Clouds Proportionate Medical Data Stream Analytics for Internet of Things-Based Healthcare Systems

Priyan Malarvizhi Kumar, Choong Seon Hong, Fatemeh Afghah, Gunasekaran Manogaran, Keping Yu, Qiaozhi Hua, and Jiechao Gao

***Abstract*—Internet of Things (IoT) assisted healthcare systems are designed for providing ubiquitous access and recommendations for personal and distributed electronic health services. The heterogeneous IoT platform assists healthcare services with reliable data management through dedicated computing devices. Healthcare services' reliability depends upon the efficient handling of heterogeneous data streams due to variations and errors. A Proportionate Data Analytics (PDA) for heterogeneous healthcare data stream processing is introduced in this manuscript. This analytics method differentiates the data streams based on variations and errors for satisfying the service responses. The classification is streamlined using linear regression for segregating errors from the variations in different time intervals. The time intervals are differentiated recurrently after detecting errors in the stream's variation. This process of differentiation and classification retains a high response ratio for healthcare services through spontaneous regressions. The proposed method's performance is analyzed using the metrics accuracy, identification ratio, delivery, variation factor, and processing time.**

Manuscript received March 10, 2021; revised May 21, 2021 and June 29, 2021; accepted July 31, 2021. Date of publication August 20, 2021; date of current version March 7, 2022. This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grant Information Technology Research Center support program under Grant IITP-2021-0-00742 supervised by the IITP (Institute for Information and communications Technology Planning and Evaluation). *(Corresponding author: Choong Seon Hong.)*

Priyan Malarvizhi Kumar and Choong Seon Hong are with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, South Korea (e-mail: mkpriyan@khu.ac.kr; cshong@khu.ac.kr).

Fatemeh Afghah is with the Department of Electrical and Computer Engineering, Clemson University, SC 29634 USA (e-mail: fafghah@clemson.edu).

Gunasekaran Manogaran is with the Department of Electrical Engineering and Computer Science, Howard University, Washington, DC 20059 USA, and also with the College of Information and Electrical Engineering, Asia University Taichung 41354, Taiwan (e-mail: gunasekaran.manogar@howard.edu).

Keping Yu is with the Global Information and Telecommunication Institute, Waseda University, Tokyo 169-8050, Japan (e-mail: keping.yu@aoni.waseda.jp).

Qiaozhi Hua is with the Computer School, Hubei University of Arts and Science, Xiangyang 441000, China (e-mail: 11722@hbuas.edu.cn).

Jiechao Gao is with the Department of Computer Science, University of Virginia, Charlottesville, VA 22904 USA (e-mail: jg5ycn@virginia.edu).

Digital Object Identifier 10.1109/JBHI.2021.3106387

***Index Terms*—Data analytics, differential computing, healthcare systems, IoT, regression learning.**

I. Introduction

THE Internet of Things (IoT) paradigm is assistive in heterogeneous real-time applications, including smart healthcare and electronic-medical services. With the integration of different sensor devices, communication standards, and interaction modes, IoT assists healthcare services at ease [1]. Human-machine interaction (HMI) and device-to-device (D2D) communication modes are familiar in the healthcare systems. End-users in an electronic healthcare environment make use of Wearable Sensors (WS) in the form of head-bands or wristwatches, etc. for monitoring their health vitals [2]. The sensed vitals are transmitted to the healthcare center for storage, retrieval, and processing. The sensed information is accumulated from different WSs at different time intervals, increasing the stream size and the storage requirementI [3]. The IoT platform's pervasive communication nature provides continuous and discrete storing and updating of the sensed vitals from humans. Besides, the information from previous studies and observations is packed in Electronic Health Records (EHRs) [4]. The records are accessed and updated periodically for future use and end-user support. Healthcare applications and recommendation systems rely on the EHRs to provide on-time assistance and monitor patient/end-users' health stats [5].

The variance in healthcare data is common due to sensor and medical data and its sensing intervals. The infrastructure and connecting architectures also impose a difference in data due to different transmission, processing, and storage time [6]. These changes do not meet the intended service response or user's medical assistance. Therefore considering the significance of the healthcare data and its importance in life-saving precautions, variations need to be addressed [7]. Digital healthcare data is analyzed using the computing models and systems that are distributed in the IoT platform [8]. The computing systems rely on different analytical and visualization methods for handling sensitive healthcare data. Machine-based processing and intelligent computing augment data analytics for improving the reliability and integrity of the data streams [9]. The computing systems are capable of differentiating the variations and usefulness in the healthcare data before providing services. Healthcare services

2168-2194 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

and recommendations for the end-users are satisfied using the processed medical data and EHRs for improving the efficiency and justification of the application design [4], [10].

Medical data processing is sensitive yet requires influential computing and visualization techniques for handling EHRs and WS observations. Human-intervention-less computing solutions rely on artificial intelligence and machine learning techniques for analyzing digital data [11]. Machine learning and decision-making systems are common in IoT led platforms and workstations for providing distributed and precise solutions. These techniques influence the performance of healthcare services and computing systems by providing instantaneous and error-less analysis [12]. The computing systems receive the sensed or stored data, classify, and perform matching for precise information retrieval. The processed information is distributed as a response for the requesting users and application scenarios [13]. The computing system's prominent demand or analytical model is its response to dynamic and static medical data. That is, the computing system must handle a series of physical vitals from the WS or discrete information from the EHR to provide reliable healthcare service. This feature is demanding in designing flexible and productive healthcare management systems. Integrated with the sophisticated IoT platform, communication and data management are eased to design robust healthcare systems [14], [15]. The heterogeneous IoT platform supports health services through specialised computing devices with dependable data handling. The dependability of health services depends on the efficient management of diverse streams of data owing to changes and errors which has been considered as a significant challenge. This research introduces a PDA for the treatment of diverse health-care data streams. This analytical procedure distinguishes data streams based on changes and mistakes to meet the service answers.

II. Related Works

Zheng *et al*. [16] implemented a longitudinal data fusion approach for predictive healthcare analysis of disease progression. A data fusion approach describes the chronic disease of a patient. A sparse regularization regression approach is planned for analyzing data from the source level. Alzheimer's disease Neuroimaging Initiative (ADNI) investigates the approach for simulation and identification of disease.

For improving smart healthcare systems' performance, Alfarraj & Tolba [17] proposed an unsynchronized wearable sensor (WS) data analytics model. An unsynchronized sensor data analytics (USDA) model is designed to meet the healthcare treatments and diagnosis it for better efficiency and reduce complications. Sophisticated machine learning methods are used for analyzing the WS data with time aspects.

For improving patient assistance in the future direction, Fouad et al. [18] suggested analyzing patient health information based on IoT sensors. Accuracy of patient details is reduced in the traditional healthcare assistance system. For improving the benefits of the patient, an IoT sensor with AI is used for information forecast. The iterative golden section optimized deep belief neural network (IGDBN) is used to process the collected patient's details and produced an accurate result with less error by simple matching and Matthews's correlation coefficient.

Serhani et al. [19] designed a self-adapting cloud services orchestration for fulfilling intensive sensory data-driven IoT workflows. Some tasks like specification, composition, configuration, execution, adaptation, etc. are processed by a comprehensive architecture. It also supports integration for processing the data, data compression for fast transmission, data visualization, and customization reporting. The proposed work reduced the execution time, with less cost for optimum usage of the resource.

By considering the utilization of IoT wearable medical devices, Al-Makhadmeh & Tolba [20] proposed a classification model for heart disease prediction using the higher-order Boltzmann model. The higher-order Boltzmann deep belief neural network (HOBDBNN) is used to process the health care center's transmitted information. Based on some specific characteristics, the system's performance is estimated with the deep learning method's help. This method performs with minimal complications and helps to identify heart disease accurately.

Abdel-Basset et al. [21] considered a novel intelligent medical decision support model based on IoT and soft computing. In cloud computing operation, mobile edge computing and fog computing were used to overcome difficulties like response time, mobility, etc. Computer-assisted diagnostics and IoT architecture can help diagnose and monitor type 2 diabetes patients. Excellent reliability and robustness make this work precise for detecting mysterious data

On the internet of medical things environment, a CNN-Based model for analyzing regular health factors is implemented by Ismail et al. [22]. The disease's knowledge is predicted with the convolution neural network (CNN) model from the medical records. Pearson Correlation Coefficient is used to propose CNN-regular target recognition and detection model to overcome the complexity in CNN. Better accuracy, low computation is obtained as a result of the comparison made between different learning techniques.

To support healthcare analytics, Kakkanatt et al. [23] designed curating and integrating user-generated health data from multiple sources. The curation process considered collecting, validating, and storing data from different sources like clinical data, genomic data, microbiome data, etc. Healthcare applications are supported by curated data for visualization, analysis, etc.

Big data analytics is suggested by Razzak et al. [24] for preventive medicine. A structured and comprehensive overview is provided for analyzing data. Through traditional methodology, the prevention of disease and its conflicts are mentioned. Sophisticated data analysis algorithms are used to detect, classify, and associate specific models developed and applied to prevent disease.

A pruned random forest regression approach is implemented by Fawagreh&Gaber [25] for resource-efficient fast prediction in healthcare data analytics. On mobile devices, high accuracy and prediction are considered by healthcare applications. Based on CLUB-DRF, regression technique, fast prediction of disease, and accuracy for the proposed work are evaluated by three medical data sets from different diseases.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

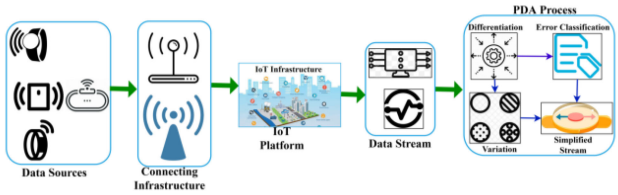


Fig. 1. PDA process

A concise temporal data representation model is proposed by Manashty et al. [26] for prediction in biomedical wearable devices. Mapped interval sequence (MIS) is used to obtain original data properties without any values. The long-term forecast is improved by Life model for time series (LMts) in real-time health prediction. The result obtained is compared with the previous method for its accuracy and other parameters.

Yang [27] modeled a medical multimedia big data analysis based on the DBN algorithm. The present architecture of IoT access multimedia data and their types are classified with the help of DBN to improve accuracy. The comparison is made with a traditional feature for better performance. Multivariate Gaussian distribution theory is used to construct and evaluate the efficiency and accuracy of the user's health status.

Based on the deep learning method, Sun et al., [28] suggested intelligent analysis of medical-big data. The model parameters are selected and optimize the proposed work through medical-big data. The process of pathological analysis from doctors has been learned automatically. This could realize the disease at an early stage, and the risk of some disease is predicted through the patient's physical examination records. It reduces the work pressure and improves doctors' work efficiency.

Kwon et al. [29] described visual analytics with interpretable and interactive recurrent neural networks (RNN) on electronic medical records. RetainVis is used for fetching information from the medical code contribution for risk prediction. For converting temporal information into useful and interactive, changes are made to the state of the art RNN model known as RETAIN. This study also guides researchers to model visual tools for RNNs.

III. Proportionate Data Analytics

Proportionate data analytics is introduced for improving the precision of health care information modeling and management. Data modeling and its value for satisfying the healthcare requirement is complex due to heterogeneous variety and periodic up-gradation. The proposed method relies on the distributed computing resources in the IoT platform for handling the variations in healthcare services. Health care data analytics improve services' precision by reducing the mean error due to varying patterns in accumulation and processing. This data analytics method tailors the complex stream (s) into its simplest form by reducing variation errors. The proposed method's process is portrayed in Fig. 1.

The process of the proposed method consists of differentiation, variation and error classification, and simplified stream. This simplified stream is the output data for healthcare services assisted by the IoT computing intelligence.

The artificial intelligences and machine learning techniques used in analysis are based on IoT computing intelligence with detailed solutions. In IoT computing intelligence platforms and work stations for the delivery of dispersed and accurate solutions, machine learning and decision making systems are widespread. These approaches improve the performance of health services and computer systems through rapid and error-free analysis. The computer systems receive, categorise and match sensed or stored input for accurate information retrieval.

Problem Definition: The stream s is acquired from different data sources with a combination of variances and healthcare data. This stream is represented as $\mathrm{s} = \mathrm{d} \pm \mathrm{v}$, here $\mathrm{d} + \mathrm{v}$ is the classification eligible stream whereas $\mathrm{d} - \mathrm{v}$ causes error. Therefore, the problem of proportionate analysis is defined as in equation (1)

$$\left.\begin{aligned} \forall\, \mathrm{s} = \mathrm{d} + \mathrm{v},\ \text{argmax}\ \alpha \text{ in all } \mathrm{s} \in \mathrm{a_t} \text{and } \mu_\mathrm{t} \\ \forall\, \mathrm{s} = \mathrm{d} - \mathrm{v},\ \text{argmin}\ \beta \text{ in all } \mathrm{s} \in \mathrm{a_t} \\ \text{such that} \\ \beta \text{ is replaced with } \alpha \text{ in } \mu_\mathrm{t} \end{aligned}\right\} \quad (1)$$

The variables α and β represent precision and error. The variables$\mathrm{a_t}$ and μ_t represents the data aggregation and update time. In this problem definition, the error observed in $\mathrm{a_t}$ is mitigated through a precise update in μ_t. Hence the $\text{argmin}\ \beta$ is to be classified in μ_t. The two processes differentiation and error classification are performed in $\mathrm{s} = \mathrm{d} + \mathrm{v}$ and $\mathrm{s} = \mathrm{d} - \mathrm{v}$ streams. The variations are observed if the data (d) aggregated does not meet requirements or irrelevant. The failure in healthcare services and its access to the users determines this factor. The following sections discuss the processes above in a detailed manner.

A. Differentiation Process

The differentiation process segregates s based on time and variations. The time-based s is used for simplified responses. On the other hand, the variations are analyzed for its error to improve analytics precision. The differentiation process operates on the variation and error by deploying regressive learning. The differentiation of s for time and variation is computed as

$$\left.\begin{aligned} \frac{\partial \mathrm{s}}{\partial^{\mathrm{a_t}}} &= \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} + \frac{\partial \mathrm{a_t}}{\mathrm{a_t}+\mu_\mathrm{t}}\ (\text{No variation})\ (\mathrm{a}) \\ \frac{\partial \mathrm{s}}{\partial \mathrm{a_t}} &= \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} - \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}\ (\text{variation})\ (\mathrm{b}) \\ \frac{\partial \mathrm{s}}{\partial \mu_\mathrm{t}} &= \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} - \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}} - \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}\ (\text{Error})\ (\mathrm{c}) \end{aligned}\right\} \quad (2)$$

The variation of a fewer data stream is directly utilized for the healthcare service distribution. The problem arises if the variation and error are jointly occurring in $\mathrm{a_t}$ and μ_t, Concurrently, Therefore, leaving out the no variation, the variation and error streams are differentiated using regressive learning. In the differentiation process, equation 2(b) exhibits both v and d as partial derivative, and hence the best fit s with less v is to be segregated at the first analysis. The differentiation is performed for $\frac{\partial \mathrm{v}}{\partial \mathrm{a_t}+1}$ and $\frac{\partial \mathrm{v}}{\partial \mu_\mathrm{t}}$ individually for augmenting error from the variation. In Figs. 2(a) and (b), regression for the above differentiation is illustrated.

The occurrence of $(\mathrm{a_t} + 1)$ and $\mathrm{a_t}$ in $\mathrm{a_t}$ and μ_t is proportional for both $(\mathrm{d} + \mathrm{v})$ and $(\mathrm{d} - \mathrm{v})$. In both instances, the best solution

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

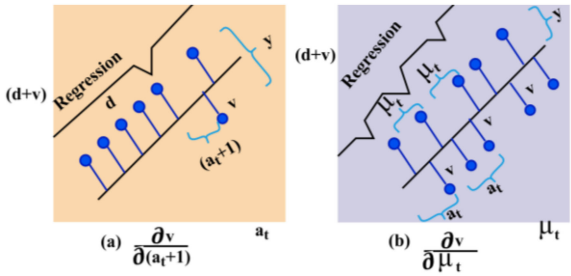


Fig. 2. Regressive differentiation

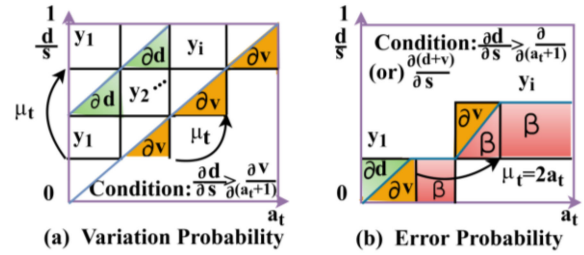


Fig. 3. Probabilities representation

(y) is computed in successive (a_t) and (a_t+1) instance for $\frac{\partial v}{\partial(a_t+1)}$. Whereas for an$a_t$ and μ_t alternates, the y for $\frac{\partial v}{\partial \mu_t}$ is to be computed. The best solution to the regressive differentiation is computed using equation (3).

$$y = \left\{ \begin{array}{l} \sum_{i=1}^{a_t} \left[\frac{d_i+v_i}{s_i}\right] + \frac{\alpha_i}{\alpha_i-\beta_i} \forall \frac{\partial v}{\partial(a_t+1)} \quad (a) \\ \sum_{i=1}^{a_t} \left[\frac{d_i-v_i}{s_i}\right] - \sum_{i=a_t+1}^{\mu_t} \frac{v_i}{d_i} - \sum_{i=1}^{\mu_t} \frac{\beta_i}{\alpha_i+\beta_i} \forall \frac{\partial v}{\partial u_t} \quad (b) \\ \text{where,} \\ \alpha = \frac{\partial d}{\partial s} \text{ and } \beta = \frac{\partial(d-v)}{\partial s} \end{array} \right\} \quad (3)$$

The alternating instances of the d and v increases the cumulative β in differena_t and hence y is useful in classifying the differentiation. The differentiation is identified using two conditions: $\frac{\partial d}{\partial s} > \frac{\partial(d-v)}{\partial s}$ and $\frac{\partial d}{\partial s} < \frac{\partial v}{\partial(a_t+1)}$. The conditions indicate a better healthcare data stream (without variation) and variations (with error). For the non-alternating $y \forall \frac{\partial v}{\partial(a_t+1)}$ and alternating y in $\frac{\partial v}{\partial u_t}$. The classification is given as in equations 4(a) and 4(b).

$$\left. \begin{array}{l} y\ (\partial d) = \sum_{i=1}^{a_t} (y_i - y_{i-1})^2 \times \frac{\alpha_i-\beta_i}{\alpha_i} \forall \frac{\partial v}{\partial(a_t+1)} \quad (a) \\ y\ (\partial v) = \sum_{i=1}^{a_t} y_i \times \beta_i - \sum_{j=a_t+1}^{\mu_t} [y_j - y_i]^2 \frac{\beta_j}{\alpha_j-\beta_j} \quad (b) \end{array} \right\} \quad (4)$$

This classification is valid until $\frac{\partial s}{\partial a_t} = \frac{\partial s}{\partial u_t}$ and hence, 4(b) is considered as a simplified stream in the distribution. The above condition fails, if a_t and μ_t are equal, and hence, equating equations 2(b) and 2(c)

$$\left. \begin{array}{l} \frac{\partial d}{\partial a_t} + \frac{\partial v}{\partial(a_t+1)} - \frac{\partial v}{\partial u_t} = \frac{\partial d}{\partial a_t} - \frac{\partial v}{\partial(a_t+1)} \\ \frac{2\partial v}{\partial(a_t+1)} = 0 \end{array} \right\} \quad (5)$$

This proves the non-occurrence of variations and hence, $\frac{\partial s}{\partial a_t} = \frac{\partial d}{\partial a_t} \cdot \frac{\partial a_t}{\partial(2a_t)}$. Retaining the further two instances of a_t, the healthcare data is allocated for the service. In this case, the condition $\frac{\partial d}{\partial s}$ is high than $\frac{\partial v}{\partial(a_t+1)}$ and $\frac{\partial(d-v)}{\partial s}$. Thus the classification is performed for $s = d + v$ (or) $s = d$ (as $v = 0$) and therefore, $y = \sum_{i=1}^{a_t} \frac{d_i}{s_i} + 1$ (as $\beta = 0$ due to $v = 0$). Similarly, $y\ (\partial d) = \sum_{i=1}^{a_t} y_i.1$ (or) $y\ (\partial d) = \sum_{i=1}^{a_t} y_i$, is the cumulative healthcare data stream. The healthcare data analysis for simple stream-based services using iteration probability and error is presented in Figs. 3(a) and (b), respectively.

The probability of the variation and error is illustrated in the above Figures. The conditions above are analyzed for $\frac{d}{s}$ for either 1 or 0 such that "1" represents no variation. Therefore, this differentiation process segregates $y = \frac{d}{s} = 1$ cases in all a_t for serving different healthcare services with the fetched data. Contrarily, for all $\frac{d}{s} \neq 1$ or $\frac{d}{s} = 0$ or < 0, error mitigation is required. The differentiation process of the d for the condition $\frac{\partial d}{\partial s} > \frac{\partial v}{\partial(a_t+1)}$ is performed throughouta_t. In ana_t is encountered, then the differentiation is verified for $y(\partial d)$ such that $\forall\ y_1 = 1$ in $i \in a_t$ instances. The differentiated healthcare data is distributed for the requesting user without further classification. It is either assimilated for independently used for satisfying the user service requirement. The variations observed in the s for $d - v$ is analyzed for consistency and relevance for all the update time intervals. This is discussed in the following section.

IV. Error Classification

The error classification is performed to prevent $\frac{\partial s}{\partial u_t}$ and the condition-failing data being disseminated for the healthcare service. The implication of healthcare data for the end-users requires non-latent and errorless dissemination. This is achievable by performing error mitigation from the varying s. The challenging task is the classification of error from the variations in s. The two differentiations $\frac{\partial s}{\partial a_t}$ and $\frac{\partial s}{\partial u_t}$ [as in equations 2(b) and 2(c)] are to be classified based on the regression and probabilistic validation. In this process, the regression learning based on the conditions (failing) $\frac{\partial d}{\partial s} < \frac{\partial(d-v)}{\partial(s)}$ and $\frac{\partial d}{\partial s} < \frac{\partial v}{\partial(a_t+1)}$ are considered. The regressive learning iterates are augmented for $\mu_t = 2at$ [Refer to Fig. 3(b)] instances and $\frac{\partial v}{\partial(a_t+1)} \neq 0$ conditions. In this error classification, the output y from 3(b), and its corresponding differentiation $y(\partial v)$ are used for regressive analysis. The classification further segregates variations and errors independently for improving the healthcare service responses. Therefore, the classification is instigated in $\partial(a_t+1)$ and $\partial(u_t)$ instances for identifying $\frac{\beta}{\alpha-\beta}$ (error) and variations $\frac{\alpha-\beta}{\alpha}$ in all $i \in a_t$ and μ_t, respectively. The different time intervals are identified as $[1,\ a_t+1]$ and $[a_t+1,\ \mu_t]$ in the linear $s \pm d$. The first interval is augmented for $s + d$ (with variations) and the second for $(s - d)$ (with errors) in this classification process. In this work, the classification using regression learning is performed and discussed in the below sections.

A. Classification 1: $s + d$ in the Interval $[1,\ a_t+1]$

Discussion 1: The classification of $s + d$ is segmented in $[1, a_t]$ and $[1,\ a_t+1]$ intervals such that the variations (or errors) are observed between $[a_t + a_t + 1]$. If this minor change is not observed, the chance for high error is feasible. Therefore, the

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

classification is performed in two intervals. However the $[1, \mathrm{a_t}]$ achieves differentiation of healthcare data without variation. Therefore, the alternating interval of $[1,\ \mathrm{a_t}+1]$ for $[\mathrm{a_t}, \mathrm{a_t}+1]$ is analyzed for identifying errors in the s. The complete sequence of s with the classification (pre-differentiation) is represented as

$$\frac{\partial \mathrm{s}}{\partial \mathrm{a_t}} = \begin{cases} \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} - \frac{\mathrm{y}(\partial \mathrm{v})}{\partial \mathrm{a_t}} - \frac{\alpha-\beta}{\alpha} \forall\ [1,\ \mathrm{a_t}+1]\ \ (\mathrm{a}) \\ \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} - \underbrace{\frac{\mathrm{y}}{\mathrm{a_t}+1}}_{\text{equation 4(b)}} - \beta\ \forall\ [\mathrm{a_t}, \mathrm{a_t}+1] \end{cases} \tag{6}$$

This sequence is feasible to occur is determined using the regression instances. The probability of the instance (ρ_v) is computed using equation (7) as

$$\rho_\mathrm{v} = \frac{1}{\left[1+\exp\left(\frac{\alpha}{\alpha-\beta} - \frac{\beta}{\alpha}\right)\right]} [\mathrm{y}\ (\partial \mathrm{v}) \left(\frac{\alpha}{\alpha-\beta}.\frac{\mathrm{a_t}}{\mathrm{a_t}+1} + \left(\frac{\alpha}{\alpha-\beta} - \frac{\beta}{\alpha}\right)\right] \tag{7}$$

This probability is analyzed for observing the occurrence of 0 and 1. If the $\rho_\mathrm{v} = 0$, then 6(c) is considered as it indicates the first occurrence of the variation. From this point, the successive instance of $(\mathrm{a_t}+1)$ is considered for providing a classification of s. Equation 6(a) and 2(b) for the variation, then

$$\left.\begin{array}{l} \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} - \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}} = \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} - \frac{\mathrm{y}(\partial \mathrm{v})}{\partial \mathrm{a_t}} - \left(\frac{\alpha-\beta}{\alpha}\right)\ \ (\mathrm{a}) \\ \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}} - \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} = \frac{\mathrm{y}(\partial \mathrm{v})}{\partial \mathrm{a_t}} + \frac{\alpha-\beta}{\alpha}\ \ (\mathrm{b}) \\ \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} = \frac{\alpha-\beta}{\alpha} \\ [\text{as } \partial \mathrm{v} \text{ at } \mathrm{u_t} = \mathrm{y}\,(\partial \mathrm{v})\,(\text{previous})\ \text{at } \mathrm{a_t}]\ \ (\mathrm{c}) \\ \beta = \left[1 + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}\right]\ \alpha\ (\mathrm{d}) \end{array}\right\} \tag{8}$$

From the above estimation, as α is high, β is estimated for the consecutive $[\mathrm{a_t}+1]$ sequence contrarily, on the other hand, for $\rho_\mathrm{v} = 1$ occurrence, equations 6(b) and 2(b),

$$\left.\begin{array}{l} \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} - \frac{\mathrm{y}}{\mathrm{a_t}+1} - \beta = \frac{\partial \mathrm{d}}{\partial \mathrm{a_t}} + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} - \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}\ \ (\mathrm{a}) \\ \beta = \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}\ \ (\mathrm{b})\,[\text{as } \mathrm{y} = \partial \mathrm{v} \text{ in } (\mathrm{a_t}+1) \end{array}\right\} \tag{9}$$

From the validations in equations (8) and (9), it is seen that the variation generates two different outputs. In the first output, β is observed as a constituent of α, and there is less error (only variation). However, the second case of β as in equation 9(b) represents the occurrence of an error in the precise time of $(\mathrm{a_t}+1)$. This classification of s helps to segregate s in $\mathrm{a_t}$ and $(\mathrm{a_t}+1)$ intervals, and the instantaneous identification of healthcare data are performed. The classification based on equation 8(d) ensures reliable and simpler distribution of healthcare services. On the other hand, $\beta = \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$ in $(\mathrm{a_t}+1)$ shows that the data is erroneous, and hence it is not adaptable for the healthcare service. This data needs to be mitigated from the service dissemination for data modeling and preventing misguided services. Classification 2 is the continuity of $\beta = \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$ under $\rho_\mathrm{v} = 1$ case that is discussed in the following section.

Classification 2: $\mathrm{s} - \mathrm{d}$ in the interval$[\mathrm{a_t}+1],\ \mu_\mathrm{t}]$

Discussion 2: From equation 9(b), the occurrence of β is observed in $\mathrm{a_t}+1$ interval, and this generates errors in the variations. The need for identifying errors is to prevent misguided healthcare information from being shared for the services. This classification helps to mitigate the error interruption in continuous healthcare services. The regression process helps to identify the errors, but it does not perform differential classification. The error classification is performed for the consecutive instances of $[\mathrm{a_t}, \mathrm{a_t}+1]$ and $[\mathrm{a_t}+1,\ \mu_\mathrm{t}]$. This classification is instigated with the output using the above interval

$$\mathrm{y}\ (\partial \mathrm{v}) = \begin{cases} (1-\beta) + \beta\frac{\partial \mathrm{d}}{\partial \mathrm{a_t}},\ \forall\ [\mathrm{a_t},\ \mathrm{a_t}+1]\ (\mathrm{a}) \\ \beta + \beta\left(\frac{\partial \mathrm{v}}{\partial \mathrm{u_t}} - \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}\right)\ \forall\ [\mathrm{a_t}+1,\ \mu_\mathrm{t}]\,(\mathrm{b}) \end{cases} \tag{10}$$

Equation 10(a) is ignored as it generates the variation and is further analyzed using the classification 1 process. But, the variation identification is processed using equation 10 (b). The simplest form of processing health data is performed by equating 10 (b) and 4 (b), respectively.

$$\left.\begin{array}{l} \mathrm{y} \times \beta - (\mathrm{y_j} - \mathrm{y_i})\ \frac{\beta}{\alpha-\beta} = \beta + \beta\left[\frac{\partial \mathrm{v}}{\partial \mathrm{u_t}} - \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}\right]\ \ (\mathrm{a}) \\ \beta\ \left[\mathrm{y} - \frac{\mathrm{y}}{\alpha-\beta}\right] = \beta\left[1 + \partial \mathrm{v}\left(\frac{1}{\partial \mathrm{u_t}} - \frac{1}{\partial(\mathrm{a_t}+1)}\right)\right] \\ [\text{as } \mathrm{y_i} \text{is variation and hence it is not an error}]\ (\mathrm{b}) \\ \mathrm{y}\ (\alpha-\beta) = \mathrm{y} \\ [\text{as } \mathrm{u_t} = \partial\,(\mathrm{a_t}+1)\,,\ \text{considering } [\mathrm{a_t}+1,\ \mathrm{u_t}]]\ \ (\mathrm{c}) \\ \beta = 1-\alpha\ (\mathrm{d}) \end{array}\right\} \tag{11}$$

From equation (11), the β is found to be high than α in $[\mathrm{a_t}+1,\ \mathrm{u_t}]$. This verification is performed in a continuous sequence of y. The change in variation from $\beta = [1 + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}]$ [equation 8(d)] to $\beta = 1-\alpha)$ (or) $\beta = \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$ represents the error, diminishing the accuracy. Therefore this healthcare data is not provided as a response to the service required. Such instances are differentiated from the data streams, and the succeeding stream is replaced for healthcare response. The overall process of the proposed PDA differentiates the streams in identifying reliable responses for the requesting services. Such classification (variation and error) also reduces the processing time of different sizes of healthcare data streams. Stream software monitoring software has been utilized for data analysis and real-time analysis of integration data. These software tools enable users to examine data between apps or via Application peripheral interface. This enables users to evaluate past events as well as exactly what is happening at this moment. Here, the stream analysis to understand better what data consumers collect and track endpoints. Users can examine data transmission between devices such as IoT Endpoints.

For the varying data stream size, the differentiation factor increases along with classification instances. The classification instances differentiate a new variation and error pair of s. Therefore, the need for differentiating $\partial \mathrm{v}$ from $\partial \mathrm{d}$ is regressively performed based onρ_v for the succeeding time instances. This requires further classification instances to differentiate $\partial \mathrm{v}$, increasing its count (Table I).

The variation and error ratio varies for the different classification instances, as illustrated in Fig. 4. It is seen that the error

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DIFFERENTIATION FACTOR UNDER CLASSIFICATIONS = 40, 80, 120 INSTANCES DIFFERENTIATION

Differentiation			
Data Stream Size (Mb)	Classification= 40	Classification= 80	Classification=120
200	0.681	0.712	0.742
400	0.723	0.763	0.769
600	0.711	0.796	0.784
800	0.778	0.812	0.845
1000	0.778	0.856	0.892
1200	0.778	0.896	0.95
1400	0.778	0.896	1
1600	0.778	0.896	1

TABLE II
CLASSIFICATION, VARIATION, AND ERROR INSTANCES

Instances			
Data Stream Size (Mb)	Classification	Variations	Error
200	10	5	1
400	73	8	9
600	93	5	5
800	104	9	2
1000	76	9	8
1200	89	16	1
1400	102	15	9
1600	120	16	10

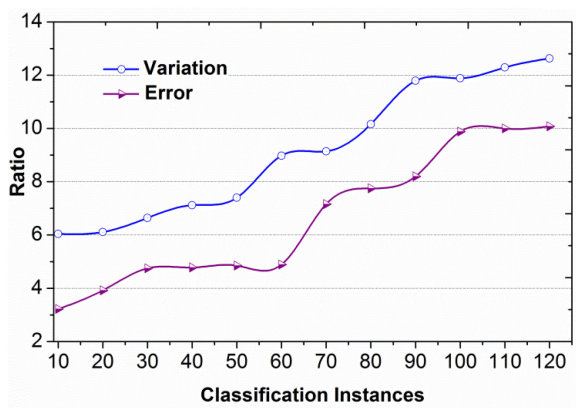


Fig. 4. Variations and error ratio under classification instances

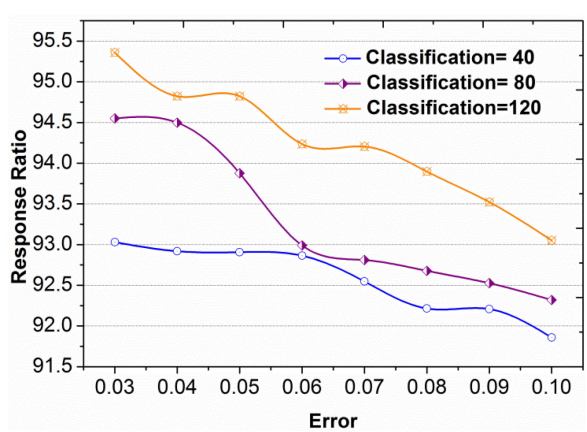


Fig. 5. Response ratio under classification instances= 40, 80, 120

is less compared to the variation due to the classification 1 and 2 processes. The β causing classification instances are identified from the $\frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$ and $(1-\alpha)$ estimations as in equation 9(b) and 11(d) for the variation in $[1,\ \mathrm{a_t}+1]$ and $[\mathrm{a_t}+1,\ \mathrm{u_t}]$ intervals. Therefore, the entire variation is not categorized under error, reducing its ratio.

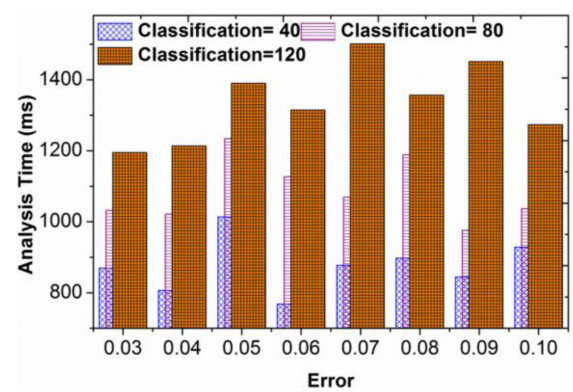


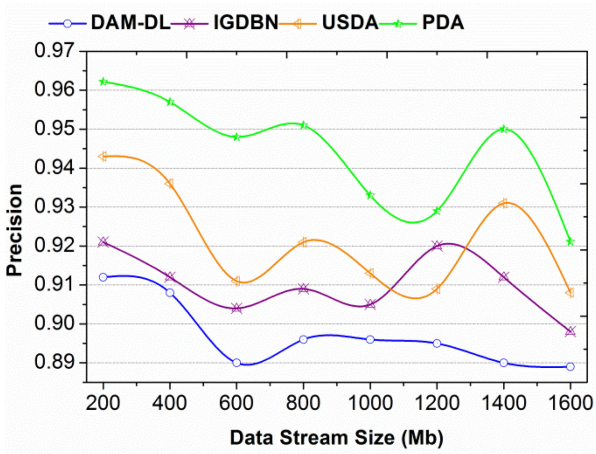
Fig. 6. Analysis time under classification instances= 40, 80, and 120

In Fig. 5, the analysis between different error factors and response ratio is performed. Though the error factor increases, the response ratio is maximized by augmenting the number of classification instances. The classification instances segregate β from ∂v, adding up the simplification of the healthcare data stream for better responses. Therefore, the error increases is balanced by the classification instances for less impact on responses.

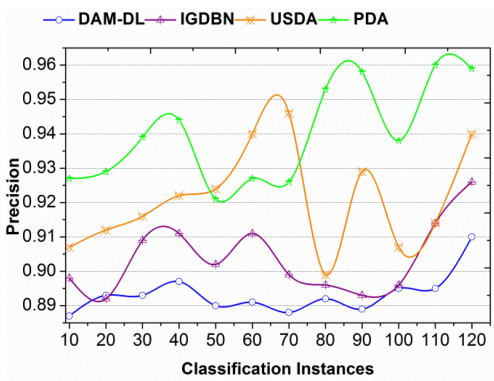
The classification, variations, and error instances of different data stream sizes are tabulated in Table II. It is seen that the classification increases the variations and errors for different sizes of the data stream. The number of classifications is varied in $[1,\ \mathrm{a_t}+1]$ for ∂v and $[\mathrm{a_t}+1,\ \mu_\mathrm{t}]$ for β identification. Suppose, a "s" is considered as an error, then classification separates ∂v from the β, augmenting the ration of response. Therefore, the error is less than the variation (other than differentiation) under different classifications.

A comparative study of the analysis time for error and classification instances is demonstrated in Fig. 6. As the error increases, analysis time increases. In the proposed PDA, the error is reduced by intending classifications 1 and 2 for identifying ∂v from the β, causing data streams. Therefore, the number of classification (1 and 2) is increased without generating high error for which the analysis time increases.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

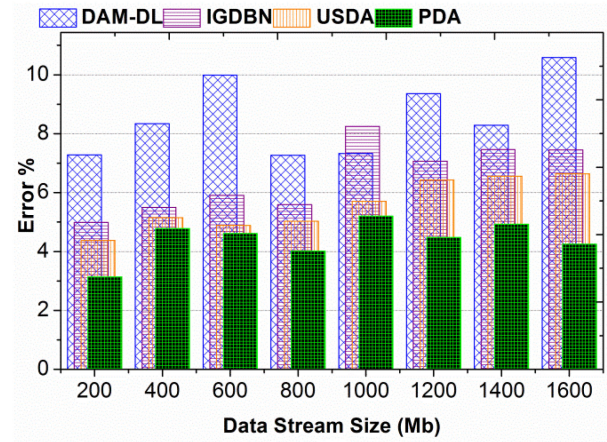


(a)

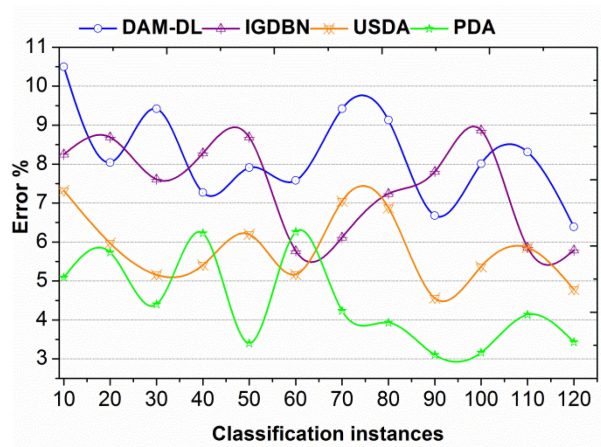


(b)

Fig. 7. (a) and (b) Precision for data stream size and classification instances



(a)



(b)

Fig. 8. (a) and (b) Error % for data stream size and classification instances

V. Discussion

The proposed PDA for healthcare services is analyzed using the observed data set [30]. This data set provides electrocardiogram observations for 12 activities of 10 human subjects. The observation sequence varies between 12 min-17 min. The information is processed for null values/ labels in the sequence, using PDA. The analysis results are measured using precision, error, analysis time, variations, and response ratio. By varying the data stream size and classification instances, the above metrics are analyzed. For verifying the performance of the proposed PDA, a comparative analysis is performed using DAM-DL [28], IG DBS [18], and USDA [17] methods.

A. Precision Analysis

Figs. 7(a) and (b) presents the precision analysis for different stream size and classification instances. The proposed PDA shows up high precision by two methods, namely differentiation, and classification. The differentiation process identifies reliable s from $(\mathrm{d}+\mathrm{v})$, fed as a simplified response to the healthcare service. The variation-less $\frac{\partial \mathrm{s}}{\partial \mathrm{a_t}}$ achieves high precision as $\alpha = \frac{\partial \mathrm{d}}{\partial \mathrm{s}}$ is the direct response to any number of healthcare requests. On the other hand, the variation based "s' consists of both high precision and error-based data in $[\mathrm{a_t},\ \mathrm{a_t}+1]$ interval. The differentiation of $\mathrm{y}(\partial \mathrm{d})$ and $\mathrm{y}(\partial \mathrm{v})$ is significant in determining the precision. Secondly, the variation probabilityρ_{v} classifies $\partial \mathrm{v}$ for μ_{t} and $(\mathrm{a_t}+1)$ time instances for improving the precision based on $(\mathrm{a_t}+1)$. Here, precision is high until $\alpha = \frac{\partial \mathrm{d}}{\partial \mathrm{s}}$ (or) $\frac{\partial \mathrm{s}}{\partial \mathrm{u_t}} = \frac{\partial \mathrm{s}}{\partial \mathrm{u_t}}$ is achieved. Similarly, in the classification process, $\frac{\partial \mathrm{d}}{\partial \mathrm{s}} > \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}$ achieves high precision in the varying s. This differentiation and classification are alternating for $\mathrm{s} = \mathrm{d} \pm \mathrm{v}$ streams in the proposed PDA, for achieving better precision.

B. Error Analysis

The error ratio is analyzed for the different data stream sizes and classification instances, demonstrated in Figs. 8(a) and (b). The drawback in error differentiation from the variation of $\mathrm{s} = \mathrm{d} - \mathrm{v}$ is effectively-identified in the proposed PDA. By identifying $\mathrm{y}(\partial \mathrm{d})$ and $\mathrm{y}(\partial \mathrm{v})$ for $\partial(\mathrm{a_t}+1)$ and $\partial \mathrm{u_t}$, respectively, helps to differentiate the classification of error and variation. The error probability is segregated from the cumulative ρ_{v} using the conditions $\frac{\partial \mathrm{d}}{\partial \mathrm{v}} < \frac{\partial(\mathrm{d}-\mathrm{v})}{\partial \mathrm{s}}$ and $\frac{\partial \mathrm{d}}{\partial \mathrm{s}} < \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}$ (failing). Post this process $\frac{\partial \mathrm{s}}{\partial \mathrm{a_t}}$ based β and α is estimated for providing sequential error classification. Classification 2 of the

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

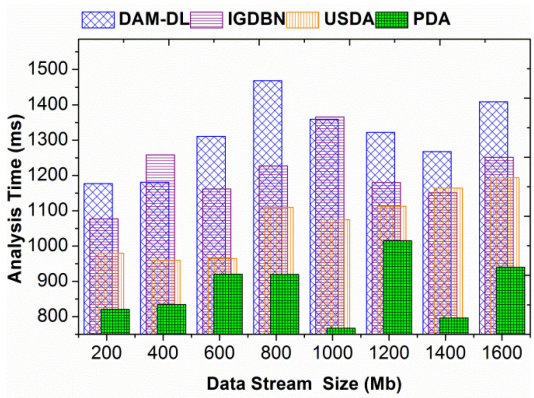


Fig. 9. Analysis time for data stream size

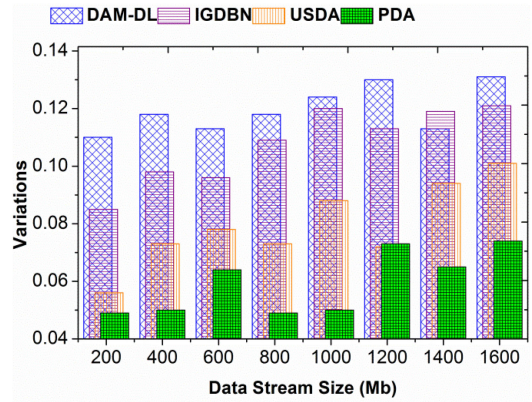


Fig. 10. Variations for data stream size

PDA process identifies β in $\mathrm{y}(\partial \mathrm{v})$ intervals. The interval $[1,\ \mathrm{a_t}]$ and $[\mathrm{a_t}+1,\ \mu_\mathrm{t}]$ and $[\mathrm{a_t}, \mathrm{a_t}+1]$ are independently analyzed for the segregation of errors and variations. The first interval is only the variation, and no error is observed. On the other hand, the other intervals generate $\frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$ and $(1-\alpha)$ as errors. This error is inherited from $\mathrm{y}(\partial \mathrm{d})$ and $\mathrm{y}(\partial \mathrm{v})$ outputs of the regression. By classifying the $\partial \mathrm{v}$ instances, the probabilistic variation from becoming an error is prevented. Therefore, the proposed method is said to achieve fewer errors for the data stream and classifications.

C. Analysis Time

The proposed PDA performs differentiation in a single instance, whereas classification 1 and 2 and the variation detection are performed concurrently. The differentiated $\mathrm{s} = \mathrm{d}+\mathrm{v}$ and $\mathrm{s} = \mathrm{d}-\mathrm{v}$ are analyzed for $\frac{\partial \mathrm{d}}{\partial \mathrm{s}}$ is the process performed in the differentiation. Contrarily, if $\mathrm{s} = \mathrm{d}-\mathrm{v}$ is observed, then $(\mathrm{s}+\mathrm{d}) \in [1,\ \mathrm{a_t}+1]$ and $(\mathrm{s}-\mathrm{d}) \in [\mathrm{a_t}+1,\ \mu_\mathrm{t}]$ are classified such that $\frac{\partial \mathrm{d}}{\partial \mathrm{a_t}}$ achieves $\mathrm{y} = \partial \mathrm{v}$ for all the variations (classification 1). Similarly, $\mathrm{y}\ (\partial \mathrm{v}) = \beta$-based solutions, as in equation 10 (b), are also analyzed in the appropriate time intervals. These two classifications share a common time interval for identifying the variations and errors. Therefore, other than y instance of differentiation, the analysis time is confined in $[1,\ \mathrm{a_t}]$. For the other processes, the difference between $(\mathrm{a_t}+1,\ \mathrm{a_t})$ and $(\mathrm{a_t}+1,\ \mu_\mathrm{t})$ is augmented with the time for $\partial \mathrm{v}$ and β. Hence, both the processes share this delay, and it is found to be less for different data stream sizes [Refer to Fig. 9].

D. Variation Analysis

Fig. 10 presents the variations observed for different data stream sizes. It is found that the variations in the proposed PDA are less compared to the other methods. The variation is $\mathrm{s} = \mathrm{d}-\mathrm{v}$ is identified as $\frac{\partial \mathrm{s}}{\partial \mathrm{a_t}}$ for which the first differentiation generates $\mathrm{y}(\partial \mathrm{d})$ solution $\forall\ \mathrm{i} = \mathrm{a_t} + 1$ to$\mathrm{u_t}$ in $\alpha = \frac{\partial \mathrm{d}}{\partial \mathrm{s}}$ [refer to equation (3)]. In the consecutive process, $\frac{\partial \mathrm{s}}{\partial \mathrm{a_t}}$ is useful in identifying the variations through β detection and classification. The probability-based error and variation are performed for classifying different cases of $\mathrm{d} \pm \mathrm{v}$ that occur conjointly. In error classification 1, $\frac{\partial \mathrm{d}}{\partial \mathrm{a_t}}$ is further differentiated for $\beta = [1 + \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}]\ \alpha$ and $\beta = \frac{\partial \mathrm{v}}{\partial \mathrm{u_t}}$. From the above differentiation, $\mathrm{y}(\partial \mathrm{v})$ is further segregated by identifying $\beta = (1-\alpha)$ instances. This identification reduces the variation to error probability, and hence there exists differentiation and error. Therefore, the variation observed in this proposed PDA is connected to the simpler data response stream; thus, the variation is less.

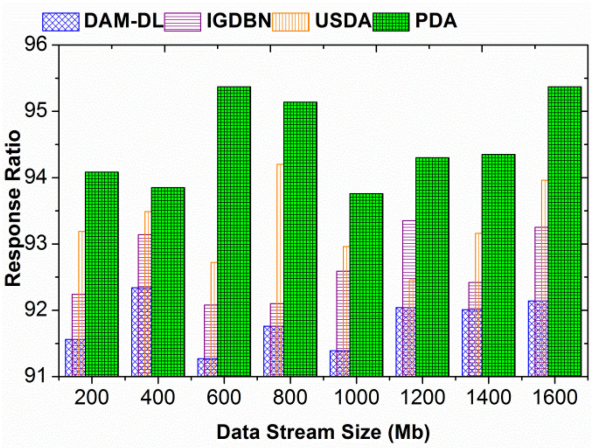
E. Response Ratio

The proposed PDA generates simplified healthcare service data streams for improving the response ratio. The useful data streams are first differentiated and then classified based on the condition of $\frac{\partial \mathrm{d}}{\partial \mathrm{s}} > \frac{\partial(\mathrm{d}-\mathrm{v})}{\partial \mathrm{s}}$ and $\frac{\partial \mathrm{d}}{\partial \mathrm{s}} < \frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)}$. The first differentiation achieves $\frac{\partial \mathrm{v}}{\partial(\mathrm{a_t}+1)} = 0$, and hence the accumulated healthcare data is sufficient for satisfying the response of the services. Different from this process, the variation classification helps to differentiate β for improving α and $\partial \mathrm{v}$ segregation. In this classification process, $\partial \mathrm{v}$ augments the simplified s in different classification instances is high, meeting the healthcare service demands. The process of differentiating $\mathrm{s} = \mathrm{d} \pm \mathrm{v}$ is unanimous for different data stream sizes using regressive learning. This is performed in a linear manner for improving simplified healthcare data responses. Therefore, the response ratio is high in the proposed PDA [Refer to Figs. 11(a) and (b)]. The comparative analysis results are tabulated in Tables III and III for different data stream sizes and classification instances.

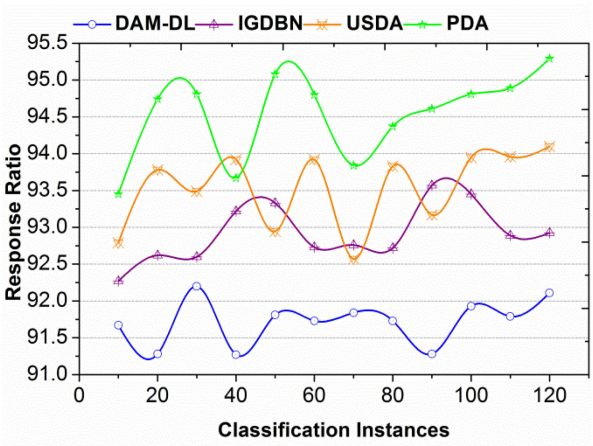
From Table III, it is seen that the proposed PDA achieves 6.83% high precision, 6.76% high response ratio. On the other hand, it achieves 11.927% less error, 26.82% less analysis time, and 13.1% fewer variations.

PDA achieves 6.7% high precision and 6.75% high response ratio for the different classification instances, and 13.28% less error.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.



(a)



(b)

Fig. 11. (a) and (b) Response ratio for data stream size and classification instnaces

TABLE III
COMPARATIVE ANALYSIS FOR DATA STREAM SIZES

Metrics	DAM-DL	IGDBN	USDA	PDA
Precision	0.889	0.898	0.908	0.9211
Error %	10.59	7.45	6.64	4.251
Analysis Time (ms)	1408.53	1251.3	1193.45	939.883
Variations	0.131	0.121	0.101	0.074
Response Ratio	92.14	93.253	93.96	95.371

TABLE IV
COMPARATIVE ANALYSIS FOR CLASSIFICATION INSTANCES

Metrics	DAM-DL	IGDBN	USDA	PDA
Precision	0.91	0.926	0.94	0.959
Error %	6.39	5.79	4.78	3.432
Response Ratio	92.11	92.926	94.09	95.291

VI. CONCLUSION

Healthcare data analytics is a tedious process due to variations in accumulated data from different sources. This results in inappropriate service responses in any healthcare application; for addressing the above issue, a proportionate data analytics is introduced in this manuscript. The proposed data analytics method differentiates the healthcare data stream that is responded to for the service. The differentiation is performed with the help of regressive learning for improving the disseminating of simplified response streams. The data streams are classified for variations and errors under controlled time for prompt services. In this, predictive analysis is used linearly for accumulation and update time instances. The output of the classification is the fine-grained differentiation of data error and variations. By using the errorless variations, the healthcare service streams are strengthened for better responses. The experimental analysis shows that the PDA achieves high accuracy by reducing error, analysis time, and healthcare data stream handling variations.

REFERENCES

[1] K. T. Kadhim, A. M. Alsahlany, S. M. Wadi, and H. T. Kadhum, "An overview of patient's health status monitoring system based on Internet of Things (IoT)," *Wireless Pers. Commun.*, vol. 114, no. 1, pp. 1–28, 2020.

[2] M. M. Islam, A. Rahaman, and M. R. Islam, "Development of smart healthcare monitoring system in IoT environment," *SN Comput. Sci.*, vol. 1, no. 3, 2020, Art. no. 185.

[3] J. C. Ni, C. S. Yang, J. K. Huang, and L. C. Shiu, "Combining non-invasive wearable device and intelligent terminal in healthcare IoT," *Procedia Comput. Sci.*, vol. 154, pp. 161–166, 2019, doi: 10.1016/j.procs.2019.06.024.

[4] A. A. Mutlag, M. K. AbdGhani, N. A. Arunkumar, M. A. Mohammed, and O. Mohd, "Enabling technologies for fog computing in healthcare IoT systems," *Future Gener. Comput. Syst.*, vol. 90, pp. 62–78, 2019.

[5] F. Khalique, S. A. Khan, and I. Nosheen, "A framework for public health monitoring, analytics and research," *IEEE Access*, vol. 7, pp. 101309–101326, 2019.

[6] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: A review," *J. Supercomputing*, vol. 76, no. 3, pp. 1754–1799, 2020.

[7] K. D. Strang, "Problems with research methods in medical device big data analytics," *Int. J. Data Sci. Analytics*, vol. 9, no. 2, pp. 229–240, 2020.

[8] P. T. Chen, C. L. Lin, and W. N. Wu, "Big data management in healthcare: Adoption challenges and implications," *Int. J. Inf. Manage.*, vol. 53, 2020, Art. no. 102078, doi: 10.1016/j.ijinfomgt.2020.102078.

[9] S. Nazir *et al.*, "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95714–95733, 2020.

[10] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," *Big Data Mining Analytics*, vol. 2, no. 4, pp. 288–305, 2019.

[11] R. Zgheib, S. Kristiansen, E. Conchon, T. Plageman, V. Goebel, and R. Bastide, "A scalable semantic framework for IoT healthcare applications," *J. Ambient Intell. Humanized Comput.*," pp. 1–19, 2020.

[12] E. Adi, A. Anwar, Z. Baig, and S. Zeadally, "Machine learning and data analytics for the IoT," *Neural Comput. Appl.*, vol. 32, pp. 1–29, 2020, doi: 10.1007/s00521-020-04874-y.

[13] M. A. Morid, O. R. L. Sheng, K. Kawamoto, and S. Abdelrahman, "Learning hidden patterns from patient multivariate time series data using convolutional neural networks: A case study of healthcare cost prediction," *J. Biomed. Informat.*, vol. 111, 2020, Art. no. 103565, doi: 10.1016/j.jbi.2020.103565Get.

[14] S. S. Johnston, J. M. Morton, I. Kalsekar, E. M. Ammann, C. W. Hsiao, and J. Reps, "Using machine learning applied to real-world healthcare data for predictive analytics: An applied example in bariatric surgery," *Value Health*, vol. 22, no. 5, pp. 580–586, 2019, doi: 10.1016/j.jval.2019.01.011.

[15] M. S. Hadi, A. Q. Lawey, T. E. El-Gorashi, and J. M. Elmirghani, "Patient-centric hetnets powered by machine learning and big data analytics for 6G networks," *IEEE Access*, vol. 8, pp. 85639–85655, 2020.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.

[16] Y. Zheng and X. Hu, "Healthcare predictive analytics for disease progression: A longitudinal data fusion approach," *J. Intell. Inf. Syst.*, vol. 55, pp. 351–369, 2020, doi: 10.1007/s10844-020-00606-9.

[17] O. Alfarraj and A. Tolba, "Unsynchronized wearable sensor data analytics model for improving the performance of smart healthcare systems," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 1, pp. 1–12, 2020.

[18] H. Fouad, A. S. Hassanein, A. M. Soliman, and H. Al-Feel, "Analyzing patient health information based on IoT sensor with AI for improving patient assistance in the future direction," *Measurement*, vol. 159, 2020, Art. no. 107757, doi: 10.1016/j.measurement.2020.107757.

[19] M. A. Serhani, H. T. El-Kassabi, K. Shuaib, A. N. Navaz, B. Benatallah, and A. Beheshti, "Self-adapting cloud services orchestration for fulfilling intensive sensory data-driven IoT workflows," *Future Gener. Comput. Syst.*, vol. 108, pp. 583–597, 2020, doi: 10.1016/j.future.2020.02.066.

[20] Z. Al-Makhadmeh and A. Tolba, "Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach," *Measurement*, vol. 147, no. 2, 2019, Art. no. 106815.

[21] M. Abdel-Basset, G. Manogaran, A. Gamal, and V. Chang, "A novel intelligent medical decision support model based on soft computing and IoT," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4160–4170, May 2019.

[22] W. N. Ismail, M. M. Hassan, H. A. Alsalamah, and G. Fortino, "CNN-based health model for regular health factors analysis in Internet-of-Medical Things environment," *IEEE Access*, vol. 8, pp. 52541–52549, 2020.

[23] C. Kakkanatt, M. Benigno, V. M. Jackson, P. L. Huang, and K. Ng, "Curating and integrating user-generated health data from multiple sources to support healthcare analytics," *IBM J. Res. Develop.*, vol. 62, no. 1, pp. 2:1–2:7, 2018.

[24] M. I. Razzak, M. Imran, and G. Xu, "Big data analytics for preventive medicine," *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4417–4451, 2020.

[25] K. Fawagreh and M. M. Gaber, "Resource-efficient fast prediction in healthcare data analytics: A pruned random forest regression approach," *Computing*, vol. 102, pp. 1187–1198, 2020, doi: 10.1007/s00607-019-00785-6.

[26] A. Manashty, J. L. Thompson, and H. Soleimani, "A concise temporal data representation model for prediction in biomedical wearable devices," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1438–1445, Apr. 2019.

[27] Y. Yang, "Medical multimedia big data analysis modeling based on DBN algorithm," *IEEE Access*, vol. 8, pp. 16350–16361, 2020.

[28] H. Sun, Z. Liu, G. Wang, W. Lian, and J. Ma, "Intelligent analysis of medical big data based on deep learning," *IEEE Access*, vol. 7, pp. 142022–142037, 2019.

[29] B. C. Kwon *et al.*, "Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records," *IEEE Trans. Visual. Comput. Graph.*, vol. 25, no. 1, pp. 299–309, Jan. 2019.

[30] O. Banos, R. Garcia and A. Saez, [Online]. Available: https://archive.ics.uci.edu/ml/datasets/MHEALTH+Dataset

[31] H. Atasoy, B. N. Greenwood, and J. S. McCullough, "The digitization of patient care: A review of the effects of electronic health records on health care quality and utilization," *Annu. Rev. Public Health*, vol. 40, pp. 487–500, 2019, doi: 10.1146/annurev-publhealth-040218-044206.

Authorized licensed use limited to: Carnegie Mellon Libraries. Downloaded on April 26,2023 at 13:22:28 UTC from IEEE Xplore. Restrictions apply.