IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 November 2022, accepted 3 January 2023, date of publication 10 January 2023, date of current version 19 January 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3235969*

RESEARCH ARTICLE

# A Hybrid Approach With GAN and DP for Privacy Preservation of IIoT Data

**YAVUZ SELIM HINDISTAN AND E. FATIH YETKIN**
Department of Management Information Systems, Kadir Has University, 34083 Istanbul, Turkey
Corresponding author: Yavuz Selim Hindistan (yavuzselim.hindistan@stu.khas.edu.tr)

**ABSTRACT** There are emerging trends to use the Industrial Internet of Things (IIoT) in manufacturing and related industries. Machine Learning (ML) techniques are widely used to interpret the collected IoT data for improving the company's operational excellence and predictive maintenance. In general, ML applications require high computational resource allocation and expertise. Manufacturing companies usually transfer their IIoT data to an ML-enabled third party or a cloud system. ML applications need decrypted data to perform ML tasks efficiently. Therefore, the third parties may have unacceptable access rights during the data processing to the content of IIoT data that contains a portrait of the production process. IIoT data may include hidden sensitive features, creating information leakage for the companies. All these concerns prevent companies from sharing their IIoT data with third parties. This paper proposes a novel method based on the hybrid usage of Generative Adversarial Networks (GAN) and Differential Privacy (DP) to preserve sensitive data in IIoT operations. We aim to sustain IIoT data privacy with minimal accuracy loss without adding high additional computational costs to the overall data processing scheme. We demonstrate the efficiency of our approach with publicly available data sets and a realistic IIoT data set collected from a confectionery production process. We employed well-known privacy six assessment metrics from the literature and measured the efficiency of the proposed technique. We showed, with the help of experiments, that the proposed method preserves the privacy of the data while keeping the Linear Regression (LR) algorithms stable in terms of the R-Squared accuracy metric. The model also ensures privacy protection for hidden sensitive data. In this way, the method prevents the production of hidden sensitive data from the sub-feature sets.

**INDEX TERMS** Data privacy, differential privacy, generative adversarial networks, IIoT, privacy metrics.

## I. INTRODUCTION

There are emerging trends for monitoring industrial production processes by sensor devices. With the Industrial Internet of Things (IIoT), a massive amount of collected data is used to forecast production tools' aging or failures by Machine Learning (ML) techniques. Some examples are the detection of equipment failure, supply chain optimization, performance monitoring, and predictive maintenance [1]. IIoT is an infrastructure of software and hardware that connects the physical world of the company processes with the Internet [2]. IIoT devices typically have limited computing power and small memories [3]. In general, this data analysis is performed offline using a cloud-based service.

Transferring the data from the production place to the processing place on the cloud is a significant bottleneck. The transmitted data will be open to third eyes, which can cause privacy leaks related to the company [4], [5]. The data from IIoTs often incorporate implicit or explicit knowledge of specific production areas gained from experience. It is key to unlocking new opportunities and can reveal how industries operate. Recent events have shown apparent failures in cybersecurity issues. They are becoming a real threat to IIoTs' competitive advantage and profitability [6], [7]. In this context, ensuring that the IIoT data is handled with the highest care regarding privacy is of tremendous importance. This paper will focus on the problem arising from the processing needs of IIoT data on cloud systems.

As shown in Figure 1, the IIoT data transmission to third parties can be performed safely by applying encryption procedures. In most cases, ML applications need decrypted data

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

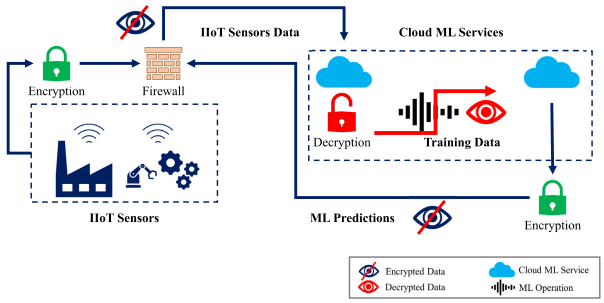


**FIGURE 1.** Conceptual representation of IIoT transmission.

to perform ML tasks efficiently. Therefore, the third parties may have unacceptable access rights during the data processing to the content of IIoT data that contains a portrait of the production process.

In this paper, we proposed a novel hybrid model to ensure the privacy of IIoT data for both direct and hidden sensitive features while preserving the accuracy of the target ML approach. Our method is based on the hybrid usage of Generative Adversarial Networks (GAN) and Differential Privacy (DP). The DP was introduced by Dwork to sustain privacy preservation with random noise and parameters [16]. Hitaj et al. [17] mentioned privacy violation cases with DP. It is insufficient to preserve privacy in shared learning against active adversaries [18]. Liu et al. [19] showed the DP ineffectiveness in data having solid correlations, such as social, medical, or mobile data. We know those explanations do not violate the DP since the DP can be parameterized. The GAN is an adversarial method that generates new data by learning the probability distribution from an original data set [20]. Liu et al. used a GAN variant to generate a synthetic copy of the sensitive data to protect its privacy [21]. We improved our model's privacy preservation abilities by combining GAN and DP methods. It is a new hybrid privacy-preserving model to maintain privacy-aware edge-based ML systems. We aim to sustain IIoT data privacy with minimal accuracy loss without adding high additional computational costs to the overall data processing scheme. We selected Linear Regression (LR) as our target ML approach for its simplicity and vast usability opportunities in real-world industrial applications. We also employed well-known privacy assessment metrics from the literature to measure the efficiency of the proposed technique. We used four data sets (wind turbine, steam production, energy efficiency, synchronous motors) in our experiments. The mentioned data are collected from several Supervisory Control and Data Acquisition (SCADA) systems. They are publicly available on a repository of community-published web sources except for the private steam production data set obtained from a food and confectionery manufacturer. We use the steam production data set to demonstrate the efficiency of the proposed method in a realistic environment. The details of the data sets and involved LR applications are given in detail in Section IV.

The main contributions of this study can be summarized as follows:

1) A novel hybrid privacy protection mechanism based on GAN and DP is developed.
2) The numerical experiments show that the proposed method preserves the privacy of the data while keeping the LR algorithms stable in terms of the *R-Squared* accuracy metric.
3) The model also ensures privacy protection for hidden sensitive data. In this way, the method prevents the production of hidden sensitive data from the sub-feature sets.
4) The efficiency of the method is tested on four different IIoT SCADA data sets: a) wind turbine (publicly available [22]), b) steam production belongs to a confectionery producer, c) energy efficiency (publicly available [23]), d) synchronous motors (publicly available [24]).

The remaining of this paper is organized as follows. The second section presents an overview of the literature on data privacy preservation, including GAN and DP implementations with IoT data protection approaches. The third section discussed the proposed methodology and its mathematical background. Furthermore, the privacy metrics used to assess the model's efficiency are defined in this section. Section IV describes the four data sets involved in the study, and the numerical results obtained from the proposed model are evaluated. In the last section, we assess the overall benefits of the proposed model and discuss our future research directions.

## II. RELATED WORK

Data privacy concerns for IoT devices have dramatically increased in recent years, in parallel with the importance of data security. The IoT data protection literature mainly focuses on infrastructure and sensor device security. Tahsien et al. grouped the possible attack types related to IoT [5]. They mentioned passive attacks as attackers collect victims' data without permission. They called privacy leakage, sharing private IoT data with others without data owners' awareness.

The cloud is mainly preferred for IoT solutions for data storing and ML operations [29]. Stout et al. prepared a survey about IoT security. They mentioned IoT devices' large-scale data mining security risks in their interactions with the cloud [30]. Although there are a lot of research and proposals for protecting the stored IoT data in the cloud, their majority focus techniques on encryption of the data in the cloud and secure connection between IoT sensors and cloud storage [4], [12], [29], [31]. The data in the cloud need to be decrypted for ML operations. IoT data owners prefer to sign Non-Disclosure Agreements (NDA) with cloud providers for legal prevention. There is a high privacy risk. Shokri et al. showed empirically that machine-learning-as-a-service platforms could lead to data leakage in their training data sets [32]. Jeong et al. proposed sending partially processed data to the cloud instead of raw IoT data.

They suggested using DL operation to ensure irreversibility [10]. Their research is one of the different approaches to IoT-Cloud Data security. It is based on partially processed neural network (NN) computations' irreversibility. They also mentioned homomorphic encryption's inconveniences.

Hitaj et al. showed that GAN usage could break DL methods, and privacy can be disclosed [17]. Another research to prevent security is the distributed/federated learning approach. Zeng et al. presented a framework for federated learning to ensure IoT data privacy and security [33]. Qi and Wei discussed that the method is not convenient for industrial IoT platforms since the data are generated at a high rate, and collecting data among clusters may create a waiting queue [12].

In the literature, several approaches exist for using encrypted data in the ML operations, such as homomorphic encryption [8] and cryptographic techniques [9]. Comparing them with standard ML techniques that process decrypted data can be very costly in computation. The homomorphic approach aims to keep the data encrypted and perform ML prediction operations on encrypted data using Deep Learning (DL) methods. Although the technique enables ML operation on the encrypted data set, it still needs decrypted data to train the pattern [8], [10]. The cryptographic techniques can be considered a secure multiparty computation. They use DL methods to perform distributed/federated learning. The construction and employment of a distributed computing environment may be too costly for an industrial production environment [11]. Sensors generate records at high speed and need to be processed fast [12]. DL models do not have strong explainability due to their black-box structure, which is essential for industrial data analysis [13]. Moreover, even in a secure cloud system, various recent hardware-centric-cyber-attack types are based on exploiting cache-based side channels such as Meltdown and Spectre. These attack types focus on extracting sensitive information from shared computational resources [14].

Besides all these potential risks regarding the IIoT data, the data may contain direct or hidden sensitive data. It is the case when some feature subset selected from the data allows the production of a set of sensitive data with a proper ML model. Malekzadeh et al. mentioned those possible threats against privacy when raw sensor data are shared with cloud-assisted applications [15]. Although their discussion is on personal wearable or portable devices, similar threats exist for industrial systems. For instance, collected raw material consumption IIoT data may disclose the data owner's production cost with profit margins and expose related product recipes. The IIoT data can also display the production area structure with possible production bottlenecks desirable for competitors, attackers, and potential investors. All these concerns prevent companies from sharing their IIoT data with third parties.

To make more clarification, we prepared this section in the subsections.

### A. DP IMPLEMENTATIONS WITH IoT DATA PROTECTION APPROACH

The DP is privacy protection by injecting some random noise. In 2006, Dwork et al. proposed a complete approach for DP [16]. It is different from a Data Anonymization (DA) technique. There is new research on DA, such as [26]. The DA generally removes individual names/identities from the data sets, creating primary vulnerabilities against linkage attacks. A linkage attack can be considered a threat. It is possible to combine the pieces of apparently anonymous data from several different sources to reveal the real identities of individuals. In 1997, Sweeney showed that the remaining data after an anonymization operation could be sufficiently disclosed by a linkage attack [27]. As shown in a related study, it is possible to identify 87% of all Americans with only three pieces of information: zip code, birthday, and gender. In 2008, It was proven that it is possible to identify people from the Netflix anonymized data published in 2006, with publicly available user data of IMDb(International Movie Database) [28].

Husnoo et al. listed privacy preservation in IoT-enabled systems in a survey in 2021 [25]. They mentioned the application and implementation of DP in IoT infrastructures. Jiang et al. [51] showed DP implementation alternatives for IIoT in a survey of 2021. In that study, DP implementations were summarized into seven groups: IoT-related industrial logistics, IoT of smart cities, IoT in bioengineering, IoT of autonomous vehicles, IoT of intelligent manufacturing, blockchain in the industrial IoT, and social network in the industrial IoT. Although each of those groups is another subject for a detailed study, we can clearly say that data protection needs have become a crucial topic, and the DP is one of the attractive approaches. The DP is still in the investigation stage for IIoT topics, especially in lowering latency to enhance efficiency [51], [52]. The following concerns related to the DP for IoT applications include integration complexity of business and industry, big data, real-time conditions, privacy optimization, and development of industrial applications [51]. Yang et al. [53] researched the implementation of DP in the IoT to confirm data utility. Liang et al. [54] suggested an edge computing-based DL model, which can relocate DL between cloud servers and edge solutions. Xu et al. [41] presented the GANobfuscator about information leakage using GAN based on DP in IoT systems. The IoT environment contains some sensitive data, and some data owners are unwilling to share the data to the cloud for proper purposes. The DP application load balancing over nodes in edge training is an open issue [51]. Feng et al. [55] presented an edge computing solution to preserve privacy. Usman et al. [56] depicted RaSEC as an intelligent framework. The DP algorithm in real-time conditions is a problem for IIoT. It requires fast computing and transmission. Langarica et al. [57] investigated fault identification in industrial motors with fast computing IIoT. One of the main problems of IIoT is preserving cloud data privacy. Hu et al. [58] mentioned the fog computing framework. Sim-

ilarly, there are alternative studies [55], [59]. Our approach as GAN and DP presents a new approach to this problem.

### B. GAN IMPLEMENTATIONS WITH IoT DATA PROTECTION APPROACH

In 2014, Goodfellow et al. designed the GAN as an adversarial process [20]. It is based on simultaneously training models G and D. G is a generative model that tries to catch the data distribution. D is a discriminative model that simultaneously attempts to guess the probability that a sample came from the training data rather than G. Some research uses GAN for privacy protection by creating a synthetic copy of the original data [34], [35], [36], [37].

Besides the GAN algorithm, several alternative ways exist to make synthetic data statistically. Ping et al. showed a Bayesian correlation among features in the data synthesizer [38]. The primary advantage of using GAN instead of other statistical models is the GAN's effectiveness in approximating the actual distribution. Xu and Veeramachaneni showed that GAN could generate high-quality synthetic data to benefit data science [39]. Xu et al. in 2019 introduced the synthetic conditional tabular GAN (CTGAN) as tabular data generation. CTGAN could learn better distributions than Bayesian solutions [40].

We can not see many samples of GAN usage for data protection in IoT. Huang et al. [60] mentioned in their study to use of the GAN for IoT data protection in medical data. The authors confirmed that a few studies exist about privacy-preserving in industrial IoT. They used Generative Adversarial Imitation Learning (GAIL) as an alternative GAN method with backward reinforcement learning. Shahid et al. [61] presented a method with GAN for IoT Network Traffic Generation. They used GAN for categorical data generation and combined it with a WGAN autoencoder.

### C. GAN AND DP HYBRID IMPLEMENTATIONS

There are some hybrid approaches for privacy preservation based on GAN and DP. However, to our knowledge, there is no such study on IoT-enabled systems [25]. For example, Xu et al. named GANobfuscator, using GAN and DP to preserve the privacy of semantic-rich data (e.g., image, text, audio, video) [41]. Liu et al. proposed a model with DP and GAN to protect sensitive data such as patient medical records [21]. Studies about hidden sensitive data mainly exist in data mining rule associations. Cheng et al. mentioned that proper mining could disclose confidential data and proposed a rule-hiding approach [42]. One of the first studies avoiding association rules uses data sanitization [43]. Jones et al. [35] proposed a method for protecting clinical medical data by combining GAN and DP.

## III. METHODOLOGY

We propose a novel approach that fits the privacy requirements for industrial IoT (IIoT) environments. The proposed model transforms the raw data obtained from IIoT sensors into a protected data set before transmitting it to third parties. We used a hybrid mechanism based on GAN and DP for this aim. The transformation process is scalable for a medium-size, affordable computational device that can be a local server located on the company premise. It needs a preliminary analysis of the data. This section will briefly introduce the building blocks of the proposed approach and evaluation metrics. Then, the algorithm of the proposed methodology will be depicted.

### A. THE GENERATIVE ADVERSARIAL NETWORKS (GAN)

GAN is defined as two neural networks that are working against each other: Generator($G$) and Discriminator($D$) [20]. GAN can generate a fake/synthetic data set ($X_G$), and this fake data set gets the same statistics similar to a given test data set ($X_O$). GAN trains simultaneously $G$ and $D$. $G$ generates a synthetic copy ($X_G$). It uses input value as random noise $z \sim p_z(z)$ and tries to create fake outputs $x \sim p_{data}(x)$ that have a similar distribution to the given data set, where $p_z$ represents the probability distribution of random noise. At the same time, $D$ learns how to differentiate real from synthetic ones. It calculates whether the probability output comes from the given data set rather than fake $G$. The GAN trains $D$ to maximize the probability of assigning the correct label to training examples and samples from $G$. Besides, the GAN also trains $G$ to minimize $\log(1 - D(G(z)))$ simultaneously. The following objective function $V$ represents these two training procedures:

$$\min_G \max_D V(D, G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

where $E$ represents the expected value.

In this paper, we used Conditional Tabular GAN (CTGAN) [40] approach to creating a synthetic copy of the original data sets. It is a synthetic data generation method with GAN for tabular data distribution. It uses mode-specific normalization, and it processes each column independently. The basic principle of CTGAN and instead of GAN is that the GAN works with the image data. We used CTGAN to create synthetic tabular data.

### B. THE DIFFERENTIAL PRIVACY (DP)

Dwork introduced the DP as a privacy preservation technique using random noise [16]. DP perturbs the data to add an uncertainty that does not affect the overall probability distribution of the original data. Since the noise level and distribution are known, they can be compensated during the statistical evaluation. The DP defines $(\epsilon, \delta)$-differential privacy where $\epsilon$ represents differential privacy, and $\delta$ is the probability of error such as disclosing information. In other words, $\delta$ defines the probability of DP failure. Assume that $\mathcal{A}$ is a randomized algorithm, $\mathbb{N}$ is a domain of all non-negative integers, including zero, and $x$ and $y$ are any two data sets such that all $x, y \in \mathbb{N}$. The algorithm $\mathcal{A}$: $\mathbb{N} \rightarrow \mathcal{R}$ with range $\mathcal{R}$ satisfies $(\epsilon, \delta)$-differential privacy for any subset of outputs

$\mathcal{S} \in \mathcal{R}$ with the following definition,

$$Pr\left[\mathcal{A}(x) \in \mathcal{S}\right] \leq exp(\epsilon) Pr\left[\mathcal{A}(y) \in \mathcal{S}\right] + \delta \tag{2}$$

where $\epsilon$ corresponds to the maximum $l_1$ distance between $x$ and $y$, such as $\|x - y\|_1 = \sum_{x,y \in \mathbb{N}} | x_i - y_i | \leq 1$.

The Laplace distribution spreads data over a more extensive range. When it is used in differential privacy, it increases anonymity [44]. Mainly, the Laplace distribution is used to add noise centered at 0 with the parameter $b = \frac{1}{\epsilon}$: $Lap(x \mid b) = \frac{1}{2b} exp(-\frac{|x|}{b}))$. The DP can be defined as a function $f : \mathbb{N} \rightarrow \mathcal{R}$ with the Laplace distribution, as given follows,

$$\mathcal{A}_f(\mathbb{N}) \triangleq f(\mathbb{N}) + Lap(x \mid b). \tag{3}$$

To perform the DP approach, in this study, we randomly added noise obtained from Laplace distribution to only sensitive features determined by the data owner. The Algorithm-2 shows the proposed model DP approach between 13 and 20 lines. We also test DP ($X_D$), where we add noise to all columns. The Algorithm-1 shows the DP ($X_D$) approach. The following subsection briefly introduces the privacy metrics used in the proposed algorithm to measure its privacy preservation performance.

**Algorithm 1** DP($X_D$)

**Require:** Original data set: $X_O \in \Re^{m \times n}$.
**Ensure:** DP data set: $X_D$.
1: **for** $j = 1$ to $n$ **do**
2:  **for** $k = 1$ to $m$ **do**
3:  $\theta =$ randn(0, 1)
4:  **if** $(\theta \geq 0.5)$ **then**
5:  $X_D(k, :) = X_O(k, :)+$ Laplace $(\epsilon)$
6:  **end if**
7:  **end for**
8: **end for**
9: **return** $X_D$

### C. PRIVACY METRICS

The proposed approach is based on creating perturbed sensitive data from the original data. Therefore, several reliable metrics are needed to evaluate how much the proposed model differs from the original. We used the following six metrics in our tests: a) Mean Square Error (MSE) error-based metrics [45], b) Normalized Variance (VAR) data similarity metrics [45], c) Directed Hausdorff (HAUS), a metric space between two data sets [46], d) Kullback-Leibler Divergence (KL) information gain or loss metrics [45], [47], e) Procrustes (PRO) a comparison of the shape of the probability distribution of given two data sets [48], and f) Pearson's Correlation Coefficient (PCC) information gain or loss metrics [45]. When the perturbed data differs enough from its origin, MSE, VAR, HAUS, KL, PRO, and PCC metrics produce a value different from zero. Similarly, when the metrics get zero value, we can say there is no privacy protection.

We prefer those six metrics due to their measurement approaches. MSE and HAUS use Euclidean to show differences and metric distances of data sets, respectively. VAR uses statistical variance and shows dispersion between two data sets. PCC shows the linear dependency of data sets. The KL uses entropy to show statistical distances, and the PRO shows shape analysis of two data sets. For example, we can add the same constant value to each item to differentiate a data set. In this case, MSE, KL, and HAUS metrics show that the new data set differs from the original. VAR, PCC, and PRO metrics show no difference. Thus, we can understand there is a difference, but it may not be enough for privacy protection. Although it makes a different data set, its original version can be easily analyzed. The six metrics help us assess the differences between data sets from a different point of view. Indeed, each metric defined in this section measures and compare a different property of the transformed data set with the original data set. Hence, if all of these metrics behave in a similar way, one can be sure that the privacy of the data set is ensured.

#### 1) MEAN SQUARED ERROR (MSE)

The MSE is a well-known statistical metric. We use it as a privacy metric to measure the Euclidean difference between the original and test data set features. MSE statistically aims to minimize the mean of squared errors. It ensures better privacy when it diverges from zero to higher values.

$$P_{MSE} = \frac{1}{n} \sum_{x \in X} \|x - y\|^2, \tag{4}$$

Here, $X$ is the original data set ($x \in X$), $Y$ is the test data set ($y \in Y$), and $n$ corresponds to the number of observations.

#### 2) NORMALIZED VARIANCE (VAR)

VAR measures the distribution difference between the $X$ and $Y$ [49]. The metric is based on the statistical variance $\sigma^2$. It shows the dispersion between $X$ and $Y$:

$$P_{VAR} = \frac{\sigma^2(X - Y)}{\sigma^2(X)} \tag{5}$$

VAR as a privacy metric ensures better privacy when it has a higher variance.

#### 3) DIRECTED HAUSDORFF (HAUS)

HAUS measures the distances of two subsets in metric space. It measures the similarity according to the corresponding data set position in metric space [46]. It equals zero when the two data sets are perfectly matched. If higher than zero, the data sets can be considered independent and have higher privacy. Note that higher values mean better privacy protection. The HAUS can be defined as the maximum set distance to the nearest point in the other set as follows,

$$P_{HAUS} = max_{x \in X}(min_{y \in Y}(d(x, y))). \tag{6}$$

Here, $x$ and $y$ are points of sets $X$ and $Y$ orderly. The $d(x, y)$ denotes any metric between these points, such as the Euclidean distance between $x$ and $y$.

#### 4) KULLBACK-LEIBLER DIVERGENCE (KL)

KL is called relative entropy, and it measures statistical distances between data sets by comparing the data set's probability distributions. Assume $P$ and $Q$ are probability distributions of the same probability space $S$,

$$P_{KL} = P_{KL}(X||Y) = \sum_{x,y \in S} P(x) \log_2(\frac{P(x)}{Q(y)}) \tag{7}$$

where $X$ and $Y$ represent, the original and test data sets, respectively. The formulation shows that the relative entropy from $Q$ to $P$ can be considered an indicator of how the distribution of $Y$ is far from the distribution of $X$. Thus, it needs to be different from zero for better privacy protection.

#### 5) PROCRUSTES(PRO)

In statistics, PRO analysis is a form of statistical shape analysis used to analyze the distribution of a set of shapes. It measures the similarity according to other's data set distribution shapes. Assume that $X$ and $Y$ are ($p$ x $k$) dimensional matrices. PRO analysis is transforming a given matrix $X$ into $Y$ such that an orthogonal transformation matrix $T$ minimizes the sum of squares of the residual matrix [48].

$$P_{PRO} = tr((XT - Y)^T(XT - Y)) \tag{8}$$

However, it can also be employed to measure privacy in terms of the variance of the residual matrix. Since the trace calculation given in (8) corresponds to that value, if it is equal to zero, we can conclude that $X$ and $Y$ are identical after the $T$ transformation, and there is no privacy protection. So, higher values for this metric mean higher privacy preservation.

#### 6) PEARSON's CORRELATION COEFFICIENT (PCC)

PCC calculates linear dependence among two variables. It can calculate linear dependence between two data sets as a privacy metric. The PCC is a privacy metric that shows the correlation between the original data set $X$ and the test data set $Y$ [45]:

$$P_{PCC} = \left| |\frac{cov(X, Y)}{\sigma X \sigma Y}| - 1 \right| \tag{9}$$

Here $cov$ means covariance, and $\sigma$ means standard deviation. When there is a linear dependency between two data sets, the PCC gives the values one or minus one. To standardize the output value of PCC, we added absolute value to the original formula, subtracted one, and retook the absolute value. In that way, the PCC shows zero when there is no privacy protection and one, meaning perfect privacy protection.

### D. PROPOSED APPROACH

Let $X_O \in \Re^{m \times n}$ represent the IIoT data collected from a manufacturing environment, where $m$ and $n$ represent sample and feature size, respectively. The primary motivation of this study is to ensure the privacy protection of ($X_O$) without compromising the accuracy (i.e., *R-Squared* score in this paper) of the ML applications. The primary implementation-related assumptions for the proposed method can be listed as follows: **a)** The data owner should explicitly determine the sensitivity of the IIoT data features due to its business, marketplace, and competitive advantages. For instance, a raw material accessible over a few vendors can be a competitive advantage, and its remaining inventory amount becomes sensitive content. Similarly, any critical production machine maintenance data, routes in production, and data that can disclose recipes are

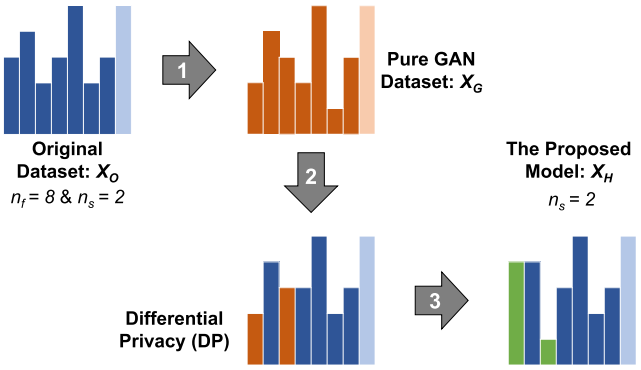


**FIGURE 2.** Illustration of the proposed approach with histograms. Step-1 corresponds the line-2, Step-2 defines the lines between 3 and 9 lines, and Step-3 represents the lines from 10 to 17 in the Algorithm 2. (Here $n_f$ means the number of features, and $n_s$ represents the number of sensitive features.)

**Algorithm 2** Hybrid Privacy-Preserving Algorithm

**Require:** Original data set: $X_O \in \Re^{m \times n}$,
The subset of selected private features: $S_\Phi$,
The number of sensitive features in original data set: $n_s$,
**Ensure:** Hybrid model data set: $X_{HGD}$,
CTGAN data set: $X_G$.

```
X_H ← X_O
X_G ← CTGAN(X_O)
for i = 1 to n do
    for j = 1 to n_s do
        if (i = j) then
            X_HG(:, i) ← X_G(:, j)
        end if
    end for
end for
for j = 1 to n_s do
    for k = 1 to m do
        θ = randn(0, 1)
        if (θ ≥ 0.5) then
            X_HGD(k, :) = X_HG(k, :)+ Laplace (ϵ)
        end if
    end for
end for
return X_HGD
```

needed to be protected. Since companies have prior knowledge of their competitive advantages, only they can define the sensitive features and prevent their disclosure. **b)** The proposed method transforms the selected features into synthetic and randomly noise-added ones. **c)** This approach only ensures privacy protection and is considered a pre-processing operation on IIoT data. Its implementation can be adapted to fit into any cost-effective computational server located in the business environment. The IIoT data are considered high-rate; in most cases, it needs to be processed in real-time or nearly real-time. The proposed approach can be realized in a pipeline with a data collection procedure. In that way, while the algorithm produces the privacy-preserved copy of the current data, edge servers can continue collecting data from the production line's sensors. **d)** The privacy-preserved data will be transmitted to a cloud environment or third parties to perform ML analysis.

The proposed model is illustrated in Figure 2, and its algorithm is given in Algorithm 2. The first step in the proposed method is to create a synthetic copy ($X_G$) of the original data ($X_O$). We run the GAN operation for the whole data set to protect the probability densities of the original data set.

Our primary motivation for using GAN is the ability of GAN to learn the distribution of the original data set. The GAN approximates the data distribution, whereas it creates a different content. Although GAN can help protect data privacy with this property, our experiments show that the CTGAN ($X_G$) application may not preserve the accuracy. Therefore it is required to design a hybrid approach that takes only the valuable part of the GAN. We perform the GAN operation with the CTGAN function, a GAN model for tabular data distribution, and can learn better distributions than Bayesian networks [40].

In the second step, the selected sensitive features from the original data set are replaced with synthetic ones. Then, the new data set ($X_{HG}$) is formed by taking the sensitive features from the synthetic data set ($X_G$) and the remaining ones from the original data set ($X_O$). At this step, the ($X_{HG}$) is the data set without a random noise model. We investigated whether this data ($X_{HG}$) alone can be enough to accomplish our targets. We added it to our test list, where the experimental results section shows the results. Then DP algorithm is implemented to the marked sensitive features to add randomized perturbation to these features. The resulting data set is shown by ($X_{HGD}$) in Algorithm 2. Thus, we not only reconstruct the sensitive columns synthetically but also add noise to them randomly with the help of the DP approach.

Another problem we want to emphasize is to prevent deducing sensitive results from a subset of ($X_O$). Our approach aims to create overall protection for privacy, and we also consider the hidden sensitive data. Different methods, such as statistical or ML, can be done by separating sensitive data from a subset of ($X_O$). For instance, it is possible to train an LR model with selected inputs from the subset of IIoT data and get predictions for different purposes by adversarial LR operations. We performed tests for hidden sensitive data and assessed the proposed model's performance in the experimental results section. The results in the related section show that the proposed method ensures overall privacy protection of the IIoT data. With the application of our proposed approach as a preprocessor, the IIoT data set or its subsets can be used only for the purpose it is prepared. It means they cannot be used for different adversarial AI operations to extract sensitive information regarding companies.

The implementation of the proposed approach is considered ($n_s = 2$) in the illustration given in Figure 2. The original data set ($X_O$) is shown in blue color. The first step is to create a synthetic copy ($X_G$) of ($X_O$) with the CTGAN function. ($X_G$) is shown in red color. There are two sensitive columns ($n_s = 2$) in the example in Figure 2. In step 2, the sensitive columns in the ($X_O$) are replaced by ($X_G$). The DP function runs over the sensitive marked columns as the final step. These steps are listed in Algorithm 2.

The complexity of Algorithm 2 is $O(n_s(n + m))$. Since the $n_s \ll n$ and $n_s \ll m$ in general, one can consider the complexity of the algorithm depends on the $\max(m, n)$. However, the pre-processing step as the CTGAN implementation covers two deep neural networks, ($G$) and ($D$) [20]. Consequently, this operation has $O(m^3)$ complexity. However, since the proposed algorithm can be implemented in a pipeline structure within the IIoT data collection procedure, the CTGAN application in the proposed approach requires few data samples. The duration of data collection and the time for performing the CTGAN operation will be overlapped.

## IV. EXPERIMENTAL RESULTS

### A. ENVIRONMENT

As a platform, we used a windows 10 operating system, the *Anaconda − Jupyter* notebook. The computer we tested has an Intel Core *i*7 2.7 GHz CPU, 16-GB RAM, and an integrated graphics card that is Intel HD graphics 620 and uses the system's memory. In the environment where we completed all the tests, we used the CTGAN function from the CTGANSynthesizer library. The DP calculations were managed by the NumPy library's random and Laplace functions. We mainly aimed to see standard IoT prediction results in tested models. We preferred to use LR, and we called it from $Sklearn.linear_model$ library. We used the Matplotlib library in our graphics. During experiments of the standard IoT tests, we tried to predict the class features, and then we measured and compared the tested models' accuracies and privacy metrics. Here we tested whether the proposed model's accuracy and privacy protection are higher than other models. When we experimented with the hidden sensitive data case, firstly, we changed the training model of the data set, then called linear regression, and tried to predict sensitive features. We tested whether the proposed model can protect predictions for sensitive components.

We implement Algorithm 2 to four SCADA data sets (wind turbine, steam production, energy efficiency, synchronous motors). In the experiments, CTGAN was run with

**TABLE 1.** Test results: wind turbine and steam production.

| | Wind Turbine | | | | | | | Steam Production | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tests** | *R-Squared* | $P_{MSE}$ | $P_{VAR}$ | $P_{HAUS}$ | $P_{KL}$ | $P_{PRO}$ | $P_{PCC}$ | *R-Squared* | $P_{MSE}$ | $P_{VAR}$ | $P_{HAUS}$ | $P_{KL}$ | $P_{PRO}$ | $P_{PCC}$ |
| Original | **0.9689** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | **0.9653** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CTGAN | 0.0035 | **0.1886** | **2.4900** | 0.4260 | **0.2943** | **0.9997** | **0.9943** | 0.0142 | 0.0645 | 2.4673 | 0.7889 | **0.0873** | **0.9977** | **0.9945** |
| DP | 0.5546 | 0.0543 | 0.4043 | 0.3215 | 0.0729 | 0.2528 | 0.1443 | 0.0442 | **0.1264** | **3.9555** | **0.8611** | 0.0782 | 0.6974 | 0.4455 |
| $X_{HGD}$ | | | | | | | | | | | | | | |
| ($n_s$=1) | 0.9685 | 0.0443 | 0.5329 | 0.3294 | 0.0571 | 0.2441 | 0.1429 | 0.9252 | 0.0127 | 0.4627 | 0.4857 | 0.0109 | 0.1996 | 0.0918 |
| ($n_s$=2) | 0.9685 | 0.0729 | 0.9500 | 0.4212 | 0.0871 | 0.4081 | 0.2871 | 0.9253 | 0.0264 | 1.4555 | 0.4909 | 0.0200 | 0.4018 | 0.1791 |
| ($n_s$=3) | 0.9685 | 0.0886 | 1.6186 | **0.4872** | 0.1071 | 0.4668 | 0.4257 | 0.9232 | 0.0482 | 1.7173 | 0.7213 | 0.0400 | 0.6655 | 0.2736 |
| ($n_s$=4) | 0.9681 | 0.1386 | 2.0100 | 0.4572 | 0.1614 | 0.6790 | 0.5643 | 0.9100 | 0.0755 | 2.0245 | 0.7397 | 0.0591 | 0.8481 | 0.3618 |
| $X_{HG}$ | | | | | | | | | | | | | | |
| ($n_s$=1) | 0.9685 | 0.0300 | 0.4643 | 0.3266 | 0.0514 | 0.2132 | 0.1414 | 0.9254 | 0.0036 | 0.2255 | 0.3622 | 0.0045 | 0.1010 | 0.0891 |
| ($n_s$=2) | 0.9685 | 0.0471 | 0.7843 | 0.3356 | 0.0771 | 0.3517 | 0.2843 | 0.9250 | 0.0064 | 0.5082 | 0.4375 | 0.0082 | 0.1871 | 0.1773 |
| ($n_s$=3) | 0.9685 | 0.0557 | 1.2871 | 0.3528 | 0.0914 | 0.4042 | 0.4243 | 0.9232 | 0.0182 | 0.6955 | 0.5105 | 0.0245 | 0.5026 | 0.2664 |
| ($n_s$=4) | 0.9682 | 0.0886 | 1.6386 | 0.3532 | 0.1414 | 0.6286 | 0.5657 | 0.9102 | 0.0309 | 0.8882 | 0.6211 | 0.0382 | 0.7387 | 0.3600 |

Note: The bold values show the max item in each column. Sensitive features (Wind Turbine: $F_2$, $F_3$, $F_0$, and $F_4$; Steam Production: $F_0$, $F_4$, $F_1$, and $F_2$).

**TABLE 2.** Test results: energy efficiency and synchronous motors.

| | Energy Efficiency | | | | | | | Synchronous Motors | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tests** | *R-Squared* | $P_{MSE}$ | $P_{VAR}$ | $P_{HAUS}$ | $P_{KL}$ | $P_{PRO}$ | $P_{PCC}$ | *R-Squared* | $P_{MSE}$ | $P_{VAR}$ | $P_{HAUS}$ | $P_{KL}$ | $P_{PRO}$ | $P_{PCC}$ |
| Original | **0.9661** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | **1.0000** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| CTGAN | 0.0525 | **0.2410** | **2.1030** | 0.8575 | **0.3120** | **0.9954** | **1.0000** | 0.0046 | 0.1620 | 2.0740 | 0.4307 | **0.1980** | **0.9964** | **0.9820** |
| DP | 0.4662 | 0.1240 | 0.5980 | **1.0668** | 0.1020 | 0.3415 | 0.1990 | 0.4044 | 0.1100 | 0.7260 | 0.3840 | 0.0090 | 0.4217 | 0.2400 |
| $X_{HGD}$ | | | | | | | | | | | | | | |
| ($n_s$=1) | 0.9660 | 0.0180 | 0.2100 | 0.4981 | 0.0230 | 0.0882 | 0.0980 | **1.0000** | 0.0580 | 0.5580 | 0.1877 | 0.0500 | 0.4169 | 0.2080 |
| ($n_s$=2) | 0.9639 | 0.0490 | 0.5080 | 0.5365 | 0.0480 | 0.2853 | 0.2000 | **1.0000** | 0.1240 | 1.1400 | 0.3883 | 0.1040 | 0.6472 | 0.4100 |
| ($n_s$=3) | 0.9643 | 0.0880 | 0.7270 | 0.6711 | 0.0830 | 0.4910 | 0.2880 | **1.0000** | 0.2060 | 1.7580 | 0.4754 | 0.1600 | 0.8457 | 0.6140 |
| ($n_s$=4) | 0.8987 | 0.1060 | 0.9990 | 0.6811 | 0.1060 | 0.5300 | 0.4040 | 0.0059 | **0.2400** | **2.2580** | **0.4907** | 0.1960 | 0.9195 | 0.7840 |
| $X_{HG}$ | | | | | | | | | | | | | | |
| ($n_s$=1) | 0.9657 | 0.0180 | 0.1880 | 0.5244 | 0.0280 | 0.0732 | 0.0980 | **1.0000** | 0.0400 | 0.4560 | 0.2050 | 0.0540 | 0.3510 | 0.2100 |
| ($n_s$=2) | 0.9638 | 0.0410 | 0.4190 | 0.6444 | 0.0510 | 0.2334 | 0.2010 | **1.0000** | 0.0760 | 0.8740 | 0.2783 | 0.0096 | 0.5433 | 0.4140 |
| ($n_s$=3) | 0.9637 | 0.0700 | 0.6290 | 0.5657 | 0.0880 | 0.4308 | 0.2980 | **1.0000** | 0.1120 | 1.2720 | 0.3312 | 0.1340 | 0.7746 | 0.6020 |
| ($n_s$=4) | 0.8987 | 0.0850 | 0.8440 | 0.6315 | 0.1170 | 0.4741 | 0.4080 | 0.0184 | 0.1360 | 0.1660 | 0.4175 | 0.1680 | 0.8832 | 0.7880 |

Note: The bold values show the max item in each column. Sensitive features (Energy E.: $F_0$, $F_7$, $F_5$, and $F_9$; Synchronous M.: $F_0$, $F_1$, $F_2$, and $F_3$.

100 epochs, and the parameter for the DP was set as $\epsilon = 0.01$. We apply traditional LR with a 0.2 ratio for splitting the data into train and test parts to solve the regression problem. The well-known *R-Squared* metric is used to evaluate the regression model accuracy. We measured the privacy protection efficiency with the six metrics explained in Section III ($P_{MSE}$, $P_{VAR}$, $P_{HAUS}$, $P_{KL}$, $P_{PRO}$, and $P_{PCC}$). The experimental setup consists of the following test scenarios. There is an additional case study to show the efficiency of the proposed algorithm on mentioned data which also has hidden sensitive features: i) Original data set ($X_O$), ii) CTGAN approach ($X_G$), iii) DP approach ($X_D$), iv) Proposed hybrid approach ($X_{HGD}$) with $n_s = 1 \ldots 4$, and finally v) Perturbed data with GAN operation ($X_{HG}$) with $n_s = 1 \ldots 4$. It is possible to extend the selection of sensitive features by setting the $n_s$ parameter. The code used in this paper is publicly available [50].

We compared the proposed method with the existing works using CTGAN synthetic copy [40] and DP (Algorithm-1). We compared R-squared as accuracy and privacy preservation. The details of the results over four data sets are shown in Tables 1 and 2. Here, as an overview, as can be seen in Figures 4 and 3, in comparison of R-squared for the LR method, the CTGAN ($X_G$) showed the lowest performance, and DP ($X_D$) showed less than 50 % performance in all four data sets. The proposed method ($X_{HGD}$) applications from $n_s = 1 \ldots 4$ showed better performance close to the accuracy values in the original data set ($X_O$). In terms of privacy preservation, CTGAN performed the best, and the proposed methods served better.

### B. PRIVACY PRESERVATION AND HIGHER ACCURACY

The first data set is the wind turbine data, where $m = 49.000$ and $n = 66$ columns were collected between 2014 and 2015. The data contains Wind Energy Converter (WEC) and various temperature measurements from inverter cabinets, front and rear bearing, rotor, and blades. In the experiments, we considered the WEC as the dependent variable. We considered 8.655 WEC-related observations until July 2014 from the whole data with the following features: ($F_0$-WEC average wind speed), ($F_1$-WEC average rotation), ($F_2$-WEC average power), ($F_3$-WEC average reactive power), ($F_4$-WEC average available power from wind), ($F_5$-WEC operating hours), and ($F_6$-WEC production). We set up an LR model to predict the ($F_6$-WEC production). The test results are shown in Table 1.

We used Algorithm 2 to get the proposed models. As shown in Table 1, the accuracy of the LR obtained from the data set ($X_{HGD}$) is comparable with the original data set (in the worst

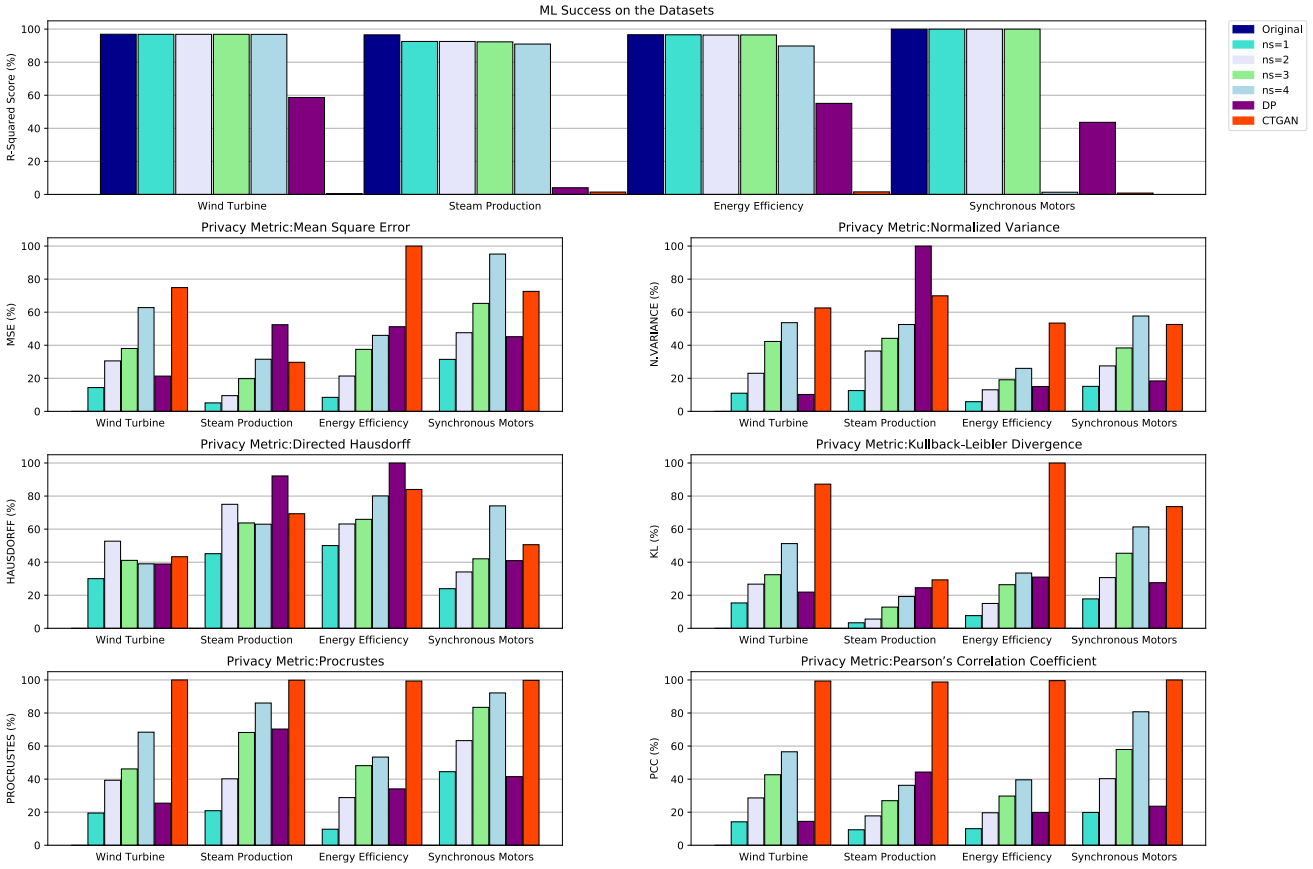


**FIGURE 3.** **Four data sets (Wind Turbine, Steam Production, Energy Efficiency, Synchronous Motors) and tested models (Original $X_O$, Proposed Model $X_{HGD}$ (with $n_s$=1 to $n_s$=4), DP ($X_D$), CTGAN ($X_G$)) *(Values are normalized)*.**

case, with $n_s$=4 *R-Squared* score, it is 0.9681). The privacy preservation capabilities of the method are highly acceptable in terms of the privacy metrics we are concerned about. The original data set ($X_O$) has the max accuracy value and no privacy protection. The proposed model ($X_{HGD}$)($n_s = 4$) has a closer accuracy value to the ($X_O$) and has a much better $P_{MSE}$ value than others. One can conclude that the proposed model sustains high LR performance. It provides better privacy protection than the other competitive approaches, such as CTGAN ($X_G$) and DP ($X_D$). The proposed hybrid approach's benefit can be observed again in Table 1. As stated before, the most crucial property of a proper privacy preservation approach for an IIoT system should be able to preserve the model accuracy while ensuring privacy protection.

On the other hand, modified data sets ($X_G$) and ($X_D$) have limited accuracy preservation abilities for the given data set. As defined in Section III, ($X_{HG}$) is the GAN-applied data set without a random noise model. We tested whether ($X_{HG}$) was enough to accomplish our targets. We completed tests ($X_{HG}$) with various $n_s$ values. As shown in Table 1, the *R-Squared* scores in ($X_{HG}$) are higher as much as the proposed model ($X_{HGD}$), where they both are close to 0.9685 for ($n_s = 1 \ldots 3$) cases. On the other hand, ($X_{HGD}$) offers higher privacy protection. For example the privacy metrics ($P_{MSE}, P_{VAR}, P_{HAUS}, P_{KL}, P_{PRO}, P_{PCC}$) for the case ($X_{HGD}$) ($n_s = 3$) are (0.0886, 1.6186, 0.4872, 0.1071, 0.4668, 0.4257) more better than ($X_{HG}$) ($n_s = 3$) which have (0.0557, 1.2871, 0.3528, 0.0914, 0.4042, 0.4243), respectively. Thus, one can conclude that applying the DP approach along with the GAN protection ($X_{HG}$) will increase the privacy protection of the proposed approach ($X_{HGD}$) for this data set. We also observed similar behaviors for the other data sets we used in our experiments.

The second data set is from a food and confectionery manufacturer's steam production unit, consisting of 2.500 observations and 11 columns. The data set contains the following features: ($F_0$-Natural gas amount), ($F_1$-Natural gas pressure), ($F_2$-Steam pressure), ($F_3$-Produced hot water amount), ($F_4$-Tank amount a raw material), ($F_5$-Steam temperature), ($F_6$-Filling water temperature), ($F_7$-Flue temperature), ($F_8$-Inlet temperature), ($F_9$-Outlet temperature), and ($F_{10}$-Steam recovery). In our experiments, we considered steam production as our dependent variable. We set up the experimental model on the LR prediction of $F_{10}$. We applied the same approach in the wind turbine test structure. Steam production test scores are shown in Table 1.

Let us explain how we set the sensitive features for this data set. The amount of natural gas usage($F_0$), gas pressure($F_1$), and steam pressure($F_2$) are related directly

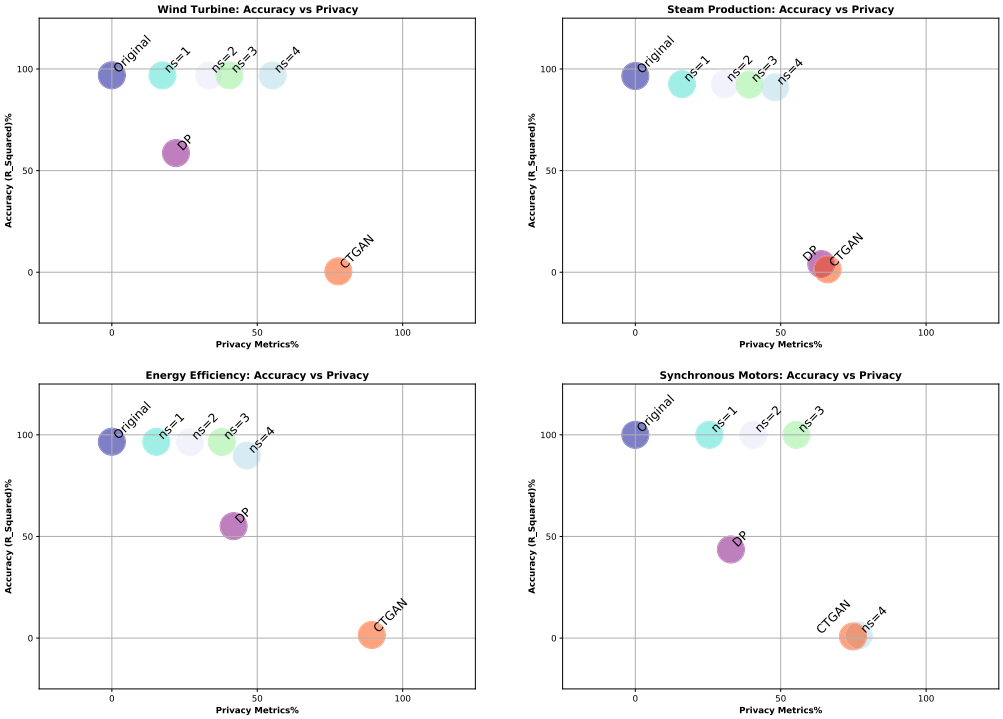


**FIGURE 4.** Relationship between privacy and accuracy. The X-axis shows the average of the six privacy metrics per data set, where zero means weak privacy. The Y-axis shows the LR accuracy in terms of R-Squared values.

to production cost, and they can give precise estimations about the manufacturer's profit margins. Moreover, the tank amount($F_4$) is related to a critical raw material. Its usage pattern can disclose sensitive information about related product recipes. Therefore, these features are selected as sensitive ones that can be helpful for third parties to predict production costs and leak the know-how about products. We prepared the test structure as the number of sensitive features and selected columns such that ($n_s = 1$):($F_0$); ($n_s = 2$):($F_0$) and ($F_4$); ($n_s = 3$):($F_0$), ($F_4$), and ($F_1$); ($n_s = 4$):($F_0$), ($F_4$), ($F_1$), and ($F_2$).

Although the LR performances are very close for different cases, there are considerable differences in the observed privacy metrics. The test on the steam production data set supports the results in the wind turbine that adding random noise to ($X_{HG}$) improved privacy and protected LR accuracy performance.

Figure 3 shows the change in the *R-Squared* scores and the privacy metrics, including the steam production data set. It can be seen that the CTGAN ($X_G$) and DP ($X_D$) obtained better privacy metrics, whereas they degraded the *R-Squared* score. We can clearly say that the ($X_G$) and ($X_D$) are successful only for privacy protection. However, it is not helpful in terms of the production of LR predictions. Hence, we can claim that ($X_G$) and ($X_D$) approaches are not good candidates for IIoT data privacy protection. The prediction and the privacy metrics support us that within ($X_{HGD}$), we could protect the sensitive features while ensuring the minimum losses on the LR accuracy.

As the last discussion, we want to depict the results arising from the analysis of the behavior of the proposed model for two additional public data sets: energy efficiency and synchronous motors. The former data set shows energy analysis with 768 observations and 9 features such as ($F_0$-Relative compactness), ($F_1$-Surface area), ($F_2$-Wall area), ($F_3$-Roof area), ($F_4$-Overall height), ($F_5$-Orientation), ($F_6$-Glazing area), ($F_7$- Glazing area distribution), ($F_8$-Cooling load) and one output variable ($F_9$-Heating load). We built an LR model to predict ($F_9$). The results of these data sets are given in Table 2. The last data set represents the real-time machine data operating environment. It has 557 rows and 5 columns: ($F_0$-Load current), ($F_1$-Power factor), ($F_2$-Power factor error), ($F_3$-Changing of excitation current of a synchronous machine), ($F_4$-Excitation current of a synchronous machine). We predicted the ($F_4$) feature with an LR function in our test environment. Table 2 shows the related results.

The most prominent observation from this last experimental setup is the unusual behavior for the case ($X_{HGD}$) with ($n_s = 4$) on synchronous motors SCADA data. As one can see from Table 2, the *R-Squared* score is (0.0059), which is worse than ($X_D$):(0.4044). Additionally, cases from ($X_{HGD}$) with ($n_s = 1$) to ($n_s = 3$) show the highest *R-Squared* score is (1.0000). The main reason for this deflection is synchronous motors data's highly correlated features. Its correlations among features show that there are dependent features. ($F_1$) correlation coefficients to ($F_2$), ($F_3$), and ($F_4$) are ($-1.0000$), ($-0.8610$), and ($-0.8610$) orderly.

Again ($F_3$) has higher correlations, such as ($F_1$):(−0.8610), ($F_2$):(0.8610), and ($F_4$):(1.0000). Indeed, the high correlation between the sensitive and non-sensitive features can make our approach fail. A possible solution for this problem is to apply a proper dimensionality reduction method before using the proposed technique.

Figure 3 gives the overall view of the proposed approach for all test cases. The figure shows the changes in accuracy values *R-Squared* and privacy metrics for all considered cases together. The first plot in Figure 3 shows *R-Squared* values for all tested data sets. It can be seen here the tested models ($X_O$), ($X_{HGD}$) with ($n_s = 1$,..,4), ($X_G$), and ($X_D$). The remaining plots show each privacy metric performance for all tested data sets with tested models. Please note that the ($X_G$) accuracy performance is deficient, whereas its privacy metrics are the max in almost all cases. The proposed model ($X_{HGD}$) ($n_s = 4$) shows better accuracy and privacy metrics performance.

Figure 4 represents the overall summary of the experimental results regarding the relationship between privacy and accuracy. In the figure, X and Y axes represent the privacy and accuracy values, respectively. In both axes, while zero depicts low accuracy and weak privacy protection, a hundred means max accuracy and maximum privacy protection. As shown in the figure, the results for the proposed model achieve both higher accuracy and privacy protection values simultaneously.

One can consider why we did not check DP only in selected features as we did for ($X_{HG}$). Firstly, we need to explain the reason for testing ($X_{HG}$). The structure of the proposed model is based on the transfer of sensitive features from ($X_G$). Since ($X_{HG}$) is a milestone in our algorithm, we preferred to see its performance without random noise addition. Secondly, in all the tests with the four data sets, we saw that the CTGAN ($X_G$) could catch better privacy protection than other methods, including our proposed model. In contrast, it could not be successful in the LR operations. Thus, picking only sensitive features from the ($X_G$) can gain us higher LR performance. On the other hand, the DP ($X_D$) could catch average or worse performance in our tests, both in privacy metrics and LR performance. Therefore we did not repeat tests with picked features from ($X_D$), and we did not change our proposed model based on DP-selected features. The DP supported our model after the GAN operation.

### C. HIDDEN SENSITIVE DATA PROTECTION

We have also tested the behavior of the proposed algorithm when the data contains hidden sensitive features. Suppose a subset of the data allows deducing sensitive information via a proper ML model. In that case, this data is considered hidden sensitive data. An attacker who has a whole data set without any privacy preservation can apply some ML algorithms to produce some predictions about a sensitive feature. For instance, even though the steam production data set's purpose is to predict $Y =$ ($F_{10}$-Steam recovery), one can change the input/output relations of the data set and produce predictions on some sensitive contents such as ($F_0$-Natural gas amount) or ($F_4$-Tank amount a raw material). One possible usage of the proposed privacy protection mechanism is protecting sensitive data from malicious prediction opportunities. Our experiments also show that the proposed model works efficiently for the above circumstances.

**TABLE 3. Case study test results.**

| | Hidden Sensitive *R-Squared* Scores. | | | |
|---|---|---|---|---|
| **Tests** | **Wind Turbine** | **Steam Production** | **Energy Efficiency** | **Synchronous Motors** |
| Original | **0.9260** | **0.6527** | **0.9912** | **0.6284** |
| $X_{HGD}$ ($n_s$=1) | 0.0010 | 0.0105 | 0.0446 | 0.0628 |
| $X_{HGD}$ ($n_s$=2) | 0.0076 | 0.0134 | 0.0440 | 0.0144 |
| $X_{HGD}$ ($n_s$=3) | 0.0079 | 0.0036 | 0.0365 | 0.0339 |
| $X_{HGD}$ ($n_s$=4) | 0.0063 | 0.0048 | 0.0342 | 0.0033 |

Note: The bold values show the max item in each column.

The purpose of the last experimental setup is to present the proposed approach's abilities to preserve the hidden sensitive data on a switched data structure. One crucial indicator of this preservation can be the low prediction accuracy values obtained from the switched ML models. One can say that LR models for predicting hidden sensitive data are not successful if the proposed method has better privacy preservation capabilities, such as proposed in this study. To demonstrate it, we have changed the input/output structures of the four data sets as the following:

1) The wind turbine regular LR model was built to predict the ($F_6$-WEC production). We changed the LR structure to get predictions about ($F_2$-WEC average power). So, we used the ($F_0$), ($F_1$), ($F_3$), ($F_4$), ($F_5$), and ($F_6$) features as input features to predict ($F_2$).
2) The steam production regular LR model was built to predict the ($F_{10}$-Steam recovery). We changed the LR structure to predict ($F_0$-Natural gas amount). We used the following input features ($F_1$), ($F_2$), ($F_3$), ($F_4$), ($F_5$), ($F_6$), ($F_7$), ($F_8$), ($F_9$), and ($F_{10}$) to predict ($F_0$).
3) The energy efficiency regular LR was for ($F_9$- Heating load) prediction. We converted it to get ($F_0$-Relative compactness) prediction.
4) The synchronous motors' regular LR structure was getting a prediction ($F_4$-Excitation current of a synchronous machine). We restructured the model to get a prediction for ($F_0$-Load current).

Table 3 shows the algorithm's accuracy for hidden sensitive data. The results for the listed data sets represent that if one has complete data without any privacy protection, the hidden data can easily be predicted for malicious purposes. Wind turbine, steam production, energy efficiency, and synchronous motors ($X_O$) model *R-Squared* scores are (0.9260), (0.6527), (0.9912), and (0.6284), respectively. The proposed model ($X_{HGD}$) with all ($n_s$) values show the lowest *R-Squared* scores for this scenario. These experiments reveal a potential privacy risk related to the hidden sensitive data in a data set. The proposed model ($X_{HGD}$) produced a limited prediction performance for the hidden sensitive data. The data sets under

the proposed privacy preservation scheme could not be used for further LR analysis.

## V. CONCLUSION

This paper introduces a novel hybrid approach that preserves IIoT data privacy without compromising the ML applications' accuracy. The experiments suggest that the proposed method has an acceptable performance on privacy preservation and LR accuracy for the IIoT data sets. Hence, the proposed approach can be used as a pre-processing scheme to preserve the critical industrial data with an acceptable computational cost. Additionally, the proposed model has a warrantable performance for protecting the hidden sensitive data embedded in IIoT data. We conducted experiments on four different data sets (Three publicly available to ensure the reproducibility of the study) to demonstrate the proposed model's success. The boost of IIoT applications will trigger more privacy preservation needs in the business. The proposed model can motivate large-scale companies to share their data with third parties with lesser concerns about privacy context. The model can operate in parallel with the data collection of an IIoT system at such a company. It provides a safer data analytics environment for the collected IIoT data set that can be handled with a higher privacy preservation mechanism. The proposed model is suitable for applying it in an efficient pipeline structure between collecting and transmitting the IIoT data. Our future work will focus on constructing an efficient pipeline mechanism for the proposed approach to optimize the algorithm's time complexities and latency.

## REFERENCES

[1] M. Fernandes, J. M. Corchado, and G. Marreiros, "Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: A systematic literature review," *Appl. Intell.*, vol. 52, no. 12, pp. 14246–14280, Sep. 2022.

[2] I. Bisio, C. Garibotto, A. Grattarola, F. Lavagetto, and A. Sciarrone, "Exploiting context-aware capabilities over the Internet of Things for industry 4.0 applications," *IEEE Netw.*, vol. 32, no. 3, pp. 101–107, May 2018.

[3] E. Dalipi, F. Van den Abeele, I. Ishaq, I. Moerman, and J. Hoebeke, "EC-IoT: An easy configuration framework for constrained IoT devices," in *Proc. IEEE 3rd World Forum Internet Things (WF-IoT)*, Dec. 2016, pp. 159–164.

[4] X. Xu, S. Fu, L. Qi, X. Zhang, Q. Liu, Q. He, and S. Li, "An IoT-oriented data placement method with privacy preservation in cloud environment," *J. Netw. Comput. Appl.*, vol. 124, pp. 148–157, Dec. 2018.

[5] S. M. Tahsien, H. Karimipour, and P. Spachos, "Machine learning based solutions for security of Internet of Things (IoT): A survey," *J. Netw. Comput. Appl.*, vol. 161, Jul. 2020, Art. no. 102630.

[6] Z. Bakhshi, A. Balador, and J. Mustafa, "Industrial IoT security threats and concerns by considering cisco and Microsoft IoT reference models," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2018, pp. 173–178.

[7] Trend Micro CISO Resource Center. (2021). *Trend Micro 2020 Annual Cybersecurity Report*. [Online]. Available: https://www.trendmicro.com

[8] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 201–210.

[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[10] H.-J. Jeong, H.-J. Lee, and S.-M. Moon, "Cloud-based machine learning for IoT devices with better privacy," in *Proc. 13th ACM Int. Conf. Embedded Softw. Companion*, 2017, pp. 1–2.

[11] G. Long, T. Shen, Y. Tan, L. Gerrard, A. Clarke, and J. Jiang, "Federated learning for privacy-preserving open innovation future on digital health," *Humanity Driven AI*, vol. 10761. Cham, Switzerland: Springer, 2021, pp. 113–133.

[12] S. Qi, Y. Lu, W. Wei, and X. Chen, "Efficient data access control with fine-grained data protection in cloud-assisted IIoT," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2886–2899, Feb. 2021.

[13] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A systematic review of explainable artificial intelligence in terms of different application domains and tasks," *Appl. Sci.*, vol. 12, no. 3, p. 1353, Jan. 2022.

[14] S. Sari, O. Demir, and G. Kucuk, "FairSDP: Fair and secure dynamic cache partitioning," in *Proc. 4th Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2019, pp. 469–474.

[15] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Privacy and utility preserving sensor-data transformations," *Pervas. Mobile Comput.*, vol. 63, Mar. 2020, Art. no. 101132.

[16] C. Dwork, "Differential privacy," in *Proc. Int. Colloq. Automata, Lang., Program.*, in Lecture Notes in Computer Science, vol. 4052, 2006, pp. 1–12.

[17] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "Deep models under the GAN: Information leakage from collaborative deep learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* New York, NY, USA: Association for Computing Machinery, Oct. 2017, pp. 603–618.

[18] C. Dwork and M. Naor, "On the difficulties of disclosure prevention in statistical databases or the case for differential privacy," *J. Privacy Confidentiality*, vol. 2, no. 1, pp. 93–107, Sep. 2010.

[19] C. Liu, S. Chakraborty, and P. Mittal, "Dependence makes you vulnerable: Differential privacy under dependent tuples," in *Proc. Netw. Distrib. Syst. Secur. Symp.*, vol. 16, 2016, pp. 21–24.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[21] Y. Liu, J. Peng, J. J. Q. Yu, and Y. Wu, "PPGAN: Privacy-preserving generative adversarial network," in *Proc. IEEE 25th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2019, pp. 985–989.

[22] W. Soontronchai. (2019). *IIoT Data of Wind Turbine*. [Online]. Available: https://www.kaggle.com/datasets/wasuratme96/iiot-data-of-wind-turbine

[23] A. Xifara. (2012). *Energy Efficiency Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Energy+efficiency

[24] R. Bayindir, and H. T. Kahraman. (2021). *Synchronous Machine Data Set Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Synchronous+Machine+Data+Set

[25] M. A. Husnoo, A. Anwar, R. K. Chakrabortty, R. Doss, and M. J. Ryan, "Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey," *IEEE Access*, vol. 9, pp. 153276–153304, 2021.

[26] A. Majeed and S. Lee, "Attribute susceptibility and entropy based data anonymization to improve users community privacy and utility in publishing data," *Appl. Intell.*, vol. 50, no. 8, pp. 2555–2574, Aug. 2020.

[27] L. Sweeney, "Maintaining patient confidentiality when sharing medical data requires a symbiotic relationship between technology and policy," *Retrieved March*, vol. 9, p. 2011, May 1997.

[28] A. Narayanan and V. Shmatikov, "How to break anonymity of the Netflix prize dataset," 2006, *arXiv:cs/0610105*.

[29] F. Samie, L. Bauer, and J. Henkel, "From cloud down to things: An overview of machine learning in Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4921–4934, Jun. 2019.

[30] W. M. S. Stout and V. E. Urias, "Challenges to securing the Internet of Things," in *Proc. IEEE Int. Carnahan Conf. Secur. Technol. (ICCST)*, Oct. 2016, pp. 1–8.

[31] P. Sharma, S. Jain, S. Gupta, and V. Chamola, "Role of machine learning and deep learning in securing 5G-driven industrial IoT applications," *Ad Hoc Netw.*, vol. 123, Dec. 2021, Art. no. 102685.

[32] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 3–18.

[33] Q. Zeng, Z. Lv, and C. Li, "FedProLs: Federated learning for IoT perception data prediction," *Appl. Intell.*, pp. 1–13, Jun. 2022.

[34] E. Choi, S. Biswal, B. Malin, J. Duke, W. Stewart, and J. Sun, "Generating multi-label discrete patient records using generative adversarial networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2017, pp. 286–305.

[35] B. K. Beaulieu-Jones, Z. S. Wu, C. Williams, R. Lee, S. P. Bhavnani, J. B. Byrd, and C. S. Greene, "Privacy-preserving generative deep neural networks support clinical data sharing," *Circulat. Cardiovasc. Qual. Outcomes*, vol. 12, no. 7, 2019, Art. no. e005122.
[36] W. Li, P. Meng, Y. Hong, and X. Cui, "Using deep learning to preserve data confidentiality," *Appl. Intell.*, vol. 50, no. 2, pp. 341–353, Feb. 2020.
[37] C. Han and R. Xue, "Differentially private GANs by adding noise to discriminator's loss," *Comput. Secur.*, vol. 107, Aug. 2021, Art. no. 102322.
[38] H. Ping, J. Stoyanovich, and B. Howe, "DataSynthesizer: Privacy-preserving synthetic datasets," in *Proc. 29th Int. Conf. Sci. Stat. Database Manage.*, Jun. 2017, pp. 1–5.
[39] L. Xu and K. Veeramachaneni, "Synthesizing tabular data using generative adversarial networks," 2018, *arXiv:1811.11264*.
[40] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–15.
[41] C. Xu, J. Ren, D. Zhang, Y. Zhang, Z. Qin, and K. Ren, "GANobfuscator: Mitigating information leakage under GAN via differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2358–2371, Sep. 2019.
[42] P. Cheng, J. F. Roddick, S.-C. Chu, and C.-W. Lin, "Privacy preservation through a greedy, distortion-based rule-hiding method," *Appl. Intell.*, vol. 44, no. 2, pp. 295–306, Mar. 2016.
[43] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V. Verykios, "Disclosure limitation of sensitive rules," in *Proc. Workshop Knowl. Data Eng. Exchange*, 1999, pp. 45–52.
[44] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.
[45] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–38, May 2019.
[46] S. Piramuthu, "Feature selection for reduction of tabular knowledge-based systems," *Inf. Technol. Manag.*, vol. 6, pp. 351–362, Oct. 2005.
[47] D. Galas, G. Dewey, J. Kunert-Graf, and N. Sakhanenko, "Expansion of the Kullback–Leibler divergence, and a new class of information metrics," *Axioms*, vol. 6, no. 4, p. 8, Apr. 2017.
[48] J. C. Gower, "Generalized Procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, Mar. 1975.
[49] S. Oliveira and O. Zaiane, "Privacy preserving clustering by data transformation," *J. Inf. Data Manage.*, vol. 1, no. 1, p. 37, 2010.
[50] Y. S. Hindistan. (2021). *Source Code*. [Online]. Available: https://github.com/yavuzselimhindistan
[51] B. Jiang, J. Li, G. Yue, and H. Song, "Differential privacy for industrial Internet of Things: Opportunities, applications, and challenges," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10430–10451, Jul. 2021.
[52] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6092–6102, Sep. 2020.
[53] X. Yang, T. Wang, X. Ren, and W. Yu, "Survey on improving data utility in differentially private sequential data publishing," *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 729–749, Oct. 2021.
[54] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward edge-based deep learning in industrial Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4329–4341, May 2020.
[55] J. Feng, L. T. Yang, R. Zhang, W. Qiang, and J. Chen, "Privacy preserving high-order bi-lanczos in cloud–fog computing for industrial applications," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7009–7018, Oct. 2022.
[56] M. Usman, A. Jolfaei, and M. A. Jan, "RaSEC: An intelligent framework for reliable and secure multilevel edge computing in industrial environments," *IEEE Trans. Ind. Appl.*, vol. 56, pp. 4543–4551, 2020.
[57] S. Langarica, C. Ruffelmacher, and F. Núñez, "An industrial internet application for real-time fault diagnosis in industrial motors," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 284–295, Jan. 2020.
[58] P. Hu, H. Ning, T. Qiu, H. Song, Y. Wang, and X. Yao, "Security and privacy preservation scheme of face identification and resolution framework using fog computing in Internet of Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1143–1155, Oct. 2017.
[59] X. Xu, R. Mo, X. Yin, M. R. Khosravi, F. Aghaei, V. Chang, and G. Li, "PDM: Privacy-aware deployment of machine-learning applications for industrial cyber–physical cloud systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5819–5828, Aug. 2021.
[60] C. Huang, S. Chen, Y. Zhang, W. Zhou, J. J. P. C. Rodrigues, and V. H. C. de Albuquerque, "A robust approach for privacy data protection: IoT security assurance using generative adversarial imitation learning," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17089–17097, Sep. 2022.
[61] M. R. Shahid, G. Blanc, H. Jmila, Z. Zhang, and H. Debar, "Generative deep learning for Internet of Things network traffic generation," in *Proc. IEEE 25th Pacific Rim Int. Symp. Dependable Comput. (PRDC)*, Dec. 2020, pp. 70–79.

**YAVUZ SELIM HINDISTAN** received the B.S. degree in physics science from Bogazici University, Turkey, in 1997, and the M.S. degree in business administration from Istanbul Bilgi University, in 2008. He is currently pursuing the Ph.D. degree in management of information systems with Kadir Has University. He has more than 25 years of professional experience in business IT-related topics. He is also an Instructor of the Management Information Systems Department, Ozyegin University, Turkey. His current research interests include data privacy, big data, finance technologies, and machine learning systems.

**E. FATIH YETKIN** received the B.S. degree in electronics engineering from Uludag University, Turkey, in 2000, and the M.S. and Ph.D. degrees in computational science and engineering (CSE) from Istanbul Technical University (ITU), Turkey, in 2003 and 2011, respectively. He was a Postdoctoral Researcher with the High-End Parallel Algorithms for Challenging Numerical Simulations (HiePACS) Team, Inria, Bordeaux, France, from 2014 to 2016. He is currently an Assistant Professor with the Management Information Systems Department, Kadir Has University, Turkey. His current research interests include applied numerical linear algebra, high-performance computing, machine learning, and dimensionality reduction.