: 93 % Average: 94% Range:(93-95)
Temp (^oC)	Max:36.87 Min:36.53 Average:36.70 Range: 36.5336.87	Max:37.06 Min:36.64 Average:36.85 Range: 36.6437.06	Max:36.98 Min:36.92 Average:36.95 Range: 36.9236.98	Max:36.99 Min:36.91 Average:36.95 Range: 36.9136.99

put compared to the IEEE802.15.4 in both indoor and outdoor communication. Note that although the IEEE 802.11 protocol outperformed the the IEEE802.15.4 on both performance parameters, the IEEE802.15.4 deployment is still a cheaper option compared to the IEEE802.11 and a more frugal option in terms of energy consumption even when using the lightweight version of the IEEE802.11 protocol often refered to as WiFi-lite.

4.3 Situation Recognition

We conducted another set of experiments to compare the two machine learning algorithms in order to select one that will be deployed as algorithm of choice for our Cyber-healthcare system. Four different performance parameters were used to compare the algorithms: Coefficient of determination, Accuracy, Runtime and the Time Complexity. The Analysis of Variance (ANOVA) method was used to evaluate the models in this paper. The most important parameter in this method is the Coefficient of determination, denoted R^2 or r^2. It indicates how well data fit a statistical model. This value ranges from 0 to 1; the value one means the data perfectly fits the model. A value less than 0.5 indicates that the data does not fit the model. Given a matrix of features $X_1 \dots X_n$, where n is the number of features and N be the number of data points or records in a dataset while $\iota \in 1 \dots n$ is the index of the ι^{th} feature. Consider that Y is the score value before training while $\hat{Y}$ is its estimated value after training and $\tilde{Y}$ is the mean of the scores before training. This paper uses the following parameters to derive our performance parameters: i) the Regression Sum of Squares SSR ii) the Error Sum of Squares SSE iii) the Total Sum of Squares TSS and the Error Mean Square MSE. They are defined by the equations

$$SSR = \sum_{\iota=1^N} (\hat{Y} - \tilde{Y})^2 \quad (1)$$

$$SSE = \sum_{i=1^N} (Y - \hat{Y})^2 \quad (2)$$

$$TSS = \sum_{i=1^N} (Y - \tilde{Y})^2 \quad (3)$$

$$MSE = SSE/(N - df) \quad (4)$$

Table 3. Situation Recognition Results.

Parameters	Multivariate Linear Regression	K-means clustering
Coefficient of determination	0.903	n.a for unsupervised learning
Accuracy (%)	90.30	n.a for unsupervised learning
Runtime (seconds)	5.01	14.22 (for only 10 clusters exponentially grows as the number of clusters increases)
Time Complexity	O(pn+kn) where p is the dimension of each observation (input), k is the number of tasks (dimension of outputs) and n is the number of observations	Big(O) for Kmeans + Big(O) for Parzen Window $O(knT)+O(n^2)$, where k is the number of clusters, ,n is the number of points and T is the number of iterations.
Recal / Detection	0.769231	n.a for unsupervised learning
Precision	0.833333	n.a for unsupervised learning
False Rate	0.6	n.a for unsupervised learning
hline		

where df is the degree of freedom; that is the number of independent variables. If the features $X_1 \dots X_n$ are independent then $df = n$ and if they are all dependent then $df = 0$. In this paper, an assumption that all the variables are dependent was made. Hence the Error Mean Square was set to $MSE = SSE/N$.

The **Coefficient of Determination** R^2, the **Accuracy** AC and **Run Time** RT are defined by the expressions

$$R^2 = (SSTO - SSE)/SSTO \quad (5)$$

$$AC = 100 * R^2 \quad (6)$$

$$RT = EET - EST \quad (7)$$

where EET and EST are respectively the execution end time and execution start time.

The **Time Complexity** (TC) quantifies the amount of time taken by an algorithm to run as a function of the length of the string representing the input. The time complexity of an algorithm is commonly expressed using big(O) notation, which excludes coefficients and lower order terms.

To avoid using healthy users as in our previous experiment, we selected for this experimentation a real patients' dataset found from an MIT website (http://www.physio.net). This dataset was used and adapted to train and compare the two different machine learning algorithms used in this paper: Multi-linear Regression and K-means Clustering. The experimental results presented in Table 4.2 reveal that the Multivariate Linear Regression (MLR) algorithm takes approximately 5 seconds to compute the Triage priority score and has a very high accuracy of approximately 90%. The K-means clustering is an unsupervised learning algorithm which is not associated with an accuracy value but has a run time of 14.22 seconds which almost the triple of the MLR algorithm.

5 Conclusion and Future Work

A Cyber-healthcare system using off-the-shelf equipment for patient prioritization was presented in this paper as a first step towards the implementation of low cost healthcare systems for the developing countries. The off-the-shelf e-Health kit used in our experimentation was tested and found ready for field deployment. Two machine learning algorithms to solving the patient prioritization problem were described and compared and the best in terms of accuracy and processing speed was selected as algorithm of choice for our system deployment. The research made so far is satisfactory even though one patient dataset was used. The research also provided enough proof that patient vital signs follow certain patterns and more information can be extracted from these patterns. It also confirmed that machine learning techniques improve the Triage Scale accuracy by learning from the dataset and taking into consideration the smallest difference between two patient records. The scoring system can be personalized by having each individual's score calculated from their own history data. This will most likely be an achievement since each individual's vital signs vary. Increasing the number of parameters does not affect negatively the performance of the prioritization algorithm but both experimental observations and verification by medical professionals are required to determine whether only vitals are enough to determine the patients' medical conditions.

The situation recognition system presented in this paper has been built on top of a communication platform that considers single hop routing to disseminate the healthcare information from their points of collection to the micro-cloud server that handles the Triage system. When considering a larger network configuration with multi-hop routing paths, multipath routing techniques such as presented in [15, 16] can be redesigned to support QoS by having different forms of healthcare data propagated over different paths form a source to a destination. The cost-based traffic engineering techniques proposed in [17–19] will also be redesigned to balance traffic over the Cyber-healthcare communication platform to increase throughput and reduce communication delays. Deploying a long distance sensor network [20, 21] to support Cyber-healthcare network deployment in the rural settings of the developing world is another key issue that needs to be addressed as future research work.

6 References

[1] "South african medical research council (mrc) mtriage: Mobile triage application," Electron, retrieved on 8–01–16, South African Medical Research Council, http://www.openmedicineproject.org/tag/mobile-apps-awards-2015/, 2015.

[2] M. Mandava, C. Lubamba, A. Ismail, H. Bagula, and A. Bagula, "Cyberhelathcare for public healthcare in the developing world," in *Proceedings of the ISCC 2016 Conference*, June 2016, pp. 14–19.

[3] E. Eskin, "Discovering genes involved in disease and the mystery of missing heritability," *Communications of the ACM*, vol. 58, no. 10, pp. 80–87, 2015.

[4] IEEE Spectrum, "Hacking the human os: Special report," pp. 31–48, June 2015.

[5] M. Hassanalieragh, A. Page, T. Soyata, G. Sharma, M. Aktas, G. Mateos, B. Kantarci, and S. Andreescu, "Health monitoring and management using internet-of-things (iot) sensing with cloud-based processing: Opportunities and challenges," in *Services Computing (SCC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 285–292.

[6] S. Earley, "The promise of healthcare analytics," *IT Professional*, vol. 17, no. 2, pp. 7–9, 2015.

[7] L. Wang and R. Ranjan, "Processing distributed internet of things data in clouds.," *IEEE Cloud Computing*, vol. 2, no. 1, pp. 76–80, 2015.

[8] A. Page, S. Hijazi, D. Askan, B. Kantarci, and T. Soyata, "Research directions in cloud-based decision support system for health monitoring using internet-of-things driven data acquisition," .

[9] A. Benharref and M.A. Serhani, "Novel cloud and soa-based framework for e-health monitoring using wireless biosensors," *IEEE journal of biomedical and health informatics*, vol. 18, no. 1, pp. 46–55, 2014.

[10] S. Babu, M. Chandini, P. Lavanya, K. Ganapathy, and V. Vaidehi, "Cloud-enabled remote health monitoring system," in *Recent Trends in Information Technology (ICRTIT), 2013 International Conference on*. IEEE, 2013, pp. 702–707.

[11] J. Augustyn, "The south african triage scale: a tool for emergency nurses," *Professional Nursing Today*, vol. 15, no. 6, pp. 24–29, 2011.

[12] Z. Zainuddin, L.K. Huong, and O. Pauline, "Reliable epileptic seizure detection using an improved wavelet neural network," *The Australasian medical journal*, vol. 6, no. 5, pp. 308–314, 2013.

[13] Q. Zhang, Y. Xie, P. Ye, and C. Pang, "Acute ischaemic stroke prediction from physiological time series patterns," *Australas Med J*, vol. 6, no. 5, pp. 280–6, 2013.

[14] Z.S. Kargar, S. Khanna, and A. Sattar, "Using prediction to improve elective surgery scheduling," *Australasia Med J*, vol. 6, pp. 287–289, 2013.

[15] a.B. Bagula, "Modelling and implementation of qos in wireless sensor networks: A multiconstrained traffic engineering model," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, no. 1, pp. 1, 2010.

[16] A.B. Bagula and A.E. Krzesinski, "Traffic engineering label switched paths in ip networks using a pre-planned flow optimization model," in *Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2001. Proceedings. Ninth International Symposium on*. IEEE, 2001, pp. 70–77.

[17] A.B. Bagula, "Hybrid traffic engineering: the least path interference algorithm," in *Proceedings of the 2004 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*. South African Institute for Computer Scientists and Information Technologists, 2004, pp. 89–96.

[18] A.B. Bagula, "Hybrid routing in next generation ip networks," *Computer communications*, vol. 29, no. 7, pp. 879–892, 2006.

[19] A.B. Bagula, "On achieving bandwidth-aware lsp/lambdasp multiplexing/separation in multi-layer networks," *IEEE Journal on Selected Areas in Communications (JSAC): Special issue on Traffic Engineering for Multi-Layer Networks*, vol. 25, no. 5, June 2007.

[20] M. Zennaro, A. Bagula, D. Gascon, and A.B. Noveleta, "Long distance wireless sensor networks: simulation vs reality," in *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions*. ACM, 2010, p. 12.

[21] A. Bagula, M. Zennaro, G. Inggs, S. Scott, and D. Gascon, "Ubiquitous sensor networking for development (usn4d): An application to pollution monitoring," *Sensors*, vol. 12, no. 1, pp. 391–414, 2012.