

Big IoT Healthcare Data Analytics Framework Based on Fog and Cloud Computing

Hamoud Alshammari*, Sameh Abd El-Ghany**,***, and Abdulaziz Shehab* **

Abstract

Throughout the world, aging populations and doctor shortages have helped drive the increasing demand for smart healthcare systems. Recently, these systems have benefited from the evolution of the Internet of Things (IoT), big data, and machine learning. However, these advances result in the generation of large amounts of data, making healthcare data analysis a major issue. These data have a number of complex properties such as high-dimensionality, irregularity, and sparsity, which makes efficient processing difficult to implement. These challenges are met by big data analytics. In this paper, we propose an innovative analytic framework for big healthcare data that are collected either from IoT wearable devices or from archived patient medical images. The proposed method would efficiently address the data heterogeneity problem using middleware between heterogeneous data sources and MapReduce Hadoop clusters. Furthermore, the proposed framework enables the use of both fog computing and cloud platforms to handle the problems faced through online and offline data processing, data storage, and data classification. Additionally, it guarantees robust and secure knowledge of patient medical data.

Keywords

Cloud Computing, Fog Computing, E-Health, Electronic Health Records, Healthcare Data Analytics, Internet of Things (IoT)

1. Introduction

While the concept of e-health is still in its infancy throughout the world, e-health has gained prominence in many countries over the last few decades. Recently, with the aid of information technology, many researchers [1-10] have been motivated to address the digital transformation of medical data. In many healthcare organizations, there are huge amounts of heterogeneous medical data within electronic health records (EHRs). EHR data must be collected, integrated, cleaned, stored, analyzed, and interpreted in efficient ways that will ensure accuracy and streamline access time [1]. In healthcare systems, making decisions demands the careful analysis of large amounts of real-time data collected from sensors and wearable devices. These decisions have a significant influence on patient health. For example, a smart city's healthcare systems could monitor their patients remotely and intervene when patients engage in activities that may lead to poor health. Many researchers utilize Internet of Things (IoT) wearable devices

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received February 11, 2020; first revision March 28, 2020; second revision July 20, 2020; accepted August 23, 2020.

Corresponding Author: Abdulaziz Shehab (abdulaziz_shehab@mans.edu.eg)

* Dept. of Computer Science, College of Science and Arts, Jouf University, Al-Jawf, Saudi Arabia (hhalshammari@ju.edu.sa, aishahab@ju.edu.sa)

** Dept. of Information Systems, College of Computer and Information Sciences, Jouf University, Al-Jawf, Saudi Arabia (saabdelwahab@ju.edu.sa)

***Dept. of Information Systems, Mansoura University, Mansoura, Egypt (Sameh_cis@mans.edu.eg, abdulaziz_shehab@mans.edu.eg)

to detect signals that may identify different diseases [2]. On the other hand, machine learning methods are becoming increasingly popular due to their excellent performance [3,4].

In the field of healthcare data analytics, there are generally four main categories: descriptive, diagnostic, predictive, and prescriptive analytics [1]. Descriptive analytics aims to identify the present status of a patient and generate such information as reports, charts, and histograms. Many analytic tools could be used to perform this type of analysis.

Diagnostic analysis depends on clustering techniques and decision trees to understand reasons for the reappearance of some specific diseases in individual patients, and it studies the occurrence of events and factors that trigger them. Predictive analytics depends on different machine learning algorithms to predict unknown events by building a suitable predictive model. Prescriptive analytics tries to make optimal decisions by proposing effective actions that lead to suitable patient treatments. The increasingly large amounts of healthcare data create an inevitable need for big data frameworks with suitable analysis tools. Significant research has been published to target this objective, focusing on areas such as infrastructure, data management, data searching, data mining, and data security.

This paper introduces a new approach to e-health, which enriches IoT, cloud computing, and fog computing for big healthcare data analysis using cognitive data mining and machine learning algorithms. Instead of centralized healthcare data processing, which is an obstacle to large-scale data analysis, fog computing helps us to perform data processing over distributed fog nodes. Furthermore, this article describes the design component analysis of the framework architecture in detail. The main objective of this work is to improve the efficiency and effectiveness of healthcare services. In this paper, we aim to provide scalability in IoT wearable devices, and minimize the communication overhead between network fog and cloud layers. The proposed e-health framework, using the best regulated analytics pipeline, will benefit from the current communication revolution to unify EHRs and enable remote medical data sharing. Additionally, the proposed framework aims to create more stable and effective communication with patients, and to upgrade healthcare services by predicting and potentially preempting illnesses through integrating EHR and sensory data obtained from 24/7 monitoring. Additionally, the proposed framework would provide timely assistive recommendations to patients. This would alleviate clinician overload, help an ageing population that needs long-term care, improve data access, and reduce medical errors. This in turn would reduce pressure on national healthcare budgets.

The paper is organized as follows: Section 2 presents a literature review and related work. The proposed framework's overall architecture and their layers are discussed in detail in Section 3. Section 5 presents remarks and discussion with a comparison between the proposed framework and some recent related frameworks. Finally, the conclusion and future work are presented in Section 5.

2. Literature Review

In the last few years, many researchers [1-10] have been inspired to write about big data and fields related to big data. At the same time, many machine learning algorithms [3,4] that can extract and analyze large amounts of data have been utilized to help decision makers. Platforms have been built to help analyze huge amounts of IoT data. Moreover, in medical big data analytics, computational issues have arisen due to the streaming nature and high dimensionality of healthcare data. Aktas et al. [5] presented

a platform for data analytics based on IoT. Their proposal was very focused on metadata only and lacks automatic indexing and partitioning. This method works best for social unstructured data. Khan et al. [4] proposed a real-time big data analytics framework. This framework focused only on data volume and velocity and considered these two aspects without supporting a variety of data sources. Shen et al. [6] proposed a classification model called oriented feature selection support vector machine (OFSSVM) for cancer prediction using gene expression data. Based on gene order, they demonstrated binary classification and multiclass classification effectiveness. However, this study lacked overall interoperability between their proposed components.

Mishu [7] proposed a framework for biomedical engineering applications using C-means clustering. Their framework has the potential to help clinicians and patients. They used the University of California–Irvine (UCI) machine learning repository and MapReduce to analyze diseases and their symptoms. Al-Khafajiy et al. [8] proposed a fog-driven IoT model based on an unsupervised deep feature learning method. The model aimed to derive a general purpose patient representation from EHRs. They used random forest classifiers trained over a 200,000 patient dataset. Alhussain [9] proposed an architectural framework of big data healthcare analytics with a focus on the importance of security and privacy issues. Images, signals, and genome data processing play key roles in healthcare data analyses. Yousefpour et al. [10] presented an analytics framework called smart health with the aim of reducing IoT service delay via fog offloading and resolving the challenges facing big healthcare data through information and communication technologies (ICT). In recent years, social networks have become a platform for data exchange through which users with common needs can share their data.

3. Materials and Methods

The proposed system will be based on multisensory devices, which are responsible for tracking known disease symptoms. Smart wearable devices will provide users with overall health data and notifications from sensors to their mobile phones. In addition, machine learning algorithms such as SVM, bagged SVM (BSVM), and deep learning models will be utilized to improve performance accuracy and time.

The research methodology used follows the common big data analytics pipeline process that starts with data acquisition, followed by data cleaning, annotation, and extraction. Thereafter, the system integrates the medical data to generate the analytical model. Finally, medical data interpretation and visualization occurs. The system promises a novel solution for the data integration problem. The proposed method efficiently addresses the data heterogeneity problem using middleware between heterogeneous data sources and MapReduce Hadoop clusters. It also utilizes recent machine learning algorithms to improve accuracy and data access time. Furthermore, it guarantees robust and secure knowledge of patient medical data.

Fog computing nodes have the benefit of utilizing resources efficiently. Fog nodes are constantly processing incoming IoT streams from wearable devices. These nodes rely on a number of virtual machines responsible for transferring the pre-processed data to the cloud computing layer for further processing [8]. On the other hand, cloud computing comprises many services such as infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and utility services. IaaS, for example, offers access to unlimited cloud storage space while PaaS is responsible for executing resource

intensive applications. SaaS manages access to different software and utility services and stores large amounts of data for remote access. Fog computing is essential for big real-time data processing collected from IoT sensors. Furthermore, fog computing is a critical step in our proposed framework as a basis for both processing and storing data in the cloud computing system. In this way, the proposed framework becomes capable of avoiding latency caused by the transport communication layer in the cloud computing system. In healthcare applications, continuous multi-wearable device sensor data streams rapidly form a massive dataset. Therefore, it is necessary to employ fog computing nodes to handle these data accurately in real time.

3.1 Framework Design Components

Fig. 1 shows the overall architecture of our proposed layered framework. It is based on healthcare big data analytics collected from IoT wearable devices with fog computing nodes and a cloud computing system. The framework layer components guarantee multiple device pre-processing, processing, integration, and analytics in real time. As shown in Fig. 1, the fog nodes are closer to the smart wearable device's physical location, thus some applications must be serviced within specific time slots. The fog nodes deliver the heavy data load to the cloud system that handles the computationally intensive application processing.

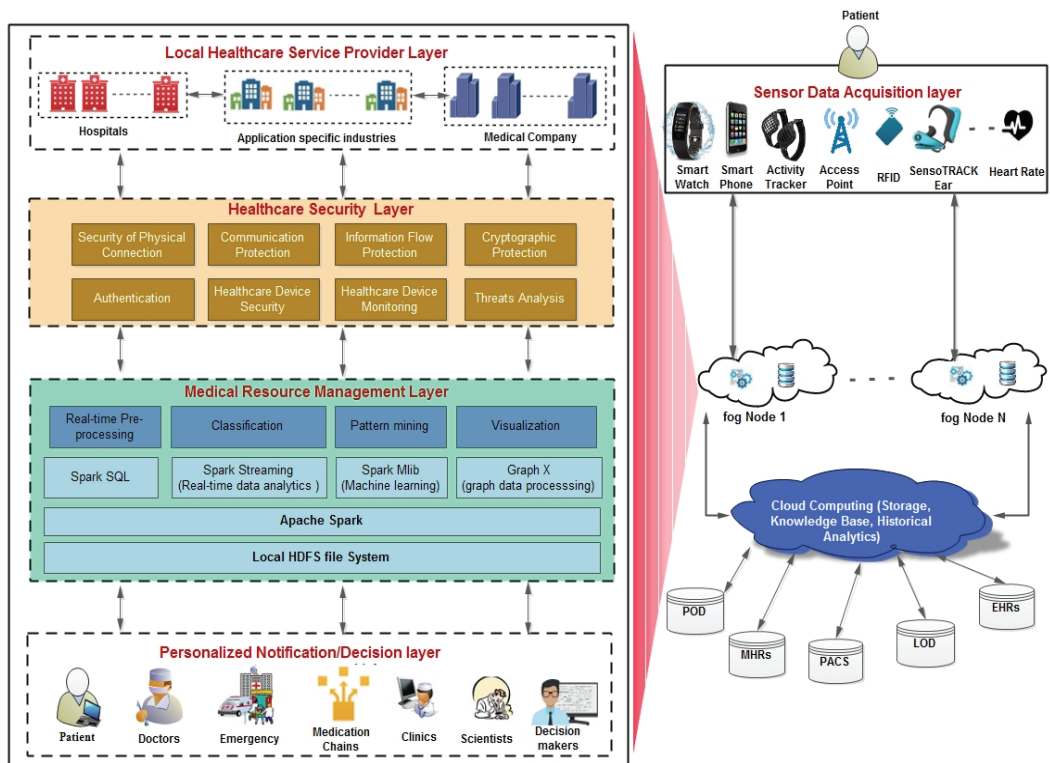


Fig. 1. The generic architecture of the proposed framework is divided into five layers: sensor data acquisition layer, local healthcare service provider layer, healthcare security layer, medical resource management layer, and personalized notification/decision layer.

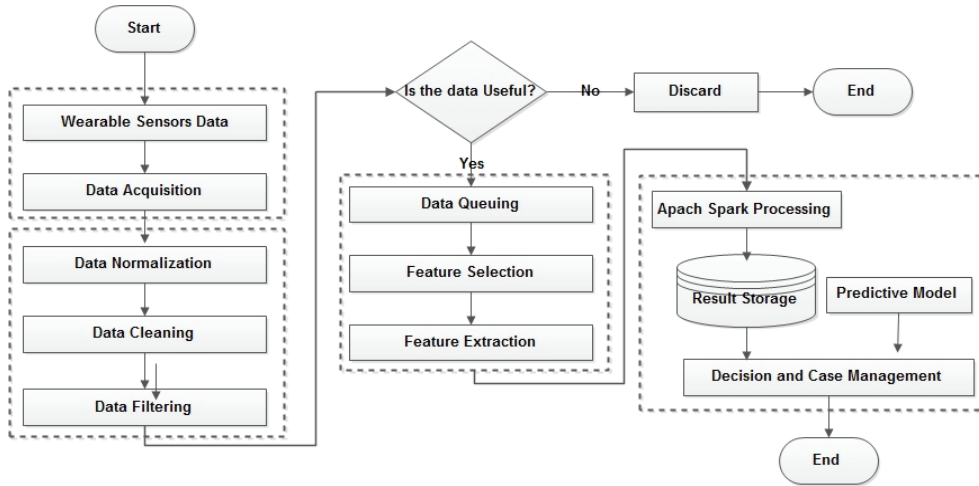


Fig. 2. Working steps of the proposed framework.

Fig. 2 shows the process flow of our proposed framework. In abstract, it contains three main blocks: (1) sensing data and pre-processing, (2) feature queuing, selection, and extraction, (3) the predictive model. Data normalization which generalizes the effect of the different data acquired from sensors could be defined in many different forms. Given the data from the sensors noted as D_S and of size n , the normalized data D_S^* is obtained by applying the normalization approach. Normalization by root mean square and normalization by maximum and minimum are given by Eqs. (1) and (2), respectively.

$$D_S^* = \frac{D_{S_i}}{\sqrt{\frac{1}{n} \sum_{j=1}^n D_{S_j}^2}}, \quad i \in [0, n] \quad (1)$$

$$D_S^* = (D_{S_i} - Min_{old}) \frac{Max_{new} - Min_{new}}{Max_{old} - Min_{old}} + Min_{new}, \quad i \in [0, n] \quad (2)$$

General feature selection algorithm is shown in Algorithm 1. Many metrics approaches are utilized for such purpose of feature selection. The Gini index, for instance, measures the inequality among values of a frequency distribution which range from 0 (complete equality) to 1 (complete inequality). The Gini coefficient is often calculated according to the Brown formula in Eq. (3) where G denotes Gini coefficient, X_i is cumulated proportion of one feature, and Y_i is cumulated proportion of target feature. Information gain (IG) is another method for feature selection which based on pioneering work of information theory. It studies the value or “information content” of messages through measuring the number of bits obtained for class prediction. The information gain $G(f)$ of a feature f is given by Eq. (4), where $\{C_i\}_{i=1}^n$ denotes the set of classes, $v \in V$ is the set of possible value of feat f which could be generalized to any number of classes.

$$G = \left| 1 - \sum_{i=1}^n (X_i - X_{i-1})(Y_i + Y_{i-1}) \right| \quad (3)$$

$$G(f) = -\sum_{i=1}^n P(c_i) \log P(c_i) + \sum_{v \in V} \sum_{i=1}^n P(f = v) P(c_i | f = v) \log P(c_i | f = v) \quad (4)$$

Algorithm 1: General feature selection algorithm**Input:** for each training vector and class value $n \leftarrow$ number of instances $m \leftarrow$ number of training instances ($m \subset n$) $a \leftarrow$ number of vectors (attributes)**Output:** Final score vector W that predicated the influence of features

```

1 Initialize all feature score. Set  $W[A] := 0.0$ ;
2   for  $i:=1$  to  $m$ 
3     Select a target instance  $R_i$  randomly
4     Find both H (nearest Hit) and M (nearest Miss)
5     for  $A:= 1$  to  $a$ 
6        $W[A] := W[A] + \frac{\text{diff}(M, A, R_i)}{m} - \frac{\text{diff}(H, A, R_i)}{m}$  // update W
7     end
8   end
9 return  $W$ 

```

3.2 Local Healthcare Service Provider Layer

Initially, hospitals, application specific industries, or medical companies are responsible for granting the stakeholders (patients, doctors, clinics, and decision makers) privileges to use the healthcare system. Healthcare data are highly sensitive information; therefore, at this layer, with the aid of the healthcare security layer, user privacy will be maintained.

3.3 Sensor Data Acquisition Layer

The IoT revolution has caused the medical wearable devices sector (and telehealth) to increase and change rapidly. According to a study conducted by Berg Insight in 2014, telehealth devices and services revenue reached \$4.5 billion, which nearly doubled by 2017 [2]. Wearable device categories and revenue are shown in Fig. 3. The priority time (T_p) performed at the sensor at local site is given according to Eq. (5) where N indicates number of sensors, T_c denotes time of entire cycle, ψ is the Poisson arrival, and T_{pac} is the Packet transmission time.

$$T_p = \psi T_c * NT_{pac} / 2(1 - \psi T_c) \quad (5)$$

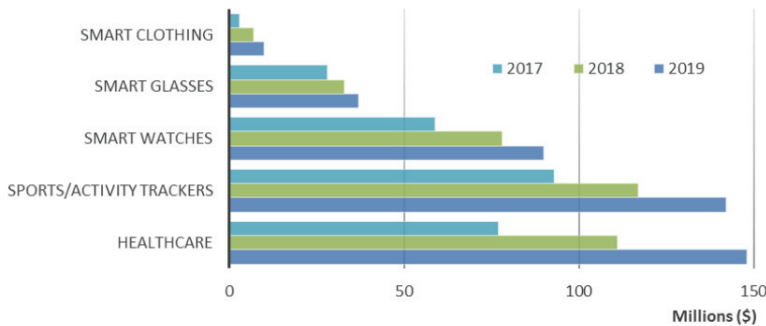


Fig. 3. Wearable device categories and revenue between 2017 and 2019. Adapted from [2].

Recently, there has been continued growth in the number of monitoring devices, such as diabetes care devices, blood pressure monitoring, and heart rate monitors, connected to a cloud data center [2]. Currently, mobile wearable health devices are becoming an important part of any monitoring system, assisting in medical prediction, detection, and diagnosis. The most popular wireless protocols utilized in medical applications are Wi-Fi, ZigBee, and LoRa (Long Range radio). The Wi-Fi protocol allows higher data throughput for low-power applications. However, the ZigBee protocol is ideal as it consumes low power with an advanced communication channel encryption standard. Finally, LoRa is a recent wireless protocol that can cover long distances at low cost with low power consumption [2].

In the proposed framework, medical wearable devices trigger data generation. These data, which are captured from different devices, in turn are delivered to healthcare IoT gateways that mediate between the devices and the fog computing layer. The data acquisition process is achieved through specific IoT protocols that link the wearable devices to the IoT gateways. Machine-to-machine (M2M) and message queuing telemetry transport (MQTT) are well-known examples of such protocols. In our proposed framework, the IoT gateways are also responsible for local processing, storage, control, and filtering of data stream functions. Furthermore, with the aid of the healthcare security layer, device connectivity integrity is ensured through policy-based access enforcement. Continuous monitoring data generates a very large amount of data because the volume, velocity, and variety characteristics are constantly growing. Medical data analytics take an extensive amount of time given the data complexity and volume and the need for preprocessing, cleaning, filtering, and classification tasks.

3.4 Fog Computing Nodes

The fog computing nodes mediate between the wearable devices and the cloud system. They accelerate the analysis process, which is mandatory in time-sensitive applications [10]. The fog nodes receive unstructured data and do not use any predefined model. They provide data pre-processing, pattern mining, classification, prediction, and visualization. The data pre-processing stage includes error, redundancy, and outlier handling. Furthermore, all received IoT data streams have to be translated into a structured model through parsing and filtration by the fog nodes. Once we have the structured model, pattern mining approaches are used on the resulting data to build correlations and association maps. The equations for average energy consumption for transmitting and receiving fog node are illustrated in both Eqs. (6) and (7), respectively. If two Fog nodes F_x and F_y need to share resources together, the power needed to propagate the data is mainly based on the distance D_{xy} between F_x and F_y nodes. The data transmitting rate at fog node F_x is given by Eq. (8) and denoted as $Rate_{F_x}^s$ and at the receiver side as $Rate_{F_y}^r$, at node F_y as given in Eq. (9).

$$En_{F_x}^s = b^\uparrow * S_i * (T_p + (D_{xy} * En_{sh})) \tag{6}$$

$$En_{F_y}^r = b^\downarrow * S_i * En_{sh} \tag{7}$$

$$Rate_{F_x}^s = \frac{1}{T_{slot}} * \log En_{F_x}^s \tag{8}$$

$$Rate_{F_y}^r = \frac{1}{T_{slot}} * \log En_{F_y}^r \tag{9}$$

3.5 The Cloud System

In the proposed framework, the core services (such as historical data analytics, storage management, authentication, authorization services, and core medical resource management infrastructure) are processed by the cloud computing layer. Additionally, large scale data mining requiring heavy computation, such as MapReduce and Apache Spark, are also implemented by the cloud system. The cloud system relies on a continuous back-end computation to keep within sight of IoT wearable medical devices and update the fog nodes.

3.6 Medical Resource Management Layer

The core of any healthcare analytics system is the efficiency of its classifier. As the collected data may be big, unbounded, or unbalanced, classification becomes increasingly complex. Furthermore, the data distribution may be unbalanced resulting in one class containing more samples compared to other classes. The most vital part of the classifier model implementation is the feature or attribute selection. Recently, deep learning algorithms such as AlexNet, LeNet, multilayer perceptron (MLP) and recurrent long short-term memory (LSTM) have the ability to deal with high dimensional data and outperform many other classic techniques in terms of efficiency, precision, and accuracy.

Apache Spark is a powerful clustering big data framework based on Hadoop developed at AMPLAB, UC Berkeley in 2010 [4]. It is distinguished by its high reliability, consistency, speed, and fault tolerance. As shown in Fig. 1, the fundamental components in Apache Spark contain four main libraries: structured query language (SQL), streaming, Spark MLlib, and GraphX [9]. In order to generate either a recommendation or specific disease diagnosis, SparkML relies on knowledge bases such as the Web Ontology Language (OWL2) and Open Biomedical and Biological Ontologies (OBO) to generate rules, ontologies, and semantic annotation. Real-time stream computation is done by Apache Spark streaming while Spark SQL deals with relational queries on the different database systems. The GraphX library provides distributed processing to manage the graph. Spark MLlib is the most important component as it contains more than 55 expandable analytical machine-learning methods to provide parallelization [8]. Its components have been developed by many researchers to benefit from big data analytics worldwide. A variety of machine-learning approaches are embedded in Apache Spark MLlib, such as classification, regression, and filtering.

3.7 Healthcare Security Layer

The healthcare security layer plays a vital role in confirming that admission privileges are accurate using pre-defined policies. The cloud system data are protected with an IoT broker that either accepts or rejects access. Registered and authenticated requests are subsequently mapped to the next layer. Data security and privacy are ensured with anomaly detection and privacy preserving data mining [5]. The IoT data received from wearable devices are vulnerable to many attacks. Therefore, a privacy protection mechanism must be employed to handle such attacks and prevent data alteration or removal. These mechanisms are not considered in this paper. To prevent malicious attacks, viruses, and any other issues that might affect user trust, the proposed framework includes physical connection security, communication protection, information flow protection, cryptographic protection, authentication, healthcare device security, healthcare device monitoring, and threat analysis modules.

The healthcare device security module ensures wearable device integrity by confirming the settings needed to perform its specified functions. These settings control different security policies of the devices and, if needed, update the structure for known vulnerabilities. The healthcare device monitoring module keeps IoT wearable devices continuously monitored. The monitoring process is done through integrity checks, service denial activities, and malicious usage detection. Physical connection layer attacks are handled by the security of physical connection module. The communication protection module with the aid of the information flow protection module encrypts the data transferred between devices, fog nodes, and the cloud system using standardization protocols and boundary protection technologies. The cryptographic protection module controls updates for all to other modules within the framework. Moreover, it manages security policies for all communication lines such as password-protected communication settings, and firewall configuration. The threat analysis module analyzes abnormal behavior by searching for malicious patterns that may cause harmful system crashes. Abnormal behavior is discovered by the system after learning normal behavior and building a rule-based set with specific abnormal events.

3.8 Personalized Notification/Decision Layer.

Finally, the overall feedback of the healthcare system is oriented towards one user through the personalized notification/decision layer. The system stakeholders are patients, doctors, clinics, data scientists, paramedics, pharmacists, and decision makers.

4. Discussion

To study the validity of the proposed framework, table 1 presents a comparison between four related frameworks and our proposed in terms of base platform, streaming type, streaming primitive, streaming source, and transmission delay. As noticed from Table 1, both [7,8] lacks real-time streaming (i.e., they only batched) while the transmission delay is high (few seconds) in both [5,7] and medium (sub seconds) in [4]. The proposed framework outperform the related frameworks in terms of latency due to dependency a delay minimizing offloading policy. It utilizes a fog layer which deliver the heavy data load to the cloud system that handles the computationally intensive application processing.

Table 1. The proposed framework compared to four related frameworks

| | Base platform | Streaming type | Streaming primitive | Streaming source | Transmission delay |
|------------------------|----------------------|-----------------------|----------------------------|-------------------------|---------------------------|
| Al-Khafajiy et al. [8] | Hadoop | batch | Key value | HDFS | N/A |
| Mishu [7] | Hadoop | batch | Key value | HDFS | High |
| Aktas et al. [5] | Opnet Simulator | N/A | N/A | N/A | High |
| Khan et al. [4] | Apache Spark | Batch/real time | DStream | HDFS | Medium |
| Proposed framework | Apache Spark | Batch/real time | DStream | HDFS | Low |

Moreover, although most recent health analytics frameworks are based on business logic tier (such as Apache Spark or Hadoop) and data access (such as Cassandra, spouts, or, HDFS), our proposed framework go further by offering some more powerful features. Compared to other related frameworks, mentioned in Table 2, the proposed framework achieve a precedence in terms of validity, sustainability,

separation of layers and fault tolerance. These frameworks are limited to images alone, not scalable and did not provide confidentiality and data integrity. Reliability, compliance, integrity, and confidentiality are important features in order to prohibit tampering, denial of service attacks, and spoofing. However, most related frameworks lacks most of these features due to either inability to remember a PIN to access, complexity in key management, or late validation timing.

Table 2. The proposed framework compared to four related frameworks in terms of 8 features

| | Validity | Sustainability | Separation of layers | Fault tolerance | Reliability | Compliance | Data integrity | Confidentiality |
|------------------------|----------|----------------|----------------------|-----------------|-------------|------------|----------------|-----------------|
| Al-Khafajiy et al. [8] | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ | ✓ | ✓ |
| Mishu [7] | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ |
| Aktas et al. [5] | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Khan et al. [4] | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ |
| Proposed framework | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ |

5. Conclusion

In this paper, we presented a framework for healthcare IoT big data analytics that is based on both fog and cloud computing. Additionally, it utilizes Apache Spark components to classify medical data using different machine-learning algorithms, and to make timely decisions. A detailed analysis and illustration of the proposed framework is presented. In general, the proposed framework could alert patients in real-time when a problem occurs, allowing them to take action when necessary. Analytics conducted at the fog side help to manage extremely large medical IoT data streams from different sources. In the future, a detailed analysis of the proposed framework will be conducted, with a focus on processing time, feature extraction, and fog computing offloading challenges.

Acknowledgement

We are indebted to the Deanship of Scientific Research at Jouf University for funding this work through General Research Project (No. 40/201).

References

- [1] S. Din and A. Paul, "Smart health monitoring and management system: toward autonomous wearable sensing for internet of things using big data analytics," *Future Generation Computer Systems*, vol. 91, pp. 611-619, 2019.
- [2] D. Dias and J. Paulo Silva Cunha, "Wearable health devices: vital sign monitoring, systems and technologies," *Sensors*, vol. 18, no. 8, article no. 2414, 2018.
- [3] J. Xie, F. R. Yu, T. Huang, R. Xie, J. Liu, C. Wang, and Y. Liu, "A survey of machine learning techniques applied to software defined networking (SDN): research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 393-430, 2018.

- [4] M. A. Khan, M. Karim, and Y. Kim, "A two-stage big data analytics framework with real world applications using spark machine learning and long short-term memory network," *Symmetry*, vol. 10, no. 10, article no. 485, 2018.
- [5] F. Aktas, C. Ceken, and Y. E. Erdemli, "IoT-based healthcare framework for biomedical applications," *Journal of Medical and Biological Engineering*, vol. 38, no. 6, pp. 966-979, 2018.
- [6] Y. Shen, C. Wu, C. Liu, Y. Wu, and N. Xiong, "Oriented feature selection SVM applied to cancer rediction in precision medicine," *IEEE Access*, vol. 6, pp. 48510-48521, 2018.
- [7] M. M. Mishu, "A patient oriented framework using big data & c-means clustering for biomedical engineering applications," in *Proceedings of 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Dhaka, Bangladesh, 2019, pp. 113-115.
- [8] M. Al-Khafajiy, L. Webster, T. Baker, and A. Waraich, "Towards fog driven IoT healthcare: challenges and framework of fog computing in healthcare," in *Proceedings of the 2nd International Conference on Future Networks and Distributed Systems*, Amman, Jordan, 2018, pp. 1-7.
- [9] T. Alhussain, "Medical big data analysis using big data tools and methods," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 4, pp. 793-795, 2018.
- [10] A. Yousefpour, G. Ishigaki, R. Gour, and J. P. Jue, "On reducing IoT service delay via fog offloading," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 998-1010, 2018.



Hamoud Alshammari <https://orcid.org/0000-0001-6843-3732>

He is an Assistance Professor and the computer sciences and Information head department at Aljurayyat College for Arts and Sciences at Aljouf University. He holds a Ph.D. in a Computer Sciences and Engineering from the University of Bridgeport, CT-USA 2016. He was the head department and computer center labs at Aljouf Technology College, then he worked as the customer relationship head department at Saudi Electricity Company at Aljouf office. He worked as a teaching assistance for graduate level at the University of Bridgeport. He is a member in Upsilon Pi Epsilon Honor Society, which recognize talent in computer science field, USA. He had attended 1-year data scientist course that is proposed by John Hopkins University, and he is a certified with many tools and skills by this course. His research area includes working on bigdata collection and analysis, Hadoop developing environment, NoSQL databases, IoT and wireless sensors systems. Also, he is working on health information systems data analysis.



Sameh Abd El-Ghany <https://orcid.org/0000-0002-5903-3048>

He received B.S. degree in 2003 from information systems department, Mansoura University, Egypt, and the M.Sc. degree in 2009 from Mansoura University, Egypt, entitled thesis "An approach for image retrieval based on mobile agent". He received Ph.D. degree from faculty of computer science and information, Mansoura University, Egypt. His research interests include information retrieval, image processing, sentiment analysis and semantic web. He is currently an Assistant Professor at the Department of Information Systems, College of Information and Computer Sciences, Jouf University, Saudi Arabia.



Abdulaziz Shehab <https://orcid.org/0000-0001-8610-7172>

He received B.Sc. degree in information systems from College of Computer Science and Information Systems, Mansoura University, Egypt, in 2004. In 2009, he obtained his M.Sc. degree in information systems department (thesis entitled “Automated essay grading system based on intelligent text mining techniques”). In 2015, he obtained his Ph.D. degree in information systems department (dissertation entitled “Efficient schemes for internet-based video delivery”). In 2015, He start working as a manager of e-courses production center, Mansoura University. Currently, he is working as an assistant professor at Jouf University, Saudi Arabia. His research interests includes text mining, neural networks, natural language processing, computer networks, multimedia communications, biometrics systems, decision support systems (DSS), Internet of Things (IoT), and related topics.