THESIS PROGRESS REPORT (2nd year)

**PhD student:** Hadi JIBBAWI

**Thesis Title:** Integrating Big Data Technologies for Securing Data Collected in VANET

**Type de thèse :** Codirection at the Université de Haute-Alsace in collaboration with the Lebanese University (LU) in Lebanon

**Directeurs et encadrants de thèse :** Prof. Abdelhafid ABOUAISSA (UHA, directeur), Dr. Lhassaane IDOUGHMAR (UHA, co-directeur), Dr. Hassan HARB (AUCE, encadrant) ), Dr. Nour CHARARA (AUCE, encadrante).

**Year of Registration :** 2nd Year

**Beginning of the thesis :** December 2020

**Planned Defense Date :** December 2023

**Keywords:** VANET ; Security ; Network Attacks, Cryptography, Blockchain, Privacy-Preserving Authentication

**Introduction**

In today's digital world, intelligent transportation system (ITS) plays a very important role in making the life of the citizens easy in every facet. ITS aims to achieve higher traffic efficiency by minimizing traffic problems and controlling unpleasant events. The ITS offers pervasive and robust services in terms of providing road and traffic safeties, reducing traffic congestion and improving traffic flow, and providing entertainment services on the vehicles, etc. Recently, the integration of smart sensing devices into the vehicles has led to a revolution in the transportation and traffic systems. As a result, we witnessed the generation of a new type of networks called as Vehicular Ad Hoc Networks (VANET). Basically, VANET is a special type mobile networks (MONET) with road routes, which depends on registration mechanism, roadside units (RSUs), and onboard units (OBUs). The OBUs are the radios that are installed in every vehicle as a transmitter to communicate with each vehicle, while RSUs are installed along the street with network devices. RSUs are used to communicate with the infrastructure and contain the network devices for dedicated short-range communication (DSRC). VANETs are classified into two categories: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. The main responsibility of VANETs is to produce effective communication; basically, the nodes require specific features to acquire information, to communicate with the neighbors, and then to take decisions based on all information collected by using sensors, cameras, global positioning system (GPS) receivers, and omnidirectional antennas.

**Problem**

Despite of its advantages, VANET is subjected to numerous risks :

1. System Performance :

The salient features of VANET (e.g., varying node density, high mobility) makes it challenging to coordinate VANETs to efficiently provide services with diverse Quality of Service (QoS) requirements. Indeed, the data collected in VANET are characterized by its massive size, high speed generation, and diversity (numerical, images, and videos). Unfortunately, such characteristics provide several challenges for data analysts and decision makers; first, the traditional data warehousing systems cannot store such amount of big data streaming. Second, data processing is another challenge in transport systems due to the huge computation power needed to handle such amount of data. Third, data analysis is a very complicated task because much of the generated data is of no interest, meaningless and redundant. Hence, the system architecture is becoming key enablers for VANETS to support inter-operation among underlying heterogeneous networks, conduct resource allocation tasks, and effectively manage a vast number of mobile nodes (or users) with heterogeneous smart devices. Recently, the Big Data technologies have been proposed as efficient solutions to overcome the Big data challenges in VANET. Subsequently, the rise of Big data technologies like Hadoop ecosystems allows to build systems based on clusters that use parallel computing. This can ensure a high scalability and reliability of the collected data as well as a fast and huge data storage. In addition, the parallel computing ensures a rapid data processing especially when the volume of data becomes bigger.

1. Secure data collection and transmission:

One of the most important challenges in VANET is security. Indeed, VANET has several weaknesses against major security attacks that violates security services such as availability, confidentiality, authentication, and data integrity. Therefore, there are several kinds of attacks that threaten VANET systems including the DDoS, forgery, jamming, impersonation, malware injection, sinkhole, sybil, and replay attacks. Hence, the VANET systems must ensure that the data are collected in a secure manner, thus none of malicious sources/events are happened. Furthermore, security during transmission means that no one should be able to read or change data during along the path to the end user. Therefore, introducing new security and data encryption methods should take a great attention from researchers when dealing with data collected in VANET. Recently, with the emergence of blockchain technology, VANET has tackle a new trend in securing the data and ensuring a high level of confidentiality for the transmitted data. The blockchain technology has led to a revolution in VANET security by providing a distributed security solutions, i.e. the case of mobile vehicles communication, rather than a centrally controlled solution. Thus, with blockchain, there will be no need for a central administrator but all the vehicles are in control of all their information and transactions. Furthermore, since VANET deals with confidential mobile information and requires quick access to information, blockchain can streamline these communicated records and enable their sharing in a secure way.

**Résumé du travail fait pendant ma première année de thèse**

* Pour l’instant, on a proposé 4 contributions :
  1. **An Efﬁcient Hadoop-Based Framework for Data Storage and Fault Recovering in Large-Scale Multimedia Sensor Networks**. Cette contribution propose un plateforme efficace et robuste basée sur Hadoop pour la collecte, le traitement et le stockage de Big Data dans réseau de capteurs multimédia sans fil. L’architecture de notre plateforme se compose de plusieurs couches (ingestion, traitement, stockage et visualisation) où chaque couche possède un ensemble des outils et, parfois, quelques algorithmes (voir Figure 1). En plus, notre plateforme peut être adapté à plusieurs applications qui génèrent des données ayant plusieurs types tels que numérique, images et vidéos. Particulièrement, on a adapté notre plateforme à l’application médicale où on a besoin de collecter de données pour chaque maladie comme numérique pour surveiller les signes vitaux (par exemple la fréquence cardiaque, la saturation en oxygène, la pression artérielle, etc.), images pour différent organes (par exemple le cerveau, les poumons, etc.) et vidéos pour enregistrer les opérations faites (par exemple la chirurgie, l'appendicectomie, etc.). On a testé notre plateforme sur des données de capteurs réels [9] de différents types tout en montrant son efficacité en termes de vitesse de traitement de stockage et de régénération des données manquantes.

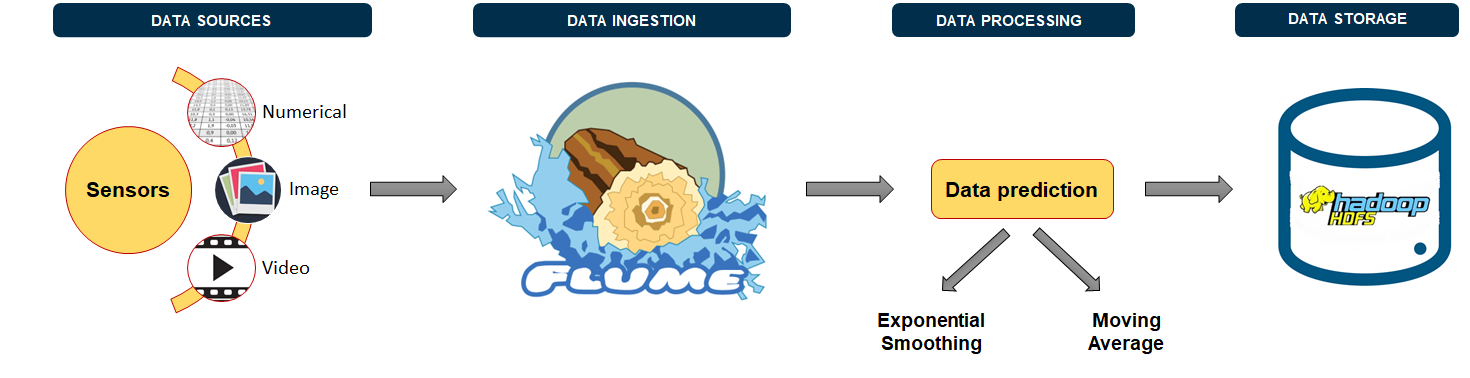


Figure 1. Hadoop-based framework for data collection and processing.

* 1. **P2D: An Efﬁcient Patient-to-Doctor Framework for Real-Time Health Monitoring and Decision Making**. Dans cet article, on propose une plateforme pour la communication entre les patients et l’équipe médicale qui permet suivre au fur et au mesures l’état des patients et de prendre une décision dans le meilleur délai. La plateforme fonctionne en deux niveaux (capteurs et coordinateur), en utilisant des donnees réelles pour la simulation [10], et propose plusieurs algorithmes qui sont décrit comme suit (voir Figure 2):

1. Niveau capteur : notre objectif est de réduire la taille de données collectées par le capteur et conserver son énergie. Cela peut faire à partir de nettoyage périodique de données collectées tout en tenant compte de la fréquence de valeurs captés afin de conserver le poids de l’information et de l’utiliser plus tard dans une analyse décisionnelle. Dans ce niveau, on propose deux algorithmes : emergency detection et adapting sampling frequency.
   * + emergency detection : cet algorithme vise à réduire la quantité de données envoyées périodiquement par chaque capteur au coordinateur en se basant sur la similarité entre les données collectées successivement. De plus, à la fin de chaque période, l’algorithme recherche un modèle de corrélation parmi les données collectées pour l’envoyer au coordinateur afin de régénérer et stocker les données pour chaque patient.
     + adapting sampling frequency : cet algorithme vise à étudier la variation de la situation du patient au cours d'un ensemble de périodes successives. Puis, il détermine la meilleure fréquence de détection du capteur en fonction du niveau de criticité du patient sans perdre l'intégrité des informations collectées.
2. Niveau coordinateur : à ce niveau-là, on propose une technique de prédiction de données qui permet d’estimer la progression de situation d’un patient en se basant sur sa situation actuelle. Nous avons utilisé la méthode de prédiction Prophet qui vise à étudier la variation de données liées aux signes vitaux d’un patient et prédire la prochaine situation.

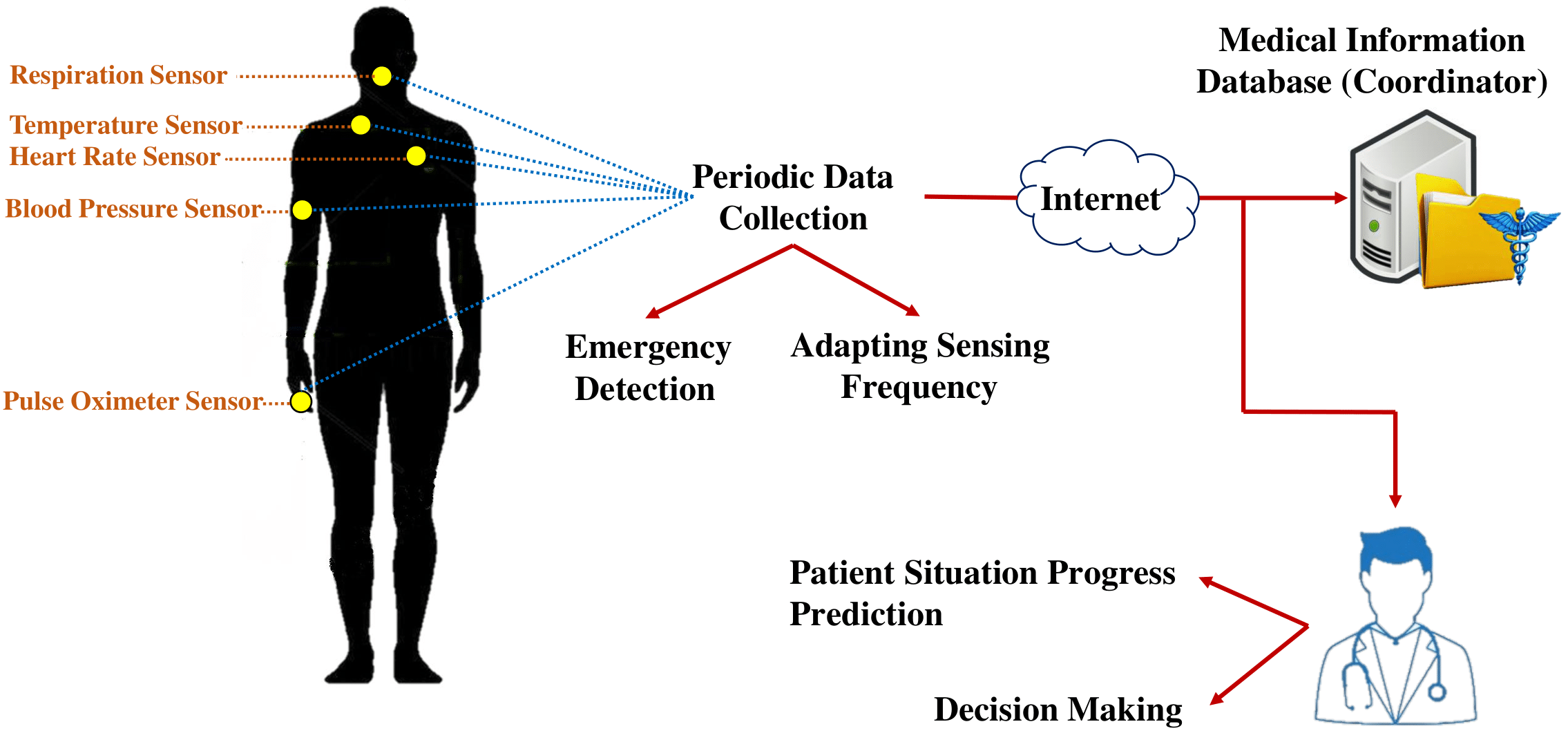


Figure 2. A multi-layer mechanism for patient monitoring and decision making in healthcare applications.

* 1. **An efficient patient classification and grouping mechanism-based framework for patients-nurse scheduling.** Cette contribution propose un mécanisme de surveillance de patients en temps réel et d’affectation des infirmières selon la criticité des patients (voir Figure 3). La première partie de notre mécanisme introduit la méthode de classification des patients qui classe les patients (entre faible, moyen et élevé) en fonction de la gravité des signes vitaux afin de suivre leur situation et de prévenir toute détérioration de la santé par une intervention médicale appropriée, et qui réduit la quantité de données transmises par chaque capteur en utilisant le principe de shapelets (voir Figure 4). Actuellement, nous travaillons dans la deuxième partie de cette mécanisme. L’objectif principale est de mieux organiser le travail des infirmières. Cette section présente un nouvel algorithme de planification qui vise à affecter les infirmières dans un hôpital aux patients selon la situation sanitaire de ses patients.

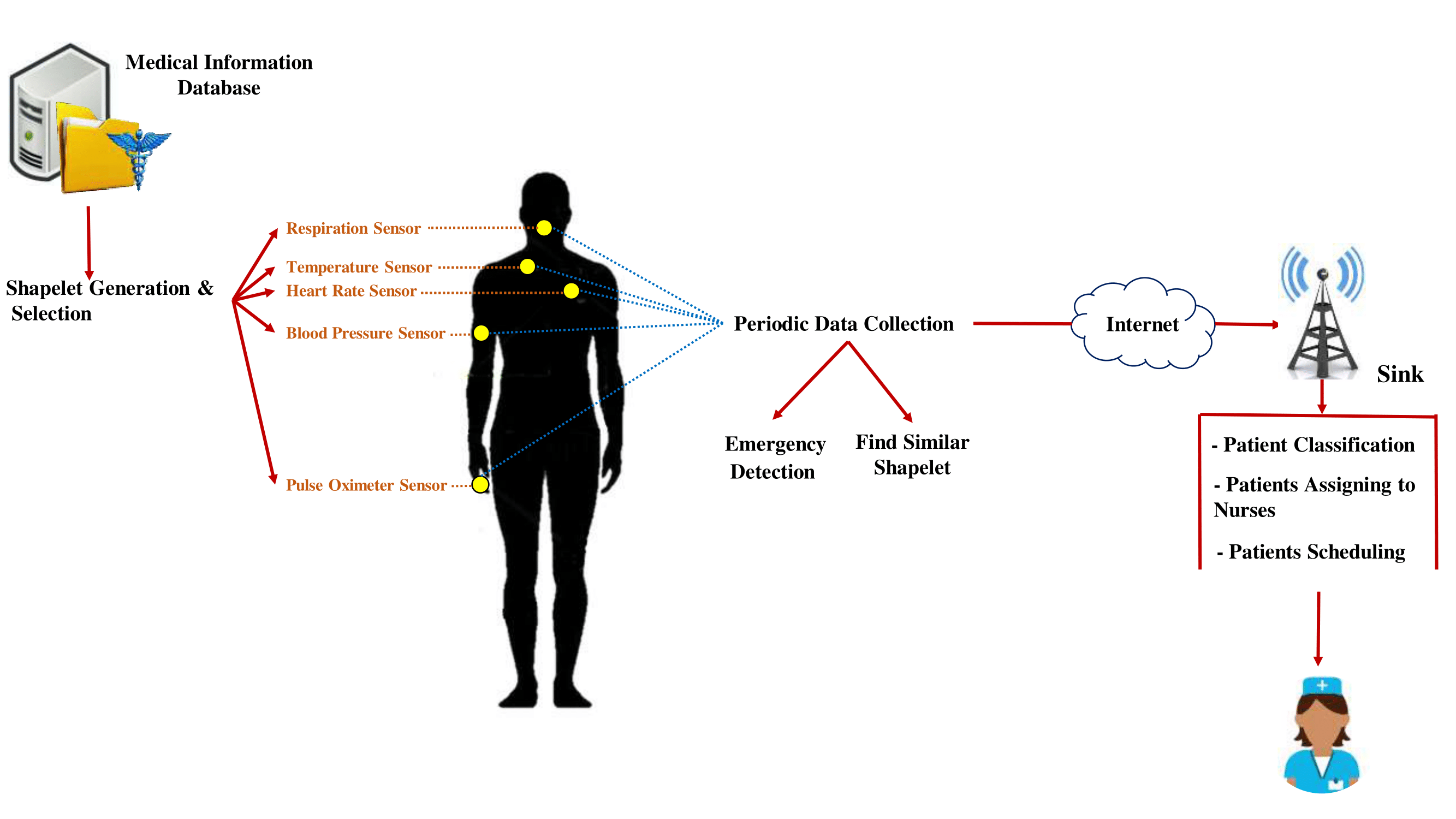
****

Figure 3. An efficient mechanism for real-time patient classification and patients-nurse scheduling.

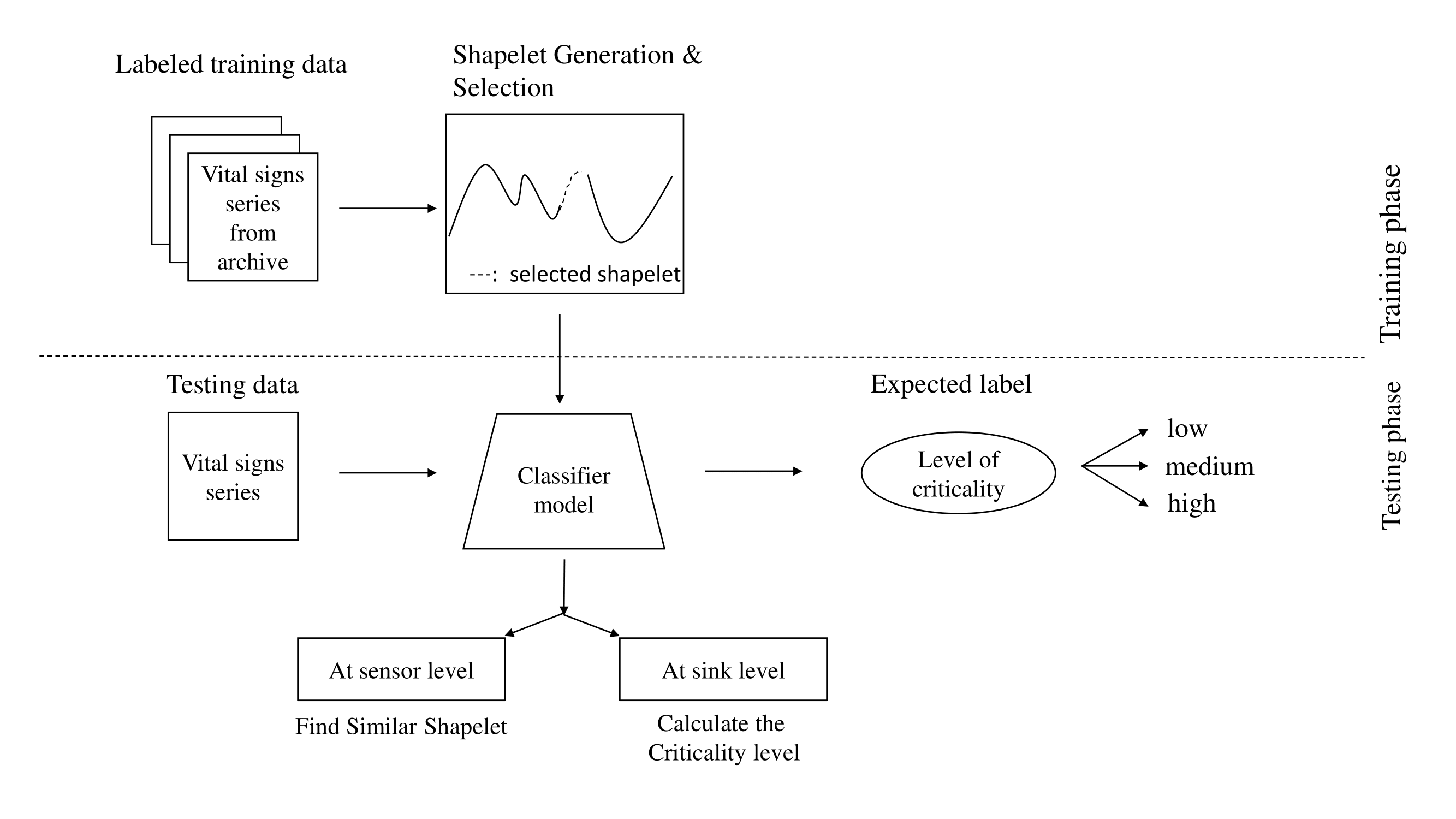


Figure 4. Patient Classification system.

* 1. **Recent Advances of Patient Monitoring in Internet of Healthcare Things: A Comparative Study.** Dans cet article, nous avons tenté d'atteindre deux objectifs principaux : Premièrement, nous présentons une étude comparative entre les récentes des techniques avancées de collecte de données dans l'internet des objets de santé. Cette comparaison est basée sur plusieurs paramètres tels que : la consommation d'énergie, la réduction des données, la surcharge sur le sink et le temps d'exécution. Par conséquent, nous travaillons dans la deuxième partie sur une nouvelle approche basée sur la fusion d'un protocole de routage dans la communication du réseau corporel sans fil. L'objectif de notre protocole est d'assurer une transmission efficace des données critiques et la conservation de l'énergie des dispositifs médicaux, grâce à la manière de déterminer la meilleure route pour atteindre le puits.

**Publications**

1. **Ghina Saad**, Hassan Harb, Abdelhafid Abouaissa, Lhassane Idoumghar and Nour Charara. “*An Efficient Hadoop-Based Framework for Data Storage and Fault Recovering in Large Scale Multimedia Sensor Networks*”. In IWCMC 2020, 16th International Wireless Communications & Mobile Computing Conference, June 15-19, 2020 - Limassol, Cyprus, pages 316–321.
2. **Ghina Saad**, Hassan Harb, Abdelhafid Abouaissa, Lhassane Idoumghar and Nour Charara. “*P2D: An Efﬁcient Patient-to-Doctor Framework for Real-Time Health Monitoring and Decision Making*”. In IEEE Sensors Journal, pp. \*-\*, 2020. Accepted manuscript, to appear.
3. **Ghina Saad**, Hassan Harb, Abdelhafid Abouaissa, Lhassane Idoumghar and Nour Charara. “*An efficient patient classification and grouping mechanism-based framework for patients-nurse scheduling*”. Submitted soon.
4. **Ghina Saad**, Hassan Harb, Abdelhafid Abouaissa, Lhassane Idoumghar and Nour Charara. “*Recent Advances of Patient Monitoring in Internet of Healthcare Things : A Comparative Study*”. Submitted soon.

**Reférences**

[1] R. Ward, “Energy-efficient data reduction techniques for eeg wireless body sensor networks,” in Qatar Foundation Annual Research Forum Volume 2013 Issue 1, vol. 2013, no. 1. Hamad bin Khalifa University Press (HBKU Press), 2013, pp. BIOP–041.

[2] Y. Qu, G. Zheng, H. Wu, B. Ji, and H. Ma, “An energy-efficient routing protocol for reliable data transmission in wireless body area networks,” Sensors, vol. 19, no. 19, p. 4238, 2019.

[3] R. Zhang and J. Yu, “Energy-efficient and reliable sleep scheduling algorithms in wbsns,” in Energy-Efficient Algorithms and Protocols for Wireless Body Sensor Networks. Springer, 2020, pp. 101–121.

[4] F. Ullah, I. U. Islam, A. H. Abdullah, and A. Khan, “Future of big data and deep learning for wireless body area networks,” pp. 53–77, 2019. [Online]. Available: https://doi.org/10.1007/978-981-13-3459-7 5.

[5] C. Habib, A. Makhoul, R. Darazi, and C. Salim, “Self-adaptive data collection and fusion for health monitoring based on body sensor networks,” IEEE transactions on Industrial Informatics, vol. 12, no. 6, pp. 2342–2352, 2016.

[6] A. Alameen and A. Gupta, “Clustering and classification based real time analysis of health monitoring and risk assessment in wireless body sensor networks,” Bio-Algorithms and Med-Systems, vol. 15, no. 4, 2019.

[7] S. Lee and Y. H. Lee, “Improving emergency department efficiency by patient scheduling using deep reinforcement learning,” Healthcare, vol. 8, no. 2, 2020. [Online]. Available: https://www.mdpi.com/ 2227-9032/8/2/77.

[8] A. Azadeh, M. Hosseinabadi Farahani, S. Torabzadeh, and M. Baghersad, “Scheduling prioritized patients in emergency department laboratories,” Computer Methods and Programs in Biomedicine, vol. 117, no. 2, pp. 61 – 70, 2014, http://www.sciencedirect.com/science/article/pii/S0169260714003198

[9] S. Madden, “Intel lab data,” http://db.csail.mit.edu/labdata/labdata.html, 2004.

[10] (2000). Mimic Database on Physionet. Accessed: Oct. 4, 2016. [Online].992 Available: <https://www.physionet.org/>.

**Abdelhafid Abouaissa**

Professeur des universités à l’UHA

Responsable de la LP Administration des Réseaux Multimédia (ARM) Email : abdelhafid.abouaissa@uha.fr

Téléphone : +33 (0)3 89 20 23 71

**Hassan Harb**

Maitre de conférences à l’Université Libanaise

Email : [hassan.harb.1@ul.edu.lb](mailto:hassan.harb.1@ul.edu.lb)

Téléphone : 03399252

signature 