computers

Article

Enhancing Workplace Safety through Personalized Environmental Risk Assessment: An AI-Driven Approach in Industry 5.0

Janaína Lemos [1], Vanessa Borba de Souza [2], Frederico Soares Falcetta [3], Fernando Kude de Almeida [4], Tânia M. Lima [1,5,*] and Pedro Dinis Gaspar [1,5,*]

1 Department of Electromechanical Engineering, University of Beira Interior, 6201-001 Covilhã, Portugal; janaina.lemos@ubi.pt
2 Postgraduate Program in Computing, Federal University of Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre 91501-970, Brazil; vanessa.borba@gmail.com
3 Laboratory Diagnostic Service, HCPA Hospital, R. Ramiro Barcelos, 2350, Porto Alegre 90035-903, Brazil; fredfalcetta@gmail.com
4 Oncology Division, Fêmina Hospital, R. Mostardeiro, 17, Porto Alegre 90430-001, Brazil; fernandokude@gmail.com
5 C-MAST—Centre for Mechanical and Aerospace Science and Technologies, 6201-001 Covilhã, Portugal
* Correspondence: tmlima@ubi.pt (T.M.L.), dinis@ubi.pt (P.D.G.)

Abstract: This paper describes an integrated monitoring system designed for individualized environmental risk assessment and management in the workplace. The system incorporates monitoring devices that measure dust, noise, ultraviolet radiation, illuminance, temperature, humidity, and flammable gases. Comprising monitoring devices, a server-based web application for employers, and a mobile application for workers, the system integrates the registration of workers' health histories, such as common diseases and symptoms related to the monitored agents, and a web-based recommendation system. The recommendation system application uses classifiers to decide the risk/no risk per sensor and crosses this information with fixed rules to define recommendations. The system generates actionable alerts for companies to improve decision-making regarding professional activities and long-term safety planning by analyzing health information through fixed rules and exposure data through machine learning algorithms. As the system must handle sensitive data, data privacy is addressed in communication and data storage. The study provides test results that evaluate the performance of different machine learning models in building an effective recommendation system. Since it was not possible to find public datasets with all the sensor data needed to train artificial intelligence models, it was necessary to build a data generator for this work. By proposing an approach that focuses on individualized environmental risk assessment and management, considering workers' health histories, this work is expected to contribute to enhancing occupational safety through computational technologies in the Industry 5.0 approach.

Keywords: occupational safety; environmental risk assessment; ai-driven monitoring; industry 5.0; personalized management; health history registration; monitoring devices; environmental parameters; decision-making; long-term safety planning

1. Introduction

Annually, approximately 2 million people die because of work-related causes worldwide [1]. According to Ncube and Kanda [2], the incidence of fatalities in workplaces varies considerably between developed and developing countries. In low- and middle-income countries, the lack of adequate Occupational Safety and Health (OSH) services and legislation contributes to the high rate of occupational diseases and accidents. The highest death rates because of work occur in agriculture, forestry, mining, and construction, particularly in companies with less than 50 employees [3].

Occupational diseases are characterized by a causal link between the damage to the worker's health and work-related exposures, such as physical, chemical, and biological agents. Occupational diseases and accidents can cause negative impacts on the quality of life of workers and their families, impact the productivity, competitiveness, and reputation of companies, and increase public costs [4,5].

Regarding approaches to improving OSH services, the concepts of Industry 4.0 and Industry 5.0 are currently being discussed. Industry 4.0 defines a new level of organization and control over the entire value chain and emphasises automation, innovation, data collection, cyber-physics systems, processes, and people to achieve efficiency, flexibility, and continuous improvement [6,7]. Industry 5.0, in turn, is a concept that reinforces the contribution of industry to society, aiming to apply research and innovative technologies to achieve human-centric and sustainable manufacturing systems. In this approach, the welfare of workers shall be at the center of the manufacturing process [8].

Concerning technologies, the terms Internet of Things (IoT), Industrial Internet of Things (IIoT), and machine learning (ML) are widely used nowadays. IoT refers to a set of technologies to provide the integration of everyday objects with the Internet and technologies that facilitate devices communicating with each other and with the cloud. IIoT stands for Industrial Internet of Things and refers to the use of IoT technologies to connect machines, equipment, and sensors in industrial applications. ML, in turn, is an area of artificial intelligence that uses data and statistical methods to train algorithms to make classifications or predictions and gradually improve their accuracy. These technologies are essential to supporting Industry 4.0 and 5.0 [6,9].

IoT, IIoT, and ML can be applied in workplaces to monitor workers' movements to prevent accidents, to identify harmful situations and their exposure to harmful agents, and to check the use of Personal Protective Equipment (PPE) [6,9]. Regarding recent research covering these topics, in the study conducted by Lemos, Gaspar, and Lima [10], a review of technologies and trends regarding OSH in Industry 4.0 was presented. A survey of Industry 4.0-compliant solutions for OSH management is shown by Jiang, Bakker, and Bartolo [11]. A systematic review of the application of immersive technologies for OSH management in the construction industry was conducted by Babalola et al. [12]. Examples of research involving smart PPE and monitoring systems are presented in several studies [13–19]. The big data collected by these systems can be analyzed to help companies both immediately identify dangerous conditions in workplaces and improve long-term decision-making in OSH.

This paper proposes a modular monitoring system to perform personalized environmental risk assessment. The system measures dust, noise, ultraviolet radiation, illuminance, temperature, and humidity and checks for the presence of flammable gases–liquefied petroleum gas (LPG), propane, hydrogen, butane, methane, and carbon monoxide. The proposed solution is composed of three parts; individual monitoring devices, a server, and a mobile application. The server hosts applications to process measurements and register employee's personal data, including diseases and symptoms related to the monitored quantities.

This research work presents tests covering the performance of different ML models considered to build a web-based recommendation system responsible for generating accurate alerts considering both the outputs of the ML, which analyzes each worker's measurements, and fixed rules for different diseases and symptoms. The set of diseases, symptoms, and rules was defined considering recent research along with the collaboration of two physicians. This approach is intended to help companies in the decision-making process, such as workers' activity planning to avoid or minimize exposure to certain agents and directing OSH actions. In the next stage of the study, the authors plan to test the system in a real work environment, in civil construction and building maintenance activities.

2. Materials and Methods

The framework is composed of the following components:

1. Monitoring devices: They must be clipped to the employees' clothes to collect measurements and send them encrypted to the server that runs in the cloud.
2. Server: Collects and processes data received from the monitoring devices. It hosts a web application to be used by companies to register employees and provide their health information regarding common diseases and symptoms related to the monitored quantities. This information is analyzed to generate alerts for the company, aiming to protect workers' health through a preventative approach. The entire history of the employee in the company regarding all the positions held is also shown by the referred web application, and a second web application shows graphs with workers' exposure data.
3. Mobile application: It can be used by workers to check their exposure data.

These components allow the system presented in this paper to provide continuous monitoring of worker exposure data, along with the assessment of health history, to help companies guide internal programs and decision-making in OSH. The system was previously described in [20,21], where initial tests regarding the monitoring device and communication with the server were conducted and where tests focused on verifying the functionalities of the web interface for employee registration were presented, especially regarding the provision of health information and the generation of corresponding alerts considering fixed rules for each health condition. The alerts generated did not use artificial intelligence in that stage. The components of the system are described at the next sections.

2.1. Monitoring Device

The monitoring device is composed of a microcontroller ESP32, battery, light emitting diodes (LEDs), GUVA-S12SD ultraviolet radiation sensor [22], KY-038 noise sensor [23], Light Dependent Resistor (LDR) [24], DHT11 temperature and humidity sensor [25], MQ-2 flammable gases sensor [26], and a Shinyei PPD42NS dust sensor [27]. The dimensions of the monitoring device are $9.5 \times 7.0 \times 4.0$ cm^3, and its weight is 250 g. Sensor specifications are shown in Table 1.

Table 1. Sensor specifications and environmental parameters.

Sensor	Specifications
DHT 11 (humidity and temperature)	Voltage: 3–5 VDC Humidity range: 20 to 90% relative humidity Temperature range: $0°$ to $50°C$ Dimensions: $23 \times 12 \times 5$ mm^3
GUVA-S12SD (UV)	Voltage: 2.5–5 V UV wave size: 240–370 nm Dimensions: 11×27 mm
KY-038 (noise)	Voltage: 4–6 VDC Frequency: 50 Hz to 20 kHz Dimensions: $37 \times 15 \times 13.7$ mm^3
LDR (illuminance)	Voltage: up to 150 VDC Spectrum: 540 nm Dimensions: diameter: 5 mm, length: 32 mm
MQ-2 (gas)	Voltage: 5 V Concentration detected: 300–10,000 ppm Dimensions: $32 \times 20 \times 15$ mm^3
Shinyei PPD42NS (dust)	Voltage: 4.75~5.75 VDC Detecting the particle diameter > 1 μm Detectable concentration: 0–28,000 particles/m^3 Dimensions: $45 \times 59 \times 22$ mm^3

The embedded application runs on the ESP32. Measurements are performed every ten minutes for almost all quantities, except for the concentration of flammable gases, which is checked every minute because when their volume is dangerous, the worker must leave the building immediately. The routines to process each monitored quantity and their limits are explained below.

Illuminance: The resistance of the LDR decreases as the illuminance increases. The ESP32 analog input pin converts the voltage (between 0 and 3.3 V) into an integer value between 0 and 4095. The embedded application classifies values as follows: less than 40 "dark" (level 1); 40 to 799 "dim" (level 2); 799 to 1999 "light" (level 3); 2000 to 3199 "bright" (level 4); greater than 3200 "very bright" (level 5) [24,28,29]. In various workplaces, such as construction sites, the workers are exposed to abrupt variations in illuminance, which can lead to accidents, for example, because of "temporary blindness" [30]. The ML model will identify abrupt changes in illumination, and based on this information, it is possible to generate alerts and suggestions to employers.

Temperature and humidity: An alarm is triggered, and a yellow LED is turned on when temperatures are lower than 10 °C or greater than 30 °C, as well as when the air's relative humidity is under 50% or above 70%. The alarm indicates that the worker must take precautions to protect his or her health, such as drinking water and staying in the shade on hot days. This approach was adopted because performance measurements by the standard [31] require expensive hardware.

- Dust: The limit is 1,415,000 particles/m^3 [32].
- Noise: The limit is 85 dBA [33].
- UV radiation is measured in Standard Erythema Dose (SED). SED is checked every ten minutes, and the total amount per day must not exceed 1.3. So, the sum is registered in a file in the ESP32 [34].
- Flammable gases: The limit is 1000 parts per million (PPM) [35].

The red LED is turned on when dust, noise, or UV radiation exceeds the referred limits. The blue LED is turned on when the gas concentration represents a risk and indicates that the worker must leave the building immediately.

Figure 1 shows the monitoring device assembled on a Printed Circuit Board (PCB) and Figure 2 shows the sensors.
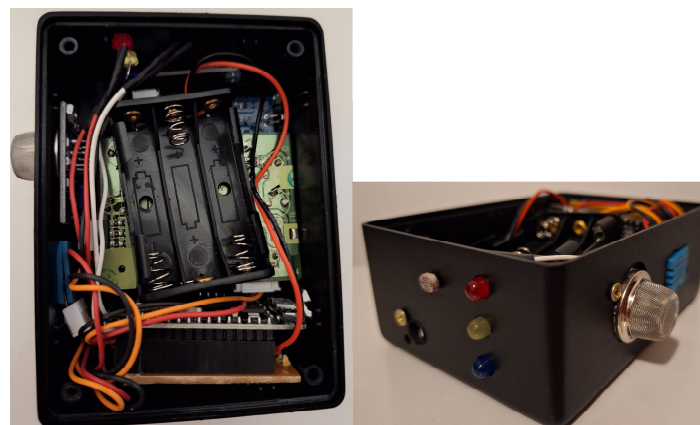
Figure 1. Monitoring device assembled on a PCB.

Monitoring devices communicate with the server using Wi-Fi and Message Queuing Telemetry Transport (MQTT) [36]. MQTT communication is secured by Transport Layer Security with a Pre-Shared Key (TLS-PSK) [37]. 256-bit long keys are shared between each device and the MQTT broker. MQTT topics are composed of the device ID together with the name of the monitored quantity, such as "6068/dust". To improve data privacy, at the employee registration stage, unique numeric identifiers are associated with unique

device identifiers. This association is useful for applications running on the server and preserves sensitive information such as names and document numbers. Therefore, if the system were used daily in a real work environment, workers' data would be stored and managed separately under company policy, which must follow local data protection law.

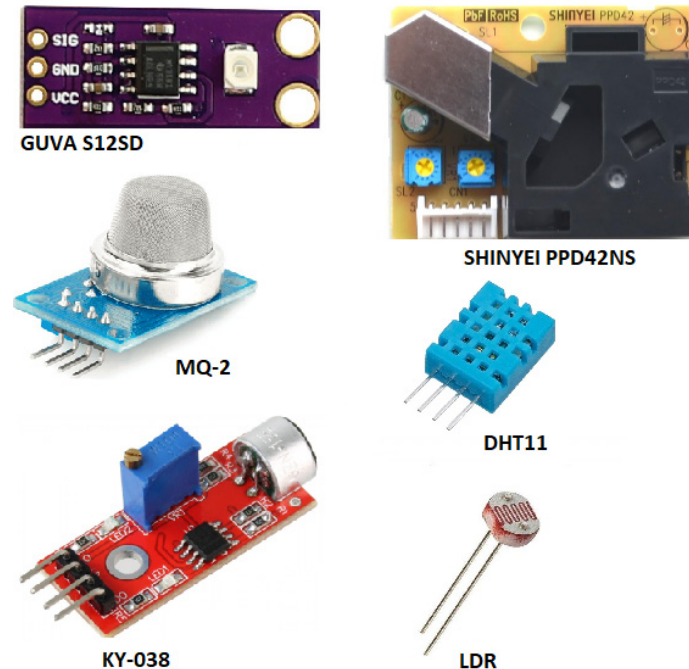**Figure 2.** Sensors.

The next section presents details regarding applications running on the server.

2.2. Server Overview

The server runs on a Virtual Private Server (VPS) provided by a web hosting company. The VPS runs the Ubuntu Server 20.04 operating system [38] and has 8 GB of RAM, 2 virtual CPUs, 100 GB of disk space, and 2 TB of bandwidth. The server runs Mosquitto MQTT broker [39] to handle the measurements sent by the monitoring devices, Telegraf MQTT client [40] to write the measurements in the InfluxDB time-series database [41], Grafana [42] to read InfluxDB and show graphics with exposure data, and MongoDB [43] to store employee information. A web application allows companies to register employees and associate them with monitoring devices, display the list of monitoring devices and the IDs of the employees who use them, display the current activity and the history of activities (if any) performed by a person in the company, and update employee information. This information is stored in MongoDB.

The web application provides a webpage on which a link to Grafana and the alerts for each employee can be found. This information is generated considering the health information stored during employee registration and can be updated anytime by the company. The alerts will be created both with fixed rules regarding health information and an ML algorithm to analyze the measurements stored in InfluxDB.

The server also hosts a mobile application backend. This application is intended for workers to check their daily exposures using a simple interface that is composed of a homepage that shows buttons to access the summary of the last 24 h for all the monitored quantities. It is possible to check the data history for 7 and 30 days as well. The application also provides information about all the monitored quantities.

2.3. Recommendation System

The recommendation system web application uses classifiers to decide the risk/no risk per sensor and crosses this information with fixed rules to define recommendations. The fixed rules list the agents monitored by the system, pre-diagnosed diseases, and symptoms that can be worsened by exposure to such agents.

Regarding disease and symptom history, two lists were prepared considering the quantities measured by the monitoring devices (temperature, humidity, dust, noise, and UV radiation). The volume of flammable gases was not considered because it is monitored, aiming to avoid accidents by indicating through the alarm that the worker must leave the place when the volume is dangerous. The variation in illuminance was also not treated in this case because it is related to occupational accidents.

The first list presents the common pre-diagnosed diseases that can be worsened by exposure to the monitored agents. The second one presents common symptoms that may result from those exposures or can be worsened by them. When registering a worker, only diseases that have been diagnosed by a physician should be marked, with or without symptoms, and in the list of symptoms only the persistent ones. To create the lists, appropriate literature was consulted, together with the collaboration of two physicians. These relationships are shown in Table 2.

The health data provided during employee registration can be updated by the employer at any time. The alerts are shown in the web application on the webpage "Exposure data and alerts" and when the health information is updated, the alert is modified. The general format of an alert is the following:

Table 2. Agents, pre-diagnosed diseases, and symptoms.

Agents	Pre-Diagnosed Diseases	Symptoms
Cold [44–47]	Asthma Hypertension Rheumatic diseases Spinal disorders Respiratory diseases Previous allergic diseases and reactions	Chest pain Cough Dyspnea (shortness of breath) Headache Hemoptysis (blood cough) Skin lesions Skin rash Weight loss Neck, low back pain, joint pain
Dust [45,48,49]	Asthma Lung cancer Respiratory diseases Previous tuberculosis infection Smoker	Chest pain Cough Dyspnea (shortness of breath) Fever Hemoptysis (blood cough) Weight loss
Heat [50,51]	Diabetes Heart disease Hypertension Hypotension Kidney disease	Chest pain Dyspnea (shortness of breath) Fainting (syncope) Headache Increased thirst Increased urinary volume Weight loss
High humidity [45,52,53]	Asthma Diabetes Respiratory diseases Previous allergic diseases and reactions	Chest pain Cough Dyspnea (shortness of breath) Hemoptysis (blood cough) Increased thirst Increased urinary volume Weight loss

Table 2. *Cont.*

Agents	Pre-Diagnosed Diseases	Symptoms
Low humidity [53–55]	Asthma Respiratory diseases Previous allergic diseases and reactions	Chest pain Cough Dyspnea (shortness of breath) Hemoptysis (blood cough) Skin lesions Skin rash
Noise [46,56,57]	Diabetes type 2 Hearing disorders Hypertension Smoker	Difficulty understanding conversation in situations with background noise Feeling that the ears are plugged up Speech or other sounds muffled after exposure to loud noise Transient tinnitus
UV radiation [58–60]	Eye disease Heart disease Skin cancer Skin diseases	Chest pain Dyspnea (shortness of breath) History of resection of skin lesions Skin lesions Skin rash

"In respect to diseases and symptoms, this person has: disease 1, disease 2, disease n. Due to reported diseases and symptoms, it is advised to avoid or reduce exposure to the following agents: agent 1, agent 2, agent n. Including other companies, the worker has performed activities of the same type and has been exposed to the same agents for n years. Attention to health is recommended".

To generate alerts regarding the monitored agents and related occupational diseases or accidents, the ML module will analyze measurements from all sensors except gas. As previously mentioned, when the gas concentration is high, an alarm is triggered by the monitoring device to indicate that the worker must leave the building immediately.

Regarding the other quantities, for each measurement classified as a risk, an entry is inserted in a table that indicates the occupational diseases related to the monitored agent. This table will be shown in the web interface, together with the alerts explained above. All agents and related occupational diseases and accidents are shown in Table 3.

Regarding the number of lines (agents), the table may be different for each worker or even for the same worker at different moments, as exposures to the agents increase or decrease.

Table 3. Agents and related occupational diseases and accidents.

Agents with Risky Exposures	Related Occupational Diseases and Accidents
Cold [44–47]	Worsening of respiratory diseases Increase in musculoskeletal disorders
Dust [45,48,49]	Pulmonary fibrosis (asbestosis) Lung cancer (due to inhalation of asbestos dust)
Heat [50,51]	Dehydration (favours the occurrence of kidney problems) Heart attack Stroke Dryness of the nasal mucosa (favours the emergence of respiratory infections)
High humidity [45,52,53]	Worsening of respiratory diseases
Low humidity [53–55]	Worsening of respiratory diseases
Noise [46,56,57]	Hearing loss Hypertension

Table 3. *Cont.*

Agents with Risky Exposures	Related Occupational Diseases and Accidents
UV radiation [58–60]	Dehydration Skin lesions Heat stroke Burns Skin cancer Photosensitization Erythema Acute inflammatory eye reactions Increased risk of cataracts Suppression of the immune system (favours the occurrence of infections and cancer)
Illuminance [28]	Abrupt variations in illuminance can cause "temporary blindness" and it can lead to accidents

This is a preventive approach in which the alerts are intended to assist OSH professionals/physicians in decision-making and to define when it is necessary to conduct health examinations, for example. The information provided by the system can be used to guide supervisors to redesign workflows aiming to minimize exposure to certain agents for a worker. The outputs of the recommendation system can also support internal health awareness campaigns in companies.

The ML module that will compose the recommendation system is presented in the next section.

2.4. Machine Learning Module

For the proposed system, the problem consists of applying ML to analyze the readings from different sensors–UV radiation, illuminance level, temperature, humidity, dust, and noise and identify whether the values read represent a risk to the worker's health.

It was not possible to find on the Internet datasets with all the sensor data needed for training an artificial intelligence model, so it was essential to build a data generator for this work. For all the quantities, we considered both the ranges that can be read by the monitoring devices, which are dependent on the capabilities of the sensors, and the exposure limits specified in current regulations. Data was generated to represent both normal and risky situations.

The behaviour of the ML module is intended to be preventive. It should identify in advance conditions that may contribute to the worker getting sick and not just conditions that do not comply with applicable regulations, as this is already done by monitoring devices/mobile applications and can be verified by employers through the web interface that integrates the system. For this reason, it will be highlighted below in the description of the data generation and classification of sensor readings that the established limits are more restrictive than the current regulations.

2.4.1. Generation of Training Data

The training data for each quantity was generated as described below:

Dust: The Shinyei PPD42NS can measure from 0 to 28,000,000 particles/m^3. The total particle limit for 8 h of work is 1415 million particles/m^3 [30]. This limit is considered for triggering alarms in the monitoring devices. However, the ML module classifies measurements above 999,999 as a risk. This approach was used because the sensor can't differentiate between several small, agglomerated particles or a slightly larger particle, so the count provided is an approximation. For this reason, values of 1,000,000 particles/m^3, when routinely obtained, indicate that special attention should be paid to the employee's health. In other words, the worker may be exposed to more dust than the sensor indicates. As explained earlier, the ML module is intended to use a preventive approach.

Illuminance: As explained earlier, the LDR sensor can determine basic illuminance changes. Values vary from 1 to 5, from dark to very bright [24,28,29]. Illuminance values were always generated in pairs (current and next values). If the difference between them is greater than 2, for example, from bright (4) to dark (1), the behaviour is classified as risky. Otherwise, the behaviour is classified as not risky.

Noise: The sound sensor can measure up to 90 dB, and the training data was generated by observing this interval. Measurements equal to or greater than 85 dB were classified as risky, while readings below this value were classified as not risky. The criteria for data classification were defined based on Brazilian and European regulations, which are very similar to noise limits in work environments [31,61,62]. Regarding the noise commonly found in civil construction activities, according to [63], noise levels at construction sites may range from 80 to 120 dBA, and construction machines are the primary sources of annoying noise. So, depending on the activity, it may be very common for construction workers to exceed the recommended limits.

Temperature and humidity: The sensor can measure temperatures from 0° to 50 °C and relative humidity from 20 to 90%. The data were generated considering the average of the months from January to August 2022 in Porto Alegre, Rio Grande do Sul, Brazil, obtained from the Brazilian National Institute of Meteorology (INMET) [64]. This location was chosen because the system will be tested in a real work environment in this region in the next stage of the project. During the work shift (from 8:00 am to 5:00 pm), random values are generated for each hour of the day in different ranges. The values generated for temperature and humidity can vary up to 4 degrees Celsius and 4%, respectively. This variation aims to reproduce what can happen when a worker frequently moves between indoors and outdoors. Being outdoors, the worker can alternate between the shadow and the sun. Different ranges of temperature and humidity were also defined for summer and winter, with hot and dry summers and cold and humid winters, as often occur in the region mentioned above. The dataset that represents a risk in the summer is mainly composed of temperature values that exceed 30 °C and percentages of humidity below 50%. In winter, the dataset that represents a risk is mainly composed of temperature values below 10 degrees Celsius and percentages of humidity that exceed 70%. Even if the ML is utilized in locations with diverse climates, it is anticipated that there will be no difficulty in recognizing risky conditions, as the temperature and humidity levels that indicate risk are precisely specified.

UV radiation: The maximum Standard Erythema Dose (SED) that can be obtained from the monitoring devices for ten minutes of sun exposure is 1.6, and the daily SED limit is 1.3. Considering this limit, values were generated from 0 to 1.6 SED. The ML module classifies values greater than 0.5 for ten minutes of sun exposure as risky because, with these environmental conditions, 30 min of work is sufficient for workers to exceed their maximum daily exposures according to WHO recommendations [32].

The artificially generated values are within the ranges measured by the sensors. When the system runs completely, with sensor readings being carried out by monitoring devices, if any sensor presents an error in its reading and provides a value outside the expected range, this value will be discarded by the embedded application that runs on the monitoring device.

Table 4 summarizes the criteria for defining the Risk class in the reading of each sensor.

Based on these criteria, a Python script was written to populate the dataset for each sensor. As emphasized above, to classify data as risky or normal, current regulations were considered. However, the approach adopted has a more preventive nature; that is, a measurement can be classified as risky when its value is close to the limit recommended by applicable standards.

To simplify the tests, the data generated for each type of sensor persisted in a PostgreSQL relational database [65], with a schema being structured for each sensor type. Each schema presents the same data structure, making information retrieval standard for the model training stage, regardless of the type of sensor. As it was not possible to identify in

the literature the class with the highest frequency of occurrence (risk or normal), it was decided to generate a balanced dataset, maintaining the same number of samples for each class. Table 5 presents the data from the datasets generated for training the models.

Table 4. Criteria for defining the risk class when reading each sensor.

Sensor	Rule	Class
Dust	Value > 999,999 particles/m^3	Risk
Humidity	Value < 50% or value > 70%	Risk
Temperature	Value < 10 °C or value > 30 °C	Risk
Illuminance	Difference between two readings > 2	Risk
Noise	Value > 84 dB	Risk
UV radiation	Value > 0.5 SED	Risk

Table 5. Dataset structure considering each sensor individually.

	Dataset	Readings	Series	Risk	Normal
	Train	12,960	2160	1080	1080
Sensor [1]	Validation	12,960	2160	1080	1080
	Test	12,960	2160	1080	1080
Total	All	38,880	6480	3240	3240

[1] Each sensor has the same structure and total number of records per class. Thus, the dust, humidity, temperature, illuminance, noise, and UV radiation schemes together add up to a total of 233,280 records.

It was defined that the input data follows the time series pattern, where each series is composed of 6 sequential sensor readings, carried out at intervals of 10 min between readings.

2.4.2. Problem Categorization and Data Variability

For all sensors, the problem can be categorized as a binary classification problem, where for a given sequence of sensor reading values (input x), the model should be able to classify whether the scenario represented by the sequence of Input data falls into one of two classes: y = {risk, normal}. Furthermore, this problem involves time series, where the order of the values read matters to solve the problem. In this case, special care is necessary when using algorithms to prepare batches of data for model training, aiming to preserve the temporal sequence of the data.

It is also necessary to increase data variability because learning the parameters of a prediction function and testing it on the same can lead the model to have a perfect score, but it would fail to predict anything useful on yet-unseen data. This situation is called overfitting. A solution to this problem is a procedure called cross-validation (CV). A test set should still be held out for final evaluation, but the validation set is no longer needed when doing a CV. In the basic approach, called k-fold, the training set is split into k smaller sets. The following procedure is followed for each of the k "folds". A model is trained using k − 1 of the folds as training data, and the resulting model is validated on the remaining part of the data (i.e., it is used as a test set to compute a performance measure such as accuracy). The performance measure reported by k-fold cross-validation is then the average of the values computed in the loop [66].

When generating the data for this work, a time series was defined as the sequence of readings within a 1-h window, carried out every 10 min. To increase data variability during model training, k-fold = 10 cross-validation techniques adapted for the problem considering the time series were used, in which the data is divided into batches of series instead of batches of values. In other words, a fold is a set of series, each containing its own sequence of 6 readings that represent an input value x for the model.

While defining criteria for the risky and normal classes for each sensor, it was found that for most sensors, testing whether the value falls within a range of values is sufficient to determine which class the sequence of readings represents. Linear and logistic regression

models perform well for identifying relationships in time series, as well as tree-based models for problems whose rules resemble if/else chains for class definition [67]. For these reasons, the following strategy was adopted:

Instead of generating a multiclass model, which tries to learn all the rules for all sensors, it was decided to create expert models for each type of sensor. As shown by Souza et al. [68], this technique has presented satisfactory results for complex problems, as it reduces the complexity, computational cost, and training time of the generated model since each expert model will have to learn a smaller set of patterns.

With the data generated to develop the models, the feature engineering stage was covered via domain knowledge, making it possible to start testing with supervised ML techniques through shallow learning models. The nature of the data presented indicates the use of models based on decision trees. Therefore, it was decided to explore these models, followed by linear and statistical models, to compare the techniques with the best results. The ML techniques applied in this work are explained in the next section.

2.4.3. Machine Learning Techniques

Shallow learning refers to the use of relatively simple models with a small number of layers or processing stages. In general, these models have a limited capacity to learn complex patterns from data, and they are often applied when the data has relatively simple patterns and the relationships between features and outcomes are straightforward. There is a vast set of shallow learning techniques, with methods based on Support Vector Machine (SVM) optimization, search (regressions and decision trees), and probabilistic (methods based on Bayes' theorem) widely used in tasks involving data classification. According to Murphy [67], the main algorithms for each one of the methods are:

- SVM: support vector machines (SVMs) are a set of supervised learning methods used for classification, regression, and outliers' detection. The kernel method seeks to find a hyperplane that best divides the data points according to the target classes. The algorithm maps the training data to a new high-dimensional representation and calculates the maximum distance between the generated hyperplanes and the closest data points for each class. In this work, two variations of the Kernel method were used; SVC (Support Vector Classification) and LinearSVC. The main differences between LinearSVC and SVC lie in the loss function used by default and in the handling of intercept regularization between these two implementations.

- Regression: this is a predictive method that analyzes the relationship between dependent (target) and independent (predictor) variables. Different algorithms are used depending on the type of target variable and/or the relationship between variables (linear or non-linear). Among them, Linear Regression, Logistics, and Bayesian stand out. In this work, the logistic regression algorithm and its variation with the L2 Ridge Classifier regularization function were tested, as well as the SGD Classifier techniques, which implement a plain stochastic gradient descent learning routine that supports different loss functions and penalties for classification. The model it fits can be controlled with the loss parameter. By default, it fits a linear SVM.

- Decision Trees: this is a hierarchical model capable of guiding decision-making about which class a given instance belongs to, resulting in a unique path from the root node to the leaf (target class). The tree model is obtained from the training data using a divide-and-conquer strategy, applied hierarchically. In this category of algorithms, Random Forest and Gradient Boosting Machines currently stand out. In this work, the decision tree-based algorithms Decision Tree Classifier and Extra Tree Classifier were tested.

- Ensemble: these methods combine the predictions of several base estimators built with a given learning algorithm to improve generalizability and robustness over a single estimator. The premise is that each model contributes a different hypothesis space, representation language, and hypothesis evaluation function. The hypothesis space generated by the final model considers optima that are closer to the global ones

and reduces the computational cost of training a single model for a complex task. Two widely used ensemble methods are gradient-boosted trees and random forests. More generally, ensemble models can be applied to any base learner beyond trees, in averaging methods such as Bagging methods, model stacking, or Voting, or in boosting, such as AdaBoost. In this work, the models Random Forest Classifier, Ada Boost Classifier, Bagging Classifier, Gradient Boosting Classifier, and XGB Classifier (Extreme Gradient Boosting) were tested.

- Statistical models: predictive algorithms based on statistical methods, divided into (1) generative, whose statistical model used is the joint probability distribution, and (2) discriminative, whose statistical model used is the conditional probability. In this work, tests were carried out using two variations of the Quadratic Discriminant Analysis (QDA) and Linear Discriminant Analysis (LDA) discriminative models. These models differ in terms of the function used to determine the decision surface that divides the data according to classes, with the first function being quadratic and the second being linear, respectively. If the problem data space presents covariance variation, the QDA model tends to present better results due to the use of a quadratic function to divide the space for each class. In this work, a widely known statistical generative classification model called Naive Bayes (Gaussian NB–a simplified implementation that considers normal distribution) was also tested. This method is part of a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features.

- Clustering: there are many ways to define classification models, including whether the model has a fixed number of parameters or whether the number of parameters increases with the amount of training data. In the first case, the models are called parametric, and in the second case, non-parametric. Parametric models have the advantage of often being faster to use but have the disadvantage of making stronger assumptions about the nature of data distributions. Non-parametric models are more flexible but often computationally intractable for large datasets. In this work, a simple and well-known parametric model, K Nearest Neighbour (KNN), was tested. This model simply searches for K points in the training set that are closest to the input point, counts how many members of each class are in this set, and returns this empirical fraction as an estimate of the sample's probability of belonging to the class.

2.4.4. Evaluation Metrics for Machine Learning Models

Methods for evaluating model performance allow comparing the impact of using different algorithms for the same task or even evaluating different parameterizations considering the same algorithm. As the models mentioned above are binary classification models, F1-score, precision, recall, and AUC and ROC curves were selected for qualitative analysis of the results of each model. These metrics are described below by [69].

- Precision: gives the proportion of correct positive predictions. It considers false positives, which are cases that were incorrectly flagged for inclusion.
- Recall: measures the proportion of actual positives that were predicted correctly. It considers false negatives, which are cases that should have been flagged for inclusion but were not.
- F1-score: combines precision and recall in a single number.
- ROC curve: the Receiver Operating Characteristic curve is a two-dimensional curve with the True Positive Rate on the vertical axis and False Positive Rate on the horizontal axis.
- AUC: the area under the ROC curve (AUC) is a global measure of the ability of a model to distinguish between classes, for example, to differentiate whether a given condition is present or not. An AUC of 0.5 represents a model without this ability, while an AUC of 1.0 represents a model with perfect discrimination ability.

2.4.5. Training Methods

For each sensor individually, the following sequence of actions was adopted:

(1) Training pipeline generation, where all desired models are loaded.
(2) Each model is trained and validated using the respective sets.
(3) Use of the k-fold = 10 cross-validation technique adapted for the problem considering time series, in which the data is divided into batches of series instead of batches of values.
(4) Collecting metrics during the training phase.
(5) For each generated model, testing and collecting metrics on the test dataset.
(6) For each algorithm, 10 models were generated, one for each test round.
(7) The results were tabulated in terms of average performance, considering the performance of each of the 10 models for each technique when classifying the same set of tests.
(8) For the models that present the best average according to the F1-measure criterion (a metric that represents the balance between recall and precision), the model with the best performance was selected among the 10. The recall metric was chosen as a tiebreaker criterion, as it is considered that the nature of this problem is that it is better to issue a false risk alert than to fail to alert a possible risk.

The algorithm that was developed to generate the synthetic data, the machine learning models, and the entire quantitative-qualitative analysis are available on the project's Github, available at: https://github.com/borbavanessa/osh_ia/tree/main/ (accessed on 6 May 2024).

3. Results

In this study, the tests aim to check the performance of the ML models for each sensor, and the alerts generated by the recommendation system that was built with the rules shown in Table 2 (agents, pre-diagnosed diseases, and symptoms) and Table 3 (agents that can cause harm and related occupational diseases).

Table 6 summarizes the performance between training and testing sets. It is possible to note that the resulting models do not present overfitting (super-adjustment to the training data), as they presented similar performance between validation and testing.

Table 6. Performance analysis per sensor.

Sensor	Dataset	F1	AUC	Notes
UV radiation	Train	1	1	All models presented a perfect performance in both the training and testing sets.
	Test	1	1	
Noise	Train	0.99 and 1	0.99 and 1	The models PassiveAggressive, RidgeClassifier, LogisticRegression, SGDClassifier, KNeighbors, LinearSVC, SVC, and LinearDiscriminantAnalysis presented 99%, while the others presented 100%.
	Test	0.99 and 1	0.99 and 1	The models ExtraTreeClassifier, Kneighbors, LinearDiscriminantAnalysis, LinearSVC, PassiveAggressive, RidgeClassifier, and SGDClassifier presented 99%, while the other models presented 100%.
Illuminance	Train	0 to 1	0 to 1	Large variation in performance, from 0 to 100%. The best models were: XGBRegressor, Kneighbors, DecisionTree, ExtraTree, SVC, BaggingClassifier, RandomForest, and GradientBoosting.
	Test	0 to 1	0 to 1	
Humidity and temperature	Train	0 to 1	0 to 1	Only RandomForest achieved 100%. XGBRegressor, DecisionTree, BaggingClassifier, and GradientBoosting presented with 99%.
	Test	<1	<1	XGBRegressor, BaggingClassifier, and RandomForest had 99%.
Dust	Train	0 to 1	0 to 1	100% for XGBRegressor, RidgeClassifer, Kneighbors, DecisionTree, ExtraTree, SVC, GaussianNB, AdaBoost, BaggingClassifier, RandomForest, GradientBoosting, and QuadraticDiscriminant.
	Test	0 to 1	0 to 1	

The tests showed that more than one algorithm presented maximum performance for the same sensor. But for all sensors, the RandomForest model presented maximum performance. Furthermore, RandomForest (an ensemble of trees) is a model with good explicability. For these reasons, it was selected to compose the classification solution for this work.

The complete performance analysis of the models per sensor is shown in Tables A1 and A2.

Figures 3–5 display the graphs with the average performance (resulting from the 10 training rounds) in terms of Precision (P), Recall (R), F-measure (F1), and ROC-AUC curve for each model considering Dust, illuminance, humidity, and temperature sensors.

The ROC curves are available on the project's Github, accessible at: https://github.com/borbavanessa/osh_ia/tree/main/ia_model/experiment_models (accessed on 6 May 2024).

Only the comparative graphs for the dust, illuminance, and humidity/temperature sensors are being displayed, whose performance between the models showed variation in the results. The other sensors presented results that achieved, on average, 100% performance for all metrics, and for this reason, they are not displayed.
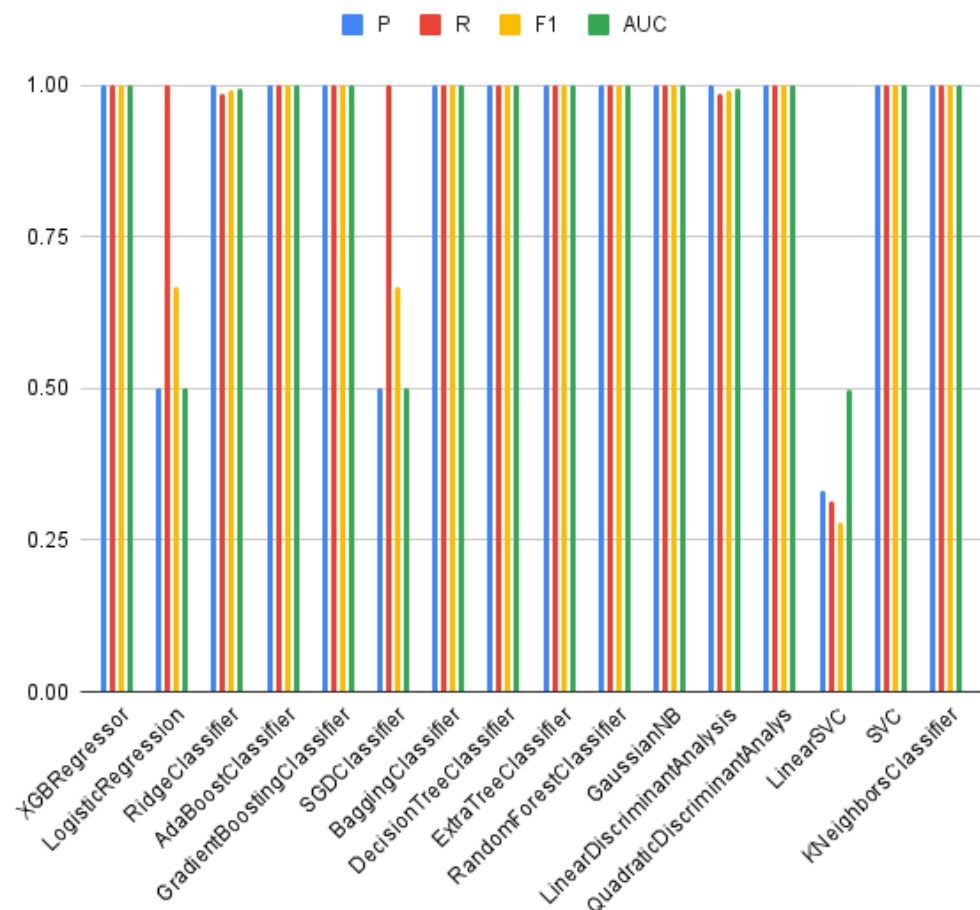

Figure 3. Precision (P), Recall (R), F-measure (F1) and ROC-AUC curve for a dust sensor.
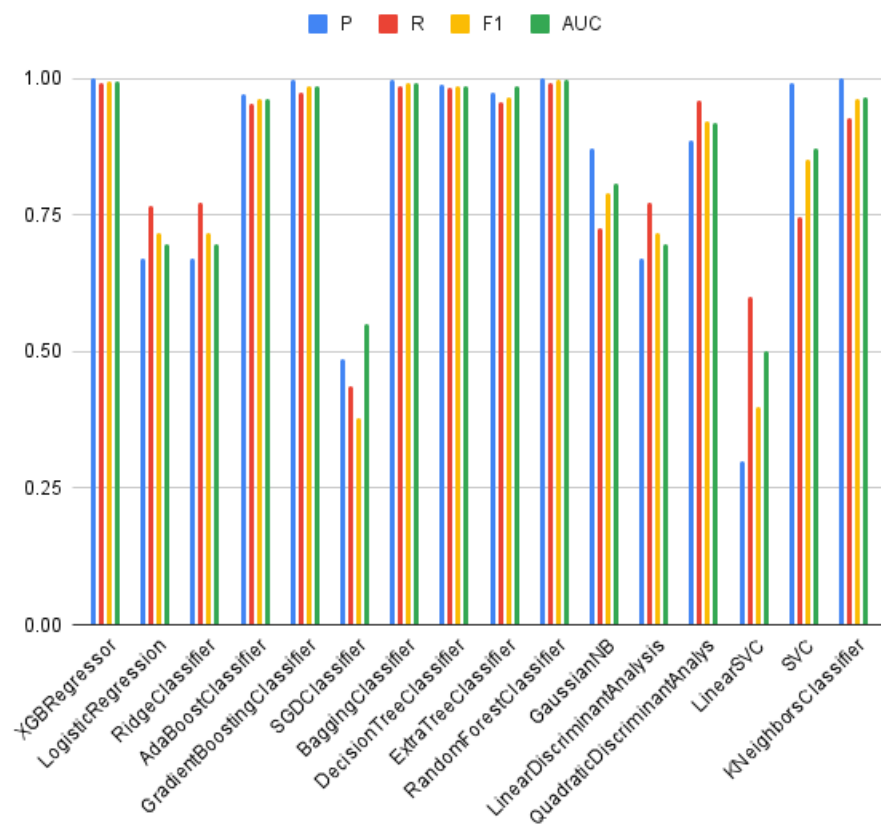
Figure 4. Precision (P), Recall (R), F-measure (F1) and ROC-AUC curve for humidity and temperature sensors.
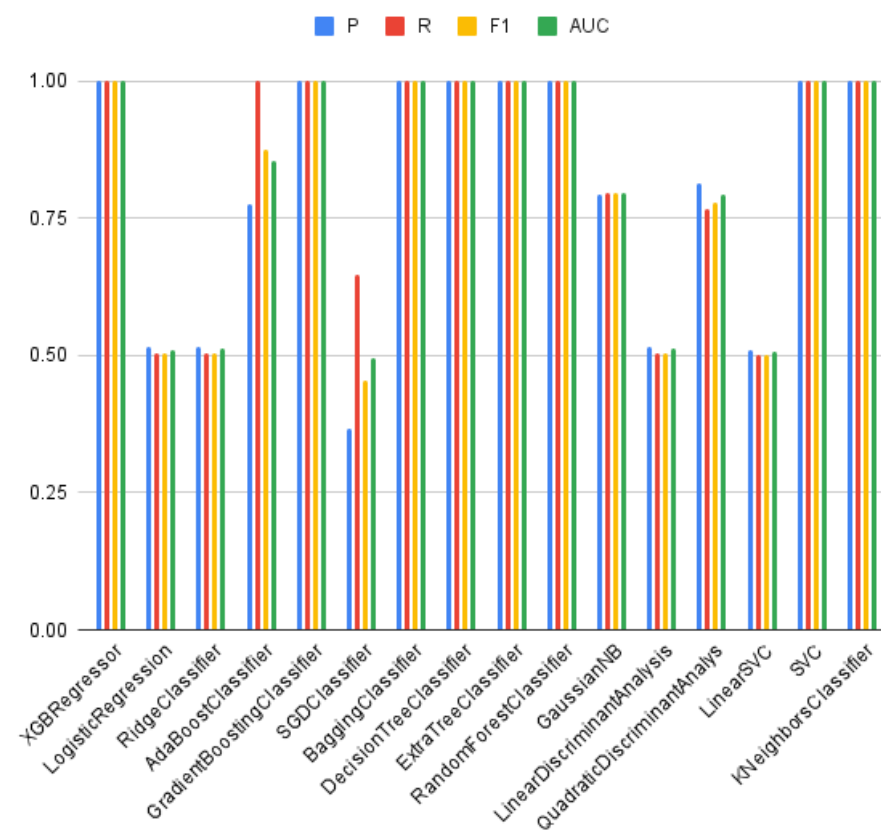

Figure 5. Precision (P), Recall (R), F-measure (F1), and ROC-AUC curve for illuminance sensor.

4. Discussion

The difference between the results was already expected since the illuminance, humidity, and temperature sensors present more complex rules for defining the risk class than the other sensors. As previously explained, illuminance is measured on a scale of 1 (dark) to 5 (very bright). The behaviour is classified as risky if the difference between two sensor readings is greater than 2.

Regarding temperature and humidity, the results obtained with the various ML models were expected to vary because they were related to one another, as humidity tends to drop as temperature increases. In the datasets, different ranges of temperature and humidity were defined for summer and winter, with hot and dry summers and cold and humid winters, as often occurs in the region where the tests in a real work environment will be conducted in the next stage of this work.

Concerning the variation in the results for the dust sensor, the rule that defines the risk class itself is not complex, but the measurements present in the datasets have high variation (from 0 to 28,000,000 particles/m^3, with values greater than 999,999 being classified as risk). In this case, the extreme values occur because of natural variability, and it is known that they have a negative impact on ML models [70]. These effects would be more significant if multiclass models were chosen instead of expert models for each sensor.

The excellent performance of the models can be justified by the strategy adopted for data pre-processing, which includes:

(1) Use of domain knowledge to format and simplify data input to the models (feature engineering stage).

(2) The selection of models and care taken with techniques to diversify training, considering their performance over time.

(3) The decision to simplify the architecture by generating expert models for each sensor instead of a multiclass model that tries to learn very different rules based on very different and uncorrelated numeric input ranges (independent events, for example, the value measured by the dust sensor does not interfere with noise sensor measurement).

The feature engineering stage can be supplied by deep learning models, but shallow learning models were chosen due to the complexity of evaluating and understanding deep learning models, accompanied by the current lack of sufficient volume of data to allow adequate training for this type of model. In addition, deep learning models are called black box models because the rules they create to decide on a class are difficult to interpret.

As mentioned in Section 3, Random Forest presented the best performance for all sensors, and for this reason, this model was selected to compose the classification solution for the proposed system.

Regarding the other components of the proposed system, considering the possibility of scaling the system for daily use in enterprises, the following points must be highlighted:

- There are more accurate and higher-cost sensor options on the market that can be considered for building more robust monitoring devices. This approach was not used in the present work because the choice of more expensive components would make the project unfeasible, considering that it was necessary to produce devices to be tested with a group of workers. However, to scale the system for everyday use in an enterprise, the sensors would need to be reevaluated, and the devices may need to use more precise sensors.
- The system can be easily extended to cover other agents, diseases, and symptoms by adding new sensors to the monitoring devices, adapting the embedded application, inserting such new data, and training the models related to the respective sensors.
- The monitoring devices for everyday use in an enterprise shall undergo tests to ensure compatibility with applicable standards. For commercial use, it would also be necessary to purchase paid versions of some of the same software used to build the prototype.

- The dimensions of the monitoring device are 9.5 × 7.0 × 4.0 cm, and its weight is 250 g. It is expected that the use of the device will not significantly interfere with the workers' routine and comfort.
- Tests in a real work environment will be conducted in the next stage of this work. During the tests, workers must use the device attached to their clothing throughout the work shift. The usability of the mobile application, monitoring device design, possible discomfort arising from its use, and data privacy concerns will be evaluated. By carrying out these tests, it is expected that other adjustment points will be verified. As mentioned in Section 2.1, communication between monitoring devices and the server is encrypted, and the server does not store names and documents. In a company, it is expected that workers' data will be stored and managed separately, following company policy and complying with local data protection laws.

As mentioned earlier, in the next stage of the study the system will be tested in a real work environment, in civil construction and building maintenance activities. In these professional activities, workers are frequently exposed to the agents monitored by the system and others, such as heavy physical exertion [71].

Regarding other possible uses relevant to the proposed system, other activities that tend to pose very significant risks to workers can be mentioned, such as agriculture and the food industry. For these scenarios, it would be appropriate to add new sensors and information about diseases and symptoms. Innovations in this type of industry and the potential of using IoT have been widely discussed [71–74].

5. Conclusions

This work has presented a monitoring system composed of devices for individual monitoring of workers' exposure to harmful agents, a server, and a mobile application.

Our previous work [21] has presented results regarding the provision of workers' health information with respect to common diseases and symptoms related to the monitored quantities in the web interface of the system. That information was used to generate alerts, considering fixed rules for each health condition.

In this work, to choose the ML model to compose the classification solution to analyze workers' exposures, tests were conducted with 16 shallow ML models. Binary classification models specialized in identifying risk/non-risk for each sensor were generated. This topology reduced the complexity and computation of training and presented satisfactory results to contribute to the proposed recommendation system. The best results were obtained with the Random Forest model. Based on the outputs of the ML model, alerts will be generated considering the diseases that can be caused by the exposures.

The web-based recommendation system that uses ML to continuously analyze worker's exposures in conjunction with fixed rules that consider previously diagnosed diseases and symptoms and generate alerts can help companies in the decision-making process. In the short term, alerts can help in workers' activity planning to avoid or minimize exposure to certain agents. In the long term, alerts can provide support for planning actions to prevent diseases, and exposure data obtained through continuous monitoring can be used in additional studies.

The proposed system can be easily extended to cover other agents, diseases, and symptoms by adding new sensors to the monitoring devices, inserting such new data, and training the models related to the respective sensors.

In the next step of the project, the entire system will be tested in a real workplace, at least for one company. At this stage, alerts generated through analysis of exposures using ML (as shown in Table 3) will also be available on the system's web interface. For this purpose, a few devices will be assembled. A group of workers will be asked both to use the monitoring devices during the entire work shift and to access the mobile application once daily. The web application will be used by a manager or OSH professional. The participants should answer questionnaires about the usability of the solution.

Appendix A

Tables A1 and A2 present the performance analysis of the models per sensor.

Table A1. Performance analysis of the models for UV and humidity and temperature sensors.

Model	UV				Humidity Temperature			
	P	R	F1	AUC	P	R	F1	AUC
XGBRegressor	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
LogisticRegression	1.00	1.00	1.00	1.00	0.67	0.77	0.72	0.70
RidgeClassifier	1.00	1.00	1.00	1.00	0.67	0.77	0.72	0.70
AdaBoostClassifier	1.00	1.00	1.00	1.00	0.97	0.95	0.96	0.96
GradientBoostingClassifier	1.00	1.00	1.00	1.00	1.00	0.97	0.99	0.99
SGDClassifier	1.00	1.00	1.00	1.00	0.49	0.44	0.38	0.55
BaggingClassifier	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
DecisionTreeClassifier	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99
ExtraTreeClassifier	1.00	1.00	1.00	1.00	0.97	0.96	0.96	0.99
RandomForestClassifier	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
GaussianNB	1.00	1.00	1.00	1.00	0.87	0.72	0.79	0.81
LinearDiscriminantAnalysis	1.00	1.00	1.00	1.00	0.67	0.77	0.72	0.70
QuadraticDiscriminantAnalysis	1.00	1.00	1.00	1.00	0.88	0.96	0.92	0.92
LinearSVC	1.00	1.00	1.00	1.00	0.30	0.60	0.40	0.50
SVC	1.00	1.00	1.00	1.00	0.99	0.75	0.85	0.87
KNeighborsClassifier	1.00	1.00	1.00	1.00	1.00	0.93	0.96	0.96

Table A2. Performance analysis of the models for noise, dust, and light sensors.

Model	Noise				Dust				Light			
	P	R	F1	AUC	P	R	F1	AUC	P	R	F1	AUC
XGBRegressor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LogisticRegression	1.00	1.00	1.00	1.00	0.50	1.00	0.67	0.50	0.51	0.50	0.50	0.51
RidgeClassifier	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.51	0.50	0.50	0.51
AdaBoostClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	1.00	0.87	0.86
GradientBoostingClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SGDClassifier	1.00	1.00	1.00	1.00	0.50	1.00	0.67	0.50	0.37	0.65	0.46	0.49
BaggingClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DecisionTreeClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A2. *Cont.*

Model	Noise				Dust				Light			
	P	R	F1	AUC	P	R	F1	AUC	P	R	F1	AUC
ExtraTreeClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RandomForestClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
GaussianNB	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.79	0.80	0.79	0.79
LinearDiscriminantAnalysis	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.51	0.50	0.50	0.51
QuadraticDiscriminantAnalysis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81	0.77	0.78	0.79
LinearSVC	0.99	1.00	1.00	1.00	0.33	0.32	0.28	0.50	0.51	0.50	0.50	0.51
SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
KNeighborsClassifier	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

References

1. WHO. Available online: https://www.who.int/news/item/17-09-2021-who-ilo-almost-2-million-people-die-from-work-related-causes-each-year (accessed on 28 October 2023).
2. Ncube, F.; Kanda, A. Current Status and the Future of Occupational Safety and Health Legislation in Low- and Middle-Income Countries. *Saf. Health Work* **2018**, *9*, 365–371. [CrossRef] [PubMed]
3. Melchior, C.; Zanini, R. Mortality per work accident: A literature mapping. *Saf. Sci.* **2019**, *114*, 72–78. [CrossRef]
4. Jilcha, K.; Kitaw, D. Industrial occupational safety and health innovation for sustainable development. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 372–380. [CrossRef]
5. Teufer, B.; Ebenberger, A.; Affengruber, L.; Kien, C.; Klerings, I.; Szelag, M.; Griebler, U. Evidence-based occupational health and safety interventions: A comprehensive overview of reviews. *BMJ Open* **2019**, *9*, e032528. [CrossRef] [PubMed]
6. Javaid, M.; Haleem, A.; Singh, R.; Rab, S.; Suman, R. Upgrading the manufacturing sector via applications of Industrial Internet of Things (IIoT). *Sens. Int.* **2021**, *2*, 100129. [CrossRef]
7. Yu, F.; Schweisfurth, T. Industry 4.0 technology implementation in SMEs—A survey in the Danish-German border region. *Int. J. Innov. Stud.* **2020**, *4*, 76–84. [CrossRef]
8. Huang, S.; Wang, B.; Li, X.; Zheng, P.; Mourtzis, D.; Wang, L. Industry 5.0 and Society 5.0—Comparison, complementation and co-evolution. *J. Manuf. Syst.* **2022**, *64*, 424–428. [CrossRef]
9. Khan, W.; Rehman, M.; Zangoti, H.; Afzal, M.; Armi, N.; Salah, K. Industrial internet of things: Recent advances, enabling technologies and open challenges. *Comput. Electr. Eng.* **2020**, *81*, 106522. [CrossRef]
10. Lemos, J.; Gaspar, P.D.; Lima, T.M. Environmental Risk Assessment and Management in Industry 4.0: A Review of Technologies and Trends. *Machines* **2022**, *10*, 702. [CrossRef]
11. Jiang, Z.; Bakker, O.; Bartolo, P. Critical Review of Industry 4.0 Technologies' Applications on Occupational Safety and Health. In Proceedings of the 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT), Istanbul, Turkey, 17–20 May 2022; pp. 1267–1272. [CrossRef]
12. Babalola, A.; Manu, P.; Cheung, C.; Yunusa-Kaltungo, A.; Bartolo, P. A systematic review of the application of immersive technologies for safety and health management in the construction sector. *J. Saf. Res.* **2023**, *85*, 66–85. [CrossRef]
13. Sánchez, M.; Sergio Rodriguez, C.; Manuel, J. Smart Protective Protection Equipment for an accessible work environment and occupational hazard prevention. In Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29–31 January 2020; pp. 581–585. [CrossRef]
14. Yang, K.; Ahn, C.; Kim, H. Deep learning-based classification of work-related physical load levels in construction. *Adv. Eng. Inform.* **2020**, *45*, 101104. [CrossRef]
15. Kim, J.; Jo, B.; Jo, J.; Kim, D. Development of an IoT-Based Construction Worker Physiological Data Monitoring Platform at High Temperatures. *Sensors* **2020**, *20*, 5682. [CrossRef] [PubMed]
16. Campero-Jurado, I.; Márquez-Sánchez, S.; Quintanar-Gómez, J.; Rodríguez, S.; Corchado, J. Smart Helmet 5.0 for Industrial Internet of Things Using Artificial Intelligence. *Sensors* **2020**, *20*, 6241. [CrossRef] [PubMed]
17. Costa, J.; Souto, E. A IoT Device for Monitoring Particulate Matter and Gaseous Pollutants in Indoor Industrial Workstations. In Proceedings of the 2022 IEEE International Conference on Consumer Electronics—Taiwan, Taipei, Taiwan, 6–8 July 2022; pp. 517–518. [CrossRef]
18. Singh, N.; Gunjan, V.; Chaudhary, G.; Kaluri, R.; Victor, N.; Lakshmanna, K. IoT enabled HELMET to safeguard the health of mine workers. *Comput. Commun.* **2022**, *193*, 1–9. [CrossRef]
19. Rajakumar, J.; Choi, J.-H. Helmet-Mounted Real-Time Toxic Gas Monitoring and Prevention System for Workers in Confined Places. *Sensors* **2023**, *23*, 1590. [CrossRef] [PubMed]
20. Lemos, J.; Gaspar, P.D.; Lima, T.M. Individual Environmental Risk Assessment and Management in Industry 4.0: An IoT-Based Model. *Appl. Syst. Innov.* **2022**, *5*, 88. [CrossRef]
21. Lemos, J.; de Souza, V.B.; Falcetta, F.S.; de Almeida, F.K.; Lima, T.M.; Gaspar, P.D. A System for Individual Environmental Risk Assessment and Management with IoT Based on the Worker's Health History. *Appl. Sci.* **2024**, *14*, 1021. [CrossRef]

22. Hew, C.; Tan, R.; Lee, C.; Hartanty, T.; Hossain, W.; Lee, Y. Development of Self Sustainable IOT Based Low Cost UV Index Monitoring Station. In Proceedings of the 2022 IEEE 8th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Melaka, Malaysia, 26–28 September 2022; pp. 36–41. [CrossRef]

23. Failing, J.M.; Abellán-Nebot, J.V.; Benavent Nácher, S.; Rosado Castellano, P.; Romero Subirón, F. A Tool Condition Monitoring System Based on Low-Cost Sensors and an IoT Platform for Rapid Deployment. *Processes* **2023**, *11*, 668. [CrossRef]

24. Marinho, F.; Carvalho, C.; Apolinário, F.; Paulucci, L. Measuring light with light-dependent resistors: An easy approach for optics experiments. *Eur. J. Phys.* **2019**, *40*, 035801. [CrossRef]

25. Jiang, B.; Huacón, C. Cloud-based smart device for environment monitoring. In Proceedings of the 2017 IEEE Conference on Technologies for Sustainability (SusTech), Phoenix, AZ, USA, 12–14 November 2017; pp. 1–6. [CrossRef]

26. Trisnawan, I.; Jati, A.; Istiqomah, N.; Wasisto, I. Detection of Gas Leaks Using The MQ-2 Gas Sensor on the Autonomous Mobile Sensor. In Proceedings of the 2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 23–24 October 2019; pp. 177–180. [CrossRef]

27. Canu, M.; Galvis, B.; Morales, R.; Ramírez, O.; Madelin, M. Understanding the Shinyei PPD24NS low-cost dust sensor. In Proceedings of the 2018 IEEE International Conference on Environmental Engineering, Milan, Italy, 12–14 March 2018; pp. 1–10. [CrossRef]

28. Zolkapli, M.; Al-Junid, S.; Othman, Z.; Manut, A.; Mohd Zulkifli, M. High-efficiency dual-axis solar tracking developement using Arduino. In Proceedings of the 2013 International Conference on Technology, Informatics, Management, Engineering and Environment, Bandung, Indonesia, 23–26 June 2013; pp. 43–47. [CrossRef]

29. ESP32 I/O. ESP 32—Light Sensor. 2018. Available online: https://esp32io.com/tutorials/esp32-light-sensor (accessed on 22 June 2023).

30. Hinze, J.; Teizer, J. Visibility-related fatalities related to construction equipment. *Saf. Sci.* **2011**, *49*, 709–718. [CrossRef]

31. *ISO 8996:2021*; Ergonomics of the Thermal Environment—Determination of Metabolic Rate. International Organization for Standardization: Geneva, Switzerland, 2021.

32. OSHA. Particulates not Otherwise Regulated, Total and Respirable Dust. 2023. Available online: https://www.osha.gov/chemicaldata/801 (accessed on 8 August 2023).

33. Ministério do Trabalho e da Solidariedade Social. Diário da República n.º 172/2006, Série I de 2006-09-06. Available online: https://data.dre.pt/eli/dec-lei/182/2006/09/06/p/dre/pt/html (accessed on 6 August 2023).

34. WHO. Global Solar UV Index: A practical Guide. A Joint Recommendation of the World Health Organization, World Meteorological Organization, United Nations Environment Programme, and the International Commission on Non-Ionizing Radiation. 2002. Available online: https://apps.who.int/iris/handle/10665/42459 (accessed on 8 August 2023).

35. NIOSH. Table of IDLH Values—L.P.G. Available online: https://www.cdc.gov/niosh/idlh/68476857.html (accessed on 12 September 2023).

36. MQTT Version 5.0. MQTT. Available online: https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html (accessed on 7 September 2023).

37. Zamfir, S.; Balan, T.; Iliescu, I.; Sandu, F. A security analysis on standard IoT protocols. In Proceedings of the 2016 International Conference on Applied and Theoretical Electricity (ICATE), Craiova, Romania, 6–8 October 2016; pp. 1–6. [CrossRef]

38. Ubuntu Server. Available online: https://ubuntu.com/download/server (accessed on 10 September 2023).

39. Eclipse. Eclipse Mosquitto. An Open Source MQTT Broker. Available online: https://mosquitto.org (accessed on 12 September 2023).

40. Telegraf. Influxdata. Available online: https://www.influxdata.com/time-series-platform/telegraf (accessed on 14 September 2023).

41. InfluxDB. Available online: https://www.influxdata.com/products/influxdb (accessed on 14 September 2023).

42. Grafana OSS. Available online: https://grafana.com/oss/grafana (accessed on 15 September 2023).

43. MongoDB. Available online: https://www.mongodb.com (accessed on 15 September 2023).

44. Burström, L.; Järvholm, B.; Nilsson, T.; Wahlström, J. Back and neck pain due to working in a cold environment: A cross-sectional study of male construction workers. *Int. Arch. Occup. Environ. Health* **2013**, *86*, 809–813. [CrossRef] [PubMed]

45. Kim, J.L.; Henneberger, P.K.; Lohman, S.; Olin, L.-C.; Dahlman-Höglund, A.; Andersson, E.; Torén, K.; Holm, M. Impact of occupational exposures on exacerbation of asthma: A population-based asthma cohort study. *BMC Pulm. Med.* **2016**, *16*, 148. [CrossRef]

46. Pettersson, H.; Olsson, D.; Järvholm, B. Occupational exposure to noise and cold environment and the risk of death due to myocardial infarction and stroke. *Int. Arch. Occup. Env. Health* **2020**, *93*, 571–575. [CrossRef] [PubMed]

47. Karthick, S.; Kermanshachi, S.; Loganathan, K. Effect of Cold Temperatures on Health and Safety of Construction Workers. In Proceedings of the Transportation Consortium of South-Central States (Tran-SET) Conference, Austin, TX, USA, 31 August 31–2 September 2022. [CrossRef]

48. Poinen-Rughooputh, S.; Rughooputh, M.S.; Guo, Y.; Rong, Y.; Chen, W. Occupational exposure to silica dust and risk of lung cancer: An updated meta-analysis of epidemiological studies. *BMC Public Health* **2016**, *16*, 1137. [CrossRef]

49. Kratzke, P.; Kratzke, R.A. Asbestos-Related Disease. *J. Radiol. Nurs.* **2018**, *37*, 21–26. [CrossRef]

50. Varghese, B.; Hansen, A.; Bi, P.; Pisaniello, D. Are workers at risk of occupational injuries due to heat exposure? A comprehensive literature review. *Saf. Sci.* **2018**, *110*, 380–392. [CrossRef]

51. Moon, J. The effect of the heatwave on the morbidity and mortality of diabetes patients; a meta-analysis for the era of the climate crisis. *Environ. Res.* **2021**, *195*, 110762. [CrossRef] [PubMed]

52. Tyrovolas, S.; Chalkias, C.; Morena, M.; Kalogeropoulos, K.; Tsakountakis, N.; Zeimbekis, A.; Gotsis, E.; Metallinos, G.; Bountziouka, V.; Lionis, C.; et al. High relative environmental humidity is associated with diabetes among elders living in Mediterranean islands. *J. Diabetes Metab. Disord.* **2014**, *13*, 25. [CrossRef]

53. Zhang, H.; Liu, S.; Chen, Z.; Zu, B.; Zhao, Y. Effects of variations in meteorological factors on daily hospital visits for asthma: A time-series study. *Environ. Res.* **2020**, *182*, 109115. [CrossRef]

54. Wang, W. Progress in the impact of polluted meteorological conditions on the incidence of asthma. *J. Thorac. Dis.* **2016**, *8*, E57–E61. [CrossRef]

55. Arcenal, K.; Carmen, M.; Garcia, R. Effects of Low Humidity and High Humidity on the Nasal Area of the People. In Proceedings of the 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, UK, 19–21 May 2023; pp. 1–6. [CrossRef]

56. Hong, O.; Kerr, M.; Poling, G.; Dhar, S. Understanding and preventing noise-induced hearing loss. *Dis.-A-Mon.* **2013**, *54*, 110–118. [CrossRef] [PubMed]

57. Lie, A.; Skogstad, M.; Johannessen, H.A.; Tynes, T.; Mehlum, I.S.; Nordby, K.-C.; Engdahl, B.; Tambs, K. Occupational noise exposure and hearing: A systematic review. *Int. Arch. Occup. Env. Health* **2016**, *89*, 351–372. [CrossRef] [PubMed]

58. Yam, J.; Kwok, A. Ultraviolet light and ocular diseases. *Int. Ophthalmol.* **2014**, *34*, 383–400. [CrossRef] [PubMed]

59. Modenese, A.; Korpinen, L.; Gobba, F. Solar Radiation Exposure and Outdoor Work: An Underestimated Occupational Risk. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2063. [CrossRef] [PubMed]

60. Bernard, J.; Gallo, R.; Krutmann, J. Photoimmunology: How ultraviolet radiation affects the immune system. *Nat. Rev. Immunol.* **2019**, *19*, 688–701. [CrossRef] [PubMed]

61. European Parliament. Directive 2003/10/EC of the European Parliament and of the Council of 6 February 2003 on the Minimum Health and Safety Requirements Regarding the Exposure of Workers to the Risks Arising from Physical Agents (Noise). Available online: https://eur-lex.europa.eu/eli/dir/2003/10/2019-07-26 (accessed on 1 November 2023).

62. Brazil. Norma Regulamentadora No. 15 (NR-15). Available online: https://www.gov.br/trabalho-e-emprego/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/comissao-tripartite-partitaria-permanente/arquivos/normas-regulamentadoras/nr-15-atualizada-2022.pdf (accessed on 4 November 2023).

63. Lee, S.C.; Kim, J.; Hong, J. Characterizing perceived aspects of adverse impact of noise on construction managers on construction sites. *Build. Environ.* **2019**, *152*, 17–27. [CrossRef]

64. INMET. Sistema Tempo. Available online: https://tempo.inmet.gov.br/TabelaEstacoes/A00 (accessed on 28 May 2023).

65. Postgresql. Available online: https://www.postgresql.org (accessed on 10 October 2023).

66. Scikit-learn. Cross-Validation: Evaluating Estimator Performance. Available online: https://scikit-learn.org/stable/modules/cross_validation.html (accessed on 20 November 2023).

67. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; The MIT Press: Cambridge, MA, USA, 2012.

68. Souza, V.; Nobre, J.; Becker, K. DAC Stacking: A Deep Learning Ensemble to Classify Anxiety, Depression, and Their Comorbidity from Reddit Texts. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 3303–3311. [CrossRef] [PubMed]

69. Zheng, A. Available online: https://www.oreilly.com/content/evaluating-machine-learning-models (accessed on 15 February 2024).

70. Measures of Variability. Available online: https://medium.com/@madhuri15/day-03-measures-of-variability-7-days-of-statistics-for-data-science-6bb7168b9300 (accessed on 26 February 2024).

71. Patel, V.; Chesmore, A.; Legner, C.M.; Pandey, S. Trends in Workplace Wearable Technologies and Connected-Worker Solutions for Next-Generation Occupational Safety, Health, and Productivity. *Adv. Intell. Syst.* **2022**, *4*, 2100099. [CrossRef]

72. Varandas, L.; Faria, J.; Gaspar, P.D.; Aguiar, M.L. Low-Cost IoT Remote Sensor Mesh for Large-Scale Orchard Monitorization. *J. Sens. Actuator Netw.* **2020**, *9*, 44. [CrossRef]

73. Gaspar, P.D.; Fernandez, C.M.; Soares, V.N.G.J.; Caldeira, J.M.L.P.; Silva, H. Development of Technological Capabilities through the Internet of Things (IoT): Survey of Opportunities and Barriers for IoT Implementation in Portugal's Agro-Industry. *Appl. Sci.* **2021**, *11*, 3454. [CrossRef]

74. Gaspar, P.D.; Soares, V.N.G.J.; Caldeira, J.M.L.P.; Andrade, L.P.; Soares, C.D. Technological modernization and innovation of traditional agri-food companies based on ICT solutions—The Portuguese case study. *J. Food Process Preserv.* **2022**, *46*, e14271. [CrossRef]