Digital Twin for Intelligent Context-Aware IoT Healthcare Systems

Haya Elayan, **Moayad Aloqaily**, *Member, IEEE*, and **Mohsen Guizani**, *Fellow, IEEE*

***Abstract*—Since the emergence of Digital and smart Healthcare, the world has hastened to apply various technologies in this field to promote better health operation and patients' well-being, increase life expectancy, and reduce healthcare cost. One promising technology and game-changer in this domain is *Digital Twin (DT)*. DT is expected to change the concept of Digital Healthcare and take this field to another level that has never seen before. DT is a virtual replica of a physical asset that reflects the current status through real-time transformed data. This paper proposes and implements an intelligent context-aware healthcare system using the DT framework. This framework is a beneficial contribution to digital healthcare and to improve healthcare operations. Accordingly, an ECG heart rhythms classifier model was built using machine learning to diagnose heart disease and detect heart problems. The implemented models successfully predicted a particular heart condition with high accuracy in different algorithms. The collected results have shown that integrating DT with the healthcare field would improve healthcare processes by bringing patients and healthcare professionals together in an intelligent, comprehensive, and scalable Health-Ecosystem. Also, implementing an ECG classifier that detects heart conditions gives the inspiration for applying ML and AI with different human body metrics for continuous monitoring and abnormalities detection. Finally, Neural-Network-based algorithms deal better with ECG data than traditional ML algorithms.**

***Index Terms*—Digital Twin, Internet of Things, Smart Healthcare, Machine Learning, ECG.**

I. Introduction

Technological developments throughout the ages have led to the emergence of new tools, techniques, and machines. These developments have contributed to the practical improvements in various fields such as manufacturing, agriculture, education, and even the health sector. One of the best examples of such technological developments is the Internet of Things (IoT).

IoT has been integrated into today's lifestyle through connecting everything to almost everything including smartphones, smart buildings, smart homes, as well as healthcare wearable devices. Moreover, IoT sensors and devices have contributed towards the improvement of healthcare systems by facilitating the health workflow, speeding the access to medical records, increasing the accuracy of collected data from different sources, sharing capabilities, as well as fighting pandemics [1] [2]. According to reports published by the US Institute of Medicine that medical errors claiming the lives of $\sim$400,000 people each year due to issues related to data inefficiency [3]. More specifically it is the inability to access a patient's medical history, missed and delayed diagnoses, or corrupted health data. IoT technological advancements have significantly influenced the Healthcare system in connecting it to the users personal device can capture, store, and notify the health institutes with the relevant health data in real-time and thereby increase effective health support and reduce the mortality rate [4].

The rise in personal health monitoring devices in the form of mobile applications or built-in sensors can actively monitor user's vital health parameters such as ECG, BP, heart rate, and the sugar level which reduces the potential errors of data recording. These devices can capture and transfer data anonymously to the cloud and compare it with historical data for symptoms of any illness or notify the appropriate health personnel (doctor, nurse, or health agent). Fewer errors mean better performance, cost, efficiency, and improvements in healthcare services where an error can literally be the difference between life and death. This is an intelligent context-aware IoT health era that is made possible by the convergence of technology and healthcare [5]. In turn, this can improve the quality of life and solve many of the challenges such as information sharing, diagnoses inefficiency, monitoring cost reduction, operations optimization, medication errors, etc.

Digital twin is the third trending technology for 2020, according to the IEEE computer society's [6]. The concept of this technology refers to a digital replica of the physical object. DT combines Artificial Intelligence (AI), Data Analytics, IoT, Virtual and Augmented Reality paired with digital and physical objects [7]. This integration allows real-time data analysis, status monitoring to head off problems before they even occur, risk management, cost reduction, and future opportunities prediction.

For healthcare systems, having a virtual replica of a patient could be an optimal solution for health promotion, increase control over health, and improve healthcare operations [8]. This simulation will help monitor the patient's current health status. Besides, it will be possible to predict the future trend using medical history, and much more. Integrating DT with healthcare to improve its processes is our motivation to design an intelligent IoT healthcare system using a DT framework. The framework has employed real-time dataset as explained later in sub-section IV-A. This framework combines IoT, data analytics, and machine learning to make the patients' virtual replica a reality, give the healthcare professionals more capabilities to control and enhance a patient's health, and take in the cooperation of patients with similar cases into the process to utilize real-life scenarios. The Framework comprises three phases:

- Processing and Prediction phase,

H. Elayan is with xAnalytics Inc., Ottawa, Canada. E-mail: haya.elayan@xanalytics.ca.

M. Aloqaily is with Faculty of Engineeting, Al Ain University, UAE. E-mail: maloqaily@ieee.org

M. Guizani is with Qatar University, Qatar. E-mail: mguizani@ieee.org

- Monitoring and Correction phase,
- Comparison phase.

Each phase is responsible for improving an aspect of healthcare operations, the patient's aspect, the healthcare professional's aspect, or other patients with similar cases.

Towards this objective, we developed a novel Electrocardiogram (ECG) classifier that diagnoses heart disease and detects heart problems. The machine learning classifier trained using real-time data of ECG rhythms collected from different patients through sensing electrodes. This classifier was chosen based on a performance evaluation comparison of five implemented models with different deep learning and traditional machine learning algorithms. The collected results are extensive and discuss the viability of the proposed framework implementation.

The remainder of this paper is structured as follows. Section II discusses the related literature work to this paper. Section III presents the framework architecture. Section IV describes the system implementation and system workflow. Section V discusses the use case results and evaluation. Section VI presents the challenges that face DT technology in healthcare sector. Finally, Section VII concludes the paper and describes directions of future work.

II. Related work

Since its inception, digital twins became useful and feasible technology, especially in healthcare applications. With the remarkable interest shown by the research community and industry in integrating DTs with healthcare in recent years, this section provides the most relevant research in this domain.

Amongst the relevant research attempts in this area, is the work proposed by Rivera *et. al.* [9]. The authors have proposed a reference model for DT healthcare systems, based on the principles of self-adaptation and autonomic computing that enables continuous monitoring and forecasting of the patient's condition. To prop their approach, they exemplified a motivational scenario in managing the diabetes chronic disease. However, they did not support it with process implementation.

On the other side, Liu *et. al.* [10] has also proposed a cloud-based DT system for elderly healthcare. They constructed a reference framework (Cloud-DTH) by combining cloud architecture with the DT Healthcare (DTH) model that was initially introduced. This combination aims to facilitate the computation and efficient management in healthcare systems. Moreover, two case studies on how the cloud-DTH mode model enables individualized healthcare have been introduced. Unfortunately, the case studies lacked evaluation of performance and results. It also didn't indicate whether AI or machine learning algorithms were used in the prediction process. The unsuccessful emerging of DT in smart healthcare systems relies on the use of autonomous machine learning algorithms to manage this process.

An application to manage hospital's services has been proposed by Karakra *et. al.* [11]. The authors have presented a hospital DT framework using discrete event simulation systems and IoT devices. The hospital's services workflow have been optimized using a predictive decision support model, using real-time data without interrupting daily activities. FlexSim HC software has been used to test the feasibility of the proposed methodology with different scenarios. The DT presence is not clear in the proposed model.

The authors in [12] have presented the idea of detecting seizures before its development using machine learning technology. This is being developed through analyzing seizure signals collected from DRE patients. A deep spiking neural-network model for the epileptic seizure detection using a surrogate gradient-based has been also developed. The complexity of the developed model has not been tested. Moreover, the integration of the different proposed models with neuromorphic chips assumes secure without the need for cloud computing. Unfortunately, their model has been trained on a small number of data that might cause an optimistic estimation of system performance.

The work presented in [13] has introduced DT for personalized healthcare to create a complete reliable and secure system. Therefore, a DT architecture for lung cancer behavior of patients under treatment which aims at providing concrete clinical information. The proposed architecture also included Generative Adversarial Networks (GANs) to allow complete anonymization of health records and obtain flexible models to generate fake patients. Likewise, the authors in [14] proposed an anonymization of patient information in the health sector using GANs to generate fake images and avoid the risk of sensitive data leakage. Several GAN systems have been trained for fake data generation. The study has pointed out that convolutional NN might help with dynamic data that needs sophisticated GAN architecture.

A wearable 2.0 healthcare system has been proposed in [15] to improve the QoE and QoS of the healthcare experience. Their proposal presents washable smart clothing (sensors, electrodes, and wires) that collect physiological data to predict emotional status using cloud-based machine intelligence. The authors of [16] have proposed a system architecture for health IoT and big data problem. It show remote health detection primary diagnostic services, smart healthcare clothing, LTE-based Tele-medicine, and emotional interaction based on robots. The relevance to digital twin and replica is missing from those studies.

Others [17], [18], and [19] used wearable devices and AI to collect and analyze human data in order to simulate human processes such as emotion recognition, recognition of user intent, user behavioral motivation understanding. Moreover, develop creative games, and help artists shape their creativity. Finally, perform sustainable health monitoring and give instructions to help users improve their health.

A cyber-physical health system for patient-centric applications and services has been introduced in [20]. The system is built on cloud and analyze big data to provide a more convenient healthcare service. The architecture encompass three layers: 1) a unified data collection layer which used to integrate public medical resources and personal health devices, 2) a multisource heterogeneous data management which is helpful for distributed storage and parallel computing, 3) a data-oriented service layer as a unified interface for the users. Several health related studies have been also proposed in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3051158, IEEE Internet of Things Journal

the literature to whether handle health record, collaborative patient-centric healthcare services delivery, edge computing healthcare service for resource management [21] [22] [23].

Others have integrated DT in fields other than healthcare, such as multimedia systems and manufacturing [24] [25]. They presented a digital twin framework replicating the processes of a real production line for Industry 4.0. DT has been placed within the interlinks of the physical system to offer extended services such as schedules order and specifies product configuration.

As seen from all of these studies, DT research began to grow, current works mainly propose theoretical frameworks and models. To date, various works mention the issue of the DT in smart healthcare systems, however applicable solutions and validation are not provided yet. In Table I, the most relevant work has been summarized.

III. Framework Architecture

The paper goal is to *propose* and *implement* a DT framework for intelligent context-aware healthcare systems to enhance patients' healthcare and improve healthcare processes. The proposed DT framework uses IoT devices, data analytics, and AI through three phases to create a patient virtual replica, enable healthcare professionals to collaborate effectively, and open doors to the cooperation of patients with similar cases.

Architecture.pdf

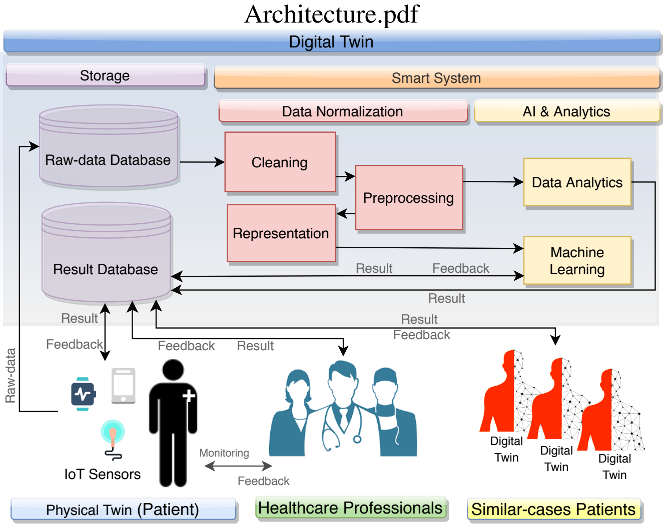


Fig. 1: System Architecture

The IoT wearable sensors capture and transfer real-time data for body metrics from the patient, as shown in Figure 1. The DT replica resides at the upper part of the system architecture with being complex in terms of handling the data. The transferred data will be cleansed, preprocessed, and transformed to be used for data analytics and building different types of machine learning models. This framework gives the ability to monitor health status continuously and early detect abnormalities. Additionally, it will enable healthcare professionals to prescribe the most right treatments, test them in a safe environment, and track responses. Moreover, design better lifestyles, and provide more effective communications with patients, thus optimizing many areas of the healthcare process.

A. Framework Phases:

As illustrated in the framework overview shown in Figure 2, the proposed framework of this paper is three folds:

1) Processing and Prediction phase: is processing and prediction. This phase is patient-centered begins with capturing the patient data using IoT wearable sensors. These sensors transfer real-time data of the human body metrics that are important to monitor the health status and help in anomalies detection. The transferred data will be stored temporarily in a cloud database responsible for raw-data storage. This data will be used by a machine learning system in the training and prediction process later. Through data analytics and machine learning capabilities, the system will build classifiers and predictive models that detect anomalies using the raw-data after cleansing, preprocessing, and representation. The machine learning model results will be stored in another scalable, secure, and immutable cloud database called Result Database. Result Database is accessible by the patient and the other framework phases' components for continuous feedback, correction, and model optimization.

2) Monitoring and Correction phase: requires intervention from the healthcare professionals from the patient domain. Healthcare professionals who provide treatments and advice based on formal training and experience will use the results of the predictive models from the Result Database. This would be along with clinical diagnosis and monitoring of the patients health status to improve his healthcare. By continuously feeding the predictive models with real-time data, that would help in detecting the body metrics anomalies, proactively monitor and identify health issues before they occur, prescribe the right treatments, and help the healthcare professionals to design a better lifestyle for the patient. The professionals can correct, verify the results, and give informative feedback besides the read permissions of the Result Database, thus optimizing the model.

3) Comparison phase: is cooperation of patients with similar cases takes place in the third phase to utilize real-life scenarios, enhance patients' DT, and thereby enhance the whole framework. By obtaining data and results of digital twins for similar cases patients, the model will be able to compare the current patient results with other patients. This process will expand the predictive models' domain by having real-life scenarios with more reliable results to improve models' accuracy. It also gives healthcare professionals the ability to make more advanced and accurate decisions. Decisions not only based on real-time data monitoring, but also rely on the history of the past, present, and prediction of the future for other patients, so that they can simulate, modify, or avoid other patients' experiences.

IV. System Implementation

In order to apply the proposed framework proposed above in a real-life scenario, in this section, a use case is presented for a patient DT that monitors real-time health status and detects

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3051158, IEEE Internet of Things Journal

TABLE I: Related Literature Comparison

Paper	Domain	Technology used	Use case scenario	real-time data	Challenges
[9]	DT Framework for Healthcare	DT, IoT	-	-	-Real-time Data Aggregation. -Integration and Interoperability. -Fidelity.
[10]	Cloud-based DT system for elderly healthcare	Digital Twin, IoT, Cloud-Computing	DT-driven health monitoring for the elderly care.	Yes	-Data fusion. -Constructing and managing the DTH models. -Model accuracy and evaluation. -Collaboration and interaction issues.
[11]	Framework for a hospital DT	DT, DES, IoT, DSS, Simulation	Implementation of the framework in four scenarios using FlexSim HC.	Yes	-Problems in real-time data-link between virtual and real-life space -Development of a virtual space that depict real-life environment
[12]	SNN for the epileptic Neuromorphic seizure detection	SNN, Digital Twin, Neuromorphic chips	Surrogate gradient-based learning algorithm.	Yes	-Securing sensitive medical data. -data safety. -Patient security. -The integration of the model on the platform.
[13]	DT for personalized healthcare	DT, GAN, AI	-	-	-
[14]	Anonymization of patient information in the health sector	GAN, CNN	Implementation of Several GAN models for fake static and fake dynamic data generation.	Yes	-Improve GAN architecture -Losses in the databases -Data organization especially images -Discriminator and generator training time.
Our Proposal	DT framework for intelligent context-aware healthcare system	DT, IoT, ML, DL	Implementation of several ECG heart rhythms classifiers	Yes	-Real-time data link -Integrate the model with the proposed framework

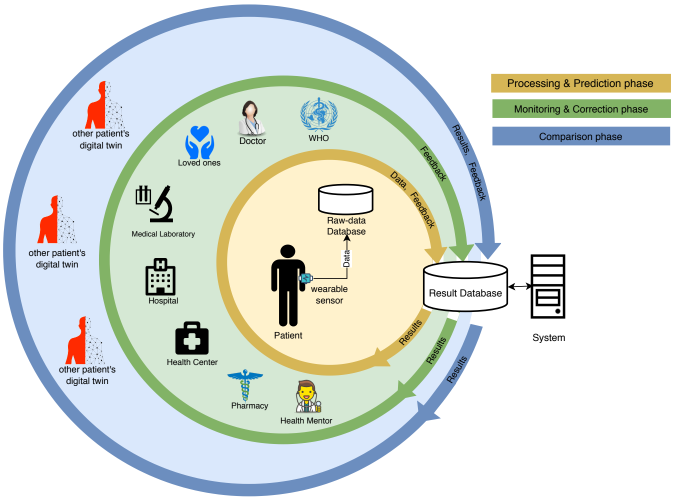


Fig. 2: Framework Overview

body metrics anomalies. The implemented use case boils down to build a machine learning model that diagnoses heart diseases and detects heart problems over predicting normal and abnormal heart rhythm. By capturing the Electrocardiogram (ECG) signals through wearable sensors or smartwatches, then convert them into a digital format. The collected data will be ready to fit into a machine learning algorithm after the preprocessing and filtering phase, as a result, a predictive model for normal and abnormal rhythms that describe a particular heart condition will be ready and in place. Accordingly, five models were built and trained using real-time ECG rhythms through various machine learning algorithms to test performance on the dataset and obtain the best accuracy. The applied algorithms were Convolutional Neural Network (CNN), Multi-layer Perceptron (MLP), Logistic Regression (LR), Long-Short-Term Memory Network (LSTM), and Support Vector Classification (SVC).

A. Dataset

The dataset used in this paper is based MIT-BIH Arrhythmia Database [26]. It contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between the years of 1975 and 1979.

Upper and Lower Signals were obtained by placing the electrodes on the chest, then the analog outputs of the playback unit were digitized at 360 Hz per signal relative to real-time using ADC (Analog-to-Digital Converter) hardware constructed at the MIT Biomedical Engineering Center and at the BIH Biomedical Engineering Laboratory. The dataset contains five classes as listed below:

- N: Normal beat
- S: Supraventricular premature beat
- V: Premature ventricular contraction
- F: Fusion of ventricular and normal beat
- Q: Unclassifiable beat

B. System Workflow

Figure 3 describes the followed workflow in the system construction process. First of all, the obtained dataset was stored in a database for easy retrieval and processing. Then, the stored data had some preprocessing operations such as cleansing, re-sampling, reshaping, and padding. After that, the data was split into two chunks; training data and testing data to be used in the Model-Building phase. The Model Building phase consists of two stages; Model Training and Model Evaluation. The Model Training phase used the training data to train the model and The Evaluation phase used the testing data to assess the model performance. This phase was repeated five times with different algorithms to find the optimal model. Finally, after saving the optimal model, the system is ready to receive new data samples for the prediction process.

V. Results and Discussion

As mentioned in section IV, five different models were built to obtain the best accuracy of the dataset, therefore,

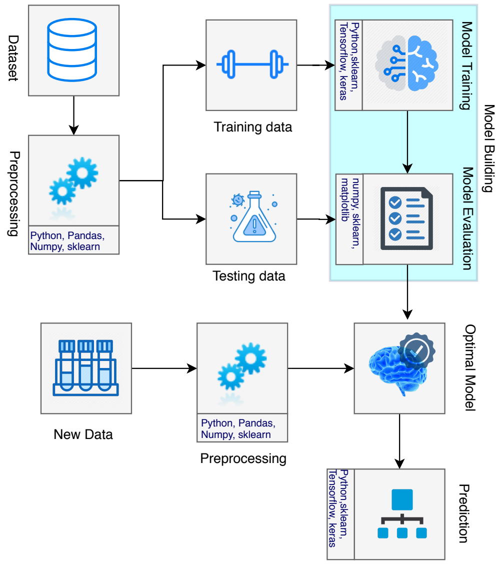


Fig. 3: System Workflow

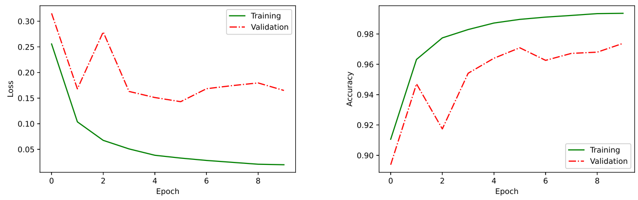
this section shows the parameters, performance, collected results and evaluation for each applied model. The experiment was carried out using Python, Sklearn library, Tensorflow and Keras. Also, other libraries were used to help in data prepossessing and results evaluation such as Pandas, Numpy, and Matplotlib. Below, the main parameters of the each used algorithm and model structure are described in detail.

First, in order to experience the potential of Neural-Network-based algorithms, the LSTM Sequential model was constructed using Long-Short-Term Memory Network and trained with a 0.01 learning rate over 10 epochs. The optimal model saved at epoch 5 with a minimum achieved validation loss of 0.1430, a 0.9709 validation accuracy, a 0.0329 training loss, and 0.9896 training accuracy.

Another Neural-Network-based model was applied, the CNN model. This model was constructed using the CNN and trained with a 0.01 learning rate over 10 epochs. The optimal model saved at epoch 4 with a minimum achieved validation loss of 0.1391, a 0.9667 validation accuracy, a 0.0331 training loss, and 0.9896 training accuracy.

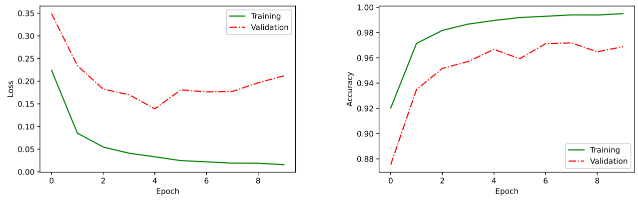
Figures 4 and 5 illustrate the performance of Loss and Accuracy for the LSTM sequential and CNN models over training epochs, respectively. Although the accuracy increases over the epochs, the loss also increases. This indicates that the models are less certain about their predictions. Accordingly, the best LSTM model was saved at epoch 5 and the best CNN model was saved at epoch 4 with the minimum achieved validation loss. The results show that the LSTM validation accuracy is higher than CNN validation accuracy but CNN validation loss is lower than LSTM validation loss. The difference between them is almost fractions and this makes their performance very close to each other.

The MLP model was constructed using the Multi-layer Perceptron algorithm and achieved 0.956 testing accuracy over 800 iterations. Also, two experiments were tested on traditional machine learning algorithms. The SVC model was constructed using the Support Vector Classification algorithm and achieved 0.756 testing accuracy on the linear kernel. The last applied model was the Logistic Regression (LR). This model was constructed using the Logistic Regression algorithm and achieved 0.676 testing accuracy over 900 iterations with saga solver.



(a) LSTM model Loss performance (b) LSTM model Accuracy performance

Fig. 4: LSTM model Loss and Accuracy performance



(a) CNN model Loss performance (b) CNN model Accuracy performance

Fig. 5: CNN model Loss and Accuracy performance

A. Evaluation

This section describes in detail the evaluation metrics used to compare the applied models performance and choose the optimal one.

1) Accuracy: as shown in equation 1 gives the percentage of correctly predicted samples which are the True Positives and True Negatives out of all data samples (True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN)). It measures how often the algorithm correctly classifies a data sample.

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

Figure 6 compares the accuracy achieved across the five models, and it is clear that the LSTM model has achieved the highest accuracy. It also shows that the Neural-Network-based algorithms perform better than other traditional algorithms at accuracy score.

2) Confusion Matrix: For the purpose that the testing dataset is imbalanced, the accuracy rate may sometimes be misleading. In order to ensure that the models perform well, tables II, III, IV, V and VI show the confusion matrix that describes the performance of each model. By Obtaining the Confusion Matrix, we were able to find the TP, FP, TN, FN values.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3051158, IEEE Internet of Things Journal

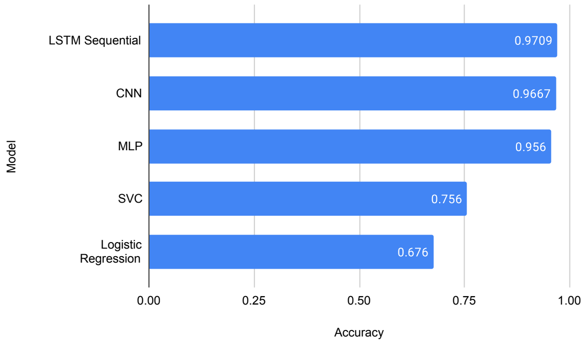


Fig. 6: Models Accuracy

a) True Positives:
A **TP** value is considered when the model correctly predicts the positive class. Figure 7(a) compares the True Positives values for each model across the five classes. Neural-Network-based models almost have similar TP values across classes and higher than SVC and Logistic Regression models values.

b) True Negatives:
A **TN** value is considered when the model correctly predicts the negative class. Figure 7(b) compares the True Negatives values for each model across the five classes. Neural-Network-based models almost have similar TN values across classes and higher than SVC and Logistic Regression models values.

c) False Positives (type I error):
A **FP** value is considered when the model incorrectly predicts the positive class. Figure 7(c) compares the False Positives values for each model across the five classes. The SVC and Logistic Regression models have a higher type I error for some classes than the rest of the models, especially in the S, V, and F classes which were predicted as S, V and F but they weren't actually from these classes.

d) False Negatives (type II error):
A **FN** value is considered when the model incorrectly predicts the negative class. Figure 7(d) compares the False Positives values for each model across the five classes. The SVC and Logistic Regression models have a higher type II error for some classes than the rest of the models, especially in the N class that consider a patient to have normal ECG rhythms, which means the model predicts a bad condition where the patient has normal beats. This explains the high False Positives for S, V and F classes in SVC and Logistic Regression models.

3) Classification Report: after obtaining the Confusion Matrix values, we were able to calculate the main classification metrics: Precision, Recall, and F1-score using Classification Report. Tables VII, VIII, IX, X and XI show the classification report to describe the complete performance for each model.

Precision as shown in equation 2 is the accuracy of positive predictions. The SVC and Logistic Regression models show a low precision scores for S, V and F classes which indicates a large number of False Positives.

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

Recall as shown in equation 3 is the fraction of positive samples that were correctly identified from the actual positives. The SVC and Logistic Regression show a lower recall score, indicating many False Negative values, compared to Neural-Network-based models, but all scores are considered good scores.

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

F1-Score as shown in equation 4 is the harmonic mean of Precision and Recall. The SVC and Logistic Regression models show a low F1-score for S, V and F classes. Also, Neural-Network-based models show a lower F1-Score for S and F compared to other classes, which means the models don't perform very well in predicting these classes.

$$F1 - score = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{4}$$

4) Macro Average and Weighted Average: **Macro Average** as shown in equation 5 computes the metric for each label, and returns the average without considering the proportion for each label in the dataset (averaging the unweighted mean per label).

$$\textbf{Macro Avg.} = \frac{\sum m}{\text{N}} \tag{5}$$

where m= metric score for each class and N= number of classes.

Weighted Average as shown in equation 6 computes the metric for each label, and returns the average considering the proportion for each label in the dataset (averaging the support-weighted mean per label).

$$\textbf{Weighted Avg.} = \sum (S \times m) \tag{6}$$

where s= percentage of samples for each class from total samples.

Figures 8, 9 and 10 illustrate a comparison of Weighted Average and Macro average between models for Precision, Recall and F1-Score .It is clear that the LSTM model always get the highest score across all metrics. Also, the Neural-Network-based models perform better across the whole dataset than traditional machine algorithms models.

5) Micro average: **Micro Average** is an averaging technique which calculates the total TP, FN, and FP when the testing dataset is imbalanced. Micro-Avg aggregates the contributions of all classes to compute the average metric. It is the same for all metrics in the Classification Report and gives the same results as Accuracy. Therefore, the same findings were considered that Neural-Network-based models performed better in classifying the dataset than other models. Figure 11 compares the Micro Average achieved across the five models.

6) ROC and AUC: **Receiver Operating Characteristic Curve** is a probability model to compare the the TP rate against FP rate at different thresholds using graphical plot (sensitivity against specificity). **Area Under the Curve** is measure the ability of a classifier to distinguish between different classes. The higher AUC close to 1 means the

classifier performs well. An AUC close to 0.50 means the classifier is guessing and has no separation capacity. The lower AUC close to 0 means that the model is 100% wrong and it predicts the opposite class.
Figure 12 illustrates the Receiver Operating Characteristic Curve for each model using the One-vs-Rest method and the legends show the area under the curve for micro average ROC. All models had an AUC score above 80 for all classes which means that all models can distinguish between classes well.

TABLE II: LSTM Model Confusion Matrix

Classes	N	S	V	F	Q
N	17,682	291	71	59	15
S	73	467	8	7	1
V	37	10	1,375	24	2
F	11	1	7	143	0
Q	14	1	5	0	1,588

TABLE III: CNN model Confusion Matrix

Classes	N	S	V	F	Q
N	17,584	327	97	69	41
S	64	476	9	5	2
V	29	10	1,370	31	8
F	10	2	8	142	0
Q	11	3	3	0	1,591

TABLE IV: MLP model Confusion Matrix

Classes	N	S	V	F	Q
N	17,430	393	165	62	68
S	93	443	14	3	3
V	50	13	1,351	26	8
F	16	4	7	134	1
Q	17	7	8	0	1,576

TABLE V: SVC model Confusion Matrix

Classes	N	S	V	F	Q
N	13,456	869	2,241	1,207	345
S	142	360	28	24	2
V	150	32	1,147	99	20
F	11	0	8	143	0
Q	69	6	57	14	1,462

TABLE VI: LR model Confusion Matrix

Classes	N	S	V	F	Q
N	11,790	2,085	2,369	1,415	459
S	125	370	31	20	10
V	158	53	1,038	153	46
F	11	0	9	142	0
Q	49	5	73	13	1,468

As a wrap-up, it was proved that the Neural-Network-Based models perform better than traditional machine learning algorithms in terms of evaluation metrics. For the Accuracy, the LSTM sequential model achieved the highest accuracy score with 0.97. Also, the Deep NN models achieved higher scores than SVC and Logistics Regression models.

The Confusion Matrices showed that the traditional algorithms (SVC and Logistics Regression) miss-classified some classes with higher False Positive and False Negative values than Neural-Network-Based algorithms.

TABLE VII: LSTM Model Classification Report

Metrics / Classes	precision	recall	f1-score
N	0.99	0.98	0.98
S	0.61	0.84	0.7
V	0.94	0.95	0.94
F	0.61	0.88	0.72
Q	0.99	0.99	0.99
macro avg	0.83	0.93	0.87
weighted avg	0.98	0.97	0.97

TABLE VIII: CNN Model Classification Report

Metrics / Classes	precision	recall	f1-score
N	0.99	0.97	0.98
S	0.58	0.86	0.69
V	0.92	0.95	0.93
F	0.57	0.88	0.69
Q	0.97	0.99	0.98
macro avg	0.81	0.93	0.86
weighted avg	0.97	0.97	0.97

TABLE IX: MLP Model Classification Report

Metrics / Classes	precision	recall	f1-score
N	0.99	0.96	0.98
S	0.52	0.8	0.63
V	0.87	0.93	0.9
F	0.6	0.83	0.69
Q	0.95	0.98	0.97
macro avg	0.79	0.9	0.83
weighted avg	0.96	0.96	0.96

TABLE X: SVC Model Classification Report

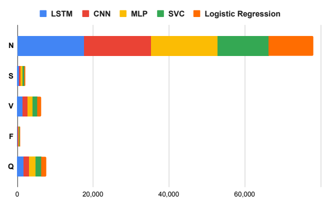
Metrics / Classes	precision	recall	f1-score
N	0.97	0.74	0.84
S	0.28	0.65	0.39
V	0.33	0.79	0.47
F	0.1	0.88	0.17
Q	0.8	0.91	0.85
macro avg	0.5	0.79	0.55
weighted avg	0.89	0.76	0.8

TABLE XI: LR Model Classification Report

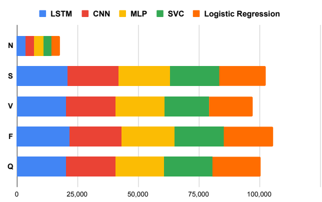
Metrics / Classes	precision	recall	f1-score
N	0.97	0.65	0.78
S	0.15	0.67	0.24
V	0.29	0.72	0.42
F	0.08	0.88	0.15
Q	0.74	0.91	0.82
macro avg	0.45	0.76	0.48
weighted avg	0.88	0.68	0.74

Precision, Recall, and F1-score metrics showed that Neural-Network-Based algorithms achieved higher scores than other models, taking into consideration the Macro and

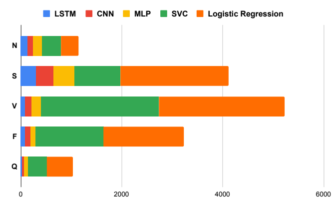
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2021.3051158, IEEE Internet of Things Journal



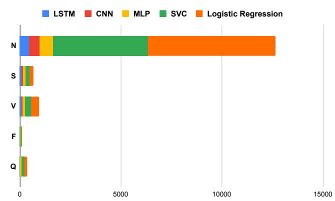
(a) TP



(b) TN



(c) FP (Type I error)



(d) FN (Type II error)

Fig. 7: Values of : TR, TN, FP, and FN

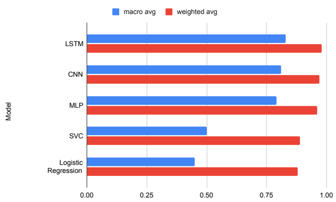


Fig. 8: **Precision** Score

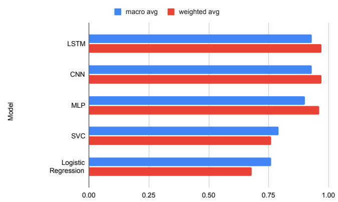


Fig. 9: **Recall** Score

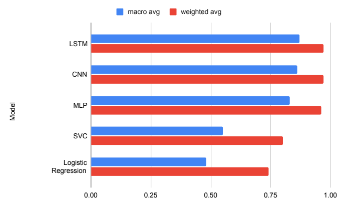


Fig. 10: **F1-Score**

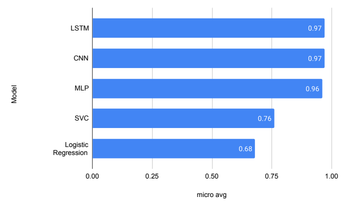
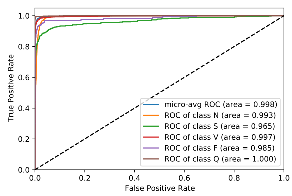
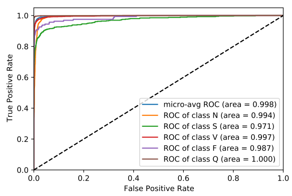


Fig. 11: Micro Average

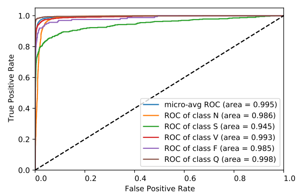
Weighted average results for each metric. Furthermore, they showed that the LSTM sequential model achieved 0.83, 0.93, and 0.87 Macro Average for Precision, Recall, and F1-score respectively and 0.98, 0.97, and 0.97 Weighted Average for Precision, Recall, and F1-score respectively, which were the highest scores across all models. Finally, the Area Under the Curve (ROC) showed that all models had a high AUC score above 80 for all classes which means all models can distinguish between classes.



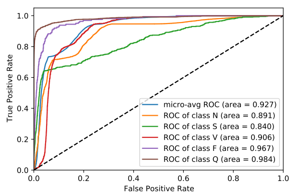
(a) LSTM Model ROC Curve



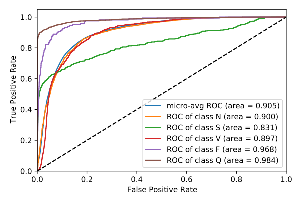
(b) CNN Model ROC Curve



(c) MLP Model ROC Curve



(d) SVC Model ROC Curve



(e) LR Model ROC Curve

Fig. 12: Models ROC Curve

VI. Challenges and Issues

A. *Trust*

The concept of the digital twin of having a virtual replica for a physical asset will always have a gap since it relies on devices to transfer the data, while these devices may be crashed or disconnected for any reason. Also, digital twins require contribution from field professionals. These professionals must be qualified and ethical to give accurate feedback, edit and preserve data. Moreover, there is a big challenge for using AI in real-world problems. How accurate are the Machine Learning models, and to which level can we trust their predictions?. Consequently, building trust at every level and for each component will contribute to building confidence around the concept of digital twin in general. This requires setting standards, raising awareness, and improving technologies, and all of this requires effort and time.

B. *Security and Privacy*

Protecting the digital twin systems from unauthorized access, abuse, modification or disclosure will be a challenge as in any other information system. As digital twin systems process large volumes of sensitive and personal data, this will make it a target for threat actors and cyber attacks. In addition, the use of the IoT devices and sensors will add more complexity in terms of implementing proper security as the traditional security controls mostly will not fit with them. Processing personal users data could raise regulatory risks. Complying with privacy regulations such as GDPR in Europe or regulations at relevant national protection laws could be a mandatory and adding more challenges when designing digital twin systems.

C. Standardization

Lack of standards is another critical challenge. This factor affects security, privacy, interactions, roles, contribution protocols, data transmission, and synchronization between the virtual and physical world. Setting global standards would help to spread the trend of digital twins more rapidly and make it a reality faster.

D. Diversity and Multi-sourcing

Another problem facing digital twins is the data diversity and their multiple sources. This occurs due to the different sources through which data is captured, also the diversity of data types. Which causes problems in processing and building machine learning models as this data is heterogeneous.

VII. Conclusion and Future Work

This paper has proposed a digital twin framework for intelligent context-aware healthcare systems. It also presents a use case for a patient's digital twin that monitors real-time health status and detects body metrics anomalies by building an ECG heart rhythms classifier model that diagnoses heart disease and detects heart problems. The conducted use case tested five different algorithms for optimum accuracy and performance. As a preliminary result, the proposed framework integrates Digital Twin with the healthcare field to improve healthcare processes. This would help in bringing patients and healthcare professionals together in an intelligent, comprehensive, and scalable Health-Ecosystem with the aim of promoting health, increasing life expectancy, reducing healthcare costs, and solving many healthcare issues and challenges. As a second result, implementing an ECG classifier that detects heart conditions gives the inspiration for applying Machine Learning and AI with human body metrics for continuous monitoring and abnormalities detection. Therefore, this would help in head off problems before they occur and improve the quality of life. Finally, Neural-Network-based algorithms deal better with ECG data than traditional ML algorithms in terms of learning evaluation metrics such as accuracy, F1-score, precision, recall, and AUC.

Our future work will focus on implementing the full system with real-time data links, extending use cases to include other body metrics, and integrating other systems into the framework.

References

[1] A. A. Hussain, O. Bouachir, F. Al-Turjman, and M. Aloqaily, "Ai techniques for covid-19," *IEEE Access*, vol. 8, pp. 128 776–128 795, 2020.

[2] B. D. Deebak, F. Al-Turjman, M. Aloqaily, and O. Alfandi, "An authentic-based privacy preservation protocol for smart e-healthcare systems in iot," *IEEE Access*, vol. 7, pp. 135 632–135 649, 2019.

[3] S. Tyagi, A. Agarwal, and P. Maheshwari, "A conceptual framework for iot-based healthcare system using cloud computing," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 2016, pp. 503–507.

[4] X. Ge, R. Zhou, and Q. Li, "5g nfv-based tactile internet for mission-critical iot services," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6150–6163, 2020.

[5] X. Zhou, W. Liang, K. I. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429–6438, 2020.

[6] "Ieee computer society's top 12 technology trends for 2020," https://www.computer.org/press-room/2019-news/ieee-computer-societys-top-12-technology-trends-for-2020.

[7] I. Al Ridhawi, S. Otoum, M. Aloqaily, and A. Boukerche, "Generalizing ai: Challenges and opportunities for plug and play ai solutions," *IEEE Network*, 2020.

[8] "What is health promotion?" https://www.who.int/news-room/q-a-detail/what-is-health-promotion.

[9] L. F. Rivera, M. Jiménez, P. Angara, N. M. Villegas, G. Tamura, and H. A. Müller, "Towards continuous monitoring in personalized healthcare through digital twins," in *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*, 2019, pp. 329–335.

[10] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, and M. J. Deen, "A novel cloud-based framework for the elderly healthcare services using digital twin," *IEEE Access*, vol. 7, pp. 49 088–49 101, 2019.

[11] A. Karakra, F. Fontanili, E. Lamine, J. Lamothe, and A. Taweel, "Pervasive computing integrated discrete event simulation for a hospital digital twin," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2018, pp. 1–6.

[12] P. S. Zarrin, R. Zimmer, C. Wenger, and T. Masquelier, "Epileptic seizure detection using a neuromorphic-compatible deep spiking neural network," in *International Work-Conference on Bioinformatics and Biomedical Engineering*. Springer, 2020, pp. 389–394.

[13] C. Angulo, L. Gonzalez-Abril, C. Raya, and J. A. Ortega, "A proposal to evolving towards digital twins in healthcare," in *International Work-Conference on Bioinformatics and Biomedical Engineering*. Springer, 2020, pp. 418–426.

[14] E. Piacentino and C. Angulo, "Generating fake data using gans for anonymizing healthcare data," in *International Work-Conference on Bioinformatics and Biomedical Engineering*. Springer, 2020, pp. 406–417.

[15] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C.-H. Youn, "Wearable 2.0: Enabling human-cloud integration in next generation healthcare systems," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 54–61, 2017.

[16] Y. Ma, Y. Wang, J. Yang, Y. Miao, and W. Li, "Big health application system based on health internet of things and big data," *IEEE Access*, vol. 5, pp. 7885–7897, 2016.

[17] M. Chen, J. Zhou, G. Tao, J. Yang, and L. Hu, "Wearable affective robot," *IEEE Access*, vol. 6, pp. 64 766–64 776, 2018.

[18] M. Chen, Y. Jiang, Y. Cao, and A. Y. Zomaya, "Creativebioman: Brain and body wearable computing based creative gaming system," *arXiv preprint arXiv:1906.01801*, 2019.

[19] M. Chen, Y. Jiang, N. Guizani, J. Zhou, G. Tao, J. Yin, and K. Hwang, "Living with i-fabric: Smart living powered by intelligent fabric and deep analytics," *IEEE Network*, 2020.

[20] 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 Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.

[21] P. Van Gorp and M. Comuzzi, "Lifelong personal health data and application software via virtual machines in the cloud," *IEEE journal of biomedical and health informatics*, vol. 18, no. 1, pp. 36–45, 2013.

[22] L. D. Serbanati, F. L. Ricci, G. Mercurio, and A. Vasilateanu, "Steps towards a digital health ecosystem," *Journal of biomedical informatics*, vol. 44, no. 4, pp. 621–636, 2011.

[23] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "An edge computing based smart healthcare framework for resource management," *Sensors*, vol. 18, no. 12, p. 4307, 2018.

[24] J. Bao, D. Guo, J. Li, and J. Zhang, "The modelling and operations for the digital twin in the context of manufacturing," *Enterprise Information Systems*, vol. 13, no. 4, pp. 534–556, 2019.

[25] M. Raza, P. M. Kumar, D. V. Hung, W. Davis, H. Nguyen, and R. Trestian, "A digital twin framework for industry 4.0 enabling next-gen manufacturing," in *2020 9th International Conference on Industrial Technology and Management (ICITM)*. IEEE, 2020, pp. 73–77.

[26] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.