System model for integration of wearable smart device data into a central health information system

PhD student Ana Koren, mag. ing. inf. et comm. techn.
(ana.koren@fer.hr)
mentor: izv. prof. dr. sc. Marko Jurčević,
University of Zagreb Faculty of Electrical Engineering and Computing

1. Introduction

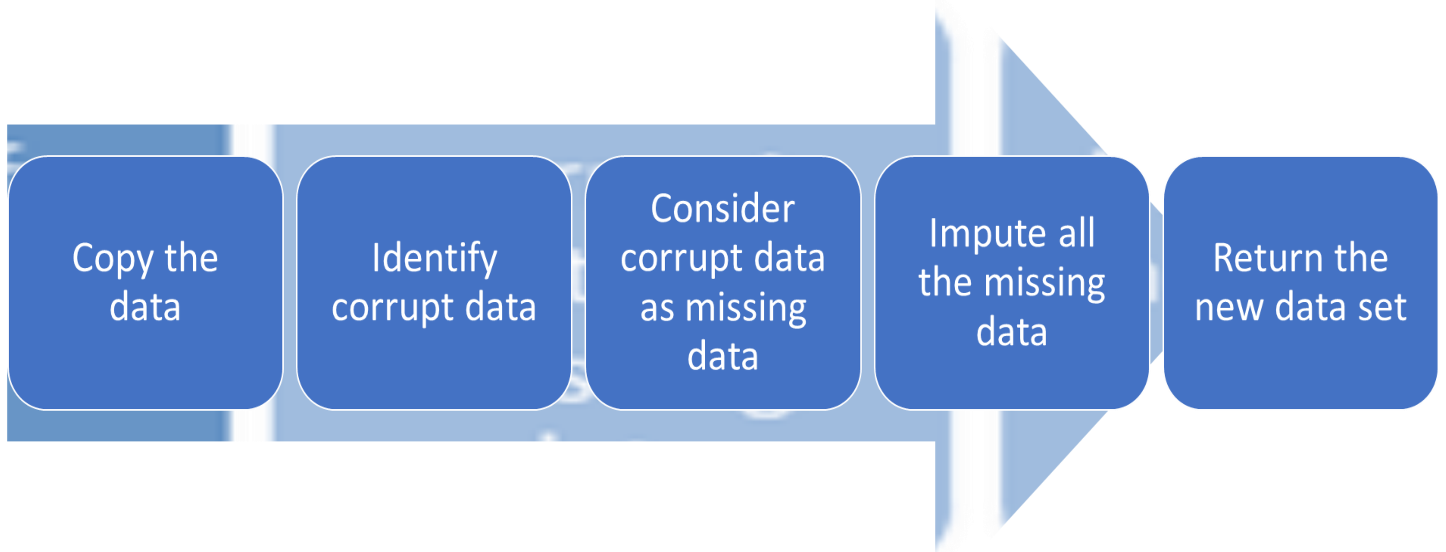
The rapid growth of wireless sensor networks and embedded devices as well as their application in the context of "smart homes" opens numerous opportunities for the development of new quality solutions, among others, in the field of eHealth. Research has begun on incorporating personalized health data collected through Internet of Things (IoMT) devices, such as smartwatches or fitness bracelets, into an electronic health record (EHR) that is part of the central health information systems of many EU countries. Advanced monitoring which IoT device offer, i.e., continuously track patient's vital signs or activities of daily living can positively influence the quality of healthcare they receive. Continuous real-time monitoring of persons health and well-being includes tracking vital signs via sensor-equipped, wearable electronic devices.

2. Problem Description

The first challenge is to bridge the differences and merge different sets of data, into a single comprehensible collection. Then, upon aggregating data it is necessary to conform it to the same standard, no matter the source. Finally, due to the sensitivity of such data, appropriate privacy guarantee is crucial. Privacy and winning patients' trust, as well as the implications of self-diagnosis without a physician's input remains one of the biggest obstacles to ensuring the success of eHealth solutions.

3. Methodology

The quality of the acquired personal health data would have to be adequate to be used in machine learning algorithm. In this case, the necessary process consists of data quality assessment and data cleaning process (shown below).



After incomplete or corrupt data is identified, an accurate missing value imputation (e.g., statistical pattern recognition, specifically, decision trees and regression) is crucial to increase the usability of data set for further use in a formal medical information system. Several data-driven models for data imputation were compared for several subjects and results show neural network and multiple linear regression performed best. As such, those algorithms are chosen for improvement.. Furthermore, the next key challenge of the research is to define a validation process to ensure the data complies with standards. This is done by:

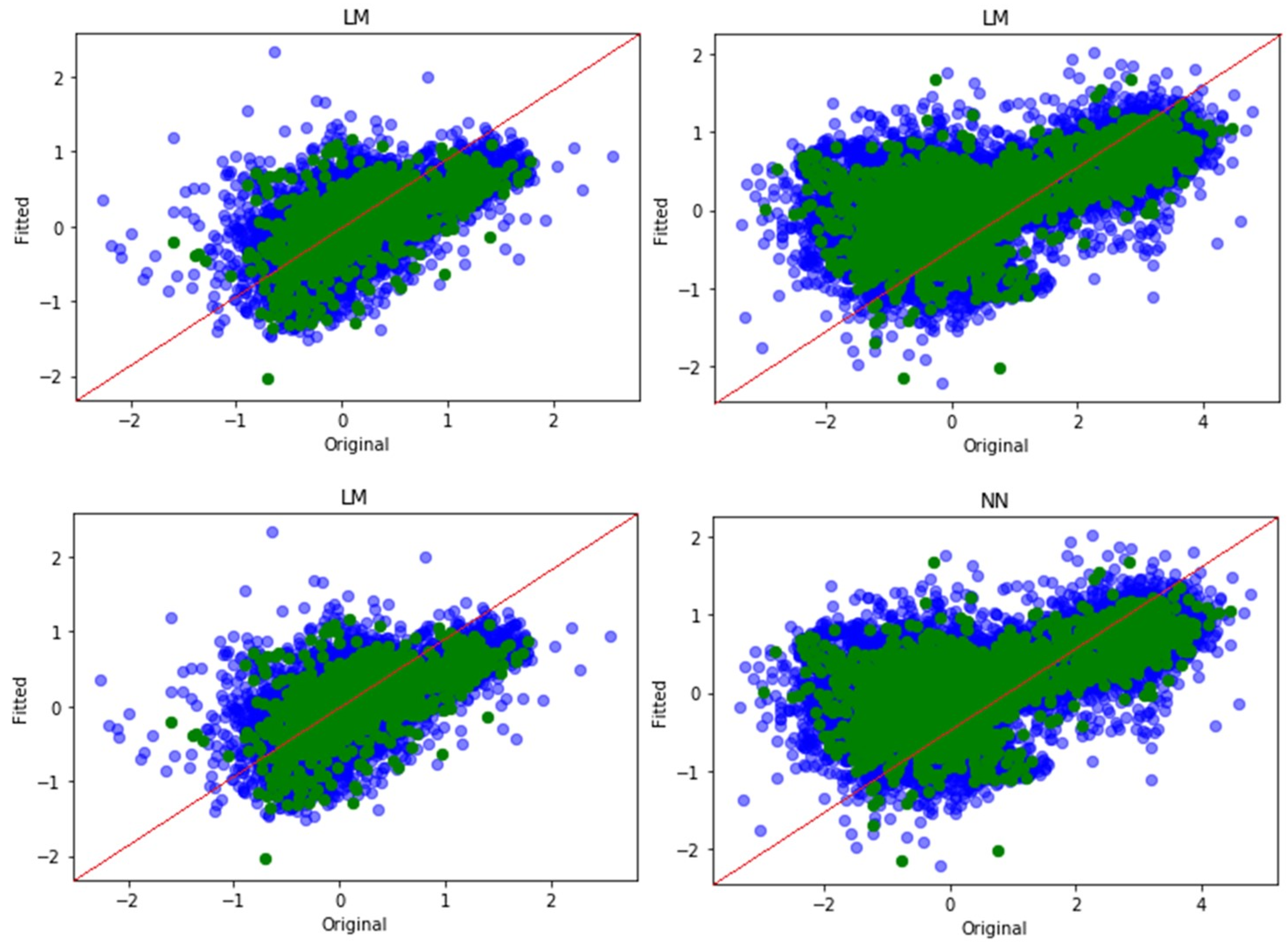
- Defining semantic constraints for healthcare datatypes
- Defining and modelling the validation process of the data collected

Interesting!
Neural networks gave best results for data imputation

Data validation process proposed uses structural schemas for syntax and semantic validation and enforces compliance to globally adopted medical standards, Health Level 7 (HL7).

4. Results

For data cleaning optimization, data was classified by intensity of activities performed and imputed part by part, yielding improvement in imputation results. Improved approach where the eHealth sensor data is first classified and then split in terms of physical activity before being imputed. By combining classification and regression algorithms, the proposed data-driven model has shown from 10 to 17 per cent better predictions.



Scatterplots for original (blue) and computed (green) values

	RMSE (left)	RRMSE (left)	Improved accuracy (%)	RMSE (right)	RRMSE (right)	Improved accuracy (%)
Multiple linear regression	0.18340	0.02728	14,29	0.05098	0.01231	9,88
Neural Network	0.17541	0.02610	17,21	0.05094	0.01230	10,0

The model focuses on EHR data integration by offering automated Schematron-based validation and is compliant with the leading industry standard, HL7.

Schematron for oxygen saturation

5. Conclusion

Improved approach where the e-Health sensor data is first classified and then split in terms of physical activity before being imputed resulted in better predictions. Compliance to regulations and standards is essential in order to use personal health data for personalized and preventive medicine. Model of data validation process was verified in a use-case study, using an existing dataset containing various relevant datatypes.

PhD Day Faculty of Electrical Engineering and Computing, 2 June 2022.