Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 14 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i14.41949

PAPER

A Learning Health-Care System for Improving Renal Health Services in Peru Using Data Analytics

Vielka Mita[1], Liliana Castillo[1], José Luis Castillo-Sequera[2], Lenis Wong[1](✉)

[1]Program Information Systems Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Perú

[2]Department of Computer Science, Universidad de Alcalá, Alcalá de Henares, Spain

pcsilewo@upc.edu.pe

ABSTRACT

The health sector around the world faces the continuous challenge of improving the services provided to patients. Therefore, digital transformation in health services plays a key role in integrating new technologies such as artificial intelligence. However, the health system in Peru has not yet taken the big step towards digitising its services, currently ranking 71st according to the World Health Organisation (WHO). This article proposes a learning health system for the management and monitoring of private health services in Peru based on the three key components of intelligent health care: (1) a health data platform (HDP); (2) intelligent technologies (IT); and (3) an intelligent health care suite (HIS). The solution consists of four layers: (1) data source, (2) data warehousing, (3) data analytics, and (4) visualization. In layer 1, all data sources are selected to create a database. The proposed learning health system is built, and the data storage is executed through the extract, transform and load (ETL) process in layer 2. In layer 3, the Kaggle dataset and the decision tree (DT) and random forest (RF) algorithms are used to predict the diagnosis of disease, resulting in the RF algorithm having the best performance. Finally, in layer 4, the intelligent health-care suite dashboards and interfaces are designed. The proposed system was applied in a clinic focused on preventing chronic kidney disease. A total of 100 patients and six kidney health experts participated. The results proved that the diagnosis of chronic kidney disease by the learning health system had a low error rate in positive diagnoses (err = 1.12%). Additionally, it was demonstrated that experts were "satisfied" with the dashboards and interfaces of the intelligent health-care suite as well as the quality of the learning health system.

KEYWORDS

chronic kidney disease, learning health-care system, random forest (RF), decision tree (DT), machine learning

1 INTRODUCTION

The World Health Organization (WHO) defines health service as any service that can contribute to the improvement of health through the diagnosis, treatment,

Mita, V., Castillo, L., Castillo-Sequera, J.L., Wong, L. (2023). A Learning Health-Care System for Improving Renal Health Services in Peru Using Data Analytics. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(14), pp. 76–97. https://doi.org/10.3991/ijoe.v19i14.41949

Article submitted 2023-06-03. Revision uploaded 2023-07-30. Final acceptance 2023-08-02.

or rehabilitation of people with health problems. This definition is not necessarily limited to medical care services [1]. The World Bank and WHO estimate that half the world's population lacks the resources to afford essential health services [2]. According to Peters et al. [3], developing countries such as those in Latin America have a disadvantage in accessing to health services compared to other countries due to factors like geographic accessibility, economics and availability [3]. Regarding the private health sector, WHO defines it as those organisations that are not directly owned or controlled by governments and are responsible for providing health services [4]. A consequence of Peru's health system problems is the lack of strategic plans for the prevention of diseases with high morbidity rates [5]. In Peru, more than 50% of the disease burden is associated with noncommunicable diseases, i.e., diseases developed by a set of risk factors and affecting all age groups [6]. According to the National Institute of Statistics and Informatics (INEI) in the "Demographic and Family Health Survey 2021", 36.9% of people over 14 years old are overweight, 25.8% suffer from obesity, 17.2% have recently excessive alcohol consumption, and 8.2% are daily smokers [7]. The probability of getting respiratory and cardiovascular diseases, cancer, or diabetes increases when a person has these risk factors [8]. Most of those diseases are preventable; therefore, health experts recommend strengthening surveillance systems on risk factors, frequent health evaluations, adopting healthy lifestyles, and avoiding tobacco use [9]. Health services are an important key for the development of a country, and, in recent years, digital transformation has had a boost in this sector due to COVID-19; however, despite the advances, Peru ranks 71st at digitization of services [10], evidencing the fact that health is not part of the 5 sectors that most implement technology in the country [11].

Throughout the last few years, many studies have emerged focused on health services [10], applied to different types of technologies [12] [13], and aiming at improving the performance of the health sector. However, most of these studies do not reflect particularities that characterise health services in Latin American countries, such as social and economic barriers to accessing health services [3].

Therefore, the learning health system (LHS) infrastructure is proposed to improve the health services performance in Peru, based on three key components for smart health care [14]: (1) the health data platform (HDP), (2) intelligent technologies (IT) and (3) the intelligent health care suite (IHS). The proposed system is developed in four layers: data dource, data storage, analytics, and visualisation.

The rest of the article comprises the following sections: Section 2 presents the related works. The details of the architecture of the proposed system are found in Section 3. In Section 4, the validation process is presented. The results and discussion are illustrated in Section 5. Finally, Section 6 presents the conclusions.

2 RELATED WORK

2.1 Architectures and frameworks

In [15] [16] [17] [18], architectures were designed with technologies applied in health services. Lo et al. [15] and Manogaran et al. [18] address a specific health service such as medical referrals and patient monitoring; Liu et al. [16] and Wang et al. [17] focus their contributions on more than one health service and health centre capacity management. Lo et al. [15] propose an architecture that integrates patient data from two health organisations to improve medical referrals using blockchain, whereas Manogaran et al. [18] have a clinical monitoring approach and use a cloud environment and sensors with IoT technology. Liu et al. [16] design their architecture assisted by a digital twin cloud to provide data availability and real-time monitoring with the implementation of a virtual copy, but Wang et al. [17] choose analytics, big data, and their different capacities for decision-making and the prediction of health indicators. In [19] [14] [20] [21] [22] frameworks are presented to understand concepts and the operation of technology applied in health services. Menear et al. [19], Gopal et al. [14], and Kopel [20] carry out a conceptual framework that gives a vision of capabilities and technologies to use in our solution as the capabilities to build health systems [19] [14], and acceptable conditions for them [20]. Sengan et al. [21] and Puri et al. [22] present frameworks explaining the technologies to be used in their contribution, such as the clinical nodes for the use of smart contracts [22] and the flow for the operation of a prediction system using machine learning [21].

2.2 Technologies applied in health services

Different technologies have been found to be applied to improve health services in healthcare systems. Most systems were built to improve one or more services, ensure the availability and security of medical data, and discover clinical information. Among those found, the most used technology is data analytics [23] [24] [25]; artificial intelligence [21] [26] [27] [28] [23] [29] and cloud [30]. Prediction models were found within the technologies in health services, mostly in articles with random forest (RF) and decision tree (DT) algorithms. Fernández-Gutiérrez et al. [23] used big data analytics and data mining techniques to identify changes in sick people's electronic medical records in primary care. Sengan et al. [21], Chen et al. [27], and Fernández-Gutiérrez et al. [23] use the DT algorithm for disease prediction to improve health services, whereas Yu et al. [28] and Kumar et al. [29] use the RF algorithm to manage treatments and early detection of diseases. On the other hand, blockchain [30] and the Internet of Things (IoT) [14] [28] [29] have been used in some studies. These techniques were generally used together with more of the mentioned technologies. Puri et al. [23] used big data analytics and data mining techniques to identify sick people's events from electronic medical records in primary care. Kumar and Devi Gandhi [29] use IoT wearable sensor devices to monitor a person's clinical values and send an alert if any value exceeds normal to make intelligent decisions.

2.3 Learning health systems

Some LHS have been applied in the health sector, such as general medicine, health care in COVID-19, oncology, and rehabilitation, among others. Zhang et al. [12] present an LHS that uses sensor data to obtain a diagnosis and control report for patients, achieving the centralization of resources, information boards, and the indirect participation of patients. Shah et al. [31] use structured data from electronic medical records for the management of the oncology specialty and the monitoring of its patients; the LHS also accepts unstructured data due to the importance of information from sources such as clinical notes and images. Chen et al. [27] carry out a disease prediction system for patients at risk in a hospital in China using machine learning algorithms, obtaining 94.8% accuracy in the prediction of the cases, as well as Williamson et al. [32], who use this technology for the COVID-19 management cases using the Open Safely platform. [33] [34] [13] present as a contribution the development of dashboards for the understanding of clinical data, surveillance data, and drug control, respectively, only using structured data. In [35], the purpose of medical research is to carry out an LHS to interact with various medical centres in Indiana. Kim and Huh et al. [36], on the other hand, apply artificial intelligence and blockchain techniques to ensure the protection and verification of medical data in electronic health records (EHRs). French et al. [37] propose an LHS for the creation of a precision rehabilitation data repository that facilitates access to systematically collected data from the EHR. In [38], they evaluate 500 political interventions to find out if they could improve the characteristics of the rapid learning health system (RLHS) in the Iranian health system during the first seven months of COVID-19, where they identify that the characteristics focused on patients and their data should be improved. Dammery et al. [39] present a case study to determine which elements of an LHS are being implemented in a university medical centre, where they identify that LHS practices could be improved in two aspects: in the use of patient data and computing. Tosteson et al. [40] propose the implementation of the coproduction learning health system (CLHS) model to transform care in an oncology programme of an academic medical centre, obtaining a set of lessons learned.

2.4 Comparison of studies

In the literature review, various studies were found that propose an LHS for different health sectors, but not all have applied the three key components for intelligent health care (HDP, IHS and IT) [14], and none have been implemented in the systems health of developing countries.

Table 1 shows that most of the studies have considered one or two key components for intelligent health care, with the exception of [31], which does consider the three key components. For example, in [27], they consider the IT component, where they apply a disease risk prediction model but do not handle data storage management (HDP) or a platform that integrates business layers (IHS). In [32], the OPENSafely system is explained as a platform built to discover the factors associated with death from COVID-19 (IHS) through a statistical analysis after creating a storage environment for medical records (HDP), but not apply IT that generate better results.

Table 1. Comparative analysis on the use of the three key components for intelligent health care

Study	Health Sector	HDP	IHS	IT
[12]	Public Health	Yes	No	Yes
[31]	Oncology	Yes	Yes	Yes
[27]	Chronic diseases	No	No	Yes
[32]	Covid 19	Yes	Yes	No
[33]	Pediatric rehabilitation	Yes	No	Yes
[34]	Biomedical	Yes	No	No
[13]	General Medicine	Yes	Yes	No

(Continued)

Table 1. Comparative analysis on the use of the three key components for intelligent health care *(Continued)*

Study	Health Sector	HDP	IHS	IT
[35]	General Medicine	Yes	No	No
[36]	General Medicine	Yes	No	Yes
[37]	Rehabilitation	Yes	Yes	No
[38]	Covid-19	Yes	Yes	No
[39]	General Medicine	Yes	No	No
[40]	Oncology	Yes	Yes	No

3 PROPOSED LEARNING HEALTH-CARE SYSTEM

The proposed learning health-care system is based on the three key components for intelligent health care, according to [14]: the HDP, IT and IHS, as shown in Figure 1. Gopal et al. [14] provide an important approach for digital transformation in health services, their outcomes, and how to set the journey for applying future technologies.

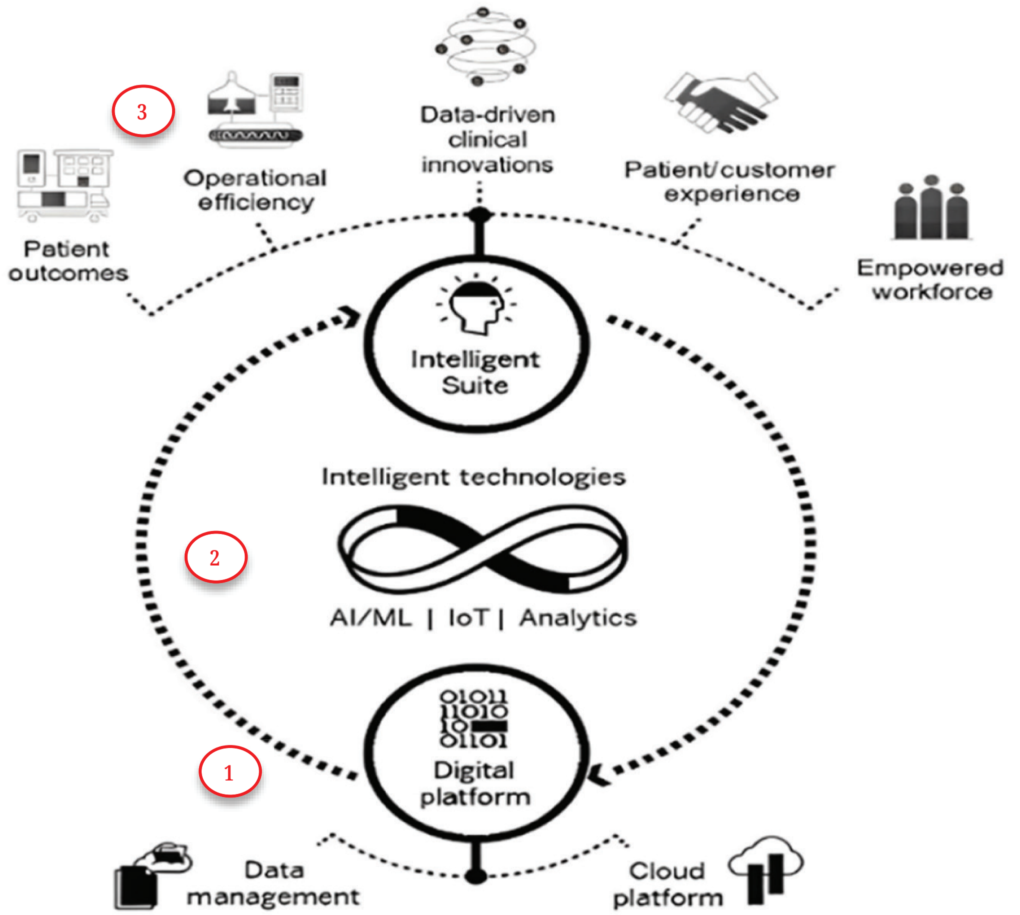


Fig. 1. The intelligent healthcare enterprise framework [14]

In addition, the architecture of big data analytics for health care is taken into account according to the proposal of [17]. The solution is made up of four layers: (1) data source, (2) data storage, (3) analytics and (4) visualisation, as shown in Figure 2. The "data source" layer represents the data source for health tracking. In layer 2, the extract, transform and load (ETL) process occurs. A variety of IT to be used in layer 3 (DT and RF algorithms) are represented after building an HDP in the "data storage" layer. Finally, in the "visualisation" layer, dashboards and interfaces (IHS) are designed that will serve to improve health services with clinical data. In order to build the proposed system, the Peru renal health service was chosen as the focus of this study due to the increase in chronic kidney disease (CKD) patients and their mortality in the last few years, according to the National Epidemiology Centre, Peru. [41].

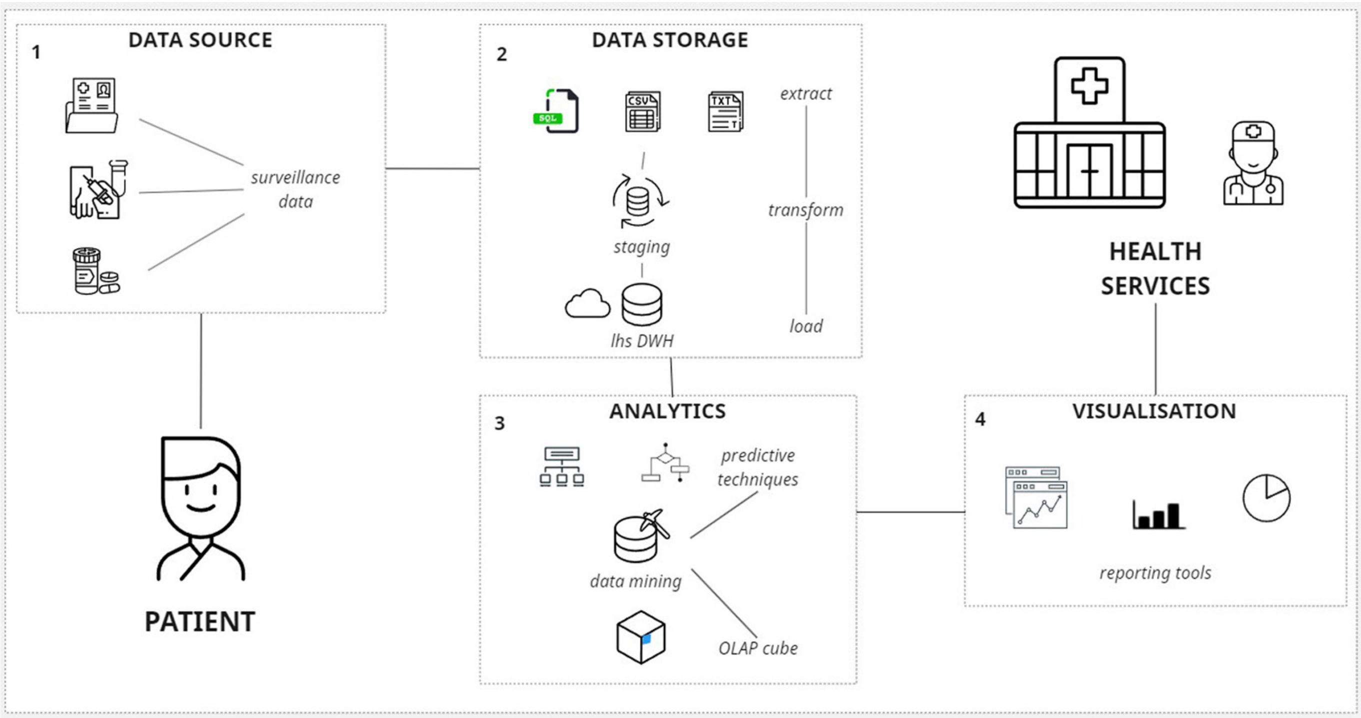


Fig. 2. Learning health system architecture

3.1 Layer 1: data source

This layer represents all necessary data sources to collect, such as patient, doctor, treatment and medicine data, to develop the system that aims to improve health services in renal clinics. Also, for the discovery of new medical information, it is necessary to have data about medical attention, surveillance, and diagnoses. Health data can be stored in flat file databases (*.xlsx,.csv*) or traditional relational databases (SQL Server, MYSQL, Oracle) for patient health records; this research focuses on renal health centres with a low technology maturity level where some or no digital system is used, which is why, in the applied case, *xlsx* and *csv* data sources were collected. In order to build LHS, the database diagram is designed to identify the relationships between 15 tables and the necessary data for the proposed solution, as shown in Figure 3.

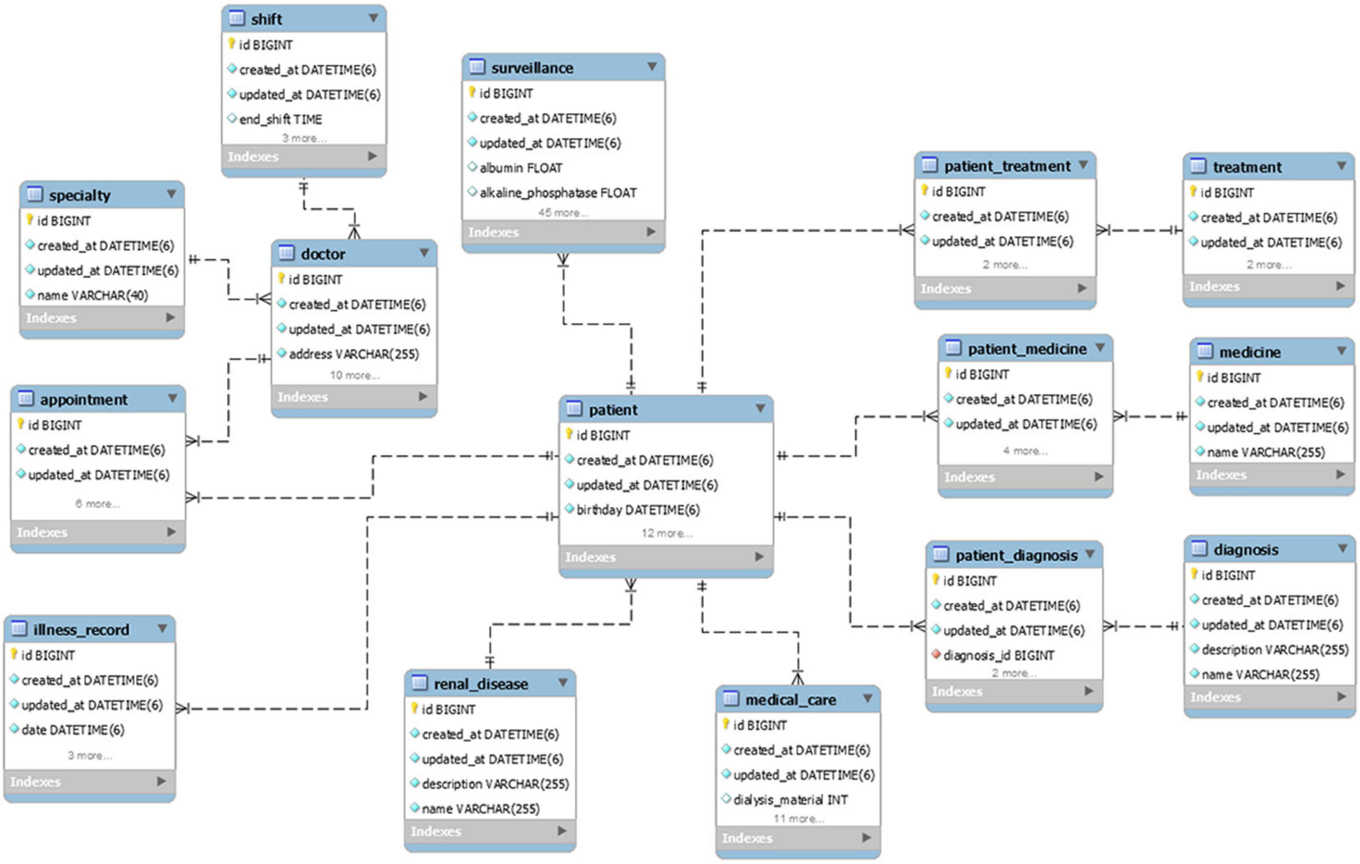


Fig. 3. LHS database diagram

3.2 Layer 2: data storage

Building LHS. The LHS is developed as a web application using Angular for the frontend and Java for the backend. The proposed system has three modules: (i) surveillance, (ii) appointments and (iii) attentions. In addition, it communicates with a prediction service hosted on a cloud platform to predict a certain disease in all patients through clinical analysis. Figure 4 shows the logical architecture of the proposed system. The "surveillance" module is in charge of storing medical analysis; the "appointments" module is integrated as it is part of administrative management; and the "attention" module is in charge of monitoring the health of patients during dialysis sessions. For the "appointments" module, a table called 'appointment' is created with variables: date, appointment status, and observations. For the "surveillance" module, another table called 'surveillance' is created that stores all the patient record data in each medical attention. With this data collection, variations of indicators such as weight, hematological data, symptoms, urological data and care received are available. Through this layer, the HDP and IHS components are fulfilled for the success of digital transformation in health [14].

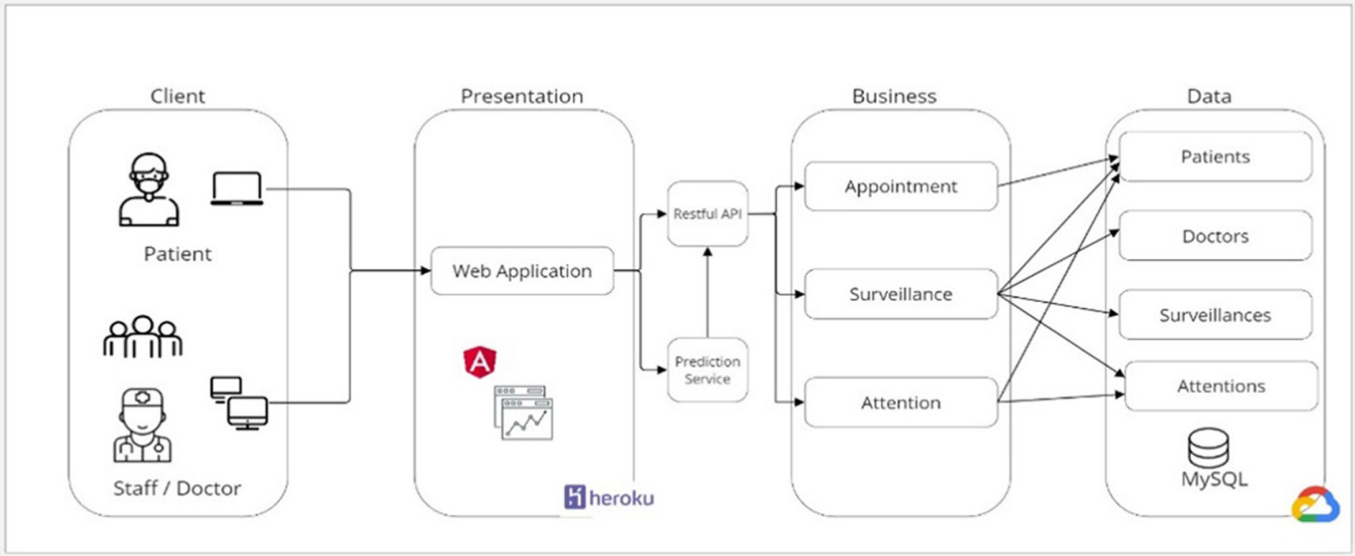


Fig. 4. Logical architecture

Design of data storage. For data storage, the ETL process was carried out in Visual Studio 2019 with SQL server data tools installed to use its SSIS Designer, which allows extracting data from different sources and origins, cleaning, modifying, and inserting it into a data warehouse [42]. In Figure 5, the ETL process carried out is represented. For the "extraction", the data that facilitates the construction of reports and predictions in LHS was selected with the DataFlow Task in the SSIS designer toolbox. In "transformation", the data and tables were customised with SQLTask using the SQL language, and some calculations were made with transformation Tasks according to the needs found. In "loading", processed data was inserted into the LHS data warehouse to be used with data analytics techniques in the next layer.

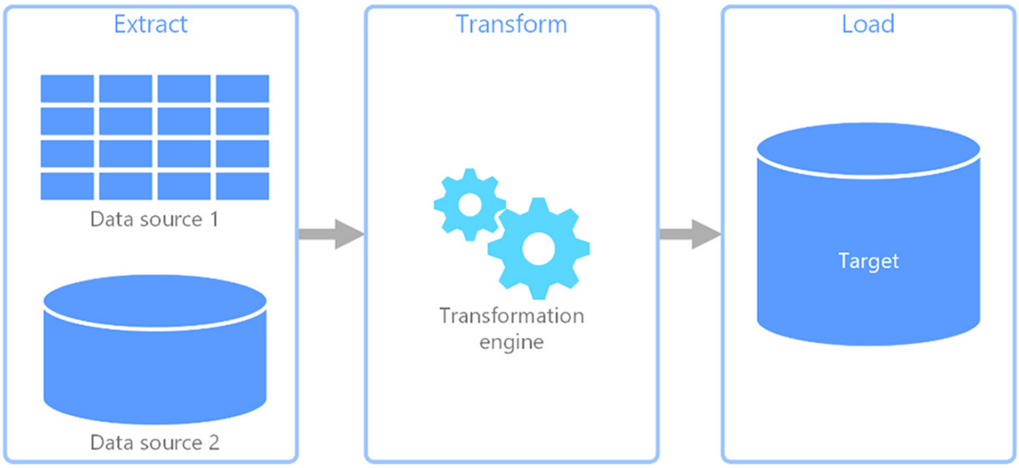


Fig. 5. ETL process [42]

3.3 Layer 3: data analytics

Predictive model. This LHS solution allows the storage of clinical patient data and facilitates the use of techniques such as machine learning (ML) for predictive analysis. For the choice of technique, a comparison of the performance of the two supervised classification algorithms most commonly used and with the best metrics in the literature was made: DT [27] and RF algorithms [28], which obtained an accuracy of 65% and 76%, respectively.

Dataset. The dataset used for the prediction model is real patients' data from India; this country and Peru are considered developing countries; therefore, their realities are similar [3]. The dataset is a *csv* flat file with numeric and text data type records; it also consists of 25 columns and 400 patients' health records with or without CKD, information collected for two months. This dataset is found on the open platform Kaggle [43]. Table 2 shows a dataset made up of 25 attributes and classified into four categories, whose attributes were also an important factor in the choice of dataset.

Table 2. Dataset taxonomy

Categories	Attributes
Patient characteristics	Age (F01)
Symptoms	Appetite (F02)
Analysis result	glucose (F03), urea (F04), creatinine (F05), sodium (F06), potassium (F07), hemoglobin (F08), leukocytes (F09), hematocrit (F10), lymphocytes (F11), albumin (F12), sugar (F13), red blood cells (F14), pus cell (F15), pus cell clumps (F16), bacteria (F17), density (F18), blood pressure (F19), classification (CKD/not CKD) (F25)
Diseases and health history	anemia (F20), hypertension (F21), diabetes (F22), heart disease (F23), pedal enema (F24).

Data was divided into training data and test data; the Pareto principle [44] was used to set the ratio at 80% and 20%, respectively, that is, 320 records as training data and 80 as test data. For this study, the quality of the data was verified with a nephrologist from Lima, who confirmed that dataset attributes are used in patient control forms, i.e., they are decisive factors for CKD prevention.

CKD prediction. This section presents the prediction model using the DT and RF algorithms in order to evaluate its performance with different validation indicators. The Python programming language is used to implement both ML algorithms. Figure 6 shows the flowchart of the DT algorithm, where the initialization of variables is first performed: the number of data processed (N), the list of data attributes (*at*), and the averages of the data to be used (*av_at*). On startup, a counter (*a*) and a list of attributes with random values for displaying the results (at [*x,y,z*]) are automatically created.

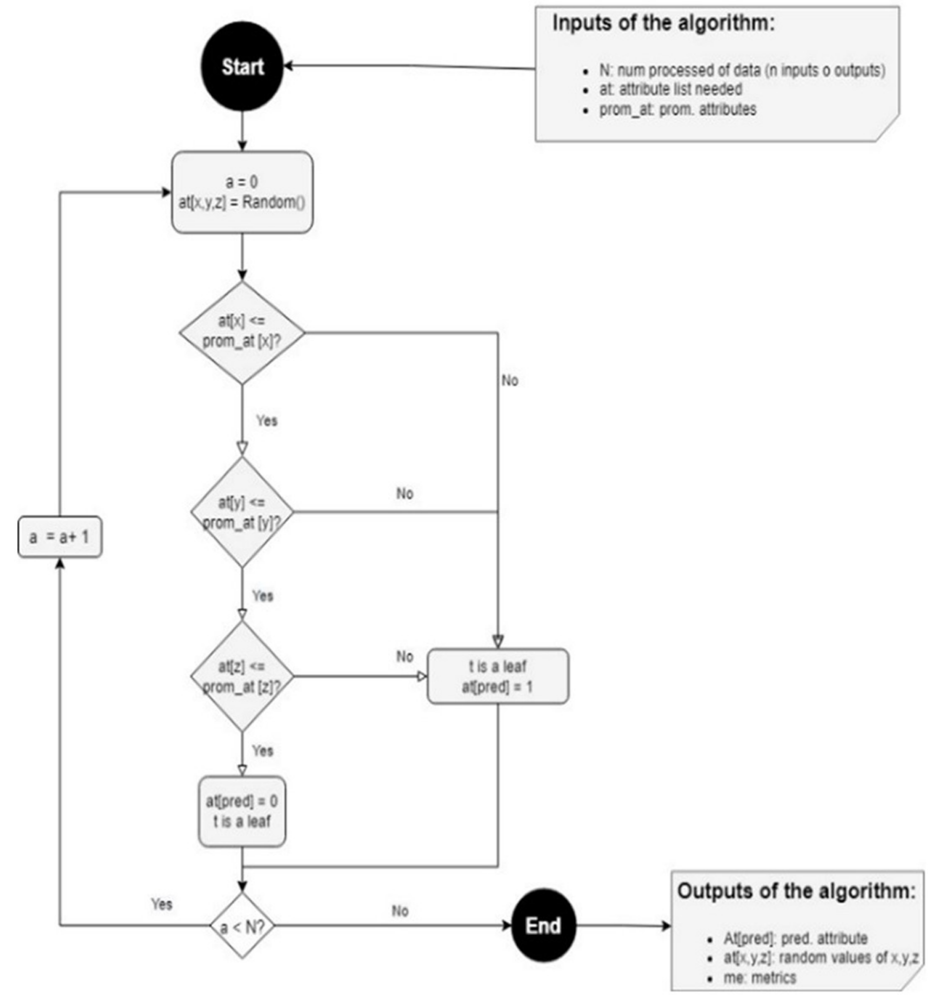


Fig. 6. DT algorithm flowchart

These results are then compared with their respective averages by means of a conditional. While each one is positive, it is passed to the prediction attribute (at [pred]) so that it has its result, which in this case can be 0 or 1. Then, these results become a leaf, causing the path of the tree to end up passing to a last conditional that asks if the counter is less than the data processed. This cycle will be repeated as many times until this condition is false, which is when the flow ends, giving as outputs the prediction attribute (at [pred]), random variables (*at*), and metrics that the algorithm has (*me*).

Figure 7 shows the flowchart of the RF algorithm, where variables are initialised: model characteristics (features), model result (labels), number of trees (t), and percentage of data that will be used in training (*p*). The flow starts by reading the data from the dataset and then separating it by attributes and labels. This algorithm is created with the training data of the attributes, labels, and number of trees. The model receives the test data, and each decision tree created determines which label belongs in the model based on the received attributes. The resulting label is determined based on the label that is returned in the results of each decision tree. The outputs of this prediction model are the prediction label (*r*) and the performance metrics.

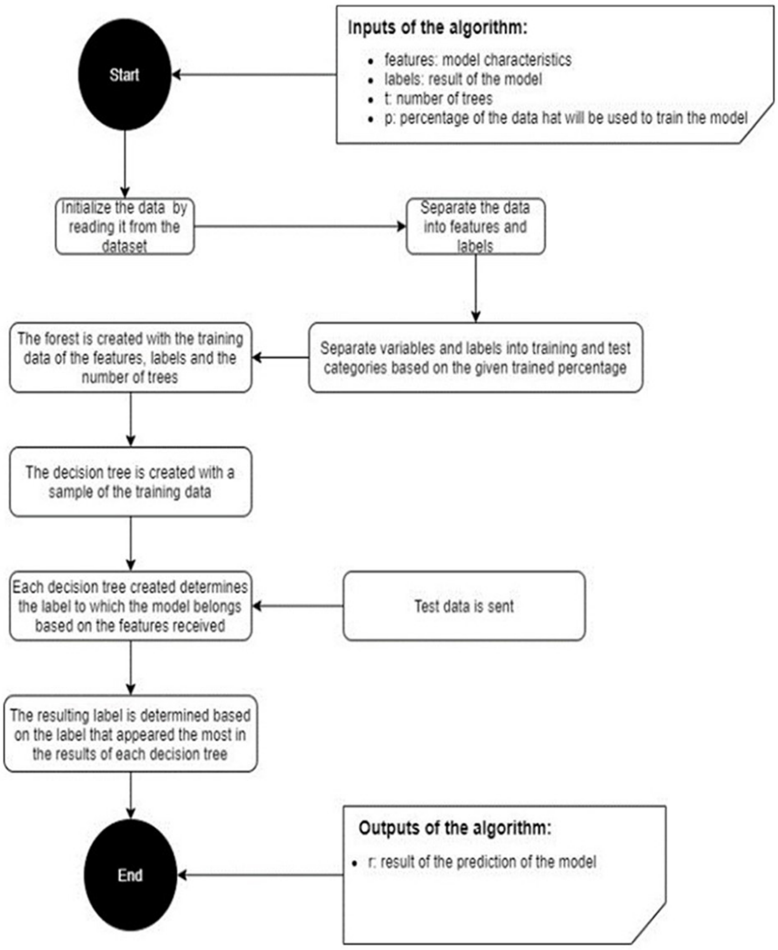


Fig. 7. RF algorithm flowchart

Evaluation metrics. To evaluate the performance of the prediction model to detect a patient with or without CKD, four metrics will be used: accuracy (Eq. 1), precision (Eq. 2), recall (Eq. 3), and F1-Score (Eq. 4). For this, a confusion matrix is used, which is a conventional method for evaluating classification algorithms [45].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2\ X\ Precisión\ X\ Recall}{Precisión + Recall} \tag{4}$$

Where,

- *TP* (True positives): number of correct predictions as records that have CKD
- *TN* (True negatives): number of correct predictions as records that do not have CKD
- *FP* (False positives): number of incorrect predictions as records that have CKD
- *FN* (False negatives): number of incorrect predictions as records that do not have CKD.

In addition, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) are used to assess the advantages and disadvantages of algorithms [46].

Algorithms result. Figure 8 shows the confusion matrix for both algorithms. The results show that the RF algorithm (Figure 8b) has a better performance than the DT algorithm (Figure 8a) since it obtained 0 *FP* and *FN*, while RF got 3 *FP* and 2 *FN*.

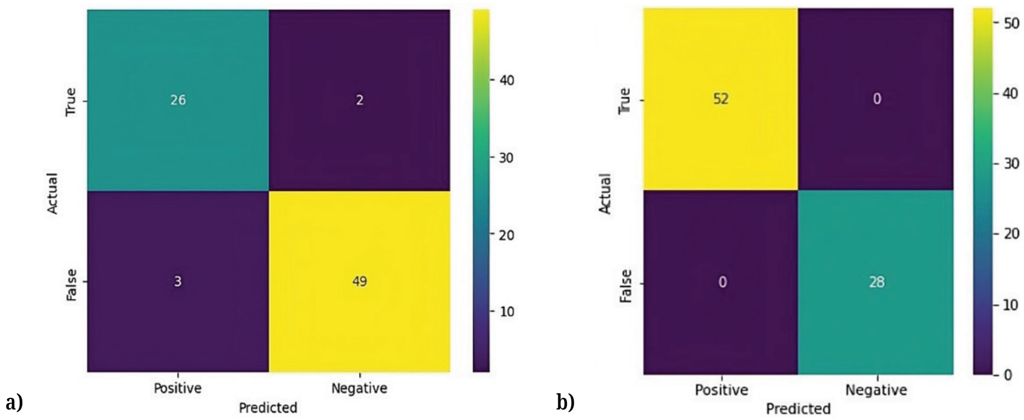


Fig. 8. Confusion matrix results of DT (a) and RF (b)

Likewise, Figure 9 represents a comparison of metrics for both algorithms. This reveals that the RF algorithm (Figure 9b) fulfills the 4 metrics evaluated with the Kaggle dataset at 100%, whereas the DT algorithm (Figure 9a) reached 85% on average.

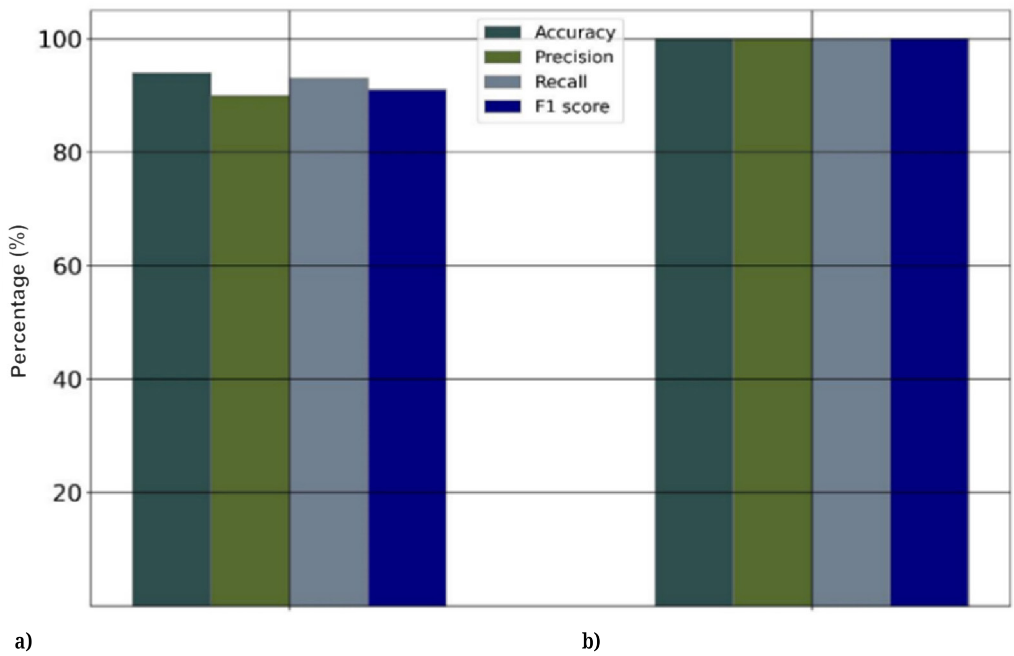
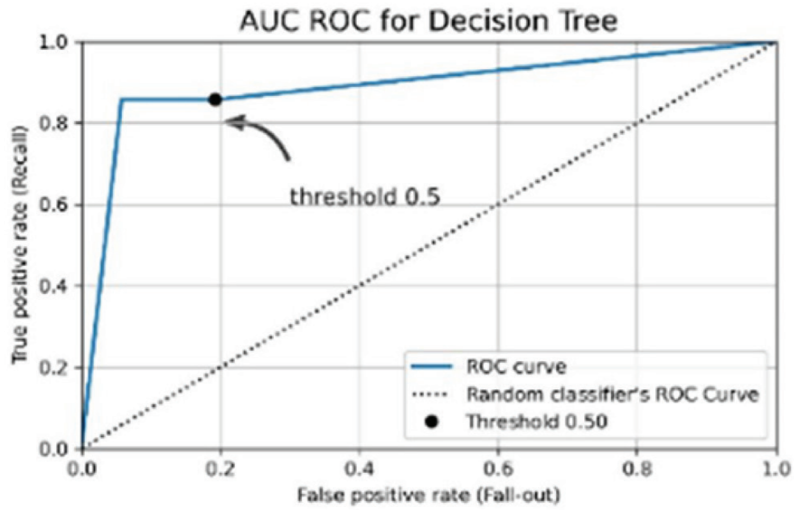
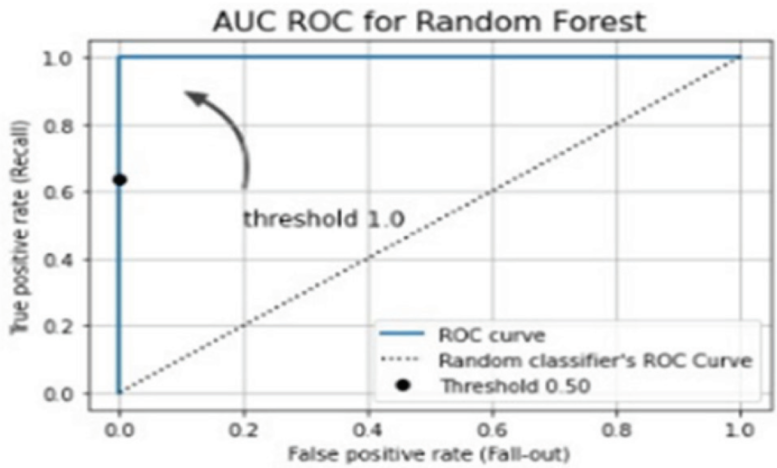


Fig. 9. Comparison of accuracy, precision, recall and F1-score of DT (a) and RF (b)

Figure 10 shows the DT and RF algorithms ROC curve comparison. It's evidenced that the RF algorithm achieved the best performance because the RF ROC curve is closer to the upper left corner (Figure 10b), unlike the DT ROC curve, which is further away (Figure 10a). This means the RF algorithm obtained a better proportion of TP (number of correct predictions as records that have CKD) over FP (number of incorrect predictions as records that have CKD). With the testing process's results, the RF algorithm is selected for the CKD prediction model to be used in learning health system.



a)



b)

Fig. 10. ROC curves, DT (a) and RF (b)

3.4 Layer 4: visualisation

In this layer, five informative dashboards are designed: (1) health indicators; (2) treatments; (3) diagnoses; (4) patients at risk and (5) efficiency control. These dashboards were made with Microsoft PowerBI and in Spanish for the use of Peruvian experts. For this purpose, a connection to the MySQL database named "LHS" is first made to access all data tables. Then charts and attributes are designed that allow an easy reading of patients' health control during all their medical attention. Consequently, dashboards are published on the Power BI App platform, which provides a link to insert dashboards in the LHS web system. Some examples of dashboards are illustrated in Figures 11, 12 and 13.

4 VALIDATION

The validation of the proposed LHS was carried out in a private health and kidney disease prevention clinic located in Lince district, in Metropolitan Lima. The clinic meets Level 1 in technological maturity according to the IS4H maturity model elaborated by the WHO [11], so the implementation of LHS is appropriate, and we know how much that can improve its health services. To validate the proposed LHS, two activities were carried out: (i) medical records analysis and (ii) experimentation with LHS. The first activity is related to control and the second is system testing. In addition, a survey was conducted among six kidney health experts (administrators, nurses and doctors) to measure the level of satisfaction with learning health system.

4.1 Medical records analysis

The clinic provided a database of 100 medical records of men and women Peruvian patients who were attended in the period from May to October 2022. The medical history evidenced that 89 patients had CKD and 11 patients did not have the disease, according to the doctor's diagnosis. This database was obtained through an anonymisation process, according to Peruvian Law N° 29733, Article 14, Paragraph 8, about the protection of personal data [47]. For example, this law indicates that the consent of the data owner is not required for the purposes of its treatment when an anonymisation or dissociation process has been applied. The anonymisation procedure for the data from the medical records was carried out as indicated in the personal data protection manual [48]. The process consisted of the following steps: (a) preparation of medical records; and (b) deletion of patients' sensitive data (names, age, address and ID number).

4.2 System experiment

The experimentation for the proposed LHS consists of the following steps: insertion of clinical data, customization of dashboards, and insertion of patient data.

Insertion of clinical data. To start the experimentation of the system, the kidney clinic professionals test the LHS modules and insert data from attention, surveillance, patients, doctors and appointment records to visualise dashboards that allow streamlining management and control processes in the clinic.

Customization of dashboards. As a second step, the required dashboards are generated with the data entered in system modules. Figure 11 shows a dashboard with information about a patient's treatment, where the recorded values and mathematical indicator KT/V (KT = urea eliminated per session, V = volume of distribution of urea in the patient) are displayed in a dialysis session. The clinic works with a minimum value of 1.30 for the KT/V indicator. The figure shows that the patient's dialysis session was favourable by obtaining 1.93. This important indicator represents the efficiency of the hemodialysis session, and kidney health centres in Peru are frequently audited with the requirement to present a report of the results of KT/V. Earlier, the report was calculated manually by the clinic medical staff. Figure 12 shows a dashboard for health indicators of a patient that contains charts with a complete follow-up of variables such as the weight before and after the dialysis was completed, the KT/V value reached, and the urea levels on specific attention dates.

Fig. 11. Patient treatment dashboard

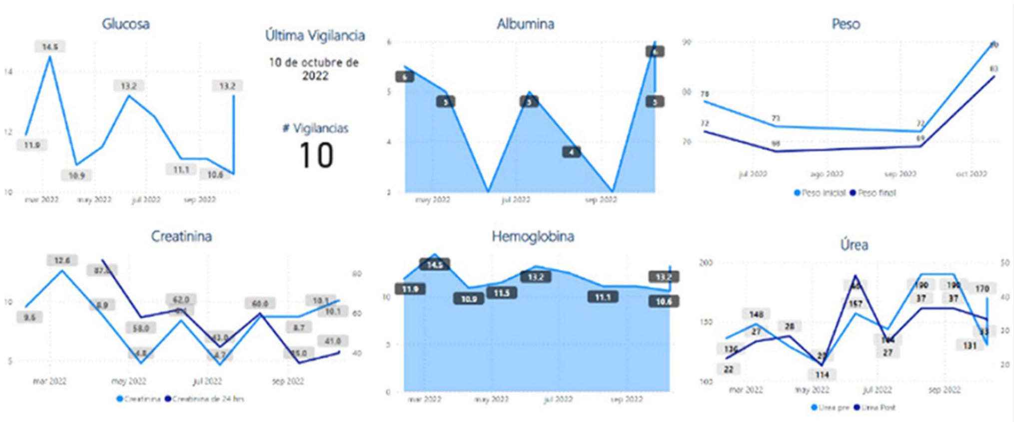


Fig. 12. Patient health indicators dashboard

Insertion of patient data. The health records of 100 patients were imported into LHS form a *CSV* file. In Figure 13, the CKD diagnosis for a patient with LHS is shown. When the data was entered, the result is "YES" (Sí), has CKD with a 75% probability of being in Stage 4.

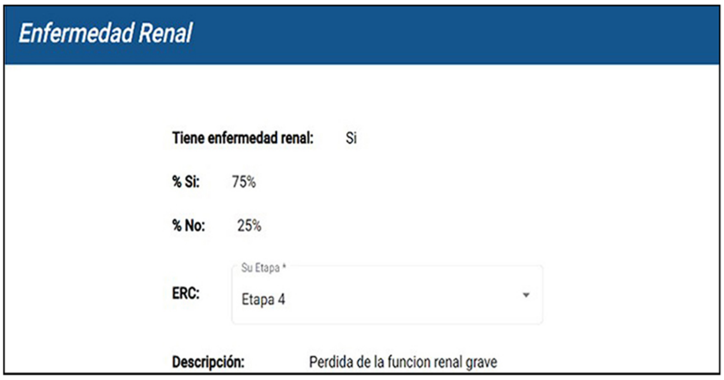


Fig. 13. Interface to obtain CKD diagnosis with LHS

4.3 Expert judgement

On the other hand, the validation through expert judgment was carried out in order to obtain their feedback for LHS. The participants were six renal health experts: two administrators, three nurses, and one doctor from renal health centers in Lima. For this aim, the following steps were taken: (1) system display; (2) use of all LHS modules by experts; and (3) taking an online survey. The experiments were performed separately for approximately one hour each. The survey was made up of seven closed-ended questions and one open question (Table 3), where questions Q1 to Q5 were asked to obtain the user's perception regarding the use of LHS; question Q6 is related to the quality attributes of LHS; question Q7 is related to the perception of the advantages of using EHR; and the open question Q8 was created with the purpose of seeking opportunities to improve the LHS. The Likert scale was applied to close-ended questions (1 = very unsatisfied, 2 = unsatisfied, 3 = neutral, 4 = satisfied and 5 = very satisfied).

Table 3. Survey questions

	Question	Type
Q1	Do you think LHS would be useful if implemented in Health Centers?	Close-Ended
Q2	Do you think LHS would improve health services?	Close-Ended
Q3	Do you consider dashboards shown would facilitate the patients' health monitoring?	Close-Ended
Q4	How satisfied are you with LHS dashboards?	Close-Ended
Q5	How would you rate the LHS?	Close-Ended
Q6	How satisfied are you with the following attributes of LHS?: reliability, usability, attractiveness, security, functionality, time behavior, stability, and content of dashboards.	Close-Ended
Q7	How would you rate EHR features compared to traditional medical records: availability, organized, updated, avoid human errors and integrated storage?	Close-Ended
Q8	What kind of difficulties did you find in LHS solution?	Open

5 RESULTS AND DISCUSSION

5.1 Experimentation

Table 4 shows the results of CKD diagnosis by doctor and LHS for 100 patients classified by genre. It evidences a variation of 1 in CKD diagnostic results for female patients.

Table 4. CKD diagnosis by doctor and LHS

Gender	CKD Diagnosis by Doctor			CKD Diagnosis by LHS		
	Yes	No	Total	Yes	No	Total
Male	42	7	49	42	7	49
Female	47	4	51	46	5	51
Total	89	11	100	88	12	100

To compare the results obtained with LHS, two metrics are used: difference in results (df) and error percentage (err), which are calculated with Eqs. (5) and (6), respectively.

Where:

- dD: diagnoses by doctor.
- $dLHS$: diagnoses by LHS.

$$df = \left| dLHS - dD \right| \quad (5)$$

$$err = \frac{|dLHS - dD|}{|dD|} * 100\% \tag{6}$$

Table 5 shows diagnoses obtained by the doctor and LHS and the results of metrics (df and err). It is observed that *err* metric in positive and negative diagnoses of CKD is low, that is, the variation between the true value and the experimental value is minimal, obtaining the lowest percentage with positive CKD diagnoses (1.12%), which indicates that the error percentage is acceptable because it is less than 10% and better than expected [27].

Table 5. CKD prediction results

CKD	Diagnosis by Doctor	Diagnosis by LHS	*df*	*err*
Yes	89	88	1	1.12%
No	11	12	1	9.09%

5.2 Expert judgement

On the other hand, Figure 14 shows the results of the survey taken by six experts on questions Q1, Q2, Q3, Q4 and Q5. It is observed that 100% of the experts gave a score of 5 (very satisfied) referring to LHS dashboards (Q4), and as for the other questions, an average score of 4.46 (satisfied) was obtained. Figure 15a shows the results on the quality attributes of LHS (Q6), and it is observed that the best qualified LHS attribute is reliability, obtaining an average score of 4.83, almost 5 (very satisfied). And Figure 15b shows the results regarding the advantages of EHR (Q7), and it is observed that the feature with the best score is the "organized aspect of the medical records", which received an average qualification of 4.67, almost 5 (indicating high satisfaction). Regarding the answers to the open question (Q8), one of the experts reported no difficulties in using LHS. Two experts recommended improving the interfaces due to the large number of fields to fill out in some modules. Other experts stated that information to be recorded must be shown with their units for correct use by staff.

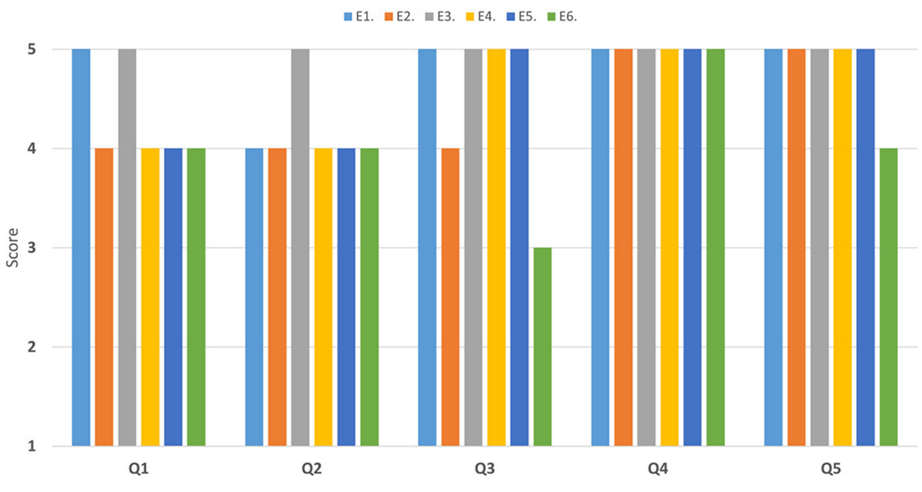
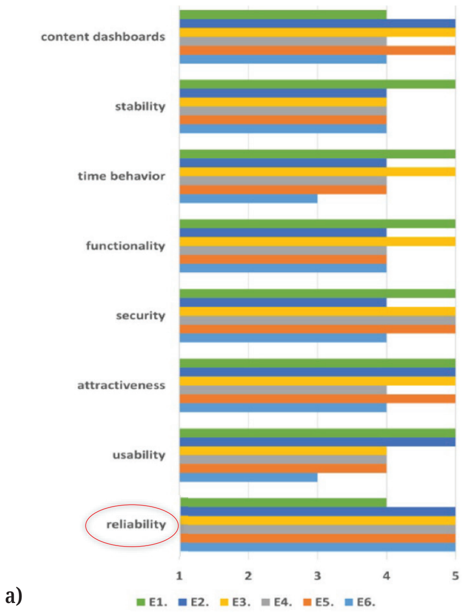


Fig. 14. Summary of responses to expert survey



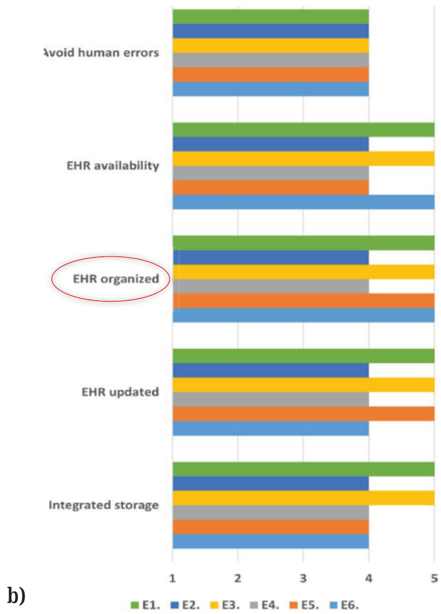


Fig. 15. LHS attributes (a) and EHR features (b) responses to expert survey

6 CONCLUSION

In the present study, a LHS was proposed for the management and monitoring of private health services in Peru based on the three key components of intelligent health care: HDP, IT and IHS. In addition, a big data analytics architecture was implemented. The solution was composed of four layers: data source, data storage, data analytics and visualisation. In the "data source" layer, all data sources were selected to create a database. The proposed LHS was developed, and the data storage was executed through the ETL process in the "data storage" layer. In the "data analytics" layer, the Kaggle dataset and the DT and RF algorithms were used to predict the diagnosis of the disease, resulting in the RF algorithm with the best performance (accuracy = 100%). Finally, in the "visualisation" layer, dashboards and IHS interfaces were developed to support clinical decision-making. To corroborate the efficacy of the proposed LHS, the validation of the study was carried out in a clinic for health and prevention of kidney diseases in Lima, Peru. The validation process consisted of: (1) entering the surveillance data of 100 patients into LHS; (2) comparing the results of LHS with the diagnosis by doctor; (3) testing all LHS functionalities by the experts; and (4) taking a survey with the experts. The CKD diagnosis by LHS had a low error rate in the positive diagnosis (err = 1.12%). In addition, with the results of the survey, it was evidenced that 100% of the experts were "very satisfied" with the use of LHS dashboards; the benefits obtained from the use of LHS had an average score of 4.4 (satisfied), and the quality of LHS obtained an average score of 4.44 (satisfied), with "reliability" being the quality attribute that stood out.

7 ACKNOWLEDGMENT

We are thankful to the Renal Prevention and Treatment Clinic staff for participating in this research and to the Research Department of the Universidad Peruana de Ciencias Aplicadas for the support provided to carry out this research work.

8 REFERENCES

[1] W. H. O., "Terminology: A glossary of technical terms on the economics and finance of health services," *World Health Organization,* 1998. https://www.euro.who.int/__data/assets/pdf_file/0014/102173/E69927.pdf

[2] W.H.O., "World Bank and WHO: Half the world lacks access to essential health services, 100 million still pushed into extreme poverty because of health expenses," *World Health Organization*, 2017. https://www.who.int/es/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses

[3] D. H. Peters, A. Garg, G. Bloom, D. G. Walker, W. R. Brieger, and M. Hafizur Rahman, "Poverty and access to health care in developing countries," *Ann. N. Y. Acad. Sci.*, vol. 1136, pp. 161–171, 2008. https://doi.org/10.1196/annals.1425.011

[4] J. Klinton, "The private health sector: An operational definition," *World Heal. Organ.*, pp. 1–5, 2020. [Online]. Available: https://www.who.int/docs/default-source/health-system-governance/private-health-sector-an-operational-definition.pdf

[5] T. P. Noticias, "Entrevista a César Loza, expresidente de la Sociedad Peruana de Nefrología," *YouTube*, Perú, 2022.

[6] PAHO, "Enfermedades no transmisibles y factores de riesgo." https://www3.paho.org/per/index.php?option=com_content&view=article&id=4110:enfermedades-no-transmisibles-y-factores-de-riesgo&Itemid=0

[7] INEI, "Perú: Enfermedades No Transmisibles y Transmisibles," 2021. https://proyectos.inei.gob.pe/endes/2021/SALUD/ENFERMEDADES_ENDES_2021.pdf

[8] WHO, "Obesidad y sobrepeso," *World Health Organization*, 2021. https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight

[9] PAHO, "Enfermedades no transmisibles." https://www3.paho.org/per/index.php?option=com_content&view=article&id=4119:enfermedades-no-transmisibles-2&Itemid=1062

[10] Gestion, "Transformación digital de la salud peruana: Cuánto se ha avanzado?" 2021. https://gestion.pe/peru/transformacion-digital-de-la-salud-peruana-cuanto-se-ha-avanzado-noticia/?ref=gesr

[11] Peru21, "Cuáles son los cinco sectores con mayor transformación digital en Perú?" 2022. https://peru21.pe/economia/cuales-son-los-cinco-sectores-con-mayor-transformacion-digital-en-peru-rmmn-noticia/

[12] Y. Zhang, M. Qiu, C. W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Syst. J.*, vol. 11, no. 1, pp. 88–95, 2017. https://doi.org/10.1109/JSYST.2015.2460747

[13] G. L. Krapohl *et al.*, "Building, scaling, and sustaining a learning health system for surgical quality improvement: A toolkit," *Learn. Heal. Syst.*, vol. 4, no. 3, pp. 1–7, 2020. https://doi.org/10.1002/lrh2.10215

[14] G. Gopal, C. Suter-Crazzolara, L. Toldo, and W. Eberhardt, "Digital transformation in healthcare – Architectures of present and future information technologies," *Clin. Chem. Lab. Med.*, vol. 57, no. 3, pp. 328–335, 2019. https://doi.org/10.1515/cclm-2018-0658

[15] Y. S. Lo, C. Y. Yang, H. F. Chien, S. S. Chang, C. Y. Lu, and R. J. Chen, "Blockchain-enabled iWellChain framework integration with the national medical referral system: Development and usability study," *J. Med. Internet Res.*, vol. 21, no. 12, 2019. https://doi.org/10.2196/13563

[16] Y. Liu *et al.*, "A novel cloud-based framework for the elderly healthcare services using digital twin," *IEEE Access*, vol. 7, pp. 49088–49101, 2019. https://doi.org/10.1109/ACCESS.2019.2909828

[17] Y. Wang, L. A. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technol. Forecast. Soc. Change*, vol. 126, pp. 3–13, 2018. https://doi.org/10.1016/j.techfore.2015.12.019

[18] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system," *Futur. Gener. Comput. Syst.*, vol. 82, pp. 375–387, 2018. https://doi.org/10.1016/j.future.2017.10.045

[19] M. Menear, M. A. Blanchette, O. Demers-Payette, and D. Roy, "A framework for value-creating learning health systems," *Heal. Res. Policy Syst.*, vol. 17, no. 1, pp. 1–14, 2019. https://doi.org/10.1186/s12961-019-0477-3

[20] R. Koppel, "The health information technology safety framework: Building great structures on vast voids," *BMJ Qual. Saf.*, vol. 25, no. 4, pp. 218–220, 2016. https://doi.org/10.1136/bmjqs-2015-004746

[21] S. Sengan, G. Kamalam, J. Vellingiri, J. Gopal, P. Velayutham, and V. Subramaniyaswamy, "Medical information retrieval systems for e-Health care records using fuzzy based machine learning model," *Microprocess. Microsyst.*, p. 103344, 2020. https://doi.org/10.1016/j.micpro.2020.103344

[22] V. Puri, A. Kataria, and V. Sharma, "Artificial intelligence-powered decentralized framework for Internet of Things in Healthcare 4.0," *Trans. Emerg. Telecommun. Technol.*, no. e4245, pp. 1–18, 2021. https://doi.org/10.1002/ett.4245

[23] F. Fernández-Gutiérrez, J. I. Kennedy, R. Cooksey, M. Atkinson, E. Choy, S. Brophy, L. Huo, and S. M. Zhou, "Mining primary care electronic health records for automatic disease phenotyping: A transparent machine learning framework," *Diagnostics*, vol. 11, no. 10, p. 1908, 2021. https://doi.org/10.3390/diagnostics11101908

[24] L. Yan, W. Huang, L. Wang, S. Feng, Y. Peng, and G. Peng, "Data-enabled digestive medicine: A new big data analytics platform," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 922–931, 2021. https://doi.org/10.1109/TCBB.2019.2951555

[25] A. Dagliati, L. Sacchi, V. Tibollo, G. Cogni, M. Teliti, A. Martinez-millana, V. Traver, D. Segagni, J. Posada, G. Fico, M. T. Arredondo, P. De Cata, L. Chiovato, and R. Bellazzi, "A dashboard-based system for supporting diabetes care," *J. Am. Med. Informatics Assoc.*, vol. 25, no. 5, pp. 538–547, 2018. https://doi.org/10.1093/jamia/ocx159

[26] A. Asadzadeh, Z. Mohammadzadeh, Z. Fathifar, S. Jahangiri-Mirshekarlou, and P. Rezaei-Hachesu, "A framework for information technology-based management against COVID-19 in Iran," *BMC Public Health*, vol. 22, no. 1, pp. 1–15, 2022. https://doi.org/10.1186/s12889-022-12781-1

[27] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, p. 8869–8879, 2017. https://doi.org/10.1109/ACCESS.2017.2694446

[28] G. Yu *et al.*, "Improving chronic disease management for children with knowledge graphs and artificial intelligence," *Expert Syst. Appl.*, vol. 201, p. 117026, 2022. https://doi.org/10.1016/j.eswa.2022.117026

[29] P. M. Kumar and U. Devi Gandhi, "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases," *Comput. Electr. Eng.*, vol. 65, pp. 222–235, 2018. https://doi.org/10.1016/j.compeleceng.2017.09.001

[30] A. Mashat and A. M. Alabdali, "QoS-aware smart phone-based user tracking application to prevent outbreak of COVID-19 in Saudi Arabia," *Comput. Intell. Neurosci.*, vol. 2022, no. iv, 2022. https://doi.org/10.1155/2022/2273910

[31] A. Shah, A. K. Stewart, A. Kolacevski, D. Michels, and R. Miller, "Building a rapid learning health care system for oncology: Why cancerlinq collects identifiable health information to achieve its vision," *J. Clin. Oncol.*, vol. 34, no. 7, pp. 756–763, 2016. https://doi.org/10.1200/JCO.2015.65.0598

[32] E. J. Williamson *et al.*, "Factors associated with COVID-19-related death using OpenSAFELY," *Nature*, vol. 584, no. 7821, pp. 430–436, 2020. https://doi.org/10.1038/s41586-020-2521-4

[33] N. Koscielniak *et al.*, "The SHOnet learning health system: Infrastructure for continuous learning in pediatric rehabilitation," *Learn. Heal. Syst.*, vol. 6, no. 3, pp. 1–8, 2022. https://doi.org/10.1002/lrh2.10305

[34] M. S. Kohn, U. Topaloglu, E. S. Kirkendall, A. Dharod, B. J. Wells, and M. Gurcan, "Creating learning health systems and the emerging role of biomedical informatics," *Learn. Heal. Syst.*, vol. 6, no. 1, pp. 1–8, 2022. https://doi.org/10.1002/lrh2.10259

[35] T. Schleyer *et al.*, "The indiana learning health system initiative: Early experience developing a collaborative, regional learning health system," *Learn. Heal. Syst.*, vol. 5, no. 3, p. 1, 2021. https://doi.org/10.1002/lrh2.10281

[36] S. K. Kim and J. H. Huh, "Artificial neural network blockchain techniques for healthcare system: Focusing on the personal health records," *Electron.*, vol. 9, no. 5, 2020. https://doi.org/10.3390/electronics9050763

[37] M. A. French *et al.*, "A learning health system infrastructure for precision rehabilitation after Stroke," *Am. J. Phys. Med. Rehabil.*, vol. 102, no. 2, pp. S56–S60, 2023. https://doi.org/10.1097/PHM.0000000000002138

[38] A. Ahmadi, L. Doshmangir, and R. Majdzadeh, "Exploring nationwide policy interventions to control COVID-19 from the perspective of the rapid learning health system approach," *Learn. Heal. Syst.*, no. e10363, pp. 1–11, 2023. https://doi.org/10.1002/lrh2.10363

[39] G. Dammery *et al.*, "The journey to a learning health system in primary care: A qualitative case study utilising an embedded research approach," *BMC Prim. Care*, vol. 24, no. 1, pp. 1–10, 2023. https://doi.org/10.1186/s12875-022-01955-w

[40] A. N. A. Tosteson *et al.*, "Harnessing the collective expertise of patients, care partners, clinical teams, and researchers through a coproduction learning health system: A case study of the dartmouth health promise partnership," *J. Ambul. Care Manage.*, vol. 46, no. 2, pp. 127–138, 2023. https://doi.org/10.1097/JAC.0000000000000460

[41] Ministerio de Salud, "Boletín Epidemiológico del Perú," 2018. https://www.dge.gob.pe/portal/docs/vigilancia/boletines/2018/16.pdf

[42] IBM, "IBM Cloud Education. ETL (Extract, Transform, Load)," 2020. https://www.ibm.com/cloud/learn/etl

[43] "UCI Machine Learning Repository: Chronic_Kidney_Disease Data Set." https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease (accessed Mar. 11, 2023).

[44] R. Kumar, K. Singh, and S. K. Jain, "Agile manufacturing: A literature review and Pareto analysis," *Int. J. Qual. Reliab. Manag.*, vol. 37, no. 3, pp. 207–222, 2020. https://doi.org/10.1108/IJQRM-12-2018-0349

[45] K. Markham, "Simple guide to confusion matrix terminology," *Data School*, 2020. https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

[46] K. Gowrishankar, V. Muthukumar, R. Sudhakara, S. Deivasigamani, and Ang, Chun Kit, "Novel SVM and K-NN classifier based machine learning technique for epileptic seizure detection," *Int. J. Online Biomed. Eng.*, vol. 19, no. 07, pp. 99–124, 2023. https://doi.org/10.3991/ijoe.v19i07.37881

[47] El Peruano, "Ley peruana N° 29733. Ley sobre la protección de datos personales. Limitaciones al consentimiento para el tratamiento de datos personales." https://www.leyes.congreso.gob.pe/Documentos/Leyes/29733.pdf

[48] Defensoría del Pueblo, "Manual de protección de datos personales. Adjuntía en los Asuntos Constitucionales." https://www.defensoria.gob.pe/wp-content/uploads/2019/11/Manual-de-Protección-de-Datos-Personales.pdf

9 AUTHORS

Vielka Mita. She is a graduate in Information Systems Engineering from Peruvian University of Applied Sciences in 2022. She is currently working as a SAP Analytics Consultant at a Technology Consultancy. She has knowledge about digital transformation, data analysis, business process and project management. She also is part of projects of SAP Business Warehouse for digital transformation in Energy and Retail companies (E-mail: u20171b752@upc.edu.pe).

Liliana Castillo. She has a degree in information systems engineering from the Peruvian University of Applied Sciences. She has knowledge about machine learning, process improvement, intern control and project management using agile methodologies. She is currently working as an internal audit analyst at a telecommunications company (E-mail: u201710373@upc.edu.pe).

José Luis Castillo-Sequera. He has a degree in computing from the National University of San Marcos, Peru, a Masters' degree in computer project management and a Masters' degree in university teaching in Spain, and Ph.D. degree in information systems, documentation, and knowledge from the University of Alcalá, Spain. He is currently a Full Professor with the Department of Computer Science, University of Alcalá. He has multiple publications at the level of high-level indexed JCR journals and is the author of books and book chapters related to the field of artificial intelligence (E-mail: jluis.castillo@uah.es).

Lenis Wong. She is a professor and researcher of Software Engineering and Information Systems Engineering at the National University of San Marcos and Peruvian University of Applied Sciences, Peru. She holds a PhD in Systems Engineering and Computer Science. MSc. in Systems and Computer Engineering with mention in Software Engineering. She is a member of the AI research group and has carried out different multidisciplinary projects where Artificial Intelligence, Software Engineering and Information Systems Engineering have been applied in education, health, medicine and healthcare. She has published several international peer-reviewed scientific articles in different multidisciplinary areas such as: ML, DL, IoT, e-Health, Software Engineering, Requirements Engineering, Cloud Computing, E-Learning, Gamification, Cyberattacks, Natural Language Processing, Networks and Blockchain Technologies (E-mail: pcsilewo@upc.edu.pe).