

Privacy Preservation for Machine Learning in IIoT Data via Manifold Learning and Elementary Row Operations

E. Fatih Yetkin^a and Tuğçe Ballı^b

Department of Management Information Systems, Kadir Has University, Istanbul, 34083, Turkey

Keywords: Privacy Preservation, IIoT, Manifold Learning, Machine Learning.

Abstract: Modern large-scale production sites are highly data-driven and need large computational power due to the amount of the data collected. Hence, relying only on in-house computing systems for computational workflows is not always feasible. Instead, cloud environments are often preferred due to their ability to provide scalable and on-demand access to extensive computational resources. While cloud-based workflows offer numerous advantages, concerns regarding data privacy remain a significant obstacle to their widespread adoption, particularly in scenarios involving sensitive data and operations. This study aims to develop a computationally efficient privacy protection (PP) approach based on manifold learning and the elementary row operations inspired from the lower-upper (LU) decomposition. This approach seeks to enhance the security of data collected from industrial environments, along with the associated machine learning models, thereby protecting sensitive information against potential threats posed by both external and internal adversaries within the collaborative computing environment.

1 INTRODUCTION

Since the advent of the Industrial Internet of Things (IIoT), companies have been able to operate more efficiently, enhancing production quality and reducing costs (Raptis et al., 2019). A well-known example includes manufacturing companies transferring their IIoT data to a ML-based third party service or a cloud system for processing to facilitate predictive maintenance (Hongxia et al., 2016). However, IIoT data often contains sensitive information related to production processes, which poses a risk of information leakage for companies (Hindistan and Yetkin, 2023).


Modern large-scale production systems are highly data-driven and need large computational power due to the amount of the data collected. Therefore, it is not always possible to use the in-house computing systems for the computational workflow. Indeed, mostly the cloud environment is preferred since it can provide a large amount of computing resources on demand (Soveizi et al., 2023). Although cloud-based workflows offer numerous advantages, security concerns remain a significant barrier to their adoption, particularly in applications involving sensitive data and operations (Varshney et al., 2019).


Outsourcing workflows or parts of them to cloud

systems result in the losing control over certain tasks and data, potentially increasing security vulnerabilities and exposing workflows to a heightened risk of malicious attacks (Soveizi et al., 2023).

The security challenges in cloud computing arise from shared infrastructure and data transmission over potentially untrusted networks, along with dynamic operational conditions which can bind to cloud services that may later reveal security vulnerabilities. Additionally, cloud providers, while adhering to protocols, may still attempt to infer sensitive information about users' data and workflow logic. There are many attempts in the literature to handle these privacy issues which are particularly important for industrial data that can possibly impact the daily life of people both digitally and physically.

To address these issues, this work extends the existing privacy preservation (PP) approaches by implementing two distinct mechanisms before sharing the data in collaborative environment: a) The feature space created within in-house computing systems will be projected onto a manifold via a nonlinear transformation to hide its physical properties while preserving its mathematical properties. However, this approach alone does not fully secure the data in untrusted environments, such as cloud platforms and also it is not a certain prevention mechanism for reconstruction or model inversion attacks, particularly for in-company intruders. b) A random reversible permutation will

^a  <https://orcid.org/0000-0003-1115-4454>

^b  <https://orcid.org/0000-0002-6509-3725>

be applied onto the input and output of the reduced model to ensure its privacy by using a appropriate seed. Perturbation mechanism, designed as a one-way function will hide the relevant information from the intruders. This mechanism will be easy to solve in forward direction but infeasible in reverse direction. The proposed approach can be interpreted as a combination of performing dimensionality reduction and adding random noise to the reduced subspace.

The rest of this paper organized is as follows. In the next section, we will discuss about related works. In methodology section, the proposed approach will be depicted. Since it is an ongoing work we will show our preliminary results with a single experimental dataset collected for predictive maintenance purposes in numerical experiments section. Finally, we will discuss the future studies in the conclusion section.

2 RELATED WORKS

The standard approach for the protection of the data privacy is called differential privacy. This approach is based on probability theory for ensuring the privacy (Dwork et al., 2014a). Even though it has several advantages, differentially private algorithms may exhibit errors that are influenced by the intrinsic dimension n of the input data, therefore Dwork et al., proposed an approach based on PCA to reduce the dimensionality of the input data (Dwork et al., 2014b). There are also many attempts in literature exploring the randomized projections as a means of implementing privacy mechanisms (Kung, 2017). Those approaches are mainly based on adding a noise matrix to the original data before the calculation of the associated covariance matrix (Blum et al., 2005). As discussed in (Hindistan and Yetkin, 2023), noise addition can still be considered a reliable privacy mechanism especially when combined with a generative adversarial network (GAN) based syntheticization mechanism. Other common approach involves using cryptographic methods such as Homomorphic Encryption (Yi et al., 2014), Garbled Circuits (Frederiksen et al., 2015) and Secret Sharing (Bogdanov et al., 2018). However, cryptographic approaches are computationally very expensive and they may not be feasible for application especially with large scale data sets (Al-Rubaie and Chang, 2019).

3 METHODOLOGY

Privacy protection schemes should rely on a specific scenario. In our case, we focused on a scenario similar to the scenario discussed in (Kung, 2017). Unlike the traditional communication based scenarios, we consider four parties in our case. The data owner, trusted third party, malicious actors from the external network, and from the collaborative computing system (or with traditional namings; Alice, Bob, Eve1, and Eve2 respectively). Here, we are aiming to convey the sensitive data from Alice to Bob by hiding its internal statistical dynamics from both Eve1 and Eve2. Therefore, we designed our experimental setup as depicted in Fig. 1. The main purpose of our study is to increase the privacy of the data collected from industrial environment while preventing the possibility of extracting additional information from the data and the model by the malicious actors from outside or inside the network. Besides, we aim to protect the model parameters from the intruders. In Fig. 1 we assume there are two possible intrusions from the perspective of data privacy: a) outside (Eve1) and b) inside (Eve2) the collaborative network environment. Overall, this mechanism is based on hiding the raw data as much as possible from the third parties including the trusted ones.

The protection mechanisms offered in this study has two phases. In the first phase, the produced feature space will be projected onto a manifold via a non-linear transformation. Since it is an ongoing study we have only used the Laplacian Eigenmaps (Belkin and Niyogi, 2003), on the other hand there are more advanced and flexible supervised mechanisms (metric learning approaches) to improve the overall privacy attack mitigation schemes, such as Large Margin Nearest Neighbours (LMNN) (Weinberger and Saul, 2009) and Neighborhood Component Analysis (NCA) (Goldberger et al., 2004).

Since the main focus of this work is to understand the effect of dimensionality reduction with elementary matrix operations on privacy, we kept the reduction order as $n - 1$ which corresponds to a single feature removal from the overall dataset. After removing one feature from the data via non-linear projection, the data will lose its physical meaning, but as already discussed in (Dwork et al., 2014b), it is not a certain prevention mechanism for model inversion attacks especially the ones from the in-company intruders. As a well-known fact, the Laplacian eigenvectors keep most of the relevant information arising from the raw data. Therefore, to ensure its privacy, as the second phase of the implementation, the input matrix of the reduced model should be perturbed. For this pertur-

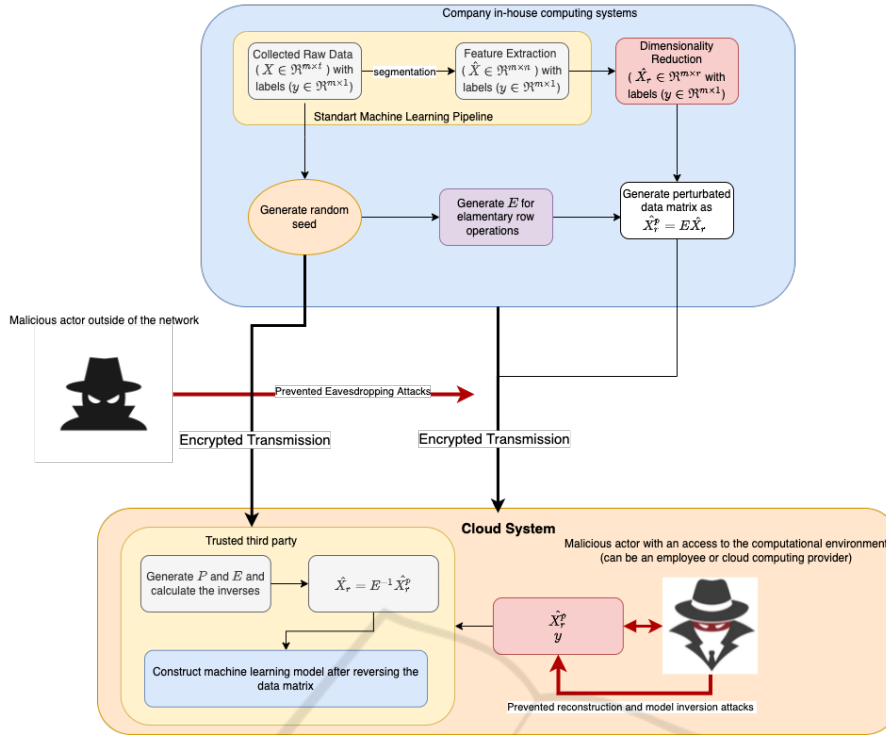


Figure 1: Illustration of the proposed approach.

bation, we followed a very basic idea arising from the well-known LU decomposition (Golub and Van Loan, 2013). This decomposition has been in use for a long time and it is also employed in several authentication mechanisms (Li et al., 2013). The LU decomposition consists of several operations based on permutation (P) and elementary (E) matrices. The most important property of both P and E matrices is that their inversion is straightforward as the matrices are orthogonal, which in turn simplifies reversal of perturbation. The mathematical model of this study is based on this property.

3.1 Proposed Approach

Let $X \in \mathfrak{R}^{m \times t}$ represent a univariate time series data collected from IIoT devices with corresponding labels as $y \in \mathfrak{R}^{m \times 1}$. In standard machine learning pipeline for time series, the features are extracted from the X with a fixed segment interval s where $t = s \times n$ and the resulting data matrix is given as $\hat{X} \in \mathfrak{R}^{m \times n}$. Once the data matrix is obtained, it is always possible to reduce its dimension via linear or non-linear dimensionality reduction methods. In our approach, we used one of the well-known manifold learning approaches, namely the Laplacian Eigenmaps (LE). This method is based on construction of a neighborhood graph of the \hat{X} , and accordingly related weighted adjacency

(W) and degree matrices (D). Then, the Laplacian of the data matrix can be defined as $L = D - W$. As discussed in (Belkin and Niyogi, 2003), it is indeed possible to construct an optimization problem to keep the geometrical properties of the data in a lower dimensional problem and the solution of this minimization problem can be determined as a set of small eigenpairs of the generalized eigenproblem defined in Eq. 1 as follows:

$$Lv = \lambda Dv \quad (1)$$

where λ is an eigenvalue and the v is the associated eigenvector. If the reduction dimension is selected as r , then $V \in \mathfrak{R}^{m \times r}$ will contain the eigenvectors associated with the r smallest eigenvalues of the generalized eigenproblem given in Eq. 1. The reduced raw data will be obtained as $\hat{X}_r = V$. In this way, the raw data will never be shared with the third parties since this transformation is irreversible without knowing the associated eigenvalues of Eq. 1.

As dimensionality reduction preserves the intrinsic behavior of the data and is proposed as a solution of the curse of dimensionality problem, it serves as a strong candidate for maximizing the utility function in privacy preservation (Dwork et al., 2014b). Even though the raw data still be stored at the in-house computational environments, data behavior is still vulnerable and intruders may use the reduced data to build malicious machine learning models to

violate data privacy. Therefore, we proposed a second mechanism which will be realized at in-house computational systems to avoid internal threats arising from collaborative environment (see Fig. 1).

One efficient way of solving a linear equation system is based on decomposition of the n -dimensional coefficient matrix A as the product of lower (L) and upper (U) triangular matrices as $A = LU$. The procedure is called the LU decomposition and start with multiplying the matrix A from left with unit lower diagonal elementary matrices to obtain an upper triangular representation of the matrix A . An elementary matrix which can be useful for this purpose is defined as an identity matrix (I) with a nonzero element in lower diagonal part and corresponds to the elementary row operations (multiplication, subtraction or addition). Formally any matrix obtained from identity matrix with a single elementary row operation is called an elementary matrix (Strang, 2000). After a successful decomposition, the matrix U can be written as,

$$E_k E_{k-1} \dots E_1 A = U \quad (2)$$

where k is the required number of the row operations to transform the matrix A into upper triangular matrix U . Then, if one multiply the both sides of Eq. 2 with the inverses of each elementary matrix in a reverse order one can obtain the LU decomposition as follows,

$$\begin{aligned} E_1^{-1} E_2^{-1} \dots E_k^{-1} E_k E_{k-1} \dots E_1 A &= E_1^{-1} E_2^{-1} \dots E_k^{-1} U \\ A &= E_1^{-1} E_2^{-1} \dots E_k^{-1} U \end{aligned}$$

where $L = E_1^{-1} E_2^{-1} \dots E_k^{-1}$. Computing the inverse of a unit lower elementary matrix E_i is an operation that is computationally very efficient. If the non-zero entry of the E_i is given as e_{kj} where k and j are row and column indices respectively, E_i^{-1} will be simply equal to the same matrix with $-e_{kj}$. We will use this property of unit lower elementary matrices to produce a set of randomly selected elementary row operations to modify the originally reduced data matrix \hat{X}_r to a \hat{X}_r^p . For implementation, randomly produced indices and a set of random elementary operation coefficients are produced at the on-site computational environment as E and perform the operation $\hat{X}_r^p = E\hat{X}_r$. This operation easily be reversed by using the inverse of matrix E . This matrix can be reproducible if the required seed is available at any trusted third party. Note that even if the elementary operations are reversed, still the original feature space will be hidden at any party. Therefore, the important benefits of using this algorithmic approach can be listed as follows:

- The raw data (X) and the original feature matrix (\hat{X}) are never necessary to be shared with any third parties.

- The perturbation with elementary row operations can be easily reversed by any trusted third party via the regeneration of matrix E with the shared seed. The seed can be shared with the trusted third party by any encrypted exchange algorithm (such as the Diffie-Hellman key exchange algorithm (Merkle, 1978)).
- Intruders from inside or outside of the collaborative environment (such as a cloud system) can only have access to \hat{X}_r^p , which prevents the construction of malicious models.

Even though the numerical results are promising, there is a need for more investigations from theoretical and numerical perspectives.

4 NUMERICAL EXPERIMENTS

4.1 Dataset

We have used an experimental dataset collected at a lab environment (Loparo, 2012). This dataset is a well-known benchmarking dataset especially for malfunction classification based on vibration signals for production environment (Smith and Randall, 2015), (Saufi et al., 2023). Experiments were carried out using a 2-horsepower Reliance Electric motor, with acceleration data collected from measurement points both proximal and distal to the motor bearings. Faults were artificially induced in the motor bearings through electro-discharge machining (EDM), with defect sizes ranging from 0.007 inches to 0.040 inches in diameter. These faults were introduced individually at the inner raceway, rolling element (ball), and outer raceway. The faulted bearings were subsequently reinstalled in the test motor, and vibration data was recorded under motor loads varying from 0 to 3 horsepower, corresponding to motor speeds between 1797 and 1720 RPM. The dataset includes three fault cases in addition to normal conditions: a) outer race, b) inner race, c) bearing. The machine learning model, therefore, can be considered as a four-class problem. The distribution of the data with respect to its labels in the original dataset is shown in Fig. 2 using t-SNE reduction (Van der Maaten and Hinton, 2008).

4.2 Data Pre-Processing Pipeline and Machine Learning Models

In the experimental setup, we have created four scenarios for comparing and evaluating the machine

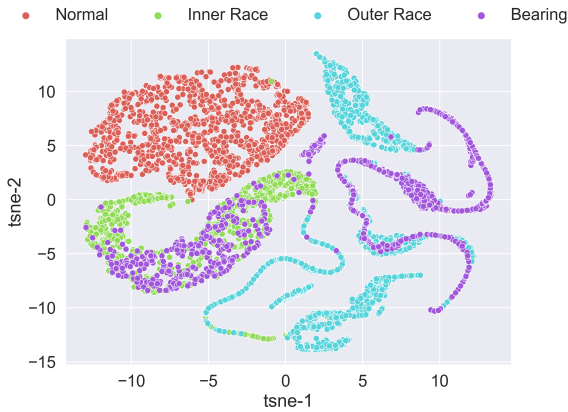


Figure 2: Distribution of the original features in 2-dimensional t-SNE projection space.

learning accuracy performance and privacy preservation properties. In the first setup, we created our baseline, which is the application of several traditional machine learning algorithms on the dataset. We considered three main types of machine learning approaches: a) linear models (logistic regression (LR) and Support Vector Machine (SVM)), b) tree-based models (Decision tree (DT), Random Forest (RF)), c) an instance-based model (k nearest neighbors (KNN)). All experiments are realized with a 10-fold cross-validation. To extract features from the X , we choose the segment length as 1000, then extract standard statistical features such as the mean, standard deviation, root mean square, variance, and median. Note that since we have restricted our discussion to the impact of the proposed approach on privacy preservation, we did not implement more sophisticated feature extraction methods.

4.3 Comparison of ML Accuracy Values: Utility Perspective

In general, any privacy preservation technique should have a balance between utility and privacy. In the first set of experiments, we compared the performance of various machine learning approaches with respect to different execution paths. The execution paths considered in these experiments can be listed as follows,

- **Case-I:** Apply ML methods with 10-fold cross-validation to the \hat{X} ,
- **Case-II:** Apply ML methods with 10-fold cross-validation to the LE applied reduced data \hat{X}_r ,
- **Case-III:** Apply ML methods with 10-fold cross-validation to the perturbed $\hat{X}^p = E\hat{X}$ for various different number of elementary row operations,

- **Case-IV:** Apply ML methods with 10-fold cross-validation to the perturbed $\hat{X}_r^p = E\hat{X}_r$ for various different number of elementary row operations,

In Figs. 3, 4, and 5, we represent the change of the accuracy values for algorithms under four execution paths discussed above. For LE application, we considered the reduction dimension as $r = 4$ without optimizing the intrinsic dimension.

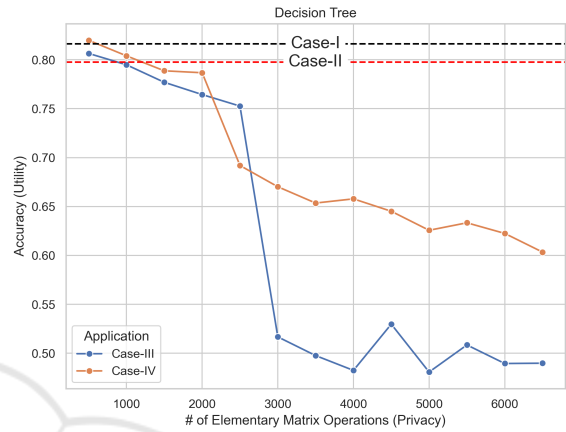


Figure 3: Accuracy behaviour of Decision Tree algorithm under different execution scenarios.

In Fig.3, the effect of low number of elementary matrix operations on utility for both Case-III and Case-IV is depicted. As we can see from the figure, Case-IV has also positive impact on utility with respect to Case-III even for high number of row operations.

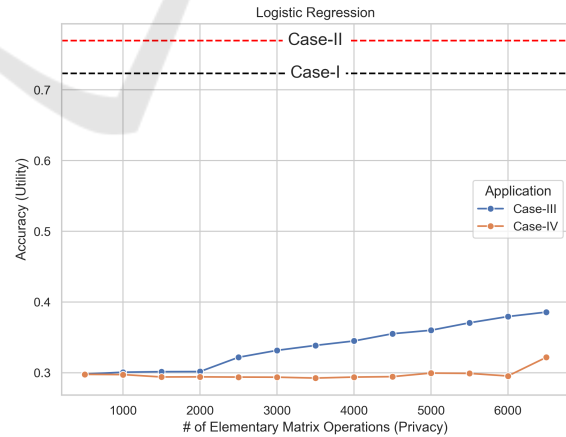


Figure 4: Accuracy behaviour of Logistic Regression algorithm under different execution scenarios.

However, in Fig.4, both Case-III and Case-IV create a dramatical decrease on accuracy of the LR algorithm. This can be expected since the LR algorithm is driven by a linear hypothesis function and

elementary row operations are malfunctioning the existing linear relations in between the data samples. The same effect can be visible for the SVM algorithm (see Table.1). The instance based algorithm, KNN,

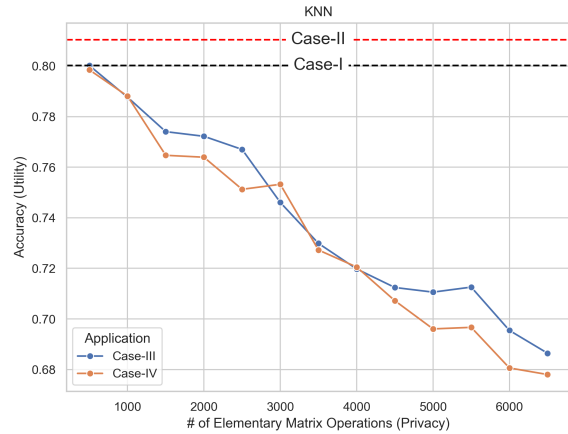


Figure 5: Accuracy behaviour of KNN algorithm under different execution scenarios.

also produced a similar behavior with the DT algorithm and it presented in Fig 5. The comparison of the whole algorithms including the RF and SVM for the Case-IV, is shown in the Table 1.

4.4 Comparison of Data Distributions: Privacy Perspective

In the second experiment, we presented the change in the data distribution with respect to the given labels by using linear (PCA) and non-linear (t-SNE) dimensionality reduction. As shown in Figs. 3 and 5, the low number of elementary row operations does not affect the utility in terms of the accuracy of the DT and KNN algorithms. Therefore, we selected the number of the elementary row operations 500 for Case-III and Case-IV and created the 2-dimensional representations of all cases with PCA and t-SNE. As clearly seen from Fig. 6, the LE method (Case-II) does not affect the data distribution; however, even for a very small number of perturbations (relatively to the data dimension where $n = 6650$), linear projections of the data change entirely in both Case-III and Case-IV. It can also be considered as an explanation of the accuracy behavior of LR and SVM algorithms with Case-III and Case-IV (see Table 1). A similar behavior can be observed from the non-linear projection of the data in Fig. 7. However, in this figure one can observe that Case-IV significantly affects data distribution while the separability of the data is still protected. Therefore, the numerical results show that Case-IV improves the privacy (in terms of distribution of the

data) while preserving its utility.

5 DISCUSSION AND FUTURE STUDIES

In this work, we proposed a hybrid mechanism based on manifold learning (LE) and the elementary row operations inspired by the well-known LU decomposition. As shown in the numerical experiments section, applying the low number of elementary row operations retains the accuracy behavior, especially for the tree and instance-based classifiers. This observation (after theoretical validation) can be used to establish a privacy preservation protocol that considers the reduction order r and the number of the elementary row operations as parameters. Although we employed accuracy as the only metric for evaluating the utility of the data, we plan to extend our study to show the balance between utility and privacy by covering several theoretical and practical aspects arising from the privacy preservation literature. Besides, in future work of this study, we will compare our proposed approach with the state-of-the-art privacy preservation techniques discussed in Section 2. Even though the numerical results depicted that the proposed approach is promising for balancing utility and privacy, several issues should be evaluated carefully. This study used visual data distributions as the primary evaluation method for privacy preservation. However, we plan to develop a solid privacy evaluation procedure based on a technical privacy metric such as entropy-based approaches defined in (Wagner and Eckhoff, 2018). Furthermore, we have used a single IIoT dataset in our experiments, and it should also be extended to a set of experiments. Lastly, we are also planning to investigate the theoretical reasons for the behavior of Case-IV to establish a solid explainable mechanism to improve the usability of the proposed approach.

6 CONCLUSION

This study aims to propose a privacy protection mechanism that can balance utility and privacy. For the implementation, the data owner selects the number of the elementary row operations as a parameter and, according to the privacy policy related to data, can share the random seed required for the reproduction of the matrix E with trusted third parties. If the utility is the priority for the selected case, then the number of the elementary row dimensions (or the number

Table 1: Comparison of the accuracy performance of several ML algorithms under various conditions.

	Original	LE	# of Elementary Matrix Operations						
			500	1500	2500	3500	4500	5500	6500
LR	0.72	0.77	0.31	0.29	0.29	0.29	0.29	0.29	0.30
KNN	0.81	0.82	0.80	0.76	0.75	0.72	0.70	0.69	0.67
SVM	0.79	0.78	0.29	0.29	0.29	0.29	0.29	0.29	0.29
DT	0.82	0.80	0.82	0.78	0.69	0.65	0.64	0.63	0.60
RF	0.83	0.80	0.82	0.79	0.76	0.74	0.74	0.71	0.69

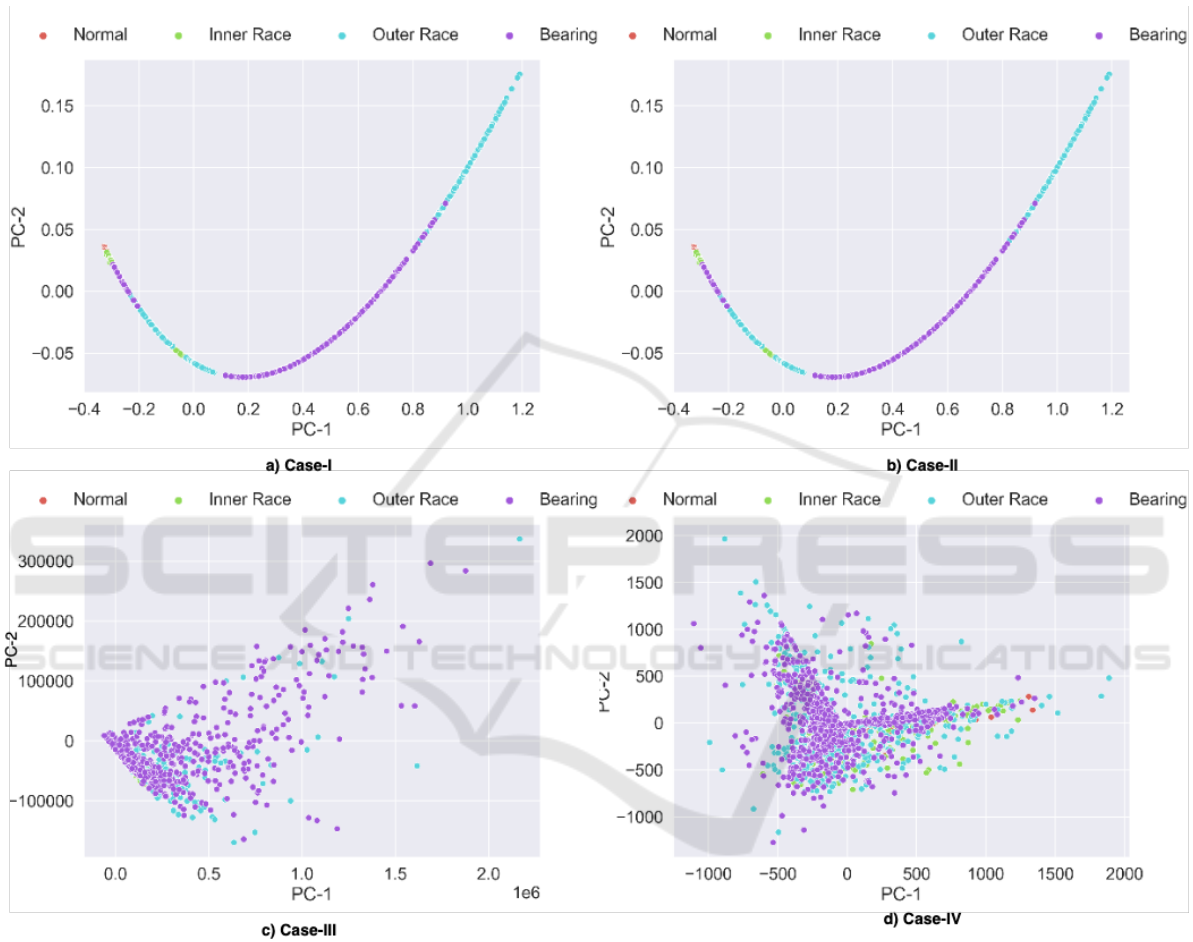


Figure 6: Distribution of the features in 2-dimensional PCA projection space with several different application scenarios.

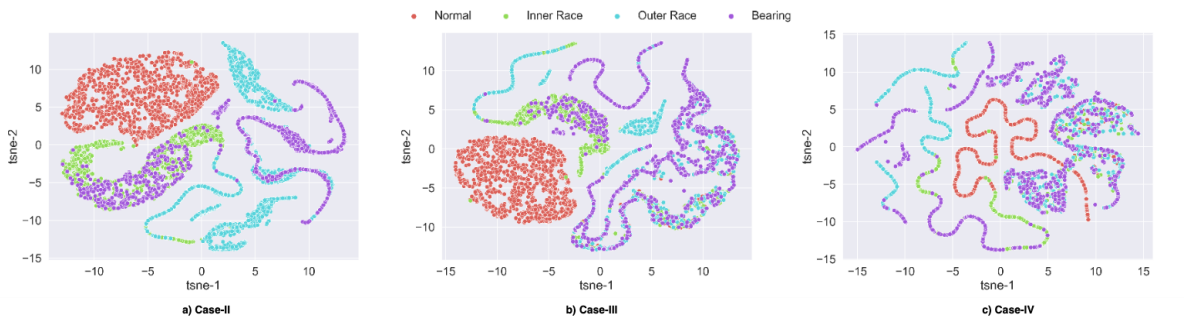


Figure 7: Distribution of the features in 2-dimensional t-SNE projection space with several different application scenarios.

of nonzero elements in the matrix E) can be set to a lower value. The numerical results are promising, and future studies of this work will focus on developing the theoretical basis of the proposed approach.

ACKNOWLEDGEMENTS

This research was supported by the European Union in the Framework of ERASMUS MUNDUS project (CyberMACS) (<https://www.cybermacs.eu>) under grant number 101082683.

REFERENCES

- Al-Rubaie, M. and Chang, J. M. (2019). Privacy-preserving machine learning: Threats and solutions. *IEEE Security & Privacy*, 17(2):49–58.
- Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.
- Blum, A., Dwork, C., McSherry, F., and Nissim, K. (2005). Practical privacy: the sulq framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138.
- Bogdanov, D., Kamm, L., Laur, S., and Sokk, V. (2018). Implementation and evaluation of an algorithm for cryptographically private principal component analysis on genomic data. *IEEE/ACM transactions on computational biology and bioinformatics*, 15(5):1427–1432.
- Dwork, C., Roth, A., et al. (2014a). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.
- Dwork, C., Talwar, K., Thakurta, A., and Zhang, L. (2014b). Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 11–20.
- Frederiksen, T. K., Nielsen, J. B., and Orlandi, C. (2015). Privacy-free garbled circuits with applications to efficient zero-knowledge. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 191–219. Springer.
- Goldberger, J., Hinton, G. E., Roweis, S., and Salakhutdinov, R. R. (2004). Neighbourhood components analysis. *Advances in neural information processing systems*, 17.
- Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. JHU press.
- Hindistan, Y. S. and Yetkin, E. F. (2023). A hybrid approach with gan and dp for privacy preservation of iiot data. *IEEE Access*, 11:5837–5849.
- Hongxia, W., Xiaohui, Y., and Ming, Y. (2016). Study on predictive maintenance strategy. *International Journal of u-and e-Service, Science and Technology*, 9(4):295–300.
- Kung, S.-Y. (2017). Compressive privacy: From information estimation theory to machine learning [lecture notes]. *IEEE Signal Processing Magazine*, 34(1):94–112.
- Li, X., Niu, J., Khan, M. K., and Wang, Z. (2013). Applying lu decomposition of matrices to design anonymity bilateral remote user authentication scheme. *Mathematical Problems in Engineering*, 2013(1):910409.
- Loparo, K. (2012). Case western reserve university bearing data center. *Bearings Vibration Data Sets, Case Western Reserve University*, pages 22–28.
- Merkle, R. C. (1978). Secure communications over insecure channels. *Communications of the ACM*, 21(4):294–299.
- Raptis, T. P., Passarella, A., and Conti, M. (2019). Data management in industry 4.0: State of the art and open challenges. *IEEE Access*, 7:97052–97093.
- Saufi, S. R., Isham, M. F., Ahmad, Z. A., and Hasan, M. D. A. (2023). Machinery fault diagnosis based on a modified hybrid deep sparse autoencoder using a raw vibration time-series signal. *Journal of Ambient Intelligence and Humanized Computing*, 14(4):3827–3838.
- Smith, W. A. and Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical systems and signal processing*, 64:100–131.
- Soveizi, N., Turkmen, F., and Karastoyanova, D. (2023). Security and privacy concerns in cloud-based scientific and business workflows: A systematic review. *Future Generation Computer Systems*, 148:184–200.
- Strang, G. (2000). *Linear algebra and its applications*.
- Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Varshney, S., Sandhu, R., and Gupta, P. (2019). Qos based resource provisioning in cloud computing environment: a technical survey. In *Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part II 3*, pages 711–723. Springer.
- Wagner, I. and Eckhoff, D. (2018). Technical privacy metrics: a systematic survey. *ACM Computing Surveys (Csur)*, 51(3):1–38.
- Weinberger, K. Q. and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research*, 10(2).
- Yi, X., Paulet, R., Bertino, E., Yi, X., Paulet, R., and Bertino, E. (2014). *Homomorphic encryption*. Springer.