Advancements on IoT and AI applied to Pneumology

Enrico Cambiaso[a,*], Sara Narteni[a,c], Ilaria Baiardini[b], Fulvio Braido[b], Alessia Paglialonga[a], Maurizio Mongelli[a]

[a] *Cnr-Istituto di Elettronica, Ingegneria dell'Informazione e delle Telecomunicazioni (CNR-IEIIT), Corso F. M. Perrone 24, 16152, Genoa, Italy*
name.surname@ieiit.cnr.it
[b] *University of Genova, Respiratory Critical Care Unit and Sleep Breathing Disorders, Genova, Italy*
fulvio.braido@unige.it
[c] *Politecnico di Torino - DAUIN Department, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy*

Abstract

The objective of this work is the design of a technological platform for remote monitoring of patients with Chronic Obstructive Pulmonary Disease (COPD). The concept of the framework is a breakthrough in the state of medical, scientific and technological art, aimed at engaging patients in the treatment plan and supporting interaction with healthcare professionals. The proposed platform is able to support a new paradigm for the management of patients with COPD, by integrating clinical data and parameters monitored in daily life using Artificial Intelligence algorithms. Therefore, the doctor is provided with a dynamic picture of the disease and its impact on lifestyle and vice versa, and can thus plan more personalized diagnostics, therapeutics, and social interventions. This strategy allows for a more effective organization of access to outpatient care and therefore a reduction of emergencies and hospitalizations because exacerbations of the disease can be better prevented and monitored. Hence, it can result in improvements in patients' quality of life and lower costs for the healthcare system.

Keywords: Internet of Things, healthcare, machine learning, intelligible analytics, statistical validation, cyber-security

*Corresponding author

1. Introduction

In the healthcare context, Chronic Obstructive Pulmonary Disease (COPD) affects about 5-10% of the adult population, although prevalence as low as 0.2% and as high as 37% has been reported, depending on country, population, COPD diagnosis and classification methods [1]. COPD is an important cause of mortality, comorbidity and social impact. The remote and continuous monitoring of treatments, as well as the collection of vital parameters (such as heart rate, blood oxygenation, sleep and patient movement), are necessary to allow for immediate and punctual medical and social interventions, to reduce short- and long-term clinical consequences of the disease, and, in general terms, to improve the quality of life of patients, families and population [2, 3].

In this contribution, we introduce a novel concept for an innovative platform, called Pneulytics, to monitor and manage patients with COPD. Pneulytics can be adopted by healthcare systems, clinical centers and Pneumology departments and clinics. Platforms which includes patients' devices to retrieve data about patients' routine are not widely adopted yet, but it is assumed that, by 2025, approximately 75.44 billion of IoT devices will be online [4].

The platform is built upon collection and integration of IoT data and clinical data (available from inpatient and outpatient visits) and the analysis and aggregation of such data using artificial intelligence (AI) algorithms.

This new concept of patient monitoring requires the inclusion of new services and technologies that will need to be developed and optimized for efficacy and usability. Such technologies will contribute to a more effective patient management and therefore will support healthcare providers in defining strategies to offer care for chronic respiratory pathology. To this end, our platform not only involves a thorough patient monitoring via a set of gold-standard devices, but it also seeks to minimize the invasiveness of the data collection by including a more comfortable wearable device like a smartwatch.

The remaining of the paper is structured as follows: Section 2 reports an overview of the current healthcare scenario in terms of technologies and investments, giving a clear context to the objectives of the work. Section 3 reports the current IoT devices and machine learning approaches adopted in the healthcare context. Section 4 describes in detail the different components of the proposed framework and considerations regarding security and privacy. Section 5 details the adopted explainable AI (XAI) algorithm, i.e., the Logic Learning Machine (LLM), while Section 6 introduces a first experimental study involving the integration of several devices involved in Pneulytics platform and Section 7 discusses the effect of adversarial attacks against the proposed XAI solution. Then, Section 8 reports a preliminary study on smartwatch data. Finally, Section 9 concludes the paper and discusses possible next steps on the topic.

2. IoT in healthcare scenarios

Healthcare is a fertile ground for innovation through digital technology, which has the potential to make the health system sustainable. For example, Italian spending on health, even if population aging and increased life expectancy increase the need for care applications, is still decidedly lower than that of other countries [5].

In recent years, digital healthcare has continued its positive development trend. However, the expenditure commitment is not sufficient to bridge the overall delay in the digitization of the sector yet. Federsanità, the Italian institutional entity that organizes local health authorities and hospitals, points out that *the primary care management systems (which also include the Individual Health Card) are present in almost all of the General Medicine and Pediatricians' offices, but they are not integrated with hospital information systems, in which the spread of electronic medical records is still very limited. In both areas there is also a considerable heterogeneity of the present IT solutions.*

In particular, the technological paradigms enabling proximity and territorial assistance logic are strictly related to the use of IoT and wearables that enable remote monitoring and remote assistance services, as reported in [6, 7, 8]. Globally, we are witnessing a rapid increase in the use of connected and wearable devices both to improve the care of patients within hospitals and to speed up recovery times at the patient's home, through continuous remote monitoring of conditions and vital parameters or to delay the transition from independent living to assisted living [9]. The digital transformation is enabled by some elements, including data security, big data, and AI. In particular, the data security issue is closely related to the usage, transmission and sharing of health data. This topic is particularly felt in public and industrial sectors, because of recent recommendations introduced by the EU General Data Protection Regulation (GDPR) policies about personal data.

2.1. Digital transformation in Pneulytics

All these crucial elements are incorporated in the Pneulytics framework by combining IoT and wearable technologies with innovative data processing and AI algorithms, to extract sensitive, dynamic information. AI can be used to generate predictive models of the patients' health status. Instead, data collection from personal devices and sensors can be used to monitor the patients' vital parameters, generate alerts, and keep a detailed record of patients' history and state of health. The concept originates from the paradigm of the patient's centrality within the healthcare processes and the need to adopt innovative technologies to facilitate daily monitoring of health conditions.

The idea is to renovate the whole patients' journey experience by integrating conventional clinical procedures with the innovation brought by new technologies, which means combine daily monitoring data with data from an optimized hospital or extra-hospital healthcare. Specifically, relevant monitoring data in respiratory pathologies include vital parameters, physical activity and pharmacological compliance. Some clinical trials, although conducted on limited cases, have shown that tools for the administration of therapy integrated with ICT platforms are able to improve adherence to the treatment and reduce hospital admissions of patients suffering from COPD [10].

Clearly, personalized home therapy, constant monitoring of COPD evolution and prevention of possible exacerbations represent fundamental aspects in future clinical practice.

2.2. Aim and Scope

The main objective of the proposed framework is to create a hardware and software platform able to provide improvement of:

- treatment and assistance related to respiratory pathologies, supporting the definition of follow-up plans and promoting personal care;
- effectiveness of therapy as related to the administration of drugs through non-intrusive monitoring of the patient's behavior and adherence to therapy;
- clinical practice, making it possible to customize and optimize the staging of diseases and the areas of intervention over time.

The innovation of this platform can ease the workflow in hospital and medical contexts, especially considering that the resources allocated to public health are decreasing concerning a rapidly increasing demand. In the long term, the proposed technological solution has the potential to generate significant developments in terms of: i) new business opportunities related to infrastructures for remote monitoring, technologies and devices for advanced biomedical sensors; ii) new generation communication technologies, AI solutions for data processing and clinical decision support; and iii) improvement of the quality of care, medical treatment of COPD, and patient outcomes.

3. Survey of medical IoT platforms and machine learning approaches

3.1. Internet of Things

The state of the art of medical IoT platforms shows a variety of tools and interfaces, as briefly summarized below. For our aim, our focus is on open source approaches, highlighting progresses will lead to the proposed framework. Widespread use of open source health platforms and sensors has led to the development of simple, inexpensive and easy to use biometric devices, thus, these technologies impact not only from a medical point of view, but also in terms of business models for SMEs.

For instance, Bitalino[1] is a popular open source biomedical development platform that has a variety of biometric sensors. These sensors include an ElectroMyoGraphic sensor (EMG), a sensor for ElectroCardioGraphy (ECG), a LUX sensor (to monitor blood volume pulse data), an Electro-Dermal Activity (EDA), and an accelerometer sensor (for dynamic and biomechanical motion analysis). In addition, the platform offers an Atmega328 microcontroller for processing sensor readings and a Bluetooth module for wireless communication. The LUX sensor can be used together with a light source to monitor blood pulse data, while the accelerometer can be used in dynamic and biomechanical motion analysis. The heterogeneity of Bitalino sensors is a good reference for Pneulytics, even if we intend to specialize our analysis on a specific pathology.

Similar considerations apply to E-Health[2], an open source sensor platform

[1]More information are available at the following address: `http://www.bitalino.com`, accessed on November 2023.

[2]More information are available at the following address: `https://www.postscapes.com/open-source-e-health-sensor-platform`, accessed on November 2023.

that offers a wide range of features for detecting biological signals for open source hardware platforms. E-Health is one of the few, perhaps the only, IoT health platforms compatible with both the Arduino and Raspberry Pi architectures. E-Health focuses mainly on heart disease, creating many dedicated solutions e.g. ECG devices. Several open-source ECG sensors are available: they are not invasive and can be used comfortably at home. Their functioning is based on different mechanisms, for example, on photoplethysmography, which is frequently used in wearable devices, including fitness trackers.

Ticuro[3] is a closed system for the user and acts as a collector of a variety of commercial sensors/devices (to be purchased separately) as well as a tele-consultation platform. The ARM[4] and Kaa[5] architectures allow healthcare system integrators to establish connectivity between heterogeneous IoT devices and implement intelligent features in the devices themselves and the related software systems. Regarding the IoT protocols related to network connection (e.g., with MQTT and CoAP protocols), these architectures are of interest to Pneulytics, but still inadequate in terms of artificial intelligence engine.

An open architecture to developers and manufacturers of sensors and devices, as Mysignals[6], is particularly suited for building Pneulytics because of its flexibility.

3.2. Adopted IoT devices

Our research focuses on examining the adoption of smart health solutions to monitor various personal and environmental variables, particularly those related to quality of life and wellbeing. One key aspect we explore is Indoor Environmental Quality (IEQ), collecting different measurements to assess the indoor environment quality. To achieve this goal, we set up an intelligent monitoring system capable of observing, capturing, and processing environmental and body measurements. The integration of Internet

[3]More information are available at the following address: `https://www.reply.com/ticuro-reply`, accessed on November 2023.

[4]More information are available at the following address: `https://www.arm.com/glossary/medical-iot`, accessed on November 2023.

[5]More information are available at the following address: `https://www.kaaproject.org/healthcare`, accessed on November 2023.

[6]More information are available at the following address: `http://www.my-signals.com`, accessed on November 2023.

of Things sensors facilitates the seamless retrieval and exchange of data, thanks to the interconnected relationship between the sensors and a shared data storage platform.

Regarding the IoT sensors adopted for the platform, our emphasis has been on accessing the raw data captured by these sensors. Specifically, we have focused on two categories: wearable devices, which are connected and actively managed by the patients themselves, and environmental devices that are physically installed in the patients' homes.

Concerning wearable devices, the adoption of the H&S cloud platform enables us to access aggregated monitoring data and grant authorized access to end-users. Particularly, H&S offers services through its proprietary platform, HealthPlatform v3 - medical device CE IIA. Such platform is equipped with a data center certified ISO 27001 and ISO 13485. The data management activities adhere to GDPR regulations and comply with CE Medical Device 5/2020 standards.

Among the available devices, our choice includes (i) a dedicated smartphone with the proprietary app (Mhealth, certified IIA class) running on it, (ii) an electrocardiogram (ECG) also providing day/night movement monitoring, (iii) a pulse meter providing oximetry monitoring, (iv) a weight scale, and (v) a sphygmomanometer for blood pressure monitoring.

In our tests, we also included surveys for the patients, such as the COPD assessment test.

Instead, regarding environmental monitoring activities, for each environment (home, office, etc.), we inbluded in our setup (i) a central node receiving data from the other nodes, (ii) a physical device combining multiple sensors, (iii) a set of modules demanded to provide connectivity to analog equipment like windows or radiators, and, optionally, (iv) an outdoor weather station.

Particularly, while the central node is represented by a Raspberry Pi 4 Model B, following sensors types and models have been considered:

- temperature and humidity (Sparkfun, SI7021)
- atmospheric pressure (AZ Delivery, BMP180)
- air speed (Modern Device, Wind Sensor Rev. C)
- CO measurement (Sparkfun, MQ7)
- CO2 measurement (Sparkfun, CSS811)

- formaldehyde concentration (Seeedstudio, Grove HCHO)
- concentration of fine dust (Honeywell, HPMA115S0-XXX)
- redundancy (Bosch, BME680)
- weather station (PCE Italia, PCE FWS 20)

To conclude, regarding analog equipment monitoring, we adopted AZDelivery, ESP8266 plus ESP-01 and DHT22 plus AM2302.

With the sensors we've implemented, we can observe and process a variety of metrics related to both the patients' conditions and the surrounding environment. Moreover, through AI methodologies, the aggregation and processing of data not only allow for the integration of information from diverse components for a comprehensive analysis but also enable the identification of potential relationships among the data.

3.3. Smartwatch-based explainable data analytics

Nowadays, smartwatches are becoming very popular in smart health monitoring, being able to measure multiple indicators inherent to different domains, such as cardiovascular and respiratory health, physical activity sleep habits, all in scaled-down devices that can be comfortably wrist-worn by people, without being invasive nor requiring dedicated training [11]. Consequently, these wearables are an important data source for healthcare XAI applications, designing transparent models that can provide clinical decision-making support to users, either these are clinicians in making diagnosis, prognosis or planning therapies. The patients themselves may become more aware of their health status as well.

3.4. Artificial Intelligence

Noticeably, none of these platforms integrates AI solutions, whereas Pneulytics, by using AI on integrated patients' clinical data (e.g., spirometry, blood analysis) and sensor data, is able to support personalized healthcare.

The adoption of statistical methods in medical scenarios is widespread. In recent years, with the evolution of big data, the importance of AI in health scenario has increased. For example, [12] proposed to extract rules for pleural mesothelioma diagnosis. Malignant Pleural Mesothelioma (MPM) is a rare highly fatal tumor where the correct diagnosis of MPM is often hampered by the presence of atypical clinical symptoms: these may cause misdiagnosis with either other malignancies (especially adenocarcinomas) or benign inflammatory or infectious diseases (BD) causing pleurisies. Cytological Examination (CE) may allow to identify malignant cells, but sometimes a very high false negative proportion may be encountered due to the high prevalence of non-neoplastic cells. Moreover, in most cases a positive result from CE examination only does not allow to distinguish MPM from other malignancies [13]. Another interesting work is focused on the extraction of a simplified gene expression signature for neuroblastoma prognosis [14]. [15] instead proposed a convolutional neural network (CNN) algorithm in order to verify if deep learning could detect the COPD stage and predict Acute Respiratory Disease (ARD) events and mortality in smokers by using a training dataset composed by more or less 8000 patients. The results are interesting since the approach shows that CNN is able to identify and predict individuals with COPD. [16] implemented a decision tree forest classifier able to predict COPD based on symptoms by monitoring 16 patients for six months. Another approach is adopted in [17], where AI is used to analyze X-Ray dataset in order to recognize and locate the common disease patterns. In [18], authors implemented a recurrent neural networks (RNN) on 260K patients to predict diagnosis by performing a multilabel prediction. Artificial intelligence is also applied to predict the Parkinson, with interesting results [19].

4. Pneulytics framework

4.1. Context and approach

COPD is a common, preventable and treatable disease characterized by persistent respiratory symptoms (dyspnea, cough, expectoration) and airway obstruction due to lung damage induced for example by cigarette smoking and environmental pollutants. The course of COPD is generally progressive and is characterized by recurrent exacerbations and by the presence of concurrent conditions (e.g. cardiovascular pathology) that increase morbidity and mortality (estimated as the 3rd cause of death in 2020 by the WHO[7]).

COPD patients typically get benefit from topically administered drugs to reduce exacerbations and to relieve symptoms, reducing exercise intolerance

[7] More information is available at the following address: `https://www.who.int/healthinfo/global_burden_disease/projections/en/`, accessed on November 2023.

and increasing pulmonary function and life quality. There is still a clear discrepancy between patients' outcomes in clinical research and outcomes in real life due to poor adherence to treatments and wrong use of drug inhalers [20]. The number of critical errors in the use of inhalers is associated with increased risk of COPD exacerbations and the combination of poor adherence and misuse of inhalers may increase the risk of death up to three times [20, 21].

Technological improvements have recently led to the creation of new devices that allow remote monitoring [22, 23], patient engagement and remote interaction with the healthcare providers.

Smart inhalers are capable to record and digitize key aspects of care, such as drug intake and inhaler mode of use (inspiratory peak flow, duration of the inspiratory phase, inhaler's orientation). Thanks to these smart inhalers, doctors are able to acquire information on adherence to the treatment and on proper/improper use of the inhaler.

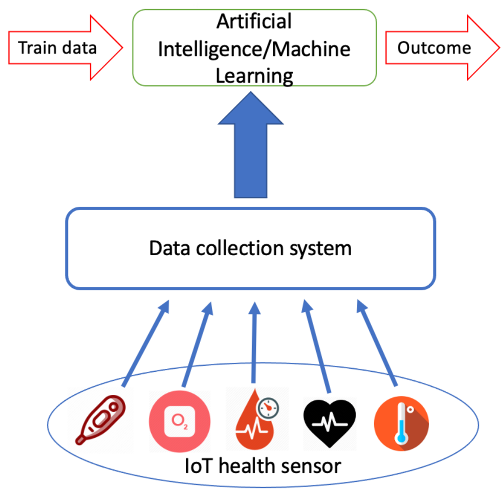


Figure 1: The concept of the Pneulytics platform

Unfortunately, the use of these devices is not widespread and currently this type of intervention is highly managed by a direct doctor-patient (or service center-patient) relationship, thus missing the opportunity to implement continuous monitoring, multivariate/multimodal clinical evaluation, remote monitoring, remote consultation on large patient cohorts. Moreover, the accessibility to the data recorded by the devices is perhaps one of the most important aspects regarding data integration.

The Pneulytics framework aims to contribute by addressing these issues and by increasing patients' involvement in the treatment plan and support interaction (including remote interaction) with the healthcare providers. The prototype hardware architecture for the collection and management of data, obtained with different IoT devices, is based on the available types of signals and communication protocols available, their size and, consequently, the processing capacity necessary for the correct application of the specific AI algorithms.

The proposed technological platform is composed of different technologies and algorithms, as the conceptual scheme in Fig. 1 illustrates.

Candidate patients will be equipped with wearable sensors that monitor relevant biomedical parameters both indoor and outdoor throughout their daily life. Additional sensors will be installed in the patients' home environment to detect relevant environmental parameters, such as air quality, humidity, temperature and pressure, and a sleep quality monitoring system. In the outdoor environment, contextual information related to the level and type of patient's activity will be collected by using consumer devices, such as smartphones or smartwatches.

Using ad-hoc AI algorithms, the platform will monitor functional and physiological parameters, such as blood oxygenation, heart rate, physical activity, sleep and lung function, as well as adherence to therapy through smart inhalers. The data collected will be also exploited to develop predictive models that will be useful to define follow-up plans and interventions.

4.2. Personal health records

The system will integrate all the data from sensors and personal devices with the medical history and clinical data, as available from the hospital databases and will be able, in future developments, to interact directly with existing systems of Personal Health Records and with information extracted from biomedical images recorded in the hospital information systems and Picture Archiving and Communication System (PACS).

Two aspects need to be addressed here: on one hand, a management and decision support system has to be designed, where all the internal and external heterogeneous data are organized and accessed. On the other hand, analysis techniques from computer vision and graphics (some of them based on AI approaches) may be applied in order to characterize and measure specific areas of biomedical images and 3D reconstructed models that are useful for the diagnosis, monitoring and follow-up of COPD. Annotation methods may also be used to code such extracted information, index 2D and 3D resources in compliance with semantic web paradigms, and finally integrated into Pneulytics platform. A similar approach has been applied to musculoskeletal pathologies in [24, 25].

4.3. Cyber-security and privacy considerations

IoT devices provide the ability to automate and enhance people's daily lives. Being a pervasive technology embedded in critical locations, the IoT phenomenon is often coupled with privacy issues: as such sensors often process sensitive information, security becomes a very critical topic. In particular, if IoT sensors are adopted to monitor and control the health parameters of patients, data security becomes extremely critical due to potential exposure to privacy leaks. For the scope of the proposed work, patients' health parameters are managed and manipulated through IT systems. For this reason, it is crucial to guarantee appropriate security and privacy, especially because of potential cyber-attacks able to steal, retrieve or infer clinical data.

In order to guarantee user data privacy, different data anonymization techniques are available, also considering ethical aspects of sensitive data management. Indeed, in literature, several anonymization algorithms are found, while some of them exploit different techniques that make the data difficult to de-anonymize [26, 27, 28, 29]. Instead, in the context of data re-identification and de-anonymization, machine learning methods can be adopted. From one side, a well-known and "classic" (unsupervised) clustering approach to data privacy is k-anonymity [30, 31, 32, 33]. In this case, the k-mean algorithm can be used for different applications: [34] adopts it to de-anonymize and extract geo-localization data from mobility traces, while [35] makes use of the k-mean to extract potentially sensitive information from social networks. Similarly, [36] adopts the k-mean to profile Facebook users, analyzing the interaction of their account, in terms of reactions, likes, or other social interactions. [37] makes instead use of the k-mean algorithm to preserve privacy when datasets are composed of different attributes, while [38] proposes a variant of the k-means algorithm to preserve the privacy of information by using as input encoded data. [39] also extends the k-means, by proposing M-Shuffle, a novel algorithm, based on k-means, to avoid information de-anonymization. By considering the same approach, a clustering approach based on k-means could be adopted to theorize a privacy breaking attack, aimed to reveal potentially sensitive information from anonymized data.

From another side, a different and more advanced (unsupervised) method could be inherited from the topic of the neural network, in order to conceptualize a novel attack against data privacy. Particularly, Restricted Boltzmann Machines (RBM) could potentially be used for data breaking purposes. Even in this case, RBM is nowadays adopted for different scopes: for instance, for simulations [40], to identify multivariate geochemical anomalies [41], or to categorize users [42]. In addition, they could be adopted similarly to its usage for user categorization: while known applications focus on the categorizations of users, for instance for marketing purposes (e.g. if a user U1 bought A, and previous users also bought B, hence, U1 may also be interested to purchase B), it is potentially possible to adopt RBM to categorize users due to their belonging of categories including other (not-anonymized) users. This is accomplished in RBM by relating each user to the belonging of one or more hidden features.

Instead, regarding security of Internet of Things systems, devices may communicate through standard networks, such as Wi-Fi or ethernet, or build a dedicated network to communicate with other sensors, called Wireless Sensor Network (WSN). In this regard, a real standard is not commonly adopted yet [43]. Currently, there are different protocols providing communication between sensors: some of them are based on pre-existing protocols (Wi-Fi, 6LowPan, MQTT or LoRa), while others provide the creation of new ad-hoc infrastructures (ZigBee, Z-Wave). Although different IoT protocols may be adopted, IoT devices are often exposed to security attacks since the data exchanged in this context are sensitive. Being exchanged information extremely sensitive due to the nature of devices and networks, the security about IoT devices and networks should be investigated in order to identify possible vulnerabilities and to protect the IoT context from them.

In order to protect IoT devices and network, well-known [44] and innovative [45, 46] attacks against IoT ad-hoc communication protocol are investigated to protect sensitive information from malicious purposes. Also, [47] considers hardware and software limitations of IoT systems, by creating a taxonomy of weaknesses. Instead, [48] analyzes the security of IoT networks by identifying crucial aspects related to common vulnerabilities, while [49] focuses on the security challenges to be addressed in the IoT field, also proposing protection solutions. Similarly, [50] focuses on security issues on environments such as healthcare, smart home or vehicles management.

In the context of this work, security and privacy aspects affecting IoT devices and networks need to be analyzed in detail to avoid possible loss of sensitive information and to ensure a secure exchange of information between the devices and the platform developed.

5. The adopted explainable AI approach

The data collected from the mentioned set of sensors is analysed via the lens of *explainable Artificial Intelligence (XAI)*. This term refers to a broad category of techniques aimed at providing intelligible interpretations to machine learning-based decisions [51], thus allowing anyone to enter their logic and increase trust in the knowledge inferred: one of the main XAI categorizations distinguishes between post-hoc XAI, where interpretations are provided to a previously trained black-box model, and transparent-by-design XAI, where the model making predictions is natively explainable. In this work, our focus is on the latter group, and in particular on rule-based models [52].

The setting considered here is that of a supervised machine learning classification task. The dataset is represented as the set $\mathcal{X} \times \mathcal{Y} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ where $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}$ is the set of input measurements for sample i and $y_i \in \mathcal{Y} = 0, \ldots, K-1$ is a label associated to it. The objective is to learn the best function $f : \mathcal{X} \longrightarrow \mathcal{Y}$ able to separate the input points according to their labels. In our study, the derivation of f is made through a specific rule-based model, called Logic Learning Machine (LLM) [53], which is the fast implementation of Switching Neural Networks.

In the LLM algorithm, the classification function f is learnt through three steps, as described in [53]. In the first phase (*discretization* and *latticization*) each input is converted into a string of binary data in a proper Boolean lattice, using the inverse only-one code binarization. All the generated strings are then concatenated into a single large string per each sample. In the second phase (*shadow clustering*) a set of binary structures, namely the *implicants*, are individuated, allowing the identification of groups of samples associated to a specific class. During the third phase (*rule generation*), all implicants are converted back to the original feature space, forming a collection of conditions, and eventually are combined into a set of if-then rules.

The LLM classifier is thus described by a set of m intelligible rules $r_k, k = 1, \ldots, m$, of the type **if** (*premise*) **then** (*consequence*), where (*premise*) is the logical conjunction (AND, $\wedge$) of d_k conditions c_{l_k}, with $l_k = 1_k, \ldots, d_k$, and (*consequence*) provides the class label $\hat{y}$ associated to the rule.

A condition c_{l_k} of rule r_k can have one of the following forms:

1. $x_{\pi(l)} > \lambda$

2. $x_{\pi(l)} \leq \mu$

3. $\lambda < x_{\pi(l)} \leq \mu$

being λ, $\mu \in \mathcal{X}$ and $\pi : \mathbb{N} \longrightarrow \mathbb{N}$ denotes the permutation of the indexes of feature vector $\mathbf{x}$ that maps rule l_k-th condition with the corresponding feature component.

For each rule generated by the model, a confusion matrix can be computed, showing true and false positives, $TP(r_k)$ and $FP(r_k)$, defined as the number of examples $(\mathbf{x}_i, y_i)$ which satisfy all the conditions in rule r_k with $\hat{y} = y_i$ and $\hat{y} \neq y_i$, respectively, and true and false negatives, $TN(r_k)$ and $FN(r_k)$, being the number of examples $(\mathbf{x}_i, y_i)$ which do not satisfy at least one condition in rule r_k, with $\hat{y} \neq y_i$ and $\hat{y} = y_i$, respectively. Useful quantities can be derived from the confusion matrix, such as the covering $C(r_k) = \frac{TP(r_k)}{TP(r_k)+FN(r_k)}$ and the error $E(r_k) = \frac{FP(r_k)}{FP(r_k)+TN(r_k)}$. The covering may also be adopted as a measures of relevance for a rule r_k; as a matter of fact, the larger is the covering, the higher is the generality and the correctness of the corresponding rule.

5.1. Feature and value ranking

Some preliminary results, reported in [54], have shown that the LLM is more performing than most of the known learning techniques of the same kind. Moreover, the computational complexity of the method is kept low thanks to the adoption of proper greedy procedures. Therefore the LLM model may be adopted also in the analysis of large datasets (i.e. having many inputs and/or examples). Notice that the LLM approach presents further interesting features such as the possibility of dealing with categorical inputs and the determination of the relevance of each variable. This last property allows the identification and elimination of redundant attributes.

Feature ranking (FR) deals with ranking the input variables based on their influence in determining the model's prediction, as calculated by an importance measure. The starting point is to compute a relevance value $R(c_{l_k})$ for a single condition, by measuring the difference in the rule error by including and excluding that condition, i.e., it holds that $R(c_{l_k}) = (E(r'_k) – E(r_k))C(r_k)$, where r'_k denotes the rule without condition c_{l_k}. As previously stated, each condition c_{l_k} refers to a specific input variable $x_{\pi(l)}$ and is satisfied by some values $\nu_{x_{\pi(l)}} \in \mathcal{X}$. The importance $R_\nu(\nu_{x_{\pi(l)}})$ of these values is given by:

$$R_\nu(\nu_{x_{\pi(l)}}) = 1 - \prod_k \left(1 - R\left(c_{l_k}\right)\right) \tag{1}$$

where the product is over rules r_k that include a condition c_{l_k} verified when $x_{\pi(l)} = \nu_{x_{\pi(l)}}$. Taking values in $[0, 1]$, it can be thought as the probability that value $\nu_{x_{\pi(l)}}$ occurs to predict $\hat{y}$. A ranking in descending order of R_ν for all possible intervals $\nu_{x_{\pi(\cdot)}}$ is referred to as *Value Ranking (VR)*.

Also, by aggregating the relevances for all different intervals $\nu_{x_{\pi(\cdot)}}$ of the variable of interest, an overall measure of its importance can be computed. Finally, the descending ordering of these importance metrics generates the feature ranking.

5.2. Classification scoring

In the inference phase, when applied to a generic input $\mathbf{x}$, the LLM model computes a score for each output class. Let us recall that any condition c_{l_k} of a rule defines a domain D_{l_k} in the input space, corresponding to an interval for feature $x_{\pi(l)}$. Let us consider the set of rules verified by $\mathbf{x}$ and predicting a label y, i.e., $\mathcal{R}^y_{\mathbf{x}}$=$\{r_k | x_{\pi(l)} \in D_{l_k}$ for each l_k and $\hat{y} = y\}$. A score for y is then defined as:

$$w_y = \frac{\sum_{r_k \in \mathcal{R}^y_{\mathbf{x}}} C(r_k)(1 - E(r_k))}{\sum_{r_k \in \mathcal{R}^y} C(r_k)(1 - E(r_k))}, \tag{2}$$

where $\mathcal{R}^y$ is the set of all rules generated for class y. A label is thus assigned to $\mathbf{x}$ by solving the following problem:

$$\hat{y} = arg \max_y w_y \tag{3}$$

6. Multi-Sensor Application

6.1. Previous results

In our previous work on the topic [6], we have shown the usage of the LLM as a prediction tool for following therapy in respiratory diseases. Patients with COPD are subjected to a monitoring period, using an inhalation tracker. The data obtained include the daily measurements taken by the inhalation tracker and patients' characteristics. The CAT[8] has been taken as a reference.

[8]More information are available in the following address: `https://www.mdcalc.com/copd-assessment-test-cat`, access on November 2023.

It is a self-report questionnaire that assesses the impact of COPD on health status. The higher the CAT score, the higher the perceived impact of the disease. Observing the LLM rules, we provided inference about adherence to treatment over time in terms of CAT.

6.2. Data Collection

Here, we show results with respect to monitoring through the devices outlined in subsection 3.2. Moreover, we show how adversarial machine learning may be used to provide fake indications that are still deemed plausible and suitable for further analysis by the clinical experts. Our first analysis regards the original dataset, without any adversarial machine learning corruption, and aims at providing the *baseline* rules generated by the LLM model.

The following quantities were daily collected for two consecutive months, and structured in a database for further LLM training[9].

`oxygen`, `body temperature`, `heart rate` (from oximeter), `heart rate master` (from sphygmomanometer), `weight`, `Body Mass Index`, `FEV1`, `PEF`, `MAP`, `diastolic blood pressure` , `systolic blood pressure`

In detail, `oxygen` is the blood oxygen saturation (i.e., the SpO2). Heart rate is measured via both the oximeter and the sphygmomanometer: the feature corresponding to the latter device is denoted by the suffix 'master' (i.e., heart rate master). The Forced Expiratory Volume in 1 second (`FEV1`) is the amount of air (in liters) that can be exhaled in the first second during forced exhalation after maximal inspiration. MAP, or mean arterial pressure, is defined as the average blood pressure in a patient's arteries during one cardiac cycle. It is considered a better indicator of perfusion to vital organs than systolic blood pressure.

Peak Expiratory Flow (`PEF`) is the maximum flow (or velocity) that can be achieved when performing a forced exhalation that is initiated after a full inspiration, measured in liters per minute or liters per second. This variable is used to set the target of the classification problem. Specifically, the dataset was labelled with *'low'* for PEF $\leq$ 400L/min or *'high'* for PEF $>$ 400 L/min. The threshold of 400 was determined under suggestion of the clinical expert.

[9]The Rulex platform has been used, `http://www.rulex.ai`

The dataset here described is available at `https://github.com/saranrt95/Medical_IoT_ML`, along with smartwatch data presented later in Section 8.

6.3. Statistical validation

Table 1: Contingency frequencies matrix between the output variable (classification) and a rule R

Contingency Matrix	**Output**	
	output	**output**
Rule R	$y\|R$	$\neg y\|R$
not Rule R	$y\|\neg R$	$\neg y\|\neg R$

We use the Fisher's Exact Test (FET) to test the statistical significance of rules obtained. The FET, indeed, is more accurate than other test of independence when the expected numbers are small, and it can be adopted to overcome the small sample size problem. The FET, in general, examines the significance of the association between two kinds of classification. In our case, we compare the distribution of a rule R, and its complementary $\neg R$, in the output classes $y, \neg y$, obtaining a contingency matrix, as in Table 1 above. Then, considering the independence between the distribution of the Output and the distribution of the rule R as null hypothesis, we compute the p-value. If the p-value is greater than 0.05, we accept the null hypothesis, otherwise, we prove the significance of rule R in detecting the output classes.

6.4. Baseline

The following rules are inferred by the LLM and validated by the FET test:

if $((heart\ rate < 74) \wedge (diastolic\ pressure > 67))$ **then** high (C=43%) (E=4.5%)
if $((FEV1 < 2.23))$ **then** low (C=41%) (E=4.7%),

where C and E denote their covering and error expressed in percentage.

Both rules provide clear indications about the status of the breath (through the PEF classification): one may argue that the knowledge extracted from them is trivial, since it puts in relation a good or bad breath performance with blood pressure and heart rate, which may be expected even by non experts. However, we highlight the following elements about the role of XAI. First of all, its adoption reveals useful to determine which of the several devices the monitoring path should focus on, by analyzing which variables more frequently occur within the rules. Moreover, it is difficult even for an expert to find the exact thresholds describing the output classes, jointly with all the input variables involved. In this perspective, XAI acts as an artificial predictor, namely, the rules map the measurements into the output class at the end of the observation period. Once the rules are available, they may be used by the medical staff at any time as predictors of the quality treatment in the near future and may drive proper decisions, such as contact the patient at home when the measurements lie in the outlined ranges of bad treatment.

7. Adversarial machine learning

In this section, we introduce further analysis with respect to adversarial machine learning (AML). Although known to experts in the sector, in recent years it had an exponential growth due to the continuous development of new machine learning applications in various sectors. As for many other fields, in the healthcare sector it plays a key role, for this reason we decided to consider possible AML attacks within Pneulytics project framework.

A very important aspect in the healthcare context is data security. An increasingly present thread in the security landscape is linked to attacks on the machine learning algorithm. Specifically, these attacks are called adversarial machine learning [55] where the aim of adversarial machine learning is to fool models by supplying deceptive input to cause a malfunction in a machine learning model. The problem is motivated by the fact that machine learning techniques were not originally designed to compete with adaptive and intelligent adversaries; therefore, in principle, the security of the entire system could be compromised by exploiting specific vulnerabilities of these algorithms, through a careful manipulation of the data that are supplied. A classic example of an adversarial machine learning attack is related to the context of image classification. The algorithm learns to classify images during the training phase based on the dataset used. An attacker could insert noise into the image to be classified (invisible to the human eye) to make the machine learning algorithm classify the image incorrectly. Several examples of image adversarial machine learning were studied and presented [56, 57, 58].

As previously described, in the healthcare sector, data is of primary importance as it can be used to predict a disease or manage remotely treatments and cures by using machine learning algorithms. An attack of this nature could lead to serious consequences as the manipulation by an attacker of the data could lead to an incorrect classification of the disease or to the administration of a drug when necessary. In the extreme case, the identification of a fatal disease only in an advanced state when the medical treatments are no longer effective.

For these reasons we decided to investigate the adversarial machine learning in our project in order to verify if the machine learning system developed in Pneulytics is able to resist the variations of the dataset and to equally correctly classify the disease. In order to achieve this results, we implemented a simple adversarial machine learning algorithm on the dataset used in Section 6.2 and compared the results. The adversarial machine learning attack works by adding a Gaussian noise $\mathcal{Z} \sim \mathcal{N}(0, 0.8)$ to the original data. We selected this small range since the variation of the dataset must be invisible (or not simply identifiable) by the statistician analyzing the data. Subsequently, the data were processed again with the techniques used in Section 5, to verify how the algorithm behaves with the dataset affected by the adversarial attack. The following rules are obtained.

if $((heart\ rate\ master \in [62, 75]) \wedge (diastolic\ pressure > 67))$ **then** high (C=77%) (E=5%)

if $((oxygen < 96) \wedge (heart\ rate \in [66, 93]) \wedge ((FEV1 \in [1.11, 2.39])))$ **then** low (C=43%) (E=4%)

if $((systolic\ pressure \in [98, 103]))$ **then** low (C=38%) (E=4%)

They are similar to the baseline in terms of features and reference intervals, in a such a way that the medical staff does not recognize the presence of an adversarial inside the machine learning engine. A subtle question naturally arises: how to set the adversarial setting in order to move the medical staff to a wrong diagnosis? This would lead to further investigation with clinicians. Moreover, how to prevent an attack like this? Is it possible to circumvent the behaviour of legitimate rules in order to understand the presence of an adversarial attack? All of these issues are argument of our ongoing research.

8. Smartwatch example

A Fitbit Versa 3 smartwatch was considered, which is equipped with many sensors, e.g., among others, GPS, red/infrared sensors for SpO2 registration, and movement sensors (3-axis accelerometer, gyroscope), which make it a very versatile instrument. Data are transmitted via Bluetooth Low Energy (BLE) technology to the dedicated smartphone app associated to the device and are stored into the Fitbit user account, remaining available for data query.

8.1. Dataset collection and clinical problem definition

From the Fitbit smartwatch, we collected one month of measurements from a subject who was following a pharmaceutical therapy for COPD. Specifically, these quantities were:

1. Heart-related daily measurements:
    - Average heart rate `HRmean` and standard deviation `HRstd`.
    - amount of time spent in different heart-rate zones, which are defined as percentage ranges of the maximum heart rate estimated for the subject (maxHR hereon): `below zone`, i.e., $< 50\%$ maxHR; `fat burn zone` (50-69% maxHR); `cardio zone` (70-84% maxHR); `peak zone` ($> 85\%$).
    - heart rate variability (HRV) during sleep: `rmssd-HRV`, `NonREM-HR` and `entropy-HRV` are values aiming at describing different aspects of the beat-to-beat intervals variations.
2. Respiratory measures during sleep:
    - Blood oxygen saturation, `average SpO2`, and its lower and upper bounds.
    - Infrared to red ratio average `IRtoRedMean` and its standard deviation `IRtoRedMeanStd`, which reflect the estimated oxygen variation via pulse oximetry principles [59].
    - Average `respiratory rate`.
3. Daily minutes of activity performed at different levels of intensity: `sedentary minutes`, `lightly active minutes`, `moderately active minutes`, `very active minutes`.

These quantities were then used to build a comprehensive dataset for a period of 32 days of observation.

The aim of our preliminary study was to assess which of the considered measurements exhibited the main variations when the involved patient was treated with 1 puff/day of the COPD drug or 2 puffs/day. To this end, we decided to disregard the activity-related measurements from the input variables, since they were related to the subject's behavior and not to its physiological status. To make the dataset suitable for supervised machine learning-based analysis, we labelled the samples with labels resembling the `therapy` followed by the patient, i.e., assigning *'1'* for 1 puff/day or *'2'* for 2 puffs/day. Lastly, we selected a rule-based binary classifier to predict the dose of therapy, specifically a Logic Learning Machine model. Due to the limited size of the dataset, statistical validation of rules was carried out through Fisher Exact Test. In addition, several random shuffles of the data were performed and a separate LLM classifier was trained on them.

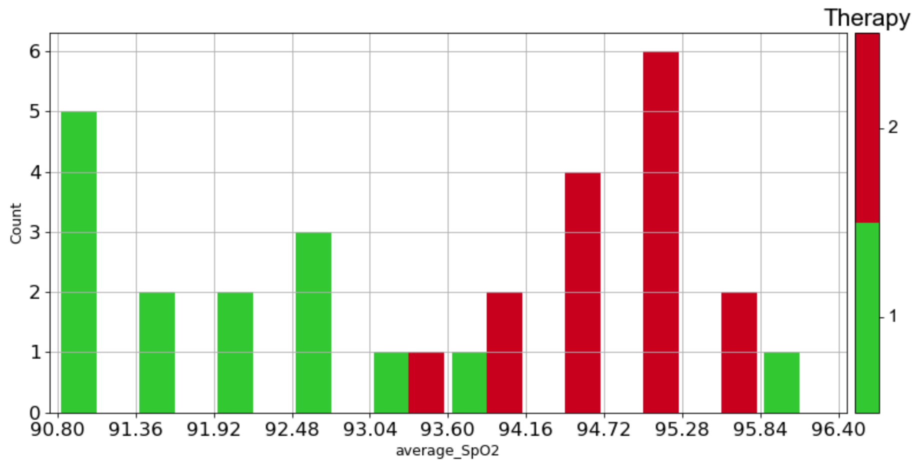
8.2. Preliminary results

As a first step in the analysis, features distributions were explored to assess the feasibility of adopting a rule-based model to discriminate the classes. As an example, we show the histograms related to the `average SpO2` and `meanHR` quantities (Fig. 2).

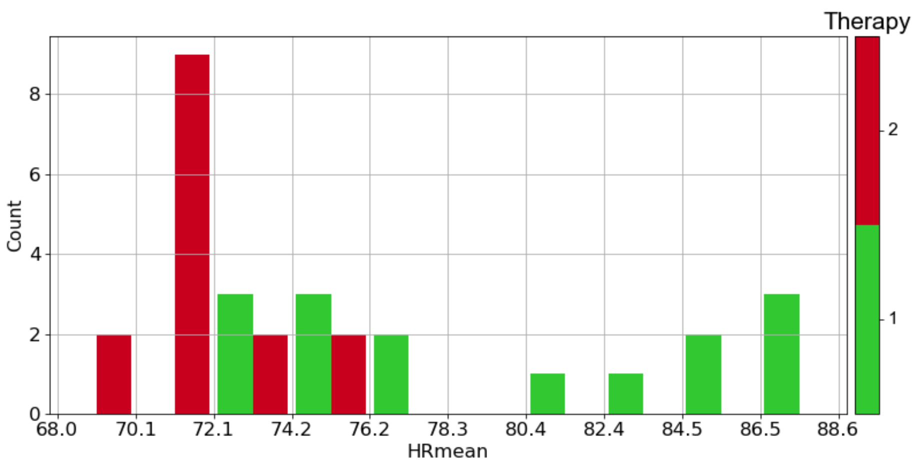
Table 2: LLM rules obtained on 5 random shuffles of the Fitbit dataset, after their statistical validation via FET test, along with their covering (C [%]) and error (r [%])

Shuffle	**Validated Rules**	C [%]	E [%]
1	**if** $averageSpO2 \leq 93.25$ **then** therapy = 1	77	0
	if $averageSpO2 > 92.80 \wedge 11.064 < HRstd \leq 15.040$ **then** therapy = 2	89	0
2	**if** $HRmean > 71.252 \wedge IRtoRedMean \leq 0.404$ **then** therapy = 1	100	0
	if $89.5 < LowerBoundSpO2 \leq 94.7 \wedge HRmean \leq 74.696$ **then** therapy = 2	90	0
3	**if** $rmssd\text{-}HRV \leq 46.065 \wedge HRmean > 71.83$ **then** therapy = 1	91	0
	if $averageSpO2 > 93.1 \wedge 11.249 \leq HRstd \leq 17.383$ **then** therapy = 2	91	0
4	**if** $HRmean > 71.570 \wedge IRtoRedMean \leq 0.404$ **then** therapy = 1	100	0
	if $HRmean \leq 72.551$ **then** therapy = 2	92	0
5	**if** $averageSpO2 \leq 93.25$ **then** therapy = 1	100	0
	if $averageSpO2 > 93.25$ **then** therapy = 2	91	0

The figures show that in both cases the two classes are pretty well distinguishable: also, as expected, a higher dose of therapy improves the average SpO2 and lowers the heart rate. However, specific cut-off values on these



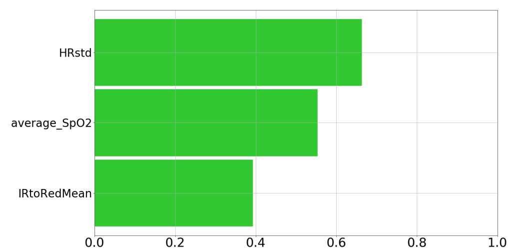
(a) Average SpO2



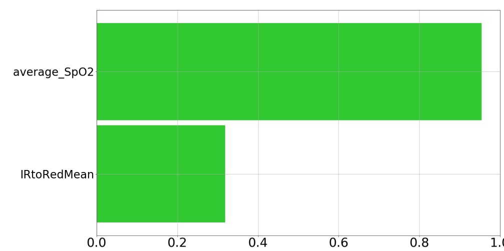
(b) Average Heart Rate

Figure 2: Class distributions for two exemplary features

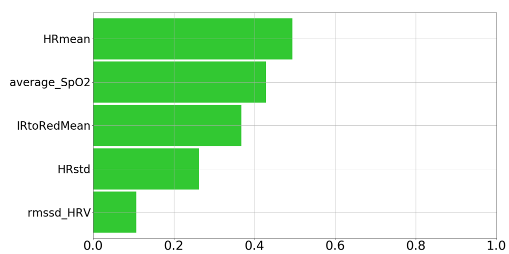
measurements and further knowledge were discovered through the usage of XAI. The LLM model was then trained on 5 random shuffles of the dataset, and, after the FET statistical validation, 2 rules were generated for each shuffle, as reported in Table 2. The average model accuracy over the shuffles reached the 74%. The feature rankings reported below give an idea about the most important variable for decision making as well as how the ranking may be sensitive to data variations (shuffles). Future research include larger time horizon of patient monitoring to achieve stable assessment of model suggestions.



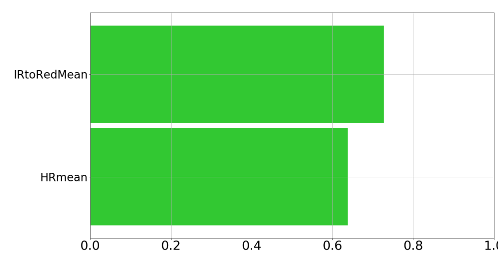
(a) Shuffle 1



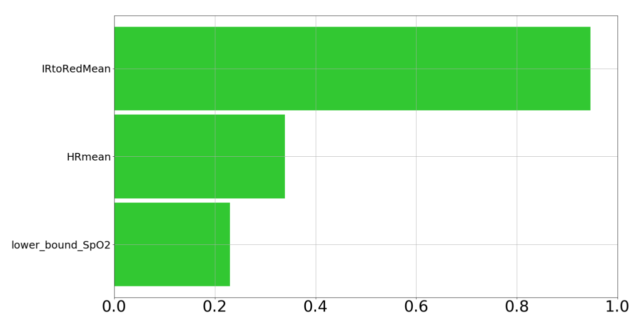
(b) Shuffle 2



(c) Shuffle 3



(d) Shuffle 4



(e) Shuffle 5

Figure 3: LLM feature rankings for the 5 data shuffles

Despite the ranking sensitivity to data variations, as well as the 15 features gien in input to the LLM, the model generated short rules, with no more than 2 conditions each. This would improve their interpretability. Overall, a few factors emerged as the most influent in predicting the therapy, namely `averageSpO2`, `HRmean` and `IRtoRedMean`. Indeed, these attributes were present in the feature ranking for at least 3 out of 5 shuffles (Fig. 3). Overall, the XAI approach gives interesting insight into the problem, thus providing promising indications for future research.

9. Conclusions and future work

In this paper, we introduce Pneulytics, a novel framework designed to use innovative technologies such as the Internet of Things (IoT) and Artificial Intelligence (AI), combined with security and privacy aspects, to collect and process heterogeneous data from environmental and wearable sensors to monitor patients' health. We believe that the approach is feasible and can be used to monitor patients' outcomes and adherence to treatment and to better understand the factors that influence individual outcomes.

Preliminary tests in [6] and in this paper show that the combination of clinical data and IoT allows to monitor the therapy and to understand the factors that influence it.

Future works may be focused on the operative development of the platform, while exploring new directions, spanning from extending the sensing scenario to the environment (e.g., how air quality may impact the treatment?) to the joint study of AI and privacy. The last relevant topic is General Data Protection Regulation (GDPR) EU regulation[10] and involves brand new approaches, as accurate statistical models of correlation may accidentally reveal more information about the patients than intended[11].

Acknowledgement

The work was supported by: Compagnia di San Paolo, scientific research call 2019 (Bando 2019-2020 per progetti di ricerca scientifica presentati da enti genovesi): project "Advances in pneumology via ICT and data analytics" (PNEULYTICS); Bando incentivazione della progettazione europea 2021 - Mission "Promoting Competitiveness" (DR nr. 3386 of 26/07/2021); LoLiPoP-IoT project, nr. 101112286 funded by European Commission.

[10]More information is available on the ICO portal, access in November 2023.

[11]More information is available at: `https://github.com/frankmcsherry/blog/blob/master/posts/2016-06-14.md`, access in November 2023.

References

[1] C. E. Rycroft, A. Heyes, L. Lanza, K. Becker, Epidemiology of chronic obstructive pulmonary disease: a literature review, International journal of chronic obstructive pulmonary disease 7 (2012) 457.

[2] J. F. van Boven, F. Lavorini, P. R. Dekhuijzen, F. Blasi, D. B. Price, G. Viegi, Urging europe to put non-adherence to inhaled respiratory medication higher on the policy agenda: a report from the first european congress on adherence to therapy, 2017.

[3] J. Virchow, G. Crompton, R. Dal Negro, S. Pedersen, A. Magnan, J. Seidenberg, P. J. Barnes, Importance of inhaler devices in the management of airway disease, Respiratory medicine 102 (2008) 10–19.

[4] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of things (iot): A vision, architectural elements, and future directions, Future generation computer systems 29 (2013) 1645–1660.

[5] M. Francese, M. Romanelli, Is there room for containing healthcare costs? an analysis of regional spending differentials in italy, The European Journal of Health Economics 15 (2014) 117–132.

[6] M. Mongelli, V. Orani, E. Cambiaso, I. Vaccari, A. Paglialonga, F. Braido, C. E. Catalano, Challenges and opportunities of iot and ai in pneumology, in: 2020 23rd Euromicro Conference on Digital System Design (DSD), 2020, pp. 285–292. doi:10.1109/DSD51259.2020.00054.

[7] Y. Wang, M. Nekovee, E. J. Khatib, R. Barco, Machine learning/ai as iot enablers, Wireless Networks and Industrial IoT: Applications, Challenges and Enablers (2021) 207–223.

[8] M. Degerli, S. Ozkan Yildirim, Enablers for iot regarding wearable medical devices to support healthy living: The five facets, in: IoT in Healthcare and Ambient Assisted Living, Springer, 2021, pp. 201–222.

[9] F. Veronese, H. Saidinejad, S. Comai, F. Salice, Elderly Monitoring and AAL for Independent Living at Home: Human Needs, Technological Issues, and Dependability, IGI Global, 2016, p. 154–181. URL: http://dx.doi.org/10.4018/978-1-4666-9530-6.ch007. doi:10.4018/978-1-4666-9530-6.ch007.

[10] F. Braido, P. Santus, A. G. Corsico, F. Di Marco, G. Melioli, N. Scichilone, P. Solidoro, Chronic obstructive lung disease "expert system": validation of a predictive tool for assisting diagnosis, International journal of chronic obstructive pulmonary disease 13 (2018) 1747.

[11] M. Masoumian Hosseini, S. T. Masoumian Hosseini, K. Qayumi, S. Hosseinzadeh, S. S. Sajadi Tabar, Smartwatches in healthcare medicine: assistance and monitoring; a scoping review, BMC Medical Informatics and Decision Making 23 (2023) 248.

[12] S. Parodi, R. Filiberti, P. Marroni, R. Libener, G. P. Ivaldi, M. Mussap, E. Ferrari, C. Manneschi, E. Montani, M. Muselli, Differential diagnosis of pleural mesothelioma using logic learning machine, BMC bioinformatics 16 (2015) S3.

[13] D. Cangelosi, F. Blengio, R. Versteeg, A. Eggert, A. Garaventa, C. Gambini, M. Conte, A. Eva, M. Muselli, L. Varesio, Logic learning machine creates explicit and stable rules stratifying neuroblastoma patients, BMC bioinformatics 14 (2013) S12.

[14] D. Cangelosi, M. Muselli, S. Parodi, F. Blengio, P. Becherini, R. Versteeg, M. Conte, L. Varesio, Use of attribute driven incremental discretization and logic learning machine to build a prognostic classifier for neuroblastoma patients, BMC bioinformatics 15 (2014) S4.

[15] G. González, S. Y. Ash, G. Vegas-Sánchez-Ferrero, J. Onieva Onieva, F. N. Rahaghi, J. C. Ross, A. Díaz, R. San José Estépar, G. R. Washko, Disease staging and prognosis in smokers using deep learning in chest computed tomography, American journal of respiratory and critical care medicine 197 (2018) 193–203.

[16] M. A. Fernandez-Granero, D. Sanchez-Morillo, A. Leon-Jimenez, An artificial intelligence approach to early predict symptom-based exacerbations of copd, Biotechnology & Biotechnological Equipment 32 (2018) 778–784.

[17] J. Monsi, J. Saji, K. Vinod, L. Joy, J. J. Mathew, Xray ai: Lung disease prediction using machine learning, International Journal of Information 8 (2019).

[18] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, J. Sun, Doctor ai: Predicting clinical events via recurrent neural networks, in: Machine Learning for Healthcare Conference, 2016, pp. 301–318.

[19] A. W. Przybyszewski, A. Śledzianowski, Parkinson's disease development prediction by c-granule computing compared to different ai methods, Journal of Information and Telecommunication (2020) 1–15.

[20] B. Cushen, I. Sulaiman, G. Greene, E. MacHale, M. Mokoka, R. B. Reilly, K. Bennett, F. Doyle, J. F. van Boven, R. W. Costello, The clinical impact of different adherence behaviors in patients with severe chronic obstructive pulmonary disease, American journal of respiratory and critical care medicine 197 (2018) 1630–1633.

[21] D. Price, D. L. Keininger, B. Viswanad, M. Gasser, S. Walda, F. S. Gutzwiller, Factors associated with appropriate inhaler use in patients with copd–lessons from the real survey [erratum], International Journal of Chronic Obstructive Pulmonary Disease 13 (2018) 2253–2254.

[22] D. Kappeler, K. Sommerer, C. Kietzig, B. Huber, J. Woodward, M. Lomax, P. Dalvi, Pulmonary deposition of fluticasone propionate/formoterol in healthy volunteers, asthmatics and copd patients with a novel breath-triggered inhaler, Respiratory medicine 138 (2018) 107–114.

[23] E. Kuipers, M. Wensing, P. de Smet, M. Teichert, Self-management research of asthma and good drug use (smaragd study): a pilot trial, International Journal of Clinical Pharmacy 39 (2017) 888–896.

[24] I. Banerjee, C. E. Catalano, G. Patané, M. Spagnuolo, Semantic annotation of 3d anatomical models to support diagnosis and follow-up analysis of musculoskeletal pathologies, International journal of computer assisted radiology and surgery 11 (2016) 707–720.

[25] A. Agibetov, E. Jiménez-Ruiz, M. Ondrésik, A. Solimando, I. Banerjee, G. Guerrini, C. E. Catalano, J. M. Oliveira, G. Patanè, R. L. Reis, et al., Supporting shared hypothesis testing in the biomedical domain, Journal of biomedical semantics 9 (2018) 9.

[26] S. Zhong, Z. Yang, R. N. Wright, Privacy-enhancing k-anonymization of customer data, in: Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2005, pp. 139–147.

[27] N. Mohammed, B. C. Fung, P. C. Hung, C.-k. Lee, Anonymizing healthcare data: a case study on the blood transfusion service, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 1285–1294.

[28] H. Lee, S. Kim, J. W. Kim, Y. D. Chung, Utility-preserving anonymization for health data publishing, BMC medical informatics and decision making 17 (2017) 104.

[29] K. El Emam, F. K. Dankar, Protecting privacy using k-anonymity, Journal of the American Medical Informatics Association 15 (2008) 627–637.

[30] L. Sweeney, k-anonymity: A model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (2002) 557–570.

[31] L. Sweeney, Achieving k-anonymity privacy protection using generalization and suppression, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (2002) 571–588.

[32] V. Ciriani, S. D. C. Di Vimercati, S. Foresti, P. Samarati, κ-anonymity, in: Secure data management in decentralized systems, Springer, 2007, pp. 323–353.

[33] A. Narayanan, V. Shmatikov, Robust de-anonymization of large sparse datasets, in: 2008 IEEE Symposium on Security and Privacy (sp 2008), 2008, pp. 111–125. doi:10.1109/SP.2008.33.

[34] S. Gambs, M.-O. Killijian, M. N. del Prado Cortez, De-anonymization attack on geolocated data, Journal of Computer and System Sciences 80 (2014) 1597–1614.

[35] H. Jiang, J. Yu, C. Hu, C. Zhang, X. Cheng, Sa framework based de-anonymization of social networks, Procedia Computer Science 129 (2018) 358–363.

[36] P. Savyan, S. M. S. Bhanu, Behaviour profiling of reactions in facebook posts for anomaly detection, in: 2017 Ninth International Conference on Advanced Computing (ICoAC), IEEE, 2017, pp. 220–226.

[37] J. Vaidya, C. Clifton, Privacy-preserving k-means clustering over vertically partitioned data, in: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 206–215.

[38] C. Biswas, D. Ganguly, D. Roy, U. Bhattacharya, Privacy preserving approximate k-means clustering, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1321–1330.

[39] S. Gayathri, P. R. Prasidha, G. Vinitha, Privacy preserving in data sets through multiple shuffle, International Journal of Research in Engineering, Science and Management 2 (2019) 839–841.

[40] L. Huang, L. Wang, Accelerated monte carlo simulations with restricted boltzmann machines, Physical Review B 95 (2017) 035105.

[41] Y. Chen, L. Lu, X. Li, Application of continuous restricted boltzmann machine to identify multivariate geochemical anomaly, Journal of Geochemical Exploration 140 (2014) 56–63.

[42] X. Liu, Y. Ouyang, W. Rong, Z. Xiong, Item category aware conditional restricted boltzmann machine based recommendation, in: International Conference on Neural Information Processing, Springer, 2015, pp. 609–616.

[43] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Computer networks 52 (2008) 2292–2330.

[44] I. Vaccari, E. Cambiaso, M. Aiello, Evaluating security of low-power internet of things networks, International Journal of Computing and Digital Systems 8 (2019) 101–114.

[45] I. Vaccari, E. Cambiaso, M. Aiello, Remotely exploiting at command attacks on zigbee networks, Security and Communication Networks 2017 (2017).

[46] I. Vaccari, M. Aiello, E. Cambiaso, Slowite, a novel denial of service attack affecting mqtt, Sensors 20 (2020) 2932.

[47] M. M. Hossain, M. Fotouhi, R. Hasan, Towards an analysis of security issues, challenges, and open problems in the internet of things, in: 2015 IEEE World Congress on Services, IEEE, 2015, pp. 21–28.

[48] H. Suo, J. Wan, C. Zou, J. Liu, Security in the internet of things: a review, in: 2012 international conference on computer science and electronics engineering, volume 3, IEEE, 2012, pp. 648–651.

[49] A. Balte, A. Kashid, B. Patil, Security issues in internet of things (iot): A survey, International Journal of Advanced Research in Computer Science and Software Engineering 5 (2015).

[50] J. S. Kumar, D. R. Patel, A survey on internet of things: Security and privacy issues, International Journal of Computer Applications 90 (2014).

[51] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion 58 (2020) 82–115.

[52] S. Narteni, V. Orani, E. Cambiaso, M. Rucco, M. Mongelli, On the intersection of explainable and reliable ai for physical fatigue prediction, IEEE Access 10 (2022) 76243–76260.

[53] M. Muselli, E. Ferrari, Coupling logical analysis of data and shadow clustering for partially defined positive boolean function reconstruction, IEEE Transactions on Knowledge and Data Engineering 23 (2011) 37–50. doi:10.1109/TKDE.2009.206.

[54] M. Muselli, Switching neural networks: A new connectionist model for classification, in: Neural Nets, Springer, 2005, pp. 23–30.

[55] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, J. D. Tygar, Adversarial machine learning, in: Proceedings of the 4th ACM workshop on Security and artificial intelligence, 2011, pp. 43–58.

[56] A. Prakash, N. Moran, S. Garber, A. DiLillo, J. Storer, Protecting jpeg images against adversarial attacks, in: 2018 Data Compression Conference, IEEE, 2018, pp. 137–146.

[57] P. McDaniel, N. Papernot, Z. B. Celik, Machine learning in adversarial settings, IEEE Security & Privacy 14 (2016) 68–72.

[58] A. Kurakin, I. J. Goodfellow, S. Bengio, Adversarial examples in the physical world, in: Artificial intelligence safety and security, Chapman and Hall/CRC, 2018, pp. 99–112.

[59] A. S. Bhogal, A. R. Mani, Pattern analysis of oxygen saturation variability in healthy individuals: Entropy of pulse oximetry signals carries information about mean oxygen saturation, Frontiers in physiology 8 (2017) 555.