A supervised ML Biometric Continuous Authentication System for Industry 4.0

Juan Manuel Espín López, Alberto Huertas Celdrán*, Francisco Esquembre,
Gregorio Martínez Pérez, and Javier G. Marín-Blázquez

***Abstract*— Continuous authentication (CA) is a promising approach to authenticate workers and avoid security breaches in the industry, especially in Industry 4.0, where most interaction between workers and devices takes place. However, introducing CA in industries raises unsolved questions regarding machine learning (ML) models: i) its precision and performance, ii) its robustness and iii) the issue about if or when to retrain the models. To answer these questions, this work explores these issues with a proposed supervised vs non-supervised ML-based CA system that uses sensors, applications statistics, or speaker data collected by the operator's devices. Experiments show supervised models with Equal Error Rates of 7.28% using sensors data, 9.29% with statistics, and 0.31% with voice, a significant improvement of 71.97%, 62.14%, and 97.08%, respectively, over unsupervised models. Voice is the most robust dimension when adding new workers, with less than 2% of false acceptance rate even if workforce size is doubled.**

***Index Terms*— Continuous Authentication, Sensors, Applications usage, Speaker recognition, ML/DL, Industry 4.0**

I. Introduction

THE Industry 4.0 ecosystem involves different agents, from factories to end customers through re-sellers. This cooperation makes it possible to improve and optimize operational processes, products, and services. Multiple stakeholders (e.g. end users) can interact within new industrial ecosystems to accelerate and customize processes of the products they order [1]. A security breach at any point of this chain might have serious consequences on industrial processes and affect various products or stakeholders. The development of Industry 4.0 requires secure, reliable, and user-friendly authentication mechanisms in the business fabric, specifically in the manufacturing industry or in factories. In this context, operator authentication is the first line of defense for accurate and secure execution of the entire system. The companies must authenticate the operators, who work in conjunction with the machines, in such a way that the authorized operation of the entire factory or a production line is guaranteed.

In most factories, operators access the interior of the factory using an identification card [2], which presents the risk of the card being stolen, or a biometric system, such as fingerprints or facial biometrics [3], where the inconvenience is that its precision can be reduced when carrying protection mechanisms. In some environments, once workers have accessed their jobs, they can perform any task with no further authentication, which may allow for unauthorized or untrained workers to do tasks they should not do. In other cases, operators do authenticate themselves multiple times by repeatedly entering a password or checking biometrics, which reduces the usability of the system and can lower the productivity. In recent research, for instance, some authors [4] propose the use of a portable authentication system through eye movement and iris authentication. But this system requires the operator to look at the screen, and is impractical when they use a voice driven device or wear safety glasses or contact lenses [5].

On the other hand, in many factories, workers manipulate machines and processes using screens, laptops, smartphones, or PDAs. Moreover, this interaction is more pronounced in those factories adapted to the Industry 4.0 revolution. In these situations, the use of continuous authentication (CA) can help to improve industry security substantially. With CA, workers properly authorized can be allowed to perform sensible operations without affecting the productivity, given the non-intrusive operation of a CA system, which will reject non-authorized ones. Moreover, a CA system can collect data from multiple sources (e. g. the movement and position of the device taken from device sensors, statistics of usage, voice, facial or iris biometrics – when available) without disturbing the worker, and react only when a possible security breach or a non-allowed operation is detected. An important fact for the work presented in this paper, is that collecting data from workers in this context does not compromise their privacy because data collection is restricted to tasks performed on factory equipment, not personal devices, and can be therefore legally collected and used.

Precisely this legal access to worker's data, only available

This research was partially funded by the Spanish Ministry of Science and Innovation, State Research Agency, FEDERfunds, through projects RTI2018-095855-B-I00, MTM2017-84079-P and PID2019-108377RB-C32, by the Swiss Federal Office for Defense Procurement (armasuisse) with the CyberTracer (CYD-C-2020003) project, and by the University of Zürich UZH.

J.M. Espín, G. Martínez and J.G. Marín-Blázquez are with the Department of Information and Communications Engineering (DIIC), University of Murcia, 30100 Murcia, Spain; e-mail: juanmanuel.espin1@um.es; gregorio@um.es; jgmarin@um.es.

A. Huertas is with Communication Systems Group (CSG), Department of Informatics (IfI), University of Zürich UZH, CH-8050 Zürich, Switzerland; email: huertas@ifi.uzh.ch.

F. Esquembre is with Department of Mathematics, University of Murcia, 30100 Murcia, Spain; email: fem@um.es.

* Corresponding author.

in a scenario such the Industry 4.0, allows the use of supervised machine learning techniques in addition to unsupervised approaches. Research on biometric authentication systems in Industry 4.0 is very limited, and it is focused on the use of simple biometrics, such as iris, fingerprints, among all. In a recent paper [6], a platform (S3 Platform) for the combined use of sensors, statistics and voice data for CA was proposed. In that work, only an unsupervised Machine Learning approach was considered, as its creators were concerned with possible legal restrictions. However,no previous work has been found using continuous authentication, both supervised or unsupervised, in this Industry scenario, which leaves the following research questions to be addressed:

1) How much a supervised machine learning approach improves versus unsupervised CA systems?
2) How robust is such a supervised system when confronting intruders in an Industry 4.0 environment?
3) Does such workforce models need to be immediately retrained if a given number of new workers is hired?

To address these questions, this paper:

- Proposes a new supervised approach of the S3 Platform to continuously authenticate the industry operators.
- Evaluates and compares the two approaches, supervised and unsupervised, to the problem. As expected, the supervised approach (binary classification) performs better than the unsupervised approach (outlier detection). A significant improvement, and its magnitude, can be observed in all the evaluated metrics.
- Evaluates the robustness of the supervised approach, segregated by the three data sources. In this experiment, several users are excluded from the training phase of the models and later evaluated as impostors/new workers. The results show that voice models are the most robust, and statistics models the least. Even if the number of new workers is similar in size as the total number of workers, that is, even if the workforce is suddenly doubled, the different data sources models still present an upper 95% error interval value for false acceptance rate of only 4.17% for sensors, 9.09% for statistics, and just 2% for voice. In other words, the current workforce models do not need to be immediately retrained every time a new worker is hired.

The paper is structured as follows. Section II analyzes previous related works, focusing on the most recent CA works. Section III describes the scenario and the S3 platform. Section IV shows the dataset and the results obtained for the different experiments. Finally, Section V draws conclusions and sketches possible future work.

II. Related work

The use of biometrics in factories adapted to Industry 4.0 is not very widespread. This fact may be due to the recent push that Industry 4.0 has experienced and the cost of coordinating research and companies willing to evaluate and develop the technology. Most of the published works on security and authentication focus on the communications, protocols, and encryption systems used between the different communication nodes of Industry 4.0, [7], [8].

Among the articles that propose biometric systems for companies adapted to Industry 4.0, the use of fingerprints and technologies related to the eye, such as eye-tracking or iris recognition, stands out. In [9], the authors propose a fingerprint-based biometric system for the manufacturing industry with a precision of 95%. Regarding iris-based authentication, in [10], the authors propose a secure and automatic payment system for small and medium-sized companies that are adapting to Industry 4.0 (they only propose the system but do not evaluate its behavior). Another solution that authenticates the user by the eye is given in [11], but now, instead of the iris, using its movement.

As can be observed, the list of works found that deal with biometric authentication in Industry 4.0 is quite limited, and even more reduced when considering only those that focus on continuous authentication.. For this reason, a more general review of works whose characteristics can be extrapolated to the industrial setting is carried out below. The characteristics that have been considered are *(i)* the use of a mobile device, smartphone, PDA, or laptop, because in manufacturing companies operators usually work with this type of devices; *(ii)* the use of one or several alternative data sources, because for some workers, or in certain situations, one or more data sources can be available and ML models can then use one or another; and *(iii)* the use of a CA system since the objective of such a system is to increase security without drastically reducing usability, nor hindering the work of the operator. A detailed summary with year, dimension, approach followed, algorithm, dataset, and results of all these works can be found in Table I.

In this context, the availability of different data sources in the same device allows building more robust solutions by combining two or more of these sources. The most common combination is the use of sensors together with another type of source. For example, sensors and statistics [12], [13], sensors and touch screen data [14], [15], or sensors, statistics, and voice in [6]. The work at [16] stands out for a design where seven different kinds of data are used. Although less common, some works also use a single source, for example, sensors in [17]–[19], or touchscreen in [20].

As seen at the beginning of this section, the list of works dealing with biometric systems in Industry 4.0, specifically in manufacturing, is minimal. The rest of the exhibited works that share characteristics of the scenario do not show the robustness of the systems to unknown users. For this reason, this paper proposes a new platform where the operator of a factory adapted to Industry 4.0 can be authenticated continuously, thanks to the information provided by the device used for operation. The system proposed below will be able to use all the sources available by sensors, application usage statistics, and voice. In addition, the robustness of the system is evaluated against unknown users.

III. Scenario and Platform

This section presents the industrial scenario considered in this work and describes the new functionality provided to the

TABLE I
COMPARISON OF WORKS WHOSE CHARACTERISTICS CAN BE EXTRAPOLATED TO THE INDUSTRY 4.0 SCENARIO.

Reference	Year	Dimension	Approach	Algorithm	Dataset	Results
[6]	2021	Sensors & App's usage statistics & voice	Unsupervised	IF + KNN + ABOD	S3-Dataset	The voice increase the accuracy when it is present
[12]	2018	App's usage statistics & sensors	Unsupervised	IF, One-class SVM, Local Outlier Factor	PD with 2 users and 5 attackers	Precision: 77% Recall: 92% Accuracy : 82.5%
[13]	2021	App's usage statistics & sensors	Supervised	XGBoost	AuthCode	The multi-device system achieved a 59.65% and 89.35% improvement in the FPR for mobile applications and mobile sensors respectively
[14]	2020	Sensors & touch-screen	Unsupervised	One-class SVM	PD with 30 users	The system achieve 79% correct user identification, 13% FAR and 11.5% EER.
[15]	2019	Sensors & touch-screen	Supervised	DNN	HMOG Dataset	The system achieve 88% accuracy and 15% EER.
[16]	2020	Seven sources	Supervised	SVM with RBF kernel	UMDAA-02	Accuracy ranging from 82.2% to 97.1%
[17]	2020	Sensors	Unsupervised	Two-Stream CNN + PCA+ One-class SVM	Private Dataset (PD) with 100 users & Brain Run	4.57% EER, 4.65% FAR and 4.48% FRR 5.71% EER, 5.87% FAR and 5.56% FRR
[18]	2021	Sensors	Unsupervised	CWGAN + One-class SVM& IF & EE & LOF	PD with 100 users	The lowest EER of 3.64% for the IF classifier.
[19]	2021	Sensors	Supervised	SVM & RF & LR	Sherlock	The authors present a system with privacy preservation with 76.85% of accuracy and 5.12ms of computation.
[20]	2021	Touch-screen	Supervised	SVM, RF, KNN	PD with 24 users	The best system achieve an AUC 0.937 and EER 10.6% using the SVM classifier.
This work	2022	Sensors & App's usage statistics & voice	Supervised Vs Unsupervised	Forest + KNN + SV	S3-Dataset	EER of 7.28%, 9.29%, and 0.31% for sensors, statistics and voice. With a improvement of 71.97%, 62.14%, and 97.08%, respectively.

existing S3 Platform. Since this work not only extends but also compares the utility and performance of [6], for the sake of self-containment, a short description of the S3 platform is provided.

A. Industrial Scenario

As stated in the Introduction, operator authentication is one of the first lines of defense against attacks in any industrial factory. It is not enough to authenticate workers at the factory entrance, but it is also necessary to authenticate them throughout their working day without interruptions and without reducing operability and efficiency. Contrary to active authentications that tend to limit the operability of workers and reduce the usability of the system, a continuous authentication system does not. Different approaches can be followed to authenticate users, both unsupervised, such as outlier detection techniques, and supervised, such as classification. In this work, binary classification has been used for the supervised approach. A model per worker helps authenticate the user. Multiclass classification has been ruled out since the appearance of new workers would imply the retraining of the only model. This implies a more significant logistical complication as compared to the binary case, where only the model of the new worker would need to be generated.

The proposed application scenario is an Industry 4.0 compliant factory, where operators work with tablets, smartphones, or PDAs, to send instructions to the production machines. In doing so, many operators interact with these devices through touch screens or voice commands. At the same time, these operators must comply with individual protection standards wearing helmets, safety glasses, and gloves. This, in turn, affects the precision of mechanisms such as facial or iris recognition enough to render them unusable for CA purposes.

After analyzing the environment, it is understood that the most viable options for data sources are: the device sensors, since these are always present when the operator holds the device, the applications statistics, present always when the operator is using them, and voice, that it is present in the cases in which the operation supports voice commands.

Among the rest of biometrics, the following have been discarded: fingerprint (due to the fact that in most factories the operators wear protective gloves), facial and iris (similarly with safety glasses and masks). The keystroke has also been discarded, because most operators do not enter data, but scan barcodes, receive instructions through the device, or simply press very few buttons. Similarly, swipe biometry has been discarded for the same reasons. Although all these biometric dimensions have been discarded in this first scenario, the platform proposed in this work is very extensible and allows adding other biometrics to the framework, with little cost.

A noteworthy difference of this work with the one proposed in [6] is that, because it is an industrial environment, and the safety and efficiency of the worker are prioritized, the operators, when they operate with the device, usually do it in a single mode way. For example, if the operator uses voice commands, it is usual that (s)he uses it with voice only, since this allows him/her to have hands free in order to perform particularly hand use intensive tasks. Therefore, it has been considered that the number of instants where the data of all the selected dimensions is available at one exact moment will be minimal. In consequence, throughout this work, the joint behavior of all the different dimensions will not be evaluated.

One final important detail. The scenario proposed enables the use of supervised machine learning models or algorithms. Since the devices belong to the company, and because the biometric information is collected during work time, the use of data from workers to improve the models of the rest does not incur into data privacy violations. In this way, without getting into legal problems, supervised methods, which are well known to consistently show better performance than unsupervised models, can be applied.

B. S3 Platform

The S3 platform is an existing CA system for smartphones that collects data from sensors, applications, and voice to create users' behavioral profiles. S3, whose architectural design of the platform is composed of a smartphone application and a framework, uses unsupervised learning techniques to calculate the authentication level of subjects interacting with their mobile devices. For each user, once there is enough data collected by the application to generate the behavioral profiles, these are extracted by the framework using outlier detection algorithms, using only the user's data. More details of the S3 platform can be found in [6]. A new extension of the S3 for this work is proposed to include several devices, not only smartphones and to support supervised algorithms. A brief description is done below. See also Fig. 1.

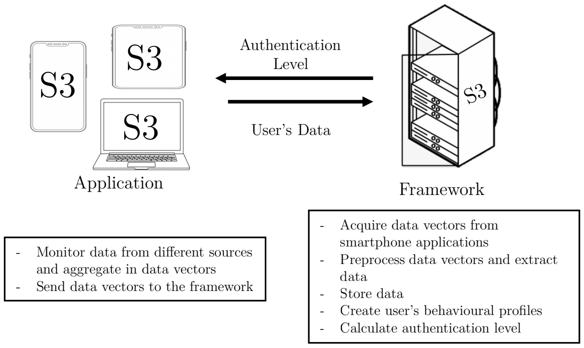
S3
S3
S3
Authentication Level
User's Data
S3
Application
Framework
- Monitor data from different sources and aggregate in data vectors
- Send data vectors to the framework
- Acquire data vectors from smartphone applications
- Preprocess data vectors and extract data
- Store data
- Create user's behavioural profiles
- Calculate authentication level

Fig. 1. Main parts and functionalities of the S3 platform.

The S3 Application was specifically designed for smartphones. However, this work extends this aspect and the application has been modified to be run on smartphones, PDAs, laptops, and computers. That is, any device from which workers operate machines and control processes. The application periodically monitors the three dimensions: sensors, statistics and voice, and collects available data. The data monitoring is carried out by the data acquisition module, through configurable time windows. For different types of data, attributes are assigned to vectors. These attributes are sent directly to the framework communication module, implemented on the server side.

The S3 Framework is the main component of the S3 Platform. In adapting to the industrial environment, its functionality has also been expanded to support supervised algorithms and enable the use of the data of all users to create each user's behavioral profile. The framework is composed of four modules. The communication module has two main functionalities; it receives vectors with data monitored by the applications and sends the authentication results to the smartphone application, once they have been computed. The data preparation module pre-processes data vectors and sends them to be saved in the platform. The storage module creates and maintains the users' datasets. Finally, the intelligent authentication module is in charge of training models containing users' behavioral profiles, and of evaluating vectors provided by smartphone applications to calculate users' authentication values in real time.

IV. Experimental Results

In this section, firstly, the dataset used is introduced. The dataset is not typical of the industrial scenario, but it can be extrapolated. Its choice and the relationship with the scenario are explained below and a brief description given. Next, a set of experiments are performed with the goal of answering the questions introduced in Section I. These experiments compare the performance of the existing (unsupervised machine learning) and new (supervised machine learning) approaches of the S3 platform, evaluate the relevance of each dimension, and assess the viability of the S3 platform as a CA system for the industrial factory.

A. Dataset

The compilation of a database in this environment is quite complex and requires cooperation and interest of an industry in the sector, which makes it extremely difficult to achieve. For this reason, for a first approach and version of the system, it was decided to use an existing, compatible dataset [21]. This dataset is valid and extrapolated to this scenario because it shares *(i)* a type of devices, smartphones, *(ii)* the data sources, *(iii)* a similar number of subjects in a warehouse, and *(iv)* actions related to the industry such as interaction with existing apps on the device. This dataset would be even more suitable if: i) it contained different user devices, ii) the time of data acquisition were limited to the working day, and iii) applications used were the same available in a company, not personal apps.

The dataset contains the behavior (sensors, statistics of applications, and voice) of 21 volunteers interacting with their smartphones for more than 60 days. The type of users is diverse. Males and females, in the age range from 18 until 70, have been considered in the dataset generation. The wide range of age is a key aspect, due to the impact of age on smartphone

TABLE II
DATABASE INFORMATION.

Characteristic	Number
Users	21
Sensors vectors	417.128
Statistics app's usage vectors	151.034
Speaker vectors	2.720
—– Call recordings	629
—– Voice messages	2.091

usage. For completeness, the content of the database is detailed below and an abbreviated explanation of how the information contained in the vectors has been calculated. Nevertheless, more specific details can be found in [21]. The vector data contained in the dataset is explained below, and Table II shows the number of vectors.

- Sensors vector. This type of vector contains data belonging to smartphone sensors (accelerometer and gyroscope) that have been acquired in a given window of five seconds. The process that extracts the sensor vector runs periodically every 20 seconds. The monitored features are:
  - Average of accelerometer and gyroscope values.
  - Maximum and minimum of accelerometer and gyroscope values.
  - Variance of accelerometer and gyroscope values.
  - Peak-to-peak (max-min) of X, Y, Z coordinates.
  - Magnitude for gyroscope and accelerometer [22].
- Statistics vector. These vectors contain data about the different applications used by the user in the last 60 seconds. Each vector of statistics is calculated every 60 seconds and contains:
  - Foreground application counters (number of different and total apps) for the last minute and the last day.
  - Most common app ID and the number of usages in the last minute and the last day.
  - ID of the currently active app.
  - ID of the last active app prior to the current one.
  - ID of the application most frequently utilized prior to the current application.
  - Bytes transmitted and received through the network interfaces.
- Speaker vector. This kind of vector is generated when the microphone is active, in a phone call, voice note, or voice command. Each time the microphone is activated, regardless of time, one speaker vector is generated. Once the application has the collected audio, it is resampled to 16 kHz, if necessary. The application then proceeds to calculate the vector that keeps the information about the speaker, this vector in the speaker recognition field is called "x-vector".

For the experiments, the dataset has been divided into two parts, train and test. To perform a fair comparison between the supervised and non-supervised system, the data of the first 14 days has been selected for training, which will generate the users' profiles, and data from day 15th up to 60th has been used for testing. The choice of the 15th day for the partition of the data is based on the results of [6], which showed that unsupervised systems needed that many days of data to generate profiles with high enough precision.

B. Experiment 1: Comparative of Supervised vs Unsupervised Methods

To answer the first research question, and assess the magnitude of the improvement that the supervised approach shows versus the unsupervised, the first experiment is done. This experiment considers a model for each user. These models are trained only with user data for unsupervised approach, and with data of other users as counterexamples for supervised. For a new data sample, the model decides if it is the user or if it is an impostor. To compare the two approaches, supervised and unsupervised, three families of models, that have both supervised and unsupervised versions, have been selected:

- Support Vector Machines: One-Class Support Vector Machine (OCSVM) for the unsupervised approach, and Support Vector Classification (SVC) for the supervised approach.
- Random Forest: Isolation Forest (IF) for the unsupervised approach, and Random Forest (RF) for the supervised approach.
- k-Nearest Neighbors: k-Nearest Neighbors Detector (kNND) for the unsupervised approach, and k-Nearest Neighbors Classifiers (KNNC) for the supervised approach.

The complexity of the algorithms is given by the parameters of each one, shown in Table III. For example, for KNN in the worst (brute force) case, it is $O(nXm)$, where n and m are the number of samples and number of dimensions in the training, respectively. For Random Forest, it is $O(TXD)$, where T and D are the size and maximum depth. And for SVM, with RBF kernel (the most used in this case), the complexity is $O(n_{SV}Xd)$, where n_{SV} is the number of support vectors and d is the input dimensionality.

Other families of algorithms could have been selected. However, these have been the final choices for this work because they are prevalent algorithms that present excellent results, as can be seen in the summary table in Section II, and have versions with similar underlying representation and learning principles for both supervised and unsupervised approaches, allowing for fair comparison. In addition, the use of neural networks has been ruled out because they require a large amount of data to obtain good results. Besides, different network architectures and learning would have to be used for each approach, autoencoders in the unsupervised one, and multilayer perceptron in the supervised one, and comparing them would be unfair. In order to facilitate independent replication, the code is available in [23].

The data vectors of the S3 dataset, split as previously described, have been used. In a preliminary step, a sweep on the main algorithms' parameters (for instance, size) was explored. A model for each user has been considered for each combination of the algorithms' parameters explored. To carry out this process, a repeated stratified K Fold cross-validation

TABLE III
LIST OF PARAMETERS EXPLORED FOR EACH ALGORITHM AND THE SELECTED VALUES.

Family	Algorithm	Params Explored	Sensors	Statistics	Voice
KNN	KNNC	Algorithm	auto	auto	auto
		leaf_size	10	10	10
		n_neighbors	10	100	100
		weights	uniform	distance	uniform
	KNND	n_neigbors	1	1	10
Forest	RF	criterion	entropy	entropy	entropy
		max_depth	100	30	30
		min_samples_split	2	2	2
		n_stimators	100	50	10
	IF	max_features	1	0.2	0.7
		max_samples	1	1	auto
		n_estimators	1	100	100
SVM	SVC	C	50	5	1
		kernel	rbf	rbf	linear
	OCSVM	kernel	rbf	rbf	rbf
		nu	0.1	0.1	0.7

TABLE IV
AVERAGE TRAINING TIME (SECONDS) FOR EACH ALGORITHM AND DIMENSION WITH THE SELECTED FINAL PARAMETERS.

Family	Algorithm	Sensors	Statistics	Voice
KNN	KNNC	0.0065	0.0614	0.0006
	KNND	0.3737	0.0410	0.0019
Forest	RF	14.5848	1.7319	0.3139
	IF	0.4705	0.2636	0.1654
SVM	SVC	813.4485	22.3902	0.0847
	OCSVM	0.4791	0.0703	0.0011

has been used with the Train set. The data has been preprocessed with a standard scaler and no further preprocessing was done. It has to be noted that, for the supervised models, the normalization extreme values have been calculated using the train data from all users, not just the user being modeled. Table III shows the algorithms' parameters where different settings were explored for each algorithm, indicating the final selected values for each of the data types. Results obtained during the parameter search for each parameter combination have been omitted due to space considerations.

Finally, after selecting the best parameters for each algorithm, the algorithms have been evaluated. To this end, the training was done with the train part complete, and the evaluation with the Test part (from day 15 onward). The average training time for each dimension and algorithm is shown in Table IV. The results for the three data types are shown in Table V. These results are global metrics for the system, all users models are evaluated together as only one.

As Table V shows, the best results for each family are obtained with the supervised algorithms, RF for sensors and statistics, and SVC for voice. This result is consistent with the general expectation that a supervised approach would work better than an unsupervised one. For each family of models, the supervised version significantly improves all metrics over the unsupervised version. Of special interest is the case of voice data and the SVM family, where the improvement is impressive.

If the behavior of each family in its two versions is analyzed individually, it can be found that KNN obtains a reduction of the EER of 47.96%, 37.48% and 94.26%, respectively. The Forest family, 71.97%, 62.14% and 88.32%. Finally, the SVC family has 71.56%, 34.11%, and 97.08%. Overall, the family that improved the most from the unsupervised to the supervised version was the Forest family, with an average reduction of 74.14%. Looking at the behavior of each type of data, one can see how voice is the data that most increases its security, improving the EER by more than 88 % for each of the families, followed by sensors and statistics.

Among these results, selecting the best for each approach, it should be noted that the EER reduction was 71.97% in sensors, going from 27.98% to 7.84%; 62.14% in statistics, dropping from 24.54% to 9.29%, and a surprising 97.08% in voice, decreasing from 10.63% to 0.31%. In the supervised approach, voice continues to be the biometric data with the highest security, as in the unsupervised approach. Meanwhile, sensors and statistics exchange their positions, sensor data being safer in the supervised version than the statistical data.

An analysis similar to that performed with the EER can be carried out with the AUC and F1 metrics present in Table V. In Fig. 2, the DET, Detection Error TradeOff, curves for each data type and for each algorithm are displayed. This figure analyzes the trade-off between the false acceptance rate (FAR) on the horizontal axis and the false reject rate (FRR) on the vertical axis. These curves are used to evaluate and compare the performance of a system for all possible thresholds. Two groups of lines can be seen in each of them, corresponding to the supervised (solid lines) and the unsupervised (dotted lines) approaches. The curves for the best system, sensors and statistics, show that, with a 2% of FAR, the system has a 20% of FRR. An interpretation of this fact could be the following: for this security level, 2% of FAR, the user could need an average of five attempts to authenticate with the sensors or statistics. However, at this same security level, voice has an FRR of near 0%, and, in most cases, the user will be authenticated at the first attempt.

These results answer the first question and support the idea that a system of this type has a place within industrial factories. But, what would happen if an unauthorized user managed to sneak into the factory? That is, how would the system behave towards unknown users? Also, every time a new worker joins, does the system need to retrain all the models of the rest of the workers? These research questions have been addressed in Experiment 2.

C. Experiment 2: Robustness of the system

The main objective of this experiment is to analyze the robustness of the system when unknown users appear in the

TABLE V

RESULTS IN THE TEST PART FOR THE BEST PARAMETERS FOR EACH ALGORITHM. HIGHLIGHTED IN BOLD THE BEST RESULTS FOR EACH METRIC ACCORDING TO THE TYPE OF DATA.

			Sensors			Statistics			Voice		
Family	**Algorithm**	**Type**	**EER ↓**	**AUC ↑**	**F1 ↑**	**EER ↓**	**AUC ↑**	**F1 ↑**	**EER ↓**	**AUC ↑**	**F1 ↑**
KNN	KNNC	Sup.	14,5691%	88,2707%	58,3670%	17,1857%	84,5600%	55,6019%	0,6147%	99,8663%	97,3884%
	KNND	Unsup.	27,9880%	78,2479%	29,8870%	27,4877%	81,7255%	31,7123%	10,6387%	94,3690%	64,6160%
Forest	RF	Sup.	**7,8462%**	**97,4835%**	**74,9495%**	**9,2923%**	**96,2081%**	**73,3675%**	1,5284%	99,8405%	94,7167%
	IF	Unsup.	34,5405%	71,6592%	26,4520%	24,5494%	83,7286%	36,5060%	13,0204%	94,3798%	61,6432%
SVM	SVC	Sup.	10,3630%	96,0787%	65,2465%	12,1353%	94,6106%	63,7220%	**0,3168%**	**99,9938%**	**98,7054%**
	OCSVM	Unsup.	36,4352%	69,2903%	14,9612%	18,4195%	88,3069%	37,8580%	42,8647%	57,1354%	24,9676%

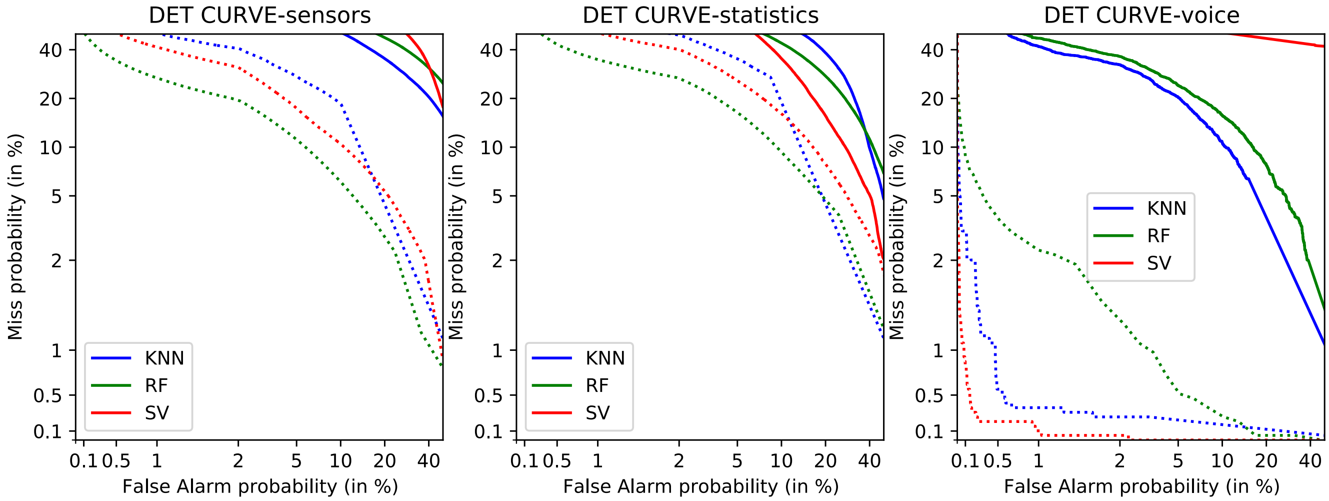

Fig. 2. Comparison of the DET curves for each of the selected model families. The unsupervised version is shown with the solid line, while the dotted line corresponds to the supervised version. The first graph is for sensors, the middle graph for statistics and the last one for voice. The colors indicate each of the families.

company (whether impostors or new employees), and answers the second and third research questions from the Introduction. To do this, available users are randomly divided into two groups, ”known” (workforce) and ”unknown” (impostors/new workers). In addition, the sizes of the groups are repeatedly changed. They range from 20 users in the workforce and one as an impostor/new worker (4%), down to 11 users in the workforce and 10 in the unknown group (47%). More divisions are not evaluated because it is not very realistic in the industrial scenario to update operative teams with more new workers than current employees in a workforce.

Once the groups are made, a model is generated for each worker (members of the known group). Next, the models are evaluated against the unknown user samples, and the metrics are calculated for the threshold of 0.5. In this experiment, one of the metrics to focus on is the FAR for unknown users. It allows the analysis of the degradation of the current trained models, when more and more impostors or new workers (data not used during training) are presented to the models. FAR shows cases wrongly classified as an actual worker and, in case it increases significantly, it will indicate that addition of new workers renders old models obsolete, and therefore they must be immediately retrained.

The data preprocessing is the same as for the first experiment. The whole setup has been carried out 20 times, each with a different combination of known (and unknown where applicable) users. The code and the pipeline followed can be found in [23]. Confidence intervals at 95% for the FRR of known users and the FAR of the samples of unknown users are calculated to be analyzed and are shown in Fig. 3.

As can be seen in Fig. 3, the FRR is progressively reduced from near 40% for sensors and statistics. For voice, the FRR is lower than 5%. This reduction of the FRR is because the models are facing fewer and fewer users and are more confident to identify the corresponding worker. FFR is not so interesting as FAR for this experiment, but it has been included to have a complete view of the system performance. The FAR is used to evaluate the system robustness by analyzing how it evolves as it faces new unknown users.

Regarding FAR, it steadily increases, linearly, as more unknown users are confronted. Looking at trends of the three data dimensions, statistics is the most degraded and voice the least, in absolute terms. The FAR obtained when only a single new unknown user appears is at most (upper confidence interval value) 1.91% for sensors, 3.49% for statistics, and 0.85% for voice, which is very promising. In any case, these

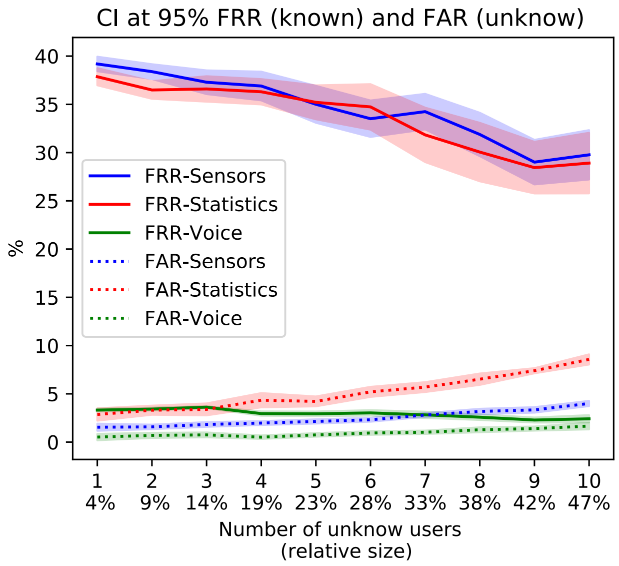

Fig. 3. Confidence intervals (at 95%) for the FRR (known users) and FAR (unknow users) at threshold 0.5.

values are below 4%, the relative size of the unknown users. This is the degradation to expect if the new user is always accepted as a legitimate worker.

In any case, the trend of FAR for unknown users of the system, as more and more unknown users are added, shows a slight linear increase. Especially with voice, with less than 2% even if the number of new workers is the same as the current workforce size. Although it is always a good practice to retrain models periodically when more user data is available, the results suggest that, when new workers are included, it is not critical to retrain the existing workers models as soon as they are hired. The retraining can be done whenever the system administrator has the computer resources available, or scheduled updates are due, instead.

V. Conclusion

Since security is a priority in an industrial 4.0 ecosystem, this work analyzes the suitability of a continuous authentication system based on three different types of data: sensors, statistics, and voice. For this, a continuous authentication system that already works, the S3 platform, has been selected, and the necessary modifications have been applied to run in this environment. Also, because the platform is deployed in an industrial environment, and data can be collected from company-owned devices during business hours with no privacy issues, supervised and unsupervised approaches are used and analyzed. In other words, in order to generate the behavior profiles of each user, the data of other users is used. For this reason, the first experiment in this work focuses on analyzing if the supervised approach will have as expected) greater precision than the unsupervised one, and quantifies how much it improves upon it. Besides, in this first experiment, a significant improvement of more than 88% is obtained for each of the different data types. This fully answers the first research question.

Next, the system robustness is evaluated, and research questions 2 and 3 are answered. For this, a second experiment is proposed in which a group of users representing unknown users (new workers/impostors) is separated. The models of known users are trained with only known user samples, and later unknown user samples are evaluated as impostors. The size of the unknown users ranges from 1 to 10 users (from 4% to $\sim 47\%$ the relative size of the workforce set). Results show that the performance degradation for supervised approaches is more intense for the statistics than for the rest, and the voice still ends up as the best data dimension. Moreover, if an unknown user appears, or our workforce increases with a single new member, the FAR with this new user is only, at most, 1.91% for sensors, 3.49% for statistics, and 0.85% for voice.

In conclusion, the results of both experiments propose this system as a suitable candidate system for the continuous authentication of workers in factories adhered to Industry 4.0. Although experiment 2 shows a reasonably high FRR, the system performance in production is not hampered. With such FRR, the user will be able to authenticate 1 out of 2 times with the sensors and statistics. With the voice, the user will always be able to do so. In the case of sensors, this implies just 40 seconds, and 2 minutes for statistics, almost negligible times for a continuous authentication system. Furthermore, the system could be configured so that, after two failed authentication attempts due to sensors or statistics, it requests an active voice authentication to confirm user identity, and discard a potential impostor attack.

Future work will be focused on trying to solve some of the limitations of the current work, and include: i) replacing the application-server distribution that forces to send the data to the server, ii) the use of algorithms oriented to improve the performance and not being restricted to particular ones for the sake of comparison value and scientific proof evidence, iii) improving privacy considerations, or iv) working on the limitations of the dataset, such as obtaining a proper industrial dataset or one with more characteristics typical of the industrial scenario.

Other work will focus on improving the robustness of the system and testing it in real conditions. For the former, two main lines will be followed. The first line investigates the optimal impostor selection mechanisms for supervised systems training. The second line of work focuses on evaluating the robustness of adversary attacks. Finally, given that the results show the viability of the implementation in a factory adapted to Industry 4.0, the necessary agreements will be sought to bring the working prototype to production, and evaluate it in a day-by-day industrial environment operation.

References

[1] P. C. Alcaraz, P. Y. Zhang, P. A. Cardenas, and P. L. Zhu, “Guest editorial: Special section on security and privacy in industry 4.0,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6530–6531, 2020.

[2] M. Potdar, E. Chang, and V. Potdar, “Applications of rfid in pharmaceutical industry,” in *2006 IEEE International Conference on Industrial Technology*, 2006, pp. 2860–2865.

[3] S. Modi and S. Elliott, "Securing the manufacturing environment using biometrics," in *Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology*, 2005, pp. 275–278.

[4] Z. Ma, Y. Yang, X. Liu, Y. Liu, S. Ma, K. Ren, and C. Yao, "Emir-auth: Eye movement and iris-based portable remote authentication for smart grid," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6597–6606, 2020.

[5] D. Osorio Roig, P. Drozdowski, C. Rathgeb, A. Morales González, E. Garea-Llano, and C. Busch, "Iris recognition in visible wavelength: Impact and automated detection of glasses," in *2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, 2018, pp. 542–546.

[6] J. M. Espín López, A. Huertas Celdrán, J. G. Marín-Blázquez, F. Esquembre, and G. Martínez Pérez, "S3: An ai-enabled user continuous authentication for smartphones based on sensors, statistics and speaker information," *Sensors*, vol. 21, no. 11, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/11/3765

[7] P. Kumar and G. S. Gaba, *Biometric-Based Robust Access Control Model for Industrial Internet of Things Applications*. John Wiley & Sons, Ltd, 2020, ch. 7, pp. 133–142. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119527978.ch7

[8] M. Golec, S. S. Gill, R. Bahsoon, and O. Rana, "Biosec: A biometric authentication framework for secure and private communication among edge devices in iot and industry 4.0," *IEEE Consumer Electronics Magazine*, vol. 11, no. 2, pp. 51–56, 2022.

[9] N. Mehdi and B. Starly, "Witness box protocol: Automatic machine identification and authentication in industry 4.0," *Computers in Industry*, vol. 123, p. 103340, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361520305741

[10] I. Islam, K. M. Munim, M. N. Islam, and M. M. Karim, "A proposed secure mobile money transfer system for sme in bangladesh: An industry 4.0 perspective," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2019, pp. 1–6.

[11] Y. Borgianni, E. Rauch, L. Maccioni, and B. G. Mark, "User experience analysis in industry 4.0 - the use of biometric devices in engineering design and manufacturing," in *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2018, pp. 192–196.

[12] J. Jorquera Valero, P. Sánchez Sánchez, L. Fernández Maimó, A. Huertas Celdrán, M. Arjona Fernández, S. De Los Santos Vílchez, and G. Martínez Pérez, "Improving the security and qoe in mobile devices through an intelligent and adaptive continuous authentication system," *Sensors*, vol. 18, no. 11, p. 3769, Nov 2018. [Online]. Available: http://dx.doi.org/10.3390/s18113769

[13] P. M. Sánchez Sánchez, L. Fernández Maimó, A. Huertas Celdrán, and G. Martínez Pérez, "Authcode: A privacy-preserving and multi-device continuous authentication architecture based on machine and deep learning," *Computers & Security*, vol. 103, p. 102168, 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404820304417

[14] Y. Barlas, O. E. Basar, Y. Akan, M. Isbilen, G. I. Alptekin, and O. D. Incel, "Dakota: Continuous authentication with behavioral biometrics in a mobile banking application," in *2020 5th International Conference on Computer Science and Engineering (UBMK)*, 2020, pp. 1–6.

[15] H. C. Volaka, G. Alptekin, O. E. Basar, M. Isbilen, and O. D. Incel, "Towards continuous authentication on mobile phones using deep learning models," *Procedia Computer Science*, vol. 155, pp. 177 – 184, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187705091930941X

[16] A. Acien, A. Morales, R. Vera-Rodriguez, and J. Fierrez, *Mobile Active Authentication based on Multiple Biometric and Behavioral Patterns*. Cham: Springer International Publishing, 2020, pp. 161–177. [Online]. Available: https://doi.org/10.1007/978-3-030-39489-9_9

[17] Y. Li, H. Hu, Z. Zhu, and G. Zhou, "Scanet: sensor-based continuous authentication with two-stream convolutional neural networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 16, no. 3, pp. 1–27, 2020.

[18] Y. Li, J. Luo, S. Deng, and G. Zhou, "Cnn-based continuous authentication on smartphones with conditional wasserstein generative adversarial network," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5447–5460, 2022.

[19] L. Hernández-Álvarez, J. M. de Fuentes, L. González-Manzano, and L. Hernández Encinas, "Smartcampp - smartphone-based continuous authentication leveraging motion sensors with privacy preservation," *Pattern Recognition Letters*, vol. 147, pp. 189–196, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865521001434

[20] L. Wang, M. S. Hossain, J. Pulfrey, and L. Lancor, "The effectiveness of zoom touchscreen gestures for authentication and identification and its changes over time," *Computers & Security*, vol. 111, p. 102462, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404821002868

[21] J. M. Espín López, A. Huertas Celdrán, J. G. Marín-Blázquez, F. Esquembre, and G. M. Pérez, "S3 dataset," https://figshare.com/articles/dataset/S3Dataset_zip/14410229/2, Apr 2021.

[22] M. Ehatishamul Haq, M. Awais Azam, U. Naeem, Y. Amin, and J. Loo, "Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing," *Journal of Network and Computer Applications*, vol. 109, pp. 24 – 35, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804518300717

[23] J. M. Espín López, "S3 platform for the industry 4.0," https://github.com/bazako/S3Platform_Industry40, 2022.

Juan M. Espín López received the M.Sc. degree in mathematics from the University of Murcia. He is currently pursuing his PhD in computer science at the University of Murcia. His research interests are focused on CA, speaker recognition, facial recognition, anti-spoofing systems and the application of machine learning and deep learning to the previous fields.

Alberto Huertas Celdrán received the M.Sc. and Ph.D. degrees in computer science from the University of Murcia, Spain. He is currently with the Communication Systems Group (CSG), Department of Informatics (IfI), University of Zürich UZH. His scientific interests include IoT, brain-computer interfaces (BCI), cybersecurity, data privacy, artificial intelligence, semantic technology, and computer networks.

Francisco Esquembre is Full Professor in the Department of Mathematics of the University of Murcia, Spain. His scientific activity is mostly devoted to mathematical modeling and computer simulation of physical and engineering phenomena, developing the Easy Java Simulations modeling tool. He is currently interested in the application of data analysis to different practical problems.

Gregorio Martinez Pérez is Full Professor in the Department of Information and Communications Engineering of the University of Murcia, Spain. His scientific activity is mainly devoted to cybersecurity and networking. He is working on different national and European IST research projects related to these topics, being Principal Investigator in most of them. He has published 160+ papers in national and international conference proceedings, magazines and journals.

Javier G. Marín-Blázquez received both Computer Science (1994) and Psychology (2012) degrees by the University of Murcia, and a M.Sc. (2001) and PhD (2003) in Artificial Intelligence by The University of Edinburgh. His research interests include: Artificial Intelligence (AI), Machine Learning, Fuzzy Systems, Soft Computing, Cybersecurity, AI for Games and Cognitive Science.