Semantic Constraints Specification and Schematron-Based Validation for Internet of Medical Things' Data

ANA KOREN[1], **MARKO JURČEVIĆ**[1], **(Member, IEEE), AND RAMJEE PRASAD**[2], **(Life Fellow, IEEE)**

[1]Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia
[2]CTIF Global Capsule, Department of Business Development and Technology, Aarhus University, 8200 Aarhus, Denmark

Corresponding author: Ana Koren (ana.koren@fer.hr)

ABSTRACT Fitness and activity trackers are hugely popular wearable devices that monitor various health-related metrics, such as step count, heartbeat rate, or even oxygen saturation. Utilizing personal health information obtained by users' personal trackers provides promising results in the fields of telemedicine and personal well-being. However, we face challenges such as data quality, privacy and compliance with standards and regulations. This paper addresses such challenges, with the focus on the last one. Semantic constraints for healthcare datatypes are defined to ensure compliance with standards, making the information medically valid and relevant. A process of semantic verification and Schematron-based validation is proposed. The validation process suggested in this paper will enable the data to be transferred and incorporated into a formal Electronic Health Record. The process is then verified using datasets containing various health-related data types. The aim is to integrate personal health data into Electronic Health Record, which forms a part of Central Health Information System. This would provide personalized medical services to patients and help physicians to make more informed decisions.

INDEX TERMS Central health information system, eHealth Services, electronic health record, internet of medical things, healthcare standards.

I. INTRODUCTION

Fitness or activity trackers are electronic wearable devices that monitor health-related metrics, e.g., steps/distance walked/ran, heartbeat rate, oxygen saturation, calories consumption, etc. Trackers, usually in a form of a wristband bracelet, can transmit data directly to a smartphone/PC application. Fortune Business Insights reports that the global market size for fitness trackers was USD 30 billion in 2019 and is projected to reach USD 92 billion by 2027 [1]. Wristband fitness trackers and smart-watches are products that lead the growth of the market; however, the wearables market also includes smart clothing and ear-worn devices. A growing number of products provide advantageous health tracking technology, which include heart rate or body temperature measurements and even blood oxygen tracking (SpO2). The integration of these functions is gaining a lot of traction with consumers, especially in the era of Coronavirus disease

2019 (COVID-19). Utilizing patient-generated health data could bring benefits to fields of medicine [2]–[4], telemedicine [5], [6], and personal well-being [6]–[8]. The information collected could mean an improved, more tailored healthcare approach as it could offer physicians constant comprehensive insight into the health of the patient. Thus, implementing the integration of personal health data, obtained through various fitness trackers' sensors into the formal Electronic Health Record (EHR) shall provide personalized medical services in accordance with standards and could be a way to offer a more personalized and consistent care, helping medical workers to make better and more informed decisions.

To achieve this aim, several requirements must be met. Main challenges are ensuring data quality, maintaining privacy and compliance with applicable standards and regulations. Research presented in this paper focuses on the latter, as data quality was addressed in previously published work [9]. On the other hand, privacy concerns are planned to be the focus of future work. Due to the complex nature of

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Lin.

healthcare data, issues related to joining disparate elements and connecting different systems arise. Achieving interoperability implies defining data structure and content for the core data types, to which then implementations would need to comply. Health data interoperability imposes standards and protocols that are used by most implementations of Central Health Information Systems worldwide. If patient-generated health data were compliant with the same standards, Central Information Health System platform applications would be able to perform analytics on personal health data and create clinically valid EHR documents.

Thus, the aim of this research is to define a semantic constraints specification for health data for common data types and to model the process for Schematron-based validation. In the following chapters, possible solutions for ensuring compliance are discussed, and reasons for the selection of Schematron-based validation are provided. Furthermore, an overview of datasets used for model verification is given, including the dataset collected for the purpose of this research. Results depict the proposed validation process in detail and its verification is provided based on a case study using the before mentioned datasets. Finally, Discussion provides conclusions and proposes future work.

II. RELATED RESEARCH

Systematic review [10] summarized the validity and reliability of some of the most popular fitness trackers and their ability to accurately estimate health-related metrics, such as steps, heartbeat rate, and sleep. Results indicated high reliability among all devices. In [11], a comparative analysis was performed between the data of the fitness tracker with photoplethysmography (measures heartbeat rate) and a certified clinical device, an electrocardiogram (ECG), which assesses heartbeat rate through the electrical activity of the heart. The analysis suggests that fitness trackers, regarding the evaluation of heartbeat rate, provide valid information for use in clinical practice. After revising results of 67 studies, [12] concludes that wearable devices meet acceptable accuracy for step count; however, a tendency to underestimate steps in controlled (test) environments and overestimate steps in free-living environments exists. Furthermore, [13] highlights the absence of a standard test protocol for the validation process. As [14] and [15] report, compared to an ambulatory electrocardiogram (ECG), two commercially available models shown to have high accuracy, giving heartbeat accuracy of less than ±10% for the 24-hour period and across all activities. Additionally, both devices were less accurate following erratic movements and increased heart rate. Other models gave similar results [16]. State-of-the-art survey on data quality has been published in [17]. Finally, case study on Quality of Experience (QoE) presents a new approach to smart wearables evaluation, suggesting that accuracy of heart rate monitoring and step count are the two most important parameters [18].

However, to be able to use personal health data obtained through sensors in wearable trackers as relevant information in a formal EHR, the data must be accurate and free of faults and errors, as this could lead to misleading conclusions and incorrect diagnoses. Fitness trackers are relatively inexpensive and easy to use; however, doubts as to their reliability and accuracy exist. The errors in sensor readings may be caused by different external sources, such as the environmental effects that affect the measurement process itself, communication channel and the transmission of signal, or due to fault of hardware as low-cost sensors have limited resources and reliability. This can result in measured signal noise, missing values, or inaccurately measured values. These are all factors that need to be addressed. Furthermore, the lack of information and validation of the algorithms used by the devices make them inadequate to be used in the medical field without additional data cleaning (as this data must be reliable and accurate) and data validation, in the sense of compliance to the standards set in the medical field, such as Health Level 7 (HL7). Regarding data quality, [19] claims, artificial neural networks can significantly improve fault detection and data analyses for wearable devices. Several studies [20]–[23] employ some form of data cleaning process for data collected from Wireless Sensor Networks (WSN); i.e., coappearance-based analysis for incorrect records, decision tree-based missing value imputation, and fault and anomaly classification. Lastly, article [9] compares various data-driven models cleaning eHealth sensor data with the goal of ensuring that the collected data is accurate, relevant and can be used in formal EHR. It also identifies multiple linear regression and neural network as best models for data imputation, which it further optimizes with result of 10-17% improvement in accuracy, depending on the person monitored (data was collected by monitoring ten volunteers, diverse in terms of sex, age, and fitness level).

The use of EHR is growing rapidly; i.e., EHR is replacing patients' paper charts [24], [25]. On the other hand, as far as transferring personal health records into a formal EHR is considered, no attempts have been implemented yet. However, [26] evaluated Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) system architecture in the use case of smart glasses as a source of personal health information by measuring the user's vital signs, i.e., temperature, heart rate, and respiration rate. One more example of this is [27], mHealth, which is "a prototype of implementing a HL7-compatible personal health record system" in a form of a mobile application which users use to access and directly communicate their medical information to healthcare service providers (also through HL7 FHIR). A similar effort is seen in [28], where Tangle is an application which serves as a "bridge", where patients can share their PHR data for physicians to access.

A Personal Health Record (PHR) contains person's health information in an electronic format and is managed by the said person. It is different from an Electronic Health Record (EHR), which is owned and stored by the healthcare provider. For the full integration of personal health data into EHR, the primary task is to overcome the discrepancies and

consolidate distinctive datasets into a unified, comprehensive collection of information and ensure data quality. This would allow highly beneficial insight and public eHealth services improvement, preventing fragmentation of health data. However, even though Tangle mentions integrating IoT sensor personal health data into EHR, no data cleaning is mentioned, and it remains unclear what is the format of said data, if they are transferred and validated, and in what way. WearableHUB [29] is a platform which collects and transforms wearable tracker's data and integrates it into a PHR. The data is said to be transformed "into a unified format according to a predefined standard", although it does not elaborate on this point further. Angel-Echo [30] is another PHR solution which helps monitor the health status of a patient by collecting data via a wristband device. OpenHealth [31] is an open-source platform for wearable health monitoring which uses machine learning algorithms to clean and transform data. Finally, mHealth4Afrika [32] mentions data validation (including sensor-collected data) in healthcare facilities in Ethiopia, Kenya, and Malawi via "formal validation sessions", which include individual or group interviews and observations, meaning it is not an automated process.

To integrate this data into EHR, data verification and validation is necessary. For this purpose, a schematron [33] can be used. Schematron is a structural schema validation language based on rules and expressed in Extensible Markup Language (XML). It is used commonly for asserting whether specified patterns are present or absent in XML trees and is capable of creating constraints in ways that other XML schema languages, such are XML Schema and Document Type Definition (DTD), cannot. This requires specific attributes, content control of certain elements by another element, or even specification of requirements between multiple XML files [34]. Schematron has been standardized by the ISO as "Information technology, Document Schema Definition Languages (DSDL), Part 3: Rule-based validation, Schematron (ISO/IEC 19757-3:2016)" [33]. Lastly, considering the said data is of a highly sensitive nature, security is a pivotal challenge. Ensuring that the patient's privacy remains undisputed is an important requisite. The transfer of personal health data to EHR data requires standardization, in this particular case, via the use of schematrons, so that it has the correct medical format and context. This information, in conjunction with the rest of EHR information (such as previous diagnoses, medication history, doctor visitations, and laboratory results), can aid medical staff in remotely screening a patient, diagnosing issues early, and improving health services.

Integrating the Healthcare Enterprise (IHE) is an initiative between healthcare professionals and industry with the goal of improving healthcare. It provides interoperability specifications, tools, and services for sharing and managing healthcare information. Its mission is to "engage clinicians, health authorities, industry, and users to develop, test, and implement standard-based solutions to vital health information needs" [35]. This is done by identifying requirements, defining standards, and providing technical guidelines and frameworks for developers to implement. IHE Technical Frameworks [36] define implementations of previously set up standards to facilitate viable and efficient systems integration, achieve warranted exchange of medical information and, thus, offer optimal patient care. Annually, following a period of public review, these are expanded and maintained on a regular basis by the IHE Technical Committees. Among others, Technical Frameworks (TF) described include those for *Anatomic pathology*, *Dental, Cardiology* and *IT infrastructure*. *IT infrastructure* describes integration profiles [37] as well as ITI transactions ITI-1 to ITI-28 [38], ITI-29 to ITI-64 [39], ITI-65 and greater [40], together with the metadata [41], [42]. Specifically, *Cardiology TF* entices the development of a range of implementation profiles, such as cardiac or intravascular imaging, resting ECG (REFW) or stress testing workflow (STRESS), many of which would benefit from the inclusion of aggregated personal health data (such as ECG information) into a formal EHR [43], [44].

Furthermore, *Quality, research and public health TF* under *Supplements for Trial Implementation* encourage initiative for the development of several implementation profiles, among them, *Aggregate Data Exchange* (ADX) [45]. ADX serves for "interoperable public health reporting of aggregate health data". Most common use cases of ADX are periodic (week, month, quarter, or annual) reports from a health facility to an administrative jurisdiction. A *Content Data Structure Creator* defines the structure of XML data that will be communicated between a *Content Creator* and *Content Consumer*, which assumes the creation of two normative message structure definition files:
- Data Structure Definition (DSD) file that is conformant to the normative schematron
- W3C XML Schema Definition (XSD) with an ISO Schematron schema, both of which must match the result generated by the normative XSLT transform from DSD to XSD and from DSD to schematron.

III. RESEARCH GOAL
The goal of the presented research is to model a data verification and validation process that would enable a standard-compliant integration of personal health data collected by wearables into EHR. A single unified EHR system within the European Union doesn't exist at this time. Rather, some countries have implemented national EHR systems in place; others have several, oftentimes overlapping, systems operating on a regional level, and there are ones having no EHR as of yet.

Examples of functional national EHR systems are Croatian and Finnish EHR systems, both of which cover healthcare facilities in public and private sectors, offering additional services to patients, such as prescription renewal or management of doctor appointments. In contrast, in Belgium, a country known for its linguistic diversity, distinct regions have separate EHR systems that do not interact with one another. Furthermore, e.g., Germany has coverage of only some of the federal states.

Thus, when integrating patient-generated data, achieving interoperability implies defining data structure and content for the core datatypes to which then implementations would need to comply. Also, health data interoperability imposes compliance to standards which are used by most implementations of EHR systems.

Data verification and validation process has to ensure the data has the adequate structure and content and that all the communication complies with existing standards.

IV. MATERIALS AND METHODS

The study was performed in several stages: 1) data collection; 2) defining the process of data parsing, data verification, and data validation; and 3) verification of the proposed process via a use-case study, using datasets containing various relevant datatypes.

A. DATA COLLECTION

To verify the validation process described in the following chapter, two datasets have been used. The PMData Dataset [46], provided from Simula Open Datasets, contains life logging and sports activity logging of sixteen persons over a five-month period using a commercial smartwatch wristband. Additionally, over the course of two months, OxyBeat [47] dataset was collected for the purpose of this research, containing heartbeat rate, as well as body temperature and oxygen saturation (SpO2). This was done to add several more datatypes, with focus on COVID-relevant datatypes and subsequently ensure a more robust use-case scenario. SpO2 specifies the percentage of oxygenated hemoglobin (hemoglobin containing oxygen) compared to the total amount of hemoglobin in the blood (oxygenated and non-oxygenated hemoglobin). Fitness tracker measures SpO2 levels using the relative reflection of red and infrared light from the blood and its variations compared to heart beats. Deoxygenated blood, which returns to the lungs via veins, is of a darker red color than the fully oxygenated blood in the arteries and arterioles. Blood oxygen saturation (SpO2) tends to fluctuate very little, even during exercise and sleep; oxygen levels in blood during the day are generally 95-100%. SpO2 during sleep is generally lower than daytime SpO2 since the total amount of air breathed drops during sleep. Typically, nighttime SpO2 values are >90%. This dataset was also collected using a commercially available smartwatch wristband. Commercial wearable activity trackers have close to 100 million active users. In the third quarter of 2020, Health Metrics Dashboard (Figure 1) was introduced to the devices, which tracks metrics like breathing rate, heart rate variability and SpO2 – all important metrics when it comes to illness detection [48].

Triaxial accelerometers in wearables capture spatial body motion. Motion data is analyzed with the use of proprietary algorithms. Identified are patterns of motion that serve to calculate health-related metrics, such as steps taken, time spent exercising or sleeping. Although devised as a consumer product for motivating individuals to exercise and promoting
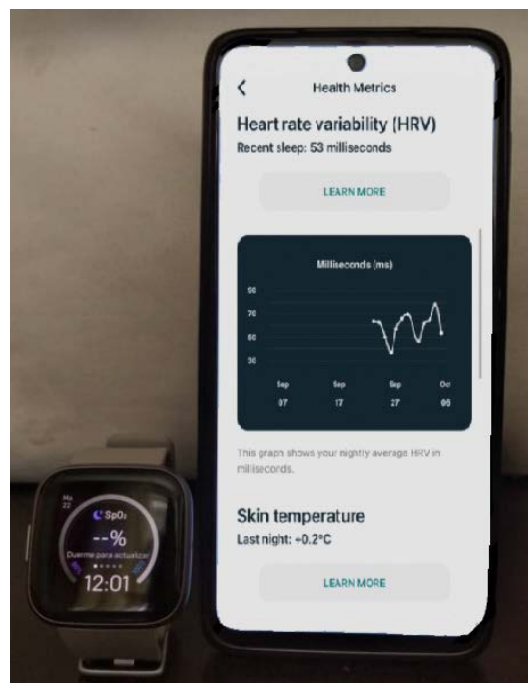

FIGURE 1. Health metrics dashboard.

FIGURE 2. An example of heartbeat rate data (JSON).

physical activity, wearables have been increasing in popularity as equipment in research as well as a support tool in doctor-patient interactions [49]. Since 2011 and up to 2020, a total of 260 clinical trials have been registered at Clinical-Trials.gov [50]. The most measured metric in the trials in question was the number of steps taken, followed by time spent in physical activity or sleep, heart-beat rate, and energy expenditure. Smart wearables, and especially devices worn on the wrist, have shown they are dependable, durable, and acceptable [14], [51].

Health-related metrics considered in this research are heartbeat rate, oxygen saturation (SpO2), and body temperature. One of the vital signs all personal trackers measure, is the heartbeat rate. Furthermore, the module for sensor data cleaning of this datatype has already been developed and in-depth described in previously published research [16]. Thus, this will also be used as the first example for the model's syntax and semantic validation. To be in line with the

HL7 standards, HL7's Structure Definition of HeartRate [52] is to be followed. The structure is derived from "observation-vitalsigns" [53]. Heartbeat rate data is exported from the tracker in the form of a JSON file, as shown below.

B. PROCESS FOR PARSING, VERIFYING, AND VALIDATING THE DATA

Heart rate readings occur every five seconds, which results in over 1.5 million readings total for the given time period in the PMData dataset. Each reading consists of a timestamp, 'bpm' field which denotes heart rate value, i.e., beats per minute, and confidence of the reading. Confidence ranges from 0 to 3, where 3 indicates that the device is extremely confident in the accuracy of the heart rate measured. Data cleaning process was employed in previous work by using a data-driven model for cleaning of eHealth data [9], which uses neural network algorithms to impute incorrect data, and has shown an improvement in accuracy between 10 and 17%.
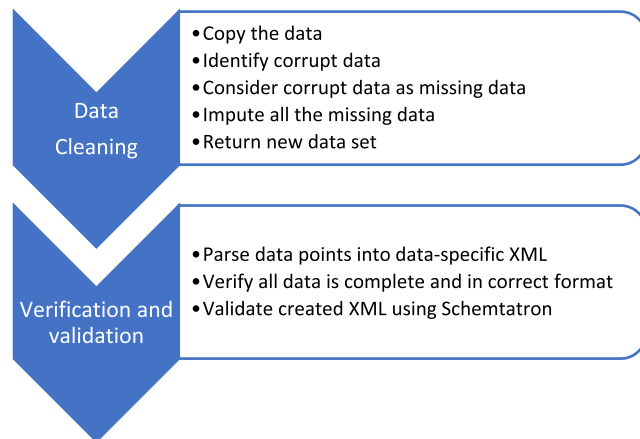

FIGURE 3. Process of data cleaning and validation.

Data cleaning process consists of copying the data, identifying corrupt data, considering corrupt data as missing data, imputing all missing data, and returning the new data set as a result. Data points selected for imputation were in this case those of confidence equal to zero. Thus, cleaned data contains original values for data points with confidence level 1-3, as well as imputed values for data points with confidence level zero. Clean data is then parsed and checked using Schematron-based validation. The process is pictured in Figure 3. The parser must be able to efficiently support large quantities of data. It should also be robust and always run without crashing. Parser was implemented in Python, using *pysimdjson* library, a SIMD-accelerated (Single instruction, multiple data) JSON parser, which can achieve speeds of up to 2.2 GB/s. It operates in two stages: the first stage processes data in batches of 64 bytes, and the second stage builds "tape representation". Detailed process is given in Figure 4 below.

The data format needed for EHR is eXtensible Markup Language (XML) document. XML documents consist of elements and attributes. For defining the allowable structure of an XML document, following schemas can be used:
- Document Type Definition (DTD),
- Microsoft XML-Data Reduced (XDR),
- or, most used, XML Schema definition language (XSD).

Initially, XML documents were only validated using schema validation. This means that if an XML document had passed schema validation, it was deemed valid. However, although schema validation ensures document structure, it cannot check conditional and integrity requirements. The proposed process for validation of personal healthcare documents (Figure 5) would first validate the document structure, followed by validation of document content and its attributes. Lastly, any existing additional constraints must also be validated.

Steps in the validation process pertaining to content and constraints validation require a rule-based schema language. Grammar-based schemas, such as XSD, simply cannot achieve necessary level of validity. Alternative options include Tree Regular Expressions for XML (TREX), Regular Language for XML Next Generation (RELAX NG) and Schematron, all of which are schema languages for XML with the ability to validate both the structure and content of XML documents. However, Schematron can also make assertions about patterns of occurrences anywhere in the document, whereas TREX and RELAX NG cannot. Another advantage of using Schematron is the possibility to nest Schematron schemas within XSD schemas. Considering this, Schematron is the most suitable schema language for the validation process presented in this work.

IHE has as its main goal integration of workflows within a healthcare setup using existing standards, for example Digital Imaging and Communications in Medicine (DICOM) and Health Level Seven (HL7). DICOM has been developed by the American College of Radiologists (ACR) and National Electrical Manufacturers Association (NEMA) and focuses on workflow of images. Complementary, HL7's focus is management of non-imaging data in hospital-based scenarios. Globally accepted, it offers protocols for sharing, managing, and administrating electronic health data [54] and enables the interoperability between electronic patient administration and practice management systems, pharmacy and billing systems, systems handling laboratory and dietary information and Electronic Health Record (EHR) systems. Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) [55] is a standard for exchanging healthcare information electronically, published by HL7. FHIR's basic building block is called Resource, and is made up of metadata, standard data, and human readable part (as pictured in Figure 6).

The HL7 FHIR specification is developing as a next-gen standard framework for the management and sharing of EHR data and health information in general. Schematron-based validation process will be defined and then verified using previously mentioned datasets. Different devices create different XML syntax to represent similar health data (e.g., heartbeat rate). For effective personal health data
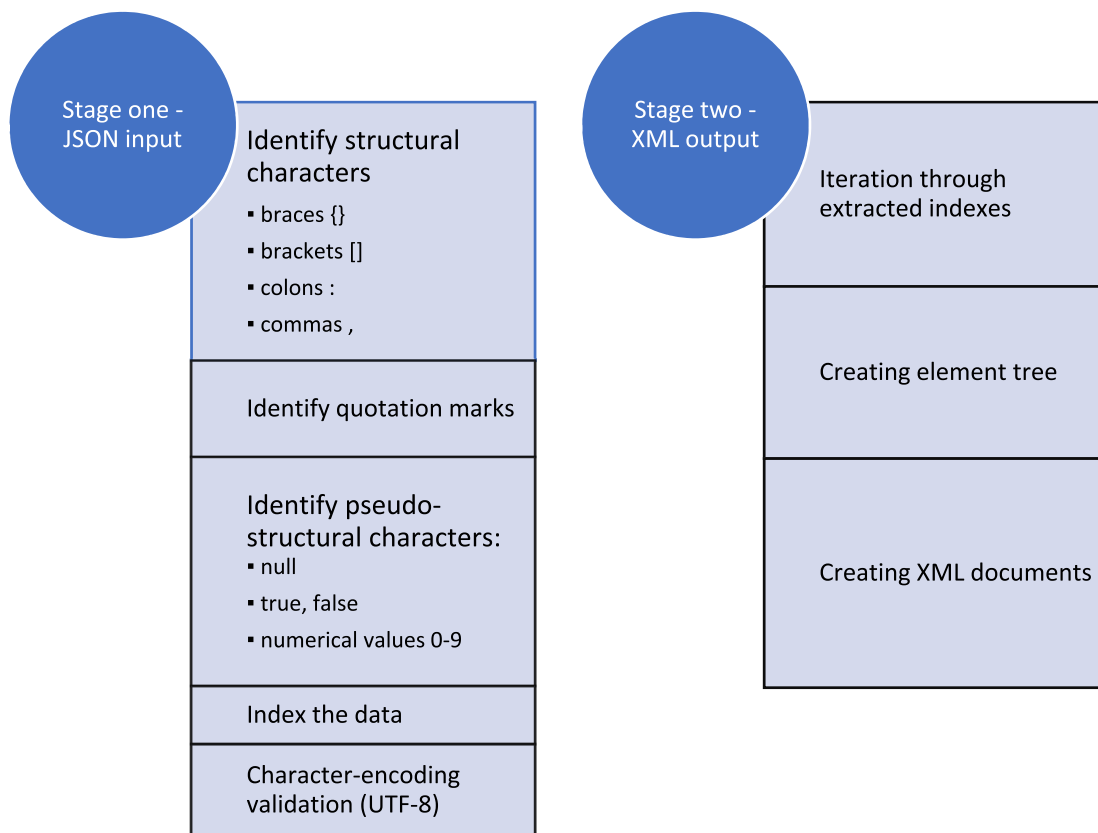
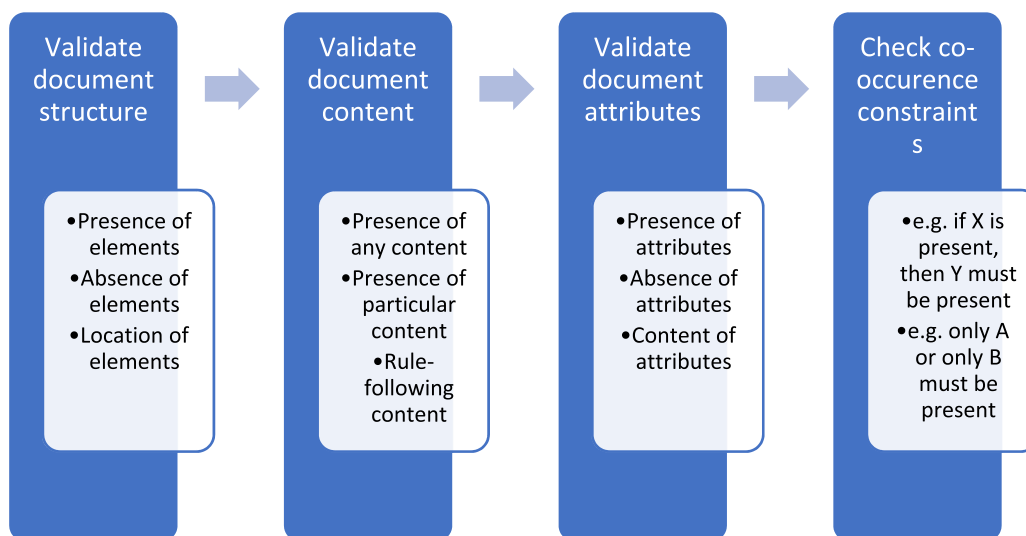**FIGURE 4.** Process of data parsing.

FIGURE 5. Expected validation process.

communication and integration, the XML documents must have correct syntax as well as be semantically meaningful, i.e., represent a complete FHIR resource. The idea is to create and maintain one Schematron document for similar health data among different personal tracking devices, thus reducing the complexity. Schematron can require the presence of some or all attributes in any given element, or it can define relationships or constrains of one element depending on another element. It allows for the implementation of complex rules and constraints needed for semantic validation. Schematron rules

```
<Patient xmlns="http://hl7.org/fhir">
    <id value="glossy"/>
    <meta>
        <lastUpdated value="2014-11-13T11:41:00+11:00"/>
    </meta>
    <text>
        <status value="generated"/>
        <div xmlns="http://www.w3.org/1999/xhtml">
            <p> Henry Levin the 7th</p>
            <p> MRN: 123456. Male, 24-Sept 1932</p>
        </div>
    </text>
    <extension url="http://example.org/StructureDefinition/trials">
        <valueCode value="renal"/>
    </extension>
    <identifier>
        <use value="usual"/>
        <type>
            <coding>
                <system value="http://terminology.hl7.org/CodeSystem/v2-0203"/>
                <code value="MR"/>
            </coding>
        </type>
        <system value="http://www.goodhealth.org/identifiers/mrn"/>
        <value value="123456"/>
    </identifier>
    <active value="true"/>
    <name>
        <family value="Levin"/>
        <given value="Henry"/>
        <suffix value="The 7th"/>
    </name>
    <gender value="male"/>
    <birthDate value="1932-09-24"/>
    <generalPractitioner>
        <reference value="Practitioner/example"/>
        <display value="Dr Adam Careful"/>
    </generalPractitioner>
    <managingOrganization>
        <reference value="Organization/2"/>
        <display value="Good Health Clinic"/>
    </managingOrganization>
</Patient>
```

Resource ID and metadata

Human-readable text (summary)

Extension URI to resource definition

Standard data: set of structured data items as described by the definition of the resource type

FIGURE 6. An example of HL7 FHIR resource.

```
<rule context="Observation">
    <assert test="@id">The element Observation must have an id attribute.</assert>
    <assert test="count(*) = 2 and count(Category) = 1 and count(Code)= 1">
        The element Observation must have the child elements Category and Code.</assert>
</rule>
```

FIGURE 7. Rule example for FHIR resource observation.

are formed using the rule element with a context attribute. The value of the attribute must match an *Xpath Expression* which selects at least one node in the document. The context attribute specifies where the assertion must be applied. From the abovementioned example, the context is fixed to the *Observation element*, so the Schematron rule with the *Observation element* being the context should look like the following (Figure 7). To fulfill all the requirements previously mentioned and comply with the standards, heart rate Schematron for data validation with the goal of its inclusion into EHR needs to contain rules provided in Figure 8.

C. PROCESS VERIFICATION: USE-CASE STUDY
The defined process was verified in a use-case study, using two datasets containing various relevant data types. The results are presented in the following section.

V. RESULTS
The goal was to research integration of personal health data into Electronic Health Record (EHR) which forms a part of Central Health Information System. To verify the data cleaning and validation process presented in this paper, a dataset was collected for the purpose of this research, containing
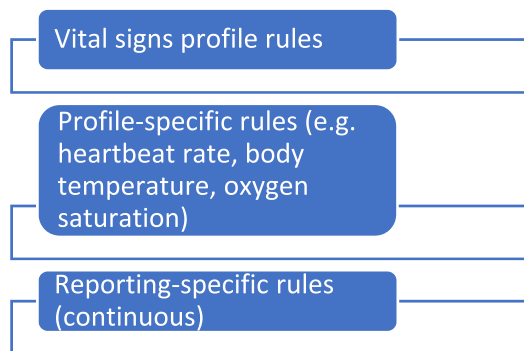
FIGURE 8. Rules hierarchy for heartbeat rate schematron.

heartbeat rate, as well as body temperature and oxygen saturation (SpO2) using a fitness tracker. Data has been cleaned by employing the data-driven model for cleaning of eHealth data [9], which uses neural network algorithms to impute incorrect data, and has previously resulted in an improvement in accuracy between 10 and 17%. To test the defined validation process in a use-case scenario, a two-stage parser has been developed. The full two months of readings for the three datatypes measured equate to around 50 MB worth of JSON files with over 2 million datapoints, which have been parsed in 58 seconds on a Ryzen 5 5600X 6-core processor (running at 3.7 GHz in 64- bit mode). As per reliability, this has been done for all data points, with a success rate of 100%.

Finally, a detailed overview of all the rules defined is given as a complete summary of the mandatory requirements below in this chapter. This facilitated the creation of Schematrons for the three datatypes. This is used in the final step of the data validation process. Heartbeat rate data is described by *FHIR HeartRate Structure Definition* [50]. Schematron for heartbeat rate type data is shown in Figure 9. Similarly, body temperature data is described by FHIR *BodyTemp Structure Definition* [56] which is to be followed. The structure is derived from *observation-vitalsigns*. Skin temperature reading data is exported in a form of a JSON file. Readings occur each minute. Each reading consists of a timestamp, temperature value and unit in which temperature is measured (in this case, degrees Celsius). Lastly, oxygen saturation in arterial blood is described by FHIR *OxygenSat Structure Definition* [57] which is to be followed. The structure is derived from *observation-vitalsigns*. Oxygen saturation is exported in a form of JSON file, consisting of *dateTime* and value attributes. *dateTime* is of the same format as in the case of heartbeat rate readings and its value is expressed in percentage. Figs. 10 and 11 show Schematrons for body temperature and oxygen saturation, respectively. Complete Summary of the Mandatory Requirements for the given datatypes is as shown below:

Observation.Code needs to have one code with:

(I) fixed value of coding system equal to loinc.org

(II) fixed coding code equal to:

    a. 8867-4 for heartbeat rate

    b. 8310-5 for body temperature

    c. 2708-6 for oxygen saturation

(III) all codes must have a system value

Also, *Observation* needs to have a value quantity or, if there is no such value, a reason for absence of data. If a value quantity exists, it needs to have:

(I) numerical value

(II) fixed value quantity system equal to unitsofmeasure.org

(III) UCUM unit code:

    a. /min (per minute) for heartbeat rate

    b. Cel (Celsius) or degF (Fahrenheit) for body temperature

    c. % (per cent) for oxygen saturation

In total, *Observation* must:

    a. contain three mandatory elements (with 4 more nested mandatory elements),

    b. support four elements,

    c. have fixed value for three (body temperature) to four (heartbeat rate, oxygen saturation) elements.

Compliance to each rule for all given data types is ensured using the corresponding Schematron. The semantic constraints framework provides the specification of data that normally has a very complex structure, while the validation tool may be used as a standalone component or can be integrated as a module into a larger data processing system. General important things to note of FHIR resource representation:

- FHIR elements are always in the namespace http://hl7.org/fhir, usually specified as default namespace on root element
- Resource names are case-sensitive
- Element names are case-sensitive and must appear in the order specified by documentation. In case of element repeating, elements must be ordered
- FHIR elements cannot be empty - they either have a value attribute, valid child element or extension
- Attributes cannot be empty
- Infrastructural elements must appear before any other defined child elements, i.e., first base resource elements, then domain resource elements.

Element *Observation resource* provides "measurements and simple assertions made about a patient, device or other subject" and is used for vital signs, height, weight, laboratory results, etc. Element *Id* contains logical ID of the resource, in this case heart rate. *Meta* is metadata about the resource, with profile references a structure profile URL to which this resource claims to conform to, *vital signs* being an example. An optional *Text* contains human-readable summary of the resource. *Category* classifies the general type of observation, whereas coding is a reference to a code unambiguously defined by a terminology system (observation-category and http://loinc.org, respectively) and identified by an existing code (vital signs and 8867-4). *Display* provides human-readable meaning of the code.

```
 1  <sch:schema xmlns:sch="http://purl.oclc.org/dsdl/schematron" queryBinding="xslt2">
 2      <sch:ns prefix="f" uri="http://hl7.org/fhir"/>
 3      <sch:ns prefix="h" uri="http://www.w3.org/1999/xhtml"/>
 4      <sch:pattern id="observation">
 5          <sch:title>Observation</sch:title>
 6          <sch:rule context="f:Observation">
 7              <sch:assert test="not(exists(f:dataAbsentReason)) or (not(exists(*[starts-with(local-name(.), 'value')])))">dataAbsentReason
 8                  SHALL only be present if Observation.value[x] is not present (inherited)</sch:assert>
 9              <sch:assert test="not(exists(f:component/f:code)) or
10                  count(for $coding in f:code/f:coding return parent::*/f:component/f:code/f:coding[f:code/@value=$coding/f:code/@value
11                  and f:system/@value=$coding/f:system/@value])=0">Component code SHALL not be same as observation code (inherited)</sch:assert>
12              <sch:assert test="f:id/@value = 'heart-rate'">Observation is not heartrate type observation</sch:assert>
13              <sch:assert test="exists(f:subject/f:reference/@value)">Patient must exist and be uniquely defined</sch:assert>
14              <sch:assert test="matches(f:effectiveDateTime/@value, '^(\d{4}-\d{2}-\d{2}T\d{2}:\d{2}:\d{2}\+\d{2}:\d{2})$')">Observation
15                  needs to have proper format for dateTime</sch:assert>
16              <sch:assert test="exists(f:valueQuantity/f:value/@value)">Observation needs to have value measured</sch:assert>
17          </sch:rule>
18      </sch:pattern>
19      <sch:pattern id="category">
20          <sch:rule context="f:Observation/f:category/f:coding/f:code">
21              <sch:assert test="@value = 'vital-signs'">Vital signs must defined by correct observation code</sch:assert>
22              <sch:assert test="count(@value) = 1">Code must exist and be uniquely defined</sch:assert>
23          </sch:rule>
24          <sch:rule context="f:Observation/f:category/f:coding/f:system">
25              <sch:assert test="@value = 'http://terminology.hl7.org/CodeSystem/observation-category'">Vital signs must
26                  defined by correct system</sch:assert>
27          </sch:rule>
28      </sch:pattern>
29      <sch:pattern id="code">
30          <sch:rule context="f:Observation/f:code/f:coding">
31              <sch:assert test="f:code/@value = '8867-4'">Heartbeat rate must defined by correct observation code</sch:assert>
32              <sch:assert test="f:system/@value = 'http://loinc.org'">Heartbeat rate must defined by correct system</sch:assert>
33              <sch:assert test="count(f:code/@value) = 1">Code must exist and be uniquely defined</sch:assert>
34          </sch:rule>
35      </sch:pattern>
36  </sch:schema>
```

FIGURE 9. Schematron for heartbeat rate type data.

Subject references the subject of the observation, i.e., the patient. *Effective[x]* has four available types: *dateTime*, *Period*, *Timing* and *instant*. In this case, effective *DateTime* is used. According to the specification, *dateTime* may represent date, date-time or partial date with the respective formats being YYYY-MM-DD or YYYY-MMDDThh: mm:ss + zz:zz and YYYY or YYYY-MM. However, for this specific use case, only *date-time* is expected and this needs to be reflected in the Schematron rules. *Value[x]* has different types available (as pictured in Figure 12) and provides information determined a result of the observation. Here, value *Quantity* is used. Value *Quantity* consists of numerical value, unit representation (e.g., bpm), system that defines coded unit form and coded form of the unit. Another (optional) parameter is comparator with possible values <(less), less or equal ($<=$), >(more) and more or equal ($>=$). Thus, after adjusting the *datetime* format, mapping data points from the wearable in this case is straightforward.

All datatypes share the rules for vital signs profile, since the FHIR *Vital Signs* profile sets minimum expectations for the *Observation resource* to record, search and fetch the vital signs (e.g., temperature, blood pressure, respiration rate, etc.). This is followed by profile-specific rules and whatever additional rules apply.

All the XMLs, produced from datapoints and compliant with the Schematrons defined in this work, are following the rules of the latest FHIR version, i.e.:
- datatypes, in XML and JSON format
- the terminology layer, i.e., *CodeSystems* and *ValueSets*
- the conformance framework (*StructureDefinition*)
- the FHIR resources, i.e., *Patient* and *Observation*

Regarding the implementation presented in this work, executable components, i.e., parser and validator were developed in Python 3 using Anaconda Spyder, opensource IDE for scientific programming and computing (data science, machine learning applications, large-scale data processing, etc.) and scientific packages NumPy, SciPy, Matplotlib, pysimdjson and pandas. XML templates and Schematrons were written in Notepad++.

Considering our research objectives, the paper provides the following contributions:
- model for validation of personal health data collected by wearable sensors with the goal of integrating the data into EHR was proposed
- semantic constraints for healthcare data types were defined in compliance with standards which were used in creating Schematron schemas corresponding to each of the datatypes

```
1  <sch:schema xmlns:sch="http://purl.oclc.org/dsdl/schematron" queryBinding="xslt2">
2    <sch:ns prefix="f" uri="http://hl7.org/fhir"/>
3    <sch:ns prefix="h" uri="http://www.w3.org/1999/xhtml"/>
4    <sch:pattern id="observation">
5      <sch:title>Observation</sch:title>
6      <sch:rule context="f:Observation">
7        <sch:assert test="not(exists(f:dataAbsentReason)) or (not(exists(*[starts-with(local-name(.), 'value')])))">dataAbsentReason
8          SHALL only be present if Observation.value[x] is not present (inherited)</sch:assert>
9        <sch:assert test="not(exists(f:component/f:code)) or
10         count(for $coding in f:code/f:coding return parent::*/f:component/f:code/f:coding[f:code/@value=$coding/f:code/@value
11         and f:system/@value=$coding/f:system/@value])=0">Component code SHALL not be same as observation code (inherited)</sch:assert>
12       <sch:assert test="f:id/@value = 'body-temperature'">Observation is not body temperature type observation</sch:assert>
13       <sch:assert test="exists(f:subject/f:reference/@value)">Patient must exist and be uniquely defined</sch:assert>
14       <sch:assert test="matches(f:effectiveDateTime/@value, '^(\d{4}-\d{2}-\d{2}T\d{2}:\d{2}:\d{2}\+\d{2}:\d{2})$')">Observation
15         needs to have proper format for dateTime</sch:assert>
16       <sch:assert test="exists(f:valueQuantity/f:value/@value)">Observation needs to have value measured</sch:assert>
17      </sch:rule>
18    </sch:pattern>
19    <sch:pattern id="category">
20      <sch:rule context="f:Observation/f:category/f:coding/f:code">
21        <sch:assert test="@value = 'vital-signs'">Vital signs must defined by correct observation code</sch:assert>
22        <sch:assert test="count(@value) = 1">Code must exist and be uniquely defined</sch:assert>
23      </sch:rule>
24      <sch:rule context="f:Observation/f:category/f:coding/f:system">
25        <sch:assert test="@value = 'http://terminology.hl7.org/CodeSystem/observation-category'">Vital signs must
26          defined by correct system</sch:assert>
27      </sch:rule>
28    </sch:pattern>
29    <sch:pattern id="code">
30      <sch:rule context="f:Observation/f:code/f:coding">
31        <sch:assert test="f:code/@value = '8310-5'">Body temperature must defined by correct observation code</sch:assert>
32        <sch:assert test="f:system/@value = 'http://loinc.org'">Body temperature must defined by correct system</sch:assert>
33        <sch:assert test="count(f:code/@value) = 1">Code must exist and be uniquely defined</sch:assert>
34      </sch:rule>
35    </sch:pattern>
36  </sch:schema>
```

FIGURE 10. Schematron for temperature type data.

```
1  <sch:schema xmlns:sch="http://purl.oclc.org/dsdl/schematron" queryBinding="xslt2">
2    <sch:ns prefix="f" uri="http://hl7.org/fhir"/>
3    <sch:ns prefix="h" uri="http://www.w3.org/1999/xhtml"/>
4    <sch:pattern id="observation">
5      <sch:title>Observation</sch:title>
6      <sch:rule context="f:Observation">
7        <sch:assert test="not(exists(f:dataAbsentReason)) or (not(exists(*[starts-with(local-name(.), 'value')])))">dataAbsentReason
8          SHALL only be present if Observation.value[x] is not present (inherited)</sch:assert>
9        <sch:assert test="not(exists(f:component/f:code)) or
10         count(for $coding in f:code/f:coding return parent::*/f:component/f:code/f:coding[f:code/@value=$coding/f:code/@value
11         and f:system/@value=$coding/f:system/@value])=0">Component code SHALL not be same as observation code (inherited)</sch:assert>
12       <sch:assert test="f:id/@value = 'satO2'">Observation is not oxygen saturation type observation</sch:assert>
13       <sch:assert test="exists(f:subject/f:reference/@value)">Patient must exist and be uniquely defined</sch:assert>
14       <sch:assert test="matches(f:effectiveDateTime/@value, '^(\d{4}-\d{2}-\d{2}T\d{2}:\d{2}:\d{2}\+\d{2}:\d{2})$')">Observation
15         needs to have proper format for dateTime</sch:assert>
16       <sch:assert test="exists(f:valueQuantity/f:value/@value)">Observation needs to have value measured</sch:assert>
17      </sch:rule>
18    </sch:pattern>
19    <sch:pattern id="category">
20      <sch:rule context="f:Observation/f:category/f:coding/f:code">
21        <sch:assert test="@value = 'vital-signs'">Vital signs must defined by correct observation code</sch:assert>
22        <sch:assert test="count(@value) = 1">Code must exist and be uniquely defined</sch:assert>
23      </sch:rule>
24      <sch:rule context="f:Observation/f:category/f:coding/f:system">
25        <sch:assert test="@value = 'http://terminology.hl7.org/CodeSystem/observation-category'">Vital signs must
26          defined by correct system</sch:assert>
27      </sch:rule>
28    </sch:pattern>
29    <sch:pattern id="code">
30      <sch:rule context="f:Observation/f:code/f:coding">
31        <sch:assert test="f:code/@value = '2708-6'">Oxygen saturation must defined by correct observation code</sch:assert>
32        <sch:assert test="f:system/@value = 'http://loinc.org'">Oxygen saturation must defined by correct system</sch:assert>
33        <sch:assert test="count(f:code/@value) = 1">Code must exist and be uniquely defined</sch:assert>
34      </sch:rule>
35    </sch:pattern>
36  </sch:schema>
```

FIGURE 11. Schematron for oxygen saturation data.

- validation process was then verified using datasets containing various health-related datatypes.

VI. DISCUSSION

Using personal healthcare data is a significant step in optimizing patient care. Despite the advances in personal and body area networks, along with the increase in official use of EHRs globally, healthcare enterprises are slow in catching on to the full potential of using personal health information and thus improving the overall quality of medical care. Especially considering the current situation, as Coronavirus disease 2019 (COVID-19) overwhelms the health care systems around the

TABLE 1. Comparative table of the performances and functionalities of different system models.

	Data collected via wearables	Data cleaning process	Data integration	Validation process	Compliance with standards
Wearable HUB	Yes	N/A	PHR	N/A	N/A
Angel-Echo	Yes	N/A	PHR	N/A	N/A
Open Health	Yes	Machine learning algorithms	PHR	N/A	N/A
mHealth	No	N/A	PHR/EHR	N/A	HL7 FHIR-based
Tangle	Yes	N/A	PHR/EHR	N/A	HL7 FHIR-based
mHealth4Afrika	Yes	N/A	PHR/EHR	Validation through human review; group interviews and observation	HL7 FHIR-based
Proposed model	Yes	Machine learning algorithm: neural network	EHR focus	Schematron validation	HL7 FHIR-based

FIGURE 12. value[x] types.

globe, personal trackers could help in monitoring coronavirus patients and help decide about hospitalization, help identify faster those who have contracted the disease, and, finally, track the progress of the pandemic.

Thus, the goal of the research presented integrates personal health record into a formal medical information system; i.e., a transformation of personal health data into adequate format and its inclusion into formal EHR. To ensure data quality, data cleaning process was described in previous work by using a data-driven model for cleaning of eHealth data, which used neural network algorithms to impute incorrect data. This approach has shown accuracy improvements of 10-17%. Parser performance and reliability has been proven to be viable with short parsing times and 100% success rate.

Compliance to every rule for all given data types has been extensively tested by purposefully attempting to validate erroneous XML files against the corresponding Schematron. The next key challenge of the research was to define a validation process to ensure the data complies with standards. This was done by:

- Defining semantic constraints for healthcare datatypes to ensure compliance to standards making the information medically valid and relevant
- Defining and modelling the validation process of the data collected which enables the data to be easily transferred and incorporated into a formal EHR. Finally, this approach was verified in a use-case study, using an existing dataset containing various relevant datatypes.

Compliance to standards is essential to use personal health data for personalized and preventive medicine. Employment of the process specified in this work enables inclusion of created information into a formal EHR, following current IHE standards for the given datatypes. Finally, a comparison of the proposed model with those mentioned earlier is given in the Table 1 below. The model focuses on EHR data integration, provides data cleaning module, offers automated Schematron-based validation and is compliant with the leading industry standard.

FUTURE WORK

Survey by Cisco found that 74% of patients worldwide were willing to allow cloud-based storage of their personal health records under certain conditions [58]. Study [59] reports that up to 75% of UK and 60% of US citizens are willing to share their anonymized personal health information,

while [60] concludes that the COVID-19 pandemic made people generally more supportive of sharing health data within the health information system. Sharing personal health information urges more caution than other types of data with similar privacy concerns (e.g., consumer spending and financial data) [61]. New participant-centered investigations show that patients are more likely to share data when they have the power to select the conditions under which said data is shared [62].

Furthermore, [63] reports that patients are somewhat comfortable sharing their health data with third-party commercial companies for patient purposes, but uncomfortable sharing it for business purposes. Still, there are many possible security risks and privacy concerns that need to be mitigated, such as data anonymization, data encryption, insecure communication channel, logging practices, etc. Future work envisions comprehensive security and privacy threat identification and analysis when integrating IoMT health data into EHR.

Furthermore, more research is necessary to see how well older trackers perform and how the quality of data changes as sensors age.

INSTITUTIONAL REVIEW BOARD STATEMENT:
The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Faculty of Electrical Engineering and Computing, University of Zagreb (January 20th, 2021).

INFORMED CONSENT STATEMENT:
Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

DATA AVAILABILITY STATEMENT:
The PMData Dataset (published at Simula Open Datasets)
https://datasets.simula.no/pmdata.
OxyBeat Dataset was collected for the purpose of this research
and is available at Github
https://github.com/korenana/oxybeatdataset.

REFERENCES

[1] *Fitness Tracker Market Size, Share & COVID-19 Impact Analysis (2020–2027)*, Fortune Bus. Insights, Maharashtra, India, Oct. 2020.

[2] P.-C. Huang, C.-C. Lin, Y.-H. Wang, and H.-J. Hsieh, "Development of health care system based on wearable devices," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 249–252, doi: 10.1109/PHM-Paris.2019.00049.

[3] R. Qian, "Healthcare on consumer-grade wearable devices," in *Proc. IEEE 3rd Int. Conf. Safe Prod. Informatization (IICSPI)*, Nov. 2020, pp. 298–303, doi: 10.1109/IICSPI51290.2020.9332400.

[4] G. Assenza, C. Fioravanti, S. Guarino, and V. Petrassi, "New perspectives on wearable devices and electronic health record systems," in *Proc. IEEE Int. Workshop Metrology Industry IoT*, Jun. 2020, pp. 740–745, doi: 10.1109/MetroInd4.0IoT48571.2020.9138170.

[5] X. Ding, D. Clifton, N. Ji, N. H. Lovell, P. Bonato, W. Chen, X. Yu, Z. Xue, T. Xiang, X. Long, K. Xu, X. Jiang, Q. Wang, B. Yin, G. Feng, and Y.-T. Zhang, "Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 48–70, 2021, doi: 10.1109/RBME.2020.2992838.

[6] K. Ueafuea, C. Boonnag, T. Sudhawiyangkul, P. Leelaarporn, A. Gulistan, W. Chen, S. C. Mukhopadhyay, T. Wilaiprasitporn, and S. Piyayotai, "Potential applications of mobile and wearable devices for psychological support during the COVID-19 pandemic: A review," *IEEE Sensors J.*, vol. 21, no. 6, pp. 7162–7178, Mar. 2021, doi: 10.1109/JSEN.2020.3046259.

[7] F. A. A. Naqiyuddin, W. Mansor, N. M. Sallehuddin, M. N. S. Mohd Johari, M. A. S. Shazlan, and A. N. Bakar, "Wearable social distancing detection system," in *Proc. IEEE Int. RF Microw. Conf. (RFM)*, Dec. 2020, pp. 1–4, doi: 10.1109/RFM50841.2020.9344786.

[8] K. Grifantini, "Tracking sleep to optimize health," *IEEE Pulse*, vol. 11, no. 5, pp. 12–16, Sep. 2020, doi: 10.1109/MPULS.2020.3022142.

[9] A. Koren, M. Jurčević, and R. Prasad, "Comparison of data-driven models for cleaning eHealth sensor data: Use case on ECG signal," *Wireless Pers. Commun.*, vol. 114, no. 2, pp. 1501–1517, Sep. 2020, doi: 10.1007/s11277-020-07435-7.

[10] K. R. Evenson, M. M. Goto, and R. D. Furberg, "Systematic review of the validity and reliability of consumer-wearable activity trackers," *Int. J. Behav. Nutrition Phys. Activity*, vol. 12, no. 1, pp. 1–22, Dec. 2015, doi: 10.1186/s12966-015-0314-1.

[11] J. Goncalves, L. Leitao, and V. Carvalho, "Analysis of the reliability and accuracy of a wearable device: Comparative study with a certified clinical device," in *Proc. 12th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Lisbon, Portugal, Jun. 2017, pp. 1–6, doi: 10.23919/CISTI.2017.7976034.

[12] L. M. Feehan, J. Geldman, E. C. Sayre, C. Park, A. M. Ezzat, J. Y. Yoo, C. B. Hamilton, and L. C. Li, "Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data," *JMIR mHealth uHealth*, vol. 6, no. 8, Aug. 2018, Art. no. e10527, doi: 10.2196/10527.

[13] L. Scalise and G. Cosoli, "Wearables for health and fitness: Measurement characteristics and accuracy," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, Houston, TX, USA, May 2018, pp. 1–6, doi: 10.1109/I2MTC.2018.8409635.

[14] B. W. Nelson and N. B. Allen, "Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study," *JMIR mHealth uHealth*, vol. 7, no. 3, Mar. 2019, Art. no. e10828, doi: 10.2196/10828.

[15] A. F. Turki, M. B. Jani, K. Ding, R. Zhang, and K. Behbehani, "An investigation of heartrate sensing accuracy by wrist-worn fitness tracking devices," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 3337–3340, doi: 10.1109/EMBC.2019.8856375.

[16] A. Russell, C. Heneghan, and S. Venkatraman, "Investigation of an estimate of daily resting heart rate using a consumer wearable device," *MedRxiv*, Oct. 2019, doi: 10.1101/19008771.

[17] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, "Data quality in Internet of Things: A state-of-the-art survey," *J. Netw. Comput. Appl.*, vol. 73, pp. 57–81, Sep. 2016.

[18] D. Pal, A. Tassanaviboon, C. Arpnikanondt, and B. Papasratorn, "Quality of experience of smart-wearables: From fitness-bands to smartwatches," *IEEE Consum. Electron. Mag.*, vol. 9, no. 1, pp. 49–53, Jan. 2020, doi: 10.1109/MCE.2019.2941462.

[19] M. Magno, M. Pritz, P. Mayer, and L. Benini, "DeepEmote: Towards multilayer neural networks in a low power wearable multi-sensors bracelet," in *Proc. 7th IEEE Int. Workshop Adv. Sensors Interfaces (IWASI)*, Vieste, Italy, Jun. 2017, pp. 32–37, doi: 10.1109/IWASI.2017.7974208.

[20] M. G. Rahman, M. Z. Islam, T. Bossomaier, and J. Gao, "CAIRAD: A co-appearance based analysis for incorrect records and attribute-values detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Brisbane, QLD, Australia, Jun. 2012, pp. 1–10.

[21] M. G. Rahman and M. Z. Islam, "FIMUS: A framework for imputing missing values using co-appearance, correlation and similarity analysis," *Knowl.-Based Syst.*, vol. 56, no. C, pp. 311–327, Jan. 2014.

[22] O. Salem, A. Guerassimov, A. Mehaoua, A. Marcus, and B. Furht, "Sensor fault and patient anomaly detection and classification in medical wireless sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 4373–4378.

[23] D. Yang, Y. Cheng, J. Zhu, D. Xue, G. Abt, H. Ye, and Y. Peng, "A novel adaptive spectrum noise cancellation approach for enhancing heartbeat rate monitoring in a wearable device," *IEEE Access*, vol. 6, pp. 8364–8375, 2018.

[24] J. Hong, P. Morris, and J. Seo, "Interconnected personal health record ecosystem using IoT cloud platform and HL7 FHIR," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Park City, UT, USA, Aug. 2017, pp. 362–367.

[25] A. Koren, M. Jurcevic, and D. Huljenic, "Requirements and challenges in integration of aggregated personal health data for inclusion into formal electronic health records (EHR)," in *Proc. 2nd Int. Colloq. Smart Grid Metrology (SMAGRIMET)*, Split, Croatia, Apr. 2019, pp. 1–5.

[26] J. Ruminski, A. Bujnowski, T. Kocejko, A. Andrushevich, M. Biallas, and R. Kistler, "The data exchange between smart glasses and healthcare information systems using the HL7 FHIR standard," in *Proc. 9th Int. Conf. Human Syst. Interact. (HSI)*, Portsmouth, U.K., Jul. 2016, pp. 525–531.

[27] Y. Aliakbarpoor, S. Comai, and G. Pozzi, "Designing a HL7 compatible personal health record for mobile devices," in *Proc. IEEE 3rd Int. Forum Res. Technol. Soc. Ind. (RTSI)*, Modena, Italy, Sep. 2017, pp. 1–6.

[28] E. Saweros and Y.-T. Song, "Connecting personal health records together with EHR using tangle," in *Proc. 20th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distributed Comput. (SNPD)*, Toyama, Japan, Jul. 2019, pp. 547–554, doi: 10.1109/SNPD.2019.8935646.

[29] Y. Chen, W. Gao, S. Wang, and S. Jiao, "Demo abstract: WearableHUB: An open pervasive wearable data fusion platform for personal health management," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2015, pp. 217–219, doi: 10.1109/PERCOMW.2015.7134027.

[30] M. Ma, M. Skubic, K. Ai, and J. Hubbard, "Angel-echo: A personalized health care application," in *Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE)*, Jul. 2017, pp. 258–259, doi: 10.1109/CHASE.2017.91.

[31] G. Bhat, R. Deb, and U. Y. Ogras, "OpenHealth: Open-source platform for wearable health monitoring," *IEEE Design Test*, vol. 36, no. 5, pp. 27–34, Oct. 2019, doi: 10.1109/MDAT.2019.2906110.

[32] M. Cunningham and P. M. Cunningham, "MHealth4Afrika pilot validation in healthcare facilities in ethiopia, Kenya and Malawi," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2019, pp. 1–8, doi: 10.1109/GHTC46095.2019.9033052.

[33] *Information Technology—Document Schema Definition Languages (DSDL)—Part 3: Rule-Based Validation—Schematron*, Standard ISO/IEC 19757-3:2016, 2016.

[34] P. Fennell, "Schematron—More useful than you'd thought," XML London, London, U.K., Jun. 2014, pp. 103–112.

[35] *Integrating Health Enterprise (IHE)*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.ihe.net/about_ihe/

[36] *IHE Resources: Technical Frameworks*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.ihe.net/resources/technical_frameworks/#IT

[37] *IHE ITI Technical Framework Volume 1 Integration Profiles*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol1.pdf

[38] *IHE ITI Technical Framework Volume 2a Transactions*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol2a.pdf

[39] *IHE ITI Technical Framework Volume 2b Transactions*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol2b.pdf

[40] *IHE ITI Technical Framework Volume 2c Transactions*, IHE International, Springfield, IL, USA, 2014.

[41] *IHE ITI Technical Framework Volume 3 Metadata*. Accessed: Feb. 15, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol3.pdf

[42] *IHE ITI Metadata Update*. Accessed: Feb. 16, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Suppl_XDS_Metadata_Update.pdf

[43] *IHE ITI Cardiology Technical Framework Resting ECG Workflow (REFW)*. Accessed: Feb. 16, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/Cardiology/IHE_CARD_Suppl_REWF.pdf

[44] *IHE ITI Cardiology Technical Framework Stress Testing Workflow (STRESS)*. Accessed: Feb. 18, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/Cardiology/IHE_CARD_Suppl_Stress.pdf

[45] *Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR)*. Accessed: Feb. 12, 2022. [Online]. Available: https://www.hl7.org/fhir

[46] V. Thambawita, S. A. Hicks, H. Borgli, H. K. Stensland, D. Jha, M. K. Svensen, S.-A. Pettersen, D. Johansen, H. D. Johansen, S. D. Pettersen, and S. Nordvang, "PMData: A sports logging dataset," in *Proc. 11th ACM Multimedia Syst. Conf.*, Feb. 2020, pp. 231–236, doi: 10.1145/3339825.3394926.

[47] *OxyBeat Dataset*. Accessed: Sep. 3, 2021. [Online]. Available: https://github.com/korenana/oxybeat-dataset

[48] (Oct. 3, 2020). *Fitbit Reports Third Quarter Results for the Three Months Ended*. [Online]. Available: https://investor.fitbit.com/press/press-releases/pressrelease-details/2020/Fitbit-Reports-Third-Quarter-Results-for-the-Three-Months-Ended-October-3-2020/default.aspx

[49] A. Mishra, A. Nieto, and S. Kitsiou, "Systematic review of mHealth interventions involving fitbit activity tracking devices," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Park City, UT, USA, Aug. 2017, p. 455, doi: 10.1109/ICHI.2017.42.

[50] (Mar. 10, 2021). *U.S. National Library of Medicine ClinicalTrials*. [Online]. Available: https://clinicaltrials.gov/

[51] D. Rosenberg, E. A. Kadokura, E. D. Bouldin, C. E. Miyawaki, C. S. Higano, and A. L. Hartzler, "Acceptability of Fitbit for physical activity tracking within clinical care among men with prostate cancer," in *Proc. AMIA Annu. Symp.*, vol. 2016, pp. 1050–1059.

[52] *FHIR Heart Rate Profile*. Accessed: Feb. 18, 2022. [Online]. Available: http://hl7.org/fhir/ Structure Definition/heartrate (published on Sat, Aug. 11, 2018 00:00+1000 as a draft by Health Level Seven International-Orders and Observations Workgroup)

[53] *Structure Definition: VitalSigns*. Accessed: Feb. 12, 2022. [Online]. Available: http://hl7.org/fhir/StructureDefinition/vitalsigns, (published on Fri, Mar. 25, 2016 00:00+1100 as a draft by Health Level Seven International-Orders and Observations Workgroup)

[54] *IHE ITI Quality, Research and Public Health Technical Framework Supplement Aggregate Data Exchange (ADX)*. Accessed: Feb. 18, 2022. [Online]. Available: https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_Suppl_ADX.pdf

[55] I. K. Indrajit and B. S. Verma, "DICOM, HL7 and IHE: A basic primer on healthcare standards for radiologists," *Indian J. Radiol. Imag.*, vol. 17, no. 2, pp. 66–68, Apr. 2007.

[56] *FHIR Body Temperature Profile*. Accessed: Feb. 19, 2022. [Online]. Available: http://hl7.org/fhir/StructureDefinition/bodytemp (published on Sat, Aug. 11, 2018 00:00+1000 as a draft by Health Level Seven International-Orders and Observations Workgroup)

[57] *FHIR Oxygen Saturation Profile*. Accessed: Feb. 19, 2022. [Online]. Available: http://hl7.org/fhir/StructureDefinition/oxygensat (published on Tue, Oct. 23, 2018 00:00+1100 as a draft by Health Level Seven International-Orders and Observations Workgroup)

[58] *Cisco Customer Experience Report for Healthcare Conducted by Insight Express*, Insight Express LLC, Norwalk, CT, USA, Feb. 2013.

[59] S. Ghafur, J. Van Dael, M. Leis, A. Darzi, and A. Sheikh, "Public perceptions on data sharing: Key insights from the U.K. and the USA," *Lancet Digit. Health*, vol. 2, no. 9, 2020, pp. e444-e446, doi: 10.1016/S2589-7500(20)30161-8.

[60] *Most Americans Want to Share and Access More Digital Health Data*, Pew Res. Center, Washington, DC, USA, Jul. 2021.

[61] R. Agarwal and C. Anderson, "The complexity of consumer willingness to disclose personal information: Unraveling health information privacy concerns," in *Proc. eHealth Initiative's 5th Annu. Conf.*, Washington, DC, USA, 2008, pp. 279–288.

[62] J. Kaye, L. Curren, N. Anderson, K. Edwards, S. M. Fullerton, N. Kanellopoulou, D. Lund, D. G. MacArthur, M. Mascalzoni, J. Shepherd, P. L. Taylor, S. F. Terry, and S. F. Winter, "From patients to partners: Participant-centric initiatives in biomedical research," *Nature Rev. Genet.*, vol. 13, no. 5, pp. 371–376, May 2012, doi: 10.1038/nrg3218.

[63] M. G. Trinidad, J. Platt, and S. L. R. Kardia, "The public's comfort with sharing health data with third-party commercial companies," *Humanities Social Sci. Commun.*, vol. 7, pp. 1–10, 2020, Art. no. 149, doi: 10.1057/s41599-020-00641-5.

• • •