Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 180 (2021) 734–743

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

Beyond federated learning: On confidentiality-critical machine learning applications in industry

Werner Zellinger[a,*], Volkmar Wieser[a], Mohit Kumar[a,b], David Brunner[a], Natalia Shepeleva[a], Rafa Gálvez[c], Josef Langer[d], Lukas Fischer[a], Bernhard Moser[a]

[a]*Software Competence Center Hagenberg GmbH (SCCH), Softwarepark 21, 4232 Hagenberg, Austria*
[b]*Faculty of Computer Science and Electrical Engineering, University of Rostock, Universitätsplatz 1, 18051 Rostock, Germany*
[c]*Computer Security and Industrial Cryptography, KU Leuven, Oude Markt 13, 3000 Leuven, Belgien*
[d]*ventopay GmbH, Softwarepark 37, 4232 Hagenberg, Austria*

Abstract

Federated machine learning frameworks, which take into account confidentiality of distributed data sources are of increasing interest in smart manufacturing. However, the scope of applicability of most such frameworks is restricted in industrial settings due to limitations in the assumptions on the data sources involved. In this work, first, we shed light on the nature of this arising gap between current federated learning and requirements in industrial settings. Our discussion aims at clarifying related notions in emerging sub-disciplines of machine learning, which are partially overlapping. Second, we envision a new confidentiality-preserving approach for smart manufacturing applications based on the more general setting of transfer learning, and envision its implementation in a module-based platform.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Industry 4.0 and Smart Manufacturing

Keywords: machine learning; federated learning; collaborative learning; transfer learning; smart manufacturing

1. Introduction

Machine learning is considered a driving force of the on-going smart factory revolution. Especially, approaches for federated learning are used to overcome problems related to confidentiality risks, on the one hand, and shortcomings in data coverage, on the other hand. The data coverage issue is tackled by improving a global model based on distributed data sources. While a single distributed source only represents a limited view of the data, the diversity across all available individual data sources will provide an enriched view allowing improved machine learning models. By making the global model available to the users the whole system learns collaboratively and a single participant takes

* Corresponding author. Tel.: +43-50-343-867.
E-mail address: werner.zellinger@scch.at

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the International Conference on Industry 4.0 and Smart Manufacturing
10.1016/j.procs.2021.01.296

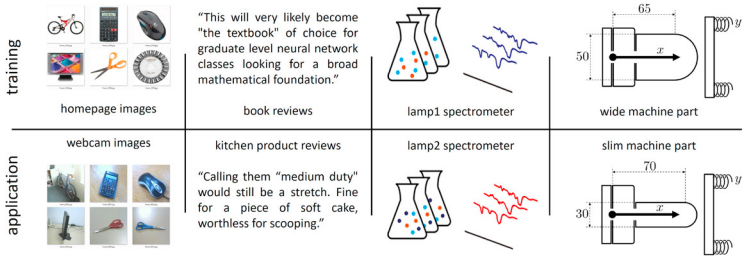

Fig. 1. Examples of deviating statistical characteristics between training and application setting.

benefit from the contribution of the others. This way the users become collaborating participants of a federated learning system. A typical application scenario arises for example from scattered users of a mobile text processing app [7].

In smart manufacturing, for example when setting up the production process of a new product of small lot size, there is also the problem of data coverage. As a rule, data is sparse or expensive to obtain. One way to overcome or, at least, to mitigate this problem is to try to learn from related tasks with a similar setting. At a first glance this resembles the situation of federated learning. But, the problem is more subtle and relies on what can be assumed about the relation between the feature spaces and the statistical characteristics of the data sources involved.

Typically, in federated learning the data source of an individual participant can be considered as a restricted view from the perspective of a presumed global data pool. In mathematical terms this translates to saying that the probability distributions of all included distributed data sources, both at training time and application time, are marginals from a joint probability distribution on a joint feature space. This assumption is often violated in industrial applications, such as fault diagnosis systems with sensors different from the training ones [41], object recognition systems based on cameras different from the training ones [21, 13], steel manufacturing machines with tool settings different from the ones of the training machines [46] and chemical measurement systems with spectrometers different from the ones the model is calibrated on [31]. See Figure 1 for illustrations.

In contrast to federated learning, the emerging research field of transfer learning aims at tackling the violation of the assumption of a presumed joint distribution. This way transfer learning allows larger deviations in the statistical characteristics of the data, including different characteristics in the training and application setting [34, 3]. While typically in federated learning scenarios also confidentiality and privacy problems are encountered, federated learning is associated with privacy-preserving collaborative learning. However, in typical transfer learning settings the aspect of protecting confidentiality remains underrepresented. In this work, therefore, we address the problem of confidentiality protection in transfer learning settings. In this context, we envision a new module-based platform for designing and implementing confidentiality-preserving transfer learning algorithms for smart manufacturing applications.

In our approach we consider how transfer-learning functionalities can be delivered while keeping data confidential from third parties. We leverage differential privacy [10] as a mechanism to provide confidentiality through the introduction of noise in the dataset. First, we exploit the composability property of differential privacy in order to obtain theoretical guarantees for the level of protection obtained by the resulting composed methods. Second, we exploit the increased robustness capabilities of transfer learning to tackle potential instability issues caused by introduction of noise.

As illustrated in Figure 2, our approach distinguishes itself from other approaches in terms of scope of applicability w.r.t. the level of required confidentiality protection versus the generality of the statistical assumptions made on the underlying problem. This way, our approach goes beyond federated learning, e.g. NVIDIA Clara [25] or Google G-board [27], by compensating for mismatch between statistical characteristics of training and application data sources. On the other hand, our approach goes beyond most frameworks in transfer learning, as e.g. Googles framework for text translation [37] or the NVIDIA Transfer Learning Toolkit [33], by taking additionally mechanisms for confidentiality protection into account.

This work is structured as follows: Section 2 describes the terms confidentiality and privacy in industrial settings, discusses related model-agnostic privacy-preservation methods and gives an overview of threat models, Section 3 reviews the concept of transfer learning and compares it to the term of federated learning, Section 4 proposes our

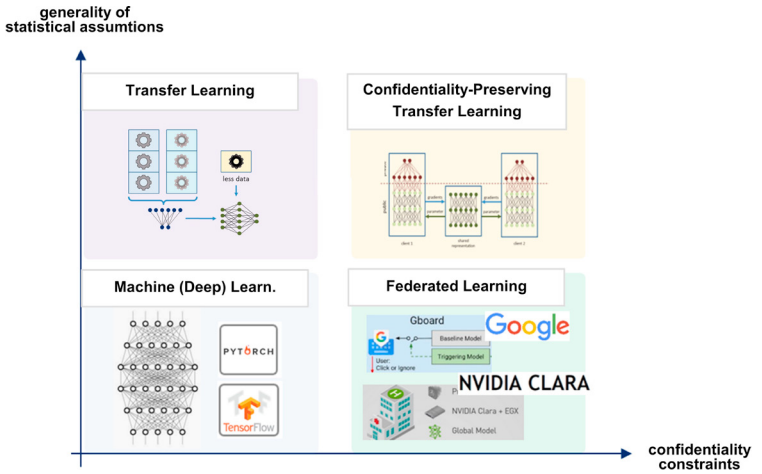

Fig. 2. Relations between confidentiality-preserving transfer learning, classical machine learning, federated learning and transfer learning w.r.t. the two dimensions criticality of confidentiality and generality of the underlying assumptions.

module-based approach to confidentiality-preserving transfer learning, and, Section 5 concludes the work and gives some future steps.

2. Confidentiality

In this section, we draw a parallel between confidentiality and privacy, present a concise overview over the existing mechanisms to enforce them, and describe a set of attacks against confidentiality applicable to federated learning systems.

Confidentiality and Privacy in Industry.. In industrial settings, machine learning can be used with data coming from two kinds of data sources: people (e.g. workers), and machines. The term *privacy* has been widely used by computer scientists and engineers as a synonym for confidentiality [20]. However, making a distinction between them is notably useful in industry. While privacy can be defined as a right (cf. GDPR [11]) entitled to individuals to protect information about them [11, 32]), confidentiality can be more broadly applied to any kind of data that must be available only to the appropriate entities. Therefore, in order for this work to also consider data about industrial devices, we will apply methods devised to protect privacy of individuals to data which must remain confidential but whose subject are not people, but machines. Note that this does not preclude the application of our platform with personal data of e.g. workers of a plant, as long as the desired privacy protection can be realized through confidentiality mechanisms.

Differential Privacy.. In this work, we focus on machine learning models with mathematical guarantees for confidentiality. One approach is to consider the concept of *differential privacy*.

Differential Privacy [8, 10] is a formalism to quantify the degree to which the confidentiality for each individual in the dataset is preserved while releasing the output of a data analysis algorithm. Differential privacy provides a guarantee that an adversary, by virtue of presence or absence of an individual's data in the dataset, would not be able to draw any conclusions about an individual from the released output of the analysis algorithm. This guarantee is achieved by means of a randomization of the data analysis process. In the context of machine learning, randomization is carried out via either adding random noise to the input or output of the machine learning algorithm or modifying the machine learning algorithm itself. However, the injection of noise would in general result in a loss of algorithm's accuracy. Therefore, design of a noise injection mechanism achieving a good trade-off between privacy and accuracy is a topic of interest [2, 18, 19, 15, 16].

Many works address the issues of differential privacy in machine learning. A general framework to provide utility guarantees for a single count query, subject to $\epsilon-$differential privacy, was studied in [18]. A similar study taking a minimax model of utility for information consumers has been made in [19]. For single real-valued query function,

a staircase-shaped probability density function was suggested in [16] for an optimal ϵ-differentially private noise adding mechanism. The approach was extended to the vector real-valued query function in [15]. For integer-valued query functions, the optimal mechanisms in (ϵ, δ)-differential privacy were studied in [16]. For single real-valued query function, the trade-off between privacy and utility in $(0, \delta)-$differential privacy was studied in [17].

Our recent work [23] has suggested a novel entropy based approach for resolving the privacy-utility trade-off for real-valued data matrices. The study in [23] derives mathematically for real-value data matrices the probability density function of noise that minimizes the expected noise magnitude together with satisfying the sufficient conditions for (ϵ, δ)-differential privacy. The optimal noise adding mechanism of [23] is used in [24] to attain differential privacy for the learning of a deep model in a distributed setting where the post-processing invariance property of differential privacy is exploited for building a global fuzzy rule-based classifier that *aggregates* the local confidentiality-preserving deep models.

Attacks Against Confidentiality in Federated Learning.. The necessity of techniques to protect confidentiality is emphasized by a growing number of attacks on machine learning models. Some of these attacks apply to traditionally trained models, but many are aimed specifically at federated learning settings. The consideration of such attacks is a key tool in confidentiality-preserving transfer learning frameworks since they provide empirical baselines for confidentiality protection.

The purpose of such attacks is often not to disrupt learning, but to extract sensitive information in the process of or after the creation of the models. In the following, we review different commonly applied attacks:

In a *model inversion attack* an attacker tries to reconstruct training samples only through black-box access to a trained model [12]. In a federated setting the attacker could be one of the participants, which allows for even more effective reconstruction since internal information on the architecture and training process is available [44, 22].

As another example, determining whether a given sample is part of the training data or not can be of interest in industry. *Membership inference attacks* show a similar susceptibility of machine learning models to attackers with additional information on their creation [39, 30, 42, 26].

In some cases, it is not the information explicitly encoded in the data, that is interesting to an attacker or competing party, but information a model implicitly takes up. *Property inference attacks* target the identification of statistical properties of the underlying data sources [28, 14]. For example consider a federated learning model for the detection of faulty industrial parts. It is not information about the type of part that is of interest, but the frequency with which it is processed that would give a competitor an edge.

Finally, an attacker that is part of a federated learning scheme might want to bias the predictions of a trained model to a certain extent, in a targeted fashion by inserting manipulated information during the training process, a technique referred to as *poisoning* [4] or *backdooring* [1].

3. Tackling Mismatch of Statistical Characteristics

In this section, we shortly review the concepts of transfer learning, multi-task learning, federated learning and confidentiality-preserving transfer learning. We describe the main differences between these concepts which arise from the statistical assumptions made on the relations between the included data sources, see e.g. Figure 3. Finally, these assumptions are summarized by Figure 4.

Transfer Learning.. In contrast to classical statistical learning, methods of transfer learning take into account data from more than one data distribution. Let us consider n *source* data distributions $\mu_1, \dots, \mu_n$ and one *target* data distribution ν, which represent different data sources. For data source examples consider different manufacturing machines [46], vibration data from bearings of different physical size [41], different chemical measurement systems [31] and 3D-video cameras [38, 47]. The goal of transfer learning[1] is to find, based on typically large source data samples from $\mu_1, \dots, \mu_n$ and typically a sparse (possibly unlabeled) target data sample from ν, a machine learning model

[1] In a probabilistic setup, $\mu_1, \dots, \mu_n, \nu$ are Borel probability measures on vector state-spaces $\mathcal{X}_1, \dots, \mathcal{X}_{n+1}$, see e.g. [35, 3]. Different transfer learning problems can be distinguished based on differences between the state-spaces, relations between the distributions and the size of given samples, e.g. *multi-task learning* (ν is joint distribution of $\mu_1, \dots, \mu_n$), *domain generalization* (no target data sample), *domain adaptation* (unlabeled target data sample), and *sample selection bias* (target distribution dominates source distributions).

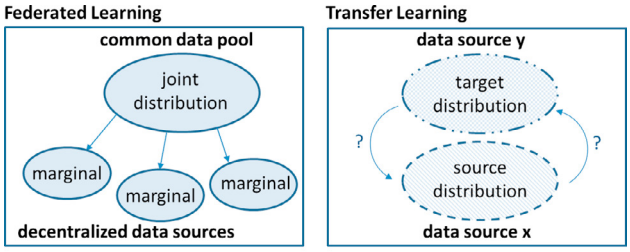

Fig. 3. While in federated learning the data sources are closely related via marginals from a presumed joint distribution, in general in a transfer learning setting, the statistical relationship between the data sources is unknown.

which will perform well on data samples drawn from the distribution ν. In the most general setting, the task, e.g. classification, to be solved on data from ν can be different from the tasks, e.g. classification, to be solved on data from $\mu_1, \dots, \mu_n$.

One key challenge is that the relation between the source distributions $\mu_1, \dots, \mu_n$ and the target distribution ν is unknown and has to be estimated based on the given information.

For example consider the following industrial problem of transfer learning.

Problem 1 (Small Lot-Size Problem). *The **task** is to estimate some product quality measure from sensor data of different manufacturing machines. Let $\mu_1, \dots, \mu_n$ be distributions of sensor data from different source manufacturing machines and ν be a distribution of sensor data from some target manufacturing machine. The **goal** is to find, based on large amounts of labeled source sensor data (where the product quality is already ranked by a human expert) from $\mu_1, \dots, \mu_n$ and small amounts of unlabeled data from μ (where the product quality is not ranked by a human), a machine learning model, which performs well on new unknown (future) data from the target manufacturing machine.*

For simplicity, let us consider Problem 1 above under the assumption that the task, i.e. ranking product quality by ten classes $1, 2, \dots, 10$, is the same for the source and target data. Then, the key challenge is that the relation between the distributions of the source sensor data and the target sensor data is unknown and has to be estimated based on a small sample of unlabeled data. In practice, often, also physical properties of the manufacturing machines are taken into account to solve Problem 1, see e.g [46].

Multi-Task Learning.. Given n data samples from respective distributions $\mu_1, \dots, \mu_n$, the goal of multi-task learning is to find a machine learning model, which performs well on each of the distributions $\mu_1, \dots, \mu_n$. The tasks on the data from $\mu_1, \dots, \mu_n$ can be different, e.g. classification with ten classes for μ_1, regression on $[1, 10]$ for μ_2 and classification with three classes for $\mu_3, \dots, \mu_n$.

Consider Problem 1 of estimating some product quality from sensor data of manufacturing machines. This problem turns into a multi-task learning problem, if we replace the goal above with goal to find, based on data from $\mu_1, \dots, \mu_n$, a model that performs well on each of the distributions $\mu_1, \dots, \mu_n$, which refer to different sensor modalities. The new target distribution ν is not taken into account in this problem setting.

Federated Learning.. Prominent examples of federated learning applications are text completion, e.g. as implemented in Google G-board [27], or tumor segmentation on data from different hospitals, as e.g. implemented in NVIDIA Clara [25]. Typically the data distributions in federated learning come from decentralized edge devices equipped with the same functionality to solve the same task, e.g. segmentation.

From the point of view of a learning task, the goal of federated learning is to find, based on data from source distributions $\mu_1, \dots, \mu_n$, a machine learning model, which performs well on data from the distributions $\mu_1, \dots, \mu_n$.

Typical application scenarios of federated learning are confidentiality critical. So that, for example, the privacy of the people, who's data is processed, is protected. This means that for sake of privacy issues, usually, problems of federated learning impose the restriction that data from $\mu_1, \dots, \mu_n$ should not be exchanged or leaked between the owners of the decentralized edge devices. This particular limitation affects the permissible choice of machine learning models.

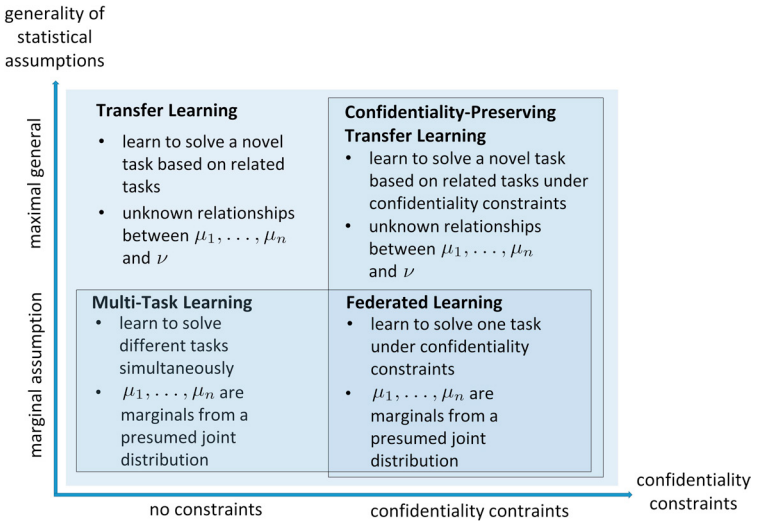

Fig. 4. Refining Figure 2 by relating the problem settings of transfer learning, multi-task learning, federated learning and confidentiality-preserving transfer learning in terms of generality of underlying statistical assumptions and imposed confidentiality constraints. Intersections of boxes refer to intersections of sets of problems, e.g. federated learning problems are multi-task learning problems and therefore also transfer learning problems.

Confidentiality-Preserving Transfer Learning.. Transfer learning can be used in scenarios where data confidentiality is important. This is especially the case if *ϵ-differential* privacy [9] constraints are uphold, see Section 2. These constraints enable models to bound the amount of information they leak about individual items in their dataset given any background information an adversary may have about it. Transfer learning algorithms which satisfy such constraints are called *confidentiality-preserving*.

Federated Learning Problems are Transfer Learning Problems.. Let us start by showing that each multi-task learning problem is also a problem of transfer learning. Therefore, let $\mu_1, \ldots, \mu_n$ be n source distributions. The requirement that a model, found based on data from $\mu_1, \ldots, \mu_n$, performs well on each of the distributions $\mu_1, \ldots, \mu_n$, is equivalent to the requirement that the model performs well on the joint distribution of $\mu_1, \ldots, \mu_n$ which we denote by ν. It can further be seen that each problem of multi-task learning is a (sub-)problem of transfer learning by recalling the transfer learning goal of finding a model which performs well on a target distribution ν. However, not every transfer learning problem can be interpreted as a problem of multi-task learning. In general in a transfer learning setting, the relationship between a source and a target data distribution might be unknown and the target data might originate from a completely new (unknown) data source possibly different from the joint distribution of $\mu_1, \ldots, \mu_n$.

Problems of federated learning are problems of multi-task learning given the same goal of finding a model which performs well on the given source distributions $\mu_1, \ldots, \mu_n$. However, not every problem of multi-task learning is a problem of federated learning. In contrast to problems of multi-task learning, problems of federated learning are based on data from decentralized edge devices or servers. That is, additional constraints on the physical system of the data sources are stated. In addition, confidentiality-constraints need to be satisfied. In multi-task learning no constraints on the physical system are stated, confidentiality-constraints are often not included, and, in *multi-task learning* different tasks can be solved on the data from $\mu_1, \ldots, \mu_n$.

As a consequence of the reasoning above, see Figure 4 for an illustrative summary, the difference between transfer learning and federated learning becomes apparent: In federated learning the distributions are closely related via marginals from a (hypothetical) joint distribution. In contrast, in general in a transfer learning setting, the relationship between a source and a target data distribution might be unknown and the target data might originate from a completely new (unknown) data source. Rather, it is the aim of transfer learning approaches to (implicitly) approximate this relationship e.g. by measuring the deviation between the probability distributions by appropriate probability distribution metrics, see Figure 3. Another crucial difference is that federated learning systems typically impose constraints on the confidentiality. On the contrary, transfer learning applications originate in the problem of reusing data in a different application context. Therefore, standard transfer learning approaches neglect confidentiality issues. This

means that transfer learning methods, which are trained on the whole data from $\mu_1, \ldots, \mu_n$ stored on one server can not be applied to problems of federated learning without taking into account confidentiality-preservation techniques as proposed in Section 2.

Combining Transfer Learning and Confidentiality-Preservation.. A decisive point in our platform is the robustness of transfer learning approaches when applied to noisy data such as generated by model-agnostic techniques for confidentiality protection as described in Section 2. We therefore focus on moment-based approaches such as proposed in [40, 36] which are distinguished by the fact that they are relatively insensitive to changes of the learning setting, see e.g. [45].

4. Module-Based Platform Approach

Ensuring high data quality, iterative model training, model validation and model deployment, are common tasks in a typical machine learning life-cycle. In an industrial environment, these tasks should be automated as much as possible to guarantee reproducibility and continuous model improvement. As a basis for our platform, we therefore rely on the TensorFlow Extended (TFX) Framework [29], which is an automated end-to-end platform for developing machine learning pipelines. The TFX framework has been developed by Google for industrial environments and contains the essential tools which serves the need of scalability and modularity.

Extension of TFX Framework. Figure 5 shows the architecture of our platform. The core of the platform consists of the following modules which are controlled by an Orchestrator: Data Confidentiality Preservation, Task Configurator, Data Preprocessor, Model Trainer, Model Evaluator, Model Confidentiality Preservation. Each of these modules is a collection of functions used for a particular step in the machine learning pipeline and can be extended with additional functionality upon the need. The Orchestrator controls operation of these modules and the order of applied sub-modules. All information about performance and outcome of such sub-modules as well as the configuration of each run is stored by the Orchestrator in the so-called Meta information layer. Apart of controlling communication between modules, the Orchestrator also takes care of the external communication with users and provides information of on-going experiments and data processing.

Our platform allows three levels of access: The User level, the DevOp level and the Developer level.

At the Developer level only developers have access to the core functionality of the platform. The core responsibility of the developers is to write, maintain and guarantee the work of the core functionality at this level. The core functionality of the platform includes all sub-modules, decisions about the storage of Meta information, decisions about the control of the machine learning flow and decisions about the information provided to the users.

At the DevOp level the access to the modules is restricted to operational. The work at this level includes preparation steps for running the ready-to-use environment for the users. Due to high flexibility of the platform such runs can be done either in an external server, e.g. Amazon Cloud, Google Cloud, or on a company side.

At the User level the usage of the platform is realized by providing data and configuration files. The users can monitor and control the process of running task via a Visualizer, e.g. stop or cancel the experiment, change parameters in the configuration and re-run the experiment. The direct communication between the users, the modules and the Orchestrator is restricted for confidentiality. In addition, the three sub-modules, Data Confidentiality Preservation, Confidential Training and Model Confidentiality Preservation are defined for further guarantees of confidentiality. This separation between the three sub-modules realizes the important separation in machine learning between data, model and learning, see e.g. [5]. As a result the user will receive a ready to deploy model.

Various problems of confidentiality-preserving transfer learning can be realized by using the platform described above. For example, let us consider a company which wants to solve a small lot-size problem as stated in Problem 1. The goal is to estimate some product quality on a target manufacturing machine hosted at a new customer by using data from machines of old source customers. This can be realized as follows All customers have access to the User level, however, without sharing any data with each other. The Developer level is completely at the company's' side. At this level, the main task of the developers is to adapt pre-defined sub-modules as e.g. Transfer Learning and Confidential Training. In addition, for confidentiality preserving of the customer data, the sub-module for Confidential Training has to be developed. The main task at the DevOp level is to find a trustworthy and secure server to run the platform, e.g. in-house or Google Cloud. This secure server can also be a separate trusted party to further increase confidentiality.

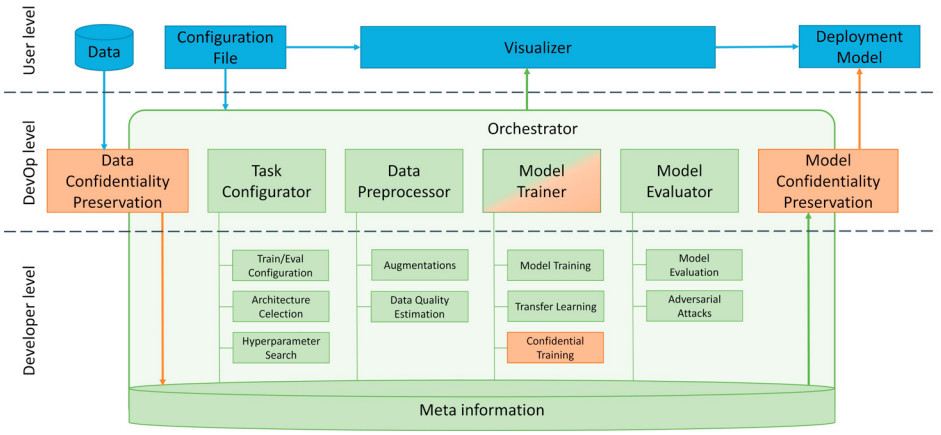

Fig. 5. End-to-end platform for confidentiality-preserving transfer learning.

We assume that the Data Confidentiality Preservation module and the Model Confidentiality Preservation module are pre-defined and privacy-preservation algorithms as e.g. described in Section 2 are implemented.

Platform Properties. The platform shows a high degree of scalability due to the generality of the statistical assumptions made on the data, see Section 3. Another advantage of the platform is its high degree of trust by allowing theoretical differential privacy guarantees. This is achieved by the mathematical composability property for algorithms showing a module-based form as envisioned above, see Section 2 and references therein. Finally, our platform enables easy model deployment in industrial settings e.g. by applying TFX with Kubeflow [6], a service dedicated specifically for this task.

5. Conclusion and Future Work

In this work, we envision a new machine learning platform for transfer learning problems under confidentiality constraints as often arise in industry. In contrast to classical platforms in federated industrial settings, our approach makes weaker assumptions on the statistical characteristics of the data. Our core idea is a module-based combination of confidentiality-preserving noise adding methods with robust transfer learning algorithms. Theoretical guarantees for differential privacy can be obtained for the applied noise adding mechanisms and extended to the platform using the composability property of sequentially applied modules.

One important future step is the empirical investigation of the performance and differential privacy of algorithms implemented in the platform. One evidence for the performance of the envisioned platform is given by recently presented algorithms which have a structure similar to the module-based approach proposed in this paper, see e.g. [43]. However, these algorithms don't apply metric-based regularization strategies as discussed in Section 3 which have been shown to perform well in smart manufacturing applications. Another future step is the investigation of the drawbacks of the combination of the transfer learning methods described in Section 3 with the confidentiality-preserving methods described in Section 2.

Acknowledgements

The research reported in this paper has been funded by the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), the Federal Ministry for Digital and Economic Affairs (BMDW), and the Province of Upper Austria in the frame of the COMET–Competence Centers for Excellent Technologies Programme and the COMET Module S3AI managed by Austrian Research Promotion Agency FFG. We further acknowledge the support of the projects AutoQual-I and PRIMAL.

References

[1] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov. How to backdoor federated learning. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2938–2948, Online, 26–28 Aug 2020. PMLR.

[2] Borja Balle and Yu-Xiang Wang. Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. *CoRR*, abs/1805.06530, 2018.

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010.

[4] Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, and Seraphin Calo. Analyzing federated learning through an adversarial lens. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 634–643, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[5] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[6] Ekaba Bisong. Kubeflow and kubeflow pipelines. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, pages 671–685. Springer, 2019.

[7] Raffaele Cioffi, Marta Travaglioni, Giuseppina Piscitelli, Antonella Petrillo, and Fabio De Felice. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, 12(2):492, 2020.

[8] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, pages 486–503, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[10] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[11] European Commission. General Data Protection Regulation.

[12] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 1322–1333, New York, NY, USA, 2015. Association for Computing Machinery.

[13] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

[14] Karan Ganju, Qi Wang, Wei Yang, Carl A. Gunter, and Nikita Borisov. Property inference attacks on fully connected neural networks using permutation invariant representations. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, page 619–633, New York, NY, USA, 2018. Association for Computing Machinery.

[15] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, Oct 2015.

[16] Q. Geng and P. Viswanath. Optimal noise adding mechanisms for approximate differential privacy. *IEEE Transactions on Information Theory*, 62(2):952–969, Feb 2016.

[17] Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Optimal noise-adding mechanism in additive differential privacy. *CoRR*, abs/1809.10224, 2018.

[18] A. Ghosh, T. Roughgarden, and M. Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[19] Mangesh Gupte and Mukund Sundararajan. Universally optimal privacy mechanisms for minimax agents. In *Proceedings of the Twenty-ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '10, pages 135–146, New York, NY, USA, 2010. ACM.

[20] S. Gürses and C. Diaz. Two tales of privacy in online social networks. 11(3):29–37.

[21] Yutaka Hirano, Christophe Garcia, Rahul Sukthankar, and Anthony Hoogs. Industry and object recognition: Applications, applied research and challenges. In *Toward category-level object recognition*, pages 49–64. Springer, 2006.

[22] Briland Hitaj, Giuseppe Ateniese, and Fernando Perez-Cruz. Deep models under the gan: Information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 603–618, New York, NY, USA, 2017. Association for Computing Machinery.

[23] Mohit Kumar, Michael Rossbory, Bernhard A. Moser, and Bernhard Freudenthaler. Deriving an optimal noise adding mechanism for privacy-preserving machine learning. In Gabriele Anderst-Kotsis, A Min Tjoa, Ismail Khalil, Mourad Elloumi, Atif Mashkoor, Johannes Sametinger, Xabier Larrucea, Anna Fensel, Jorge Martinez-Gil, Bernhard Moser, Christin Seifert, Benno Stein, and Michael Granitzer, editors, *Proceedings of the 3rd International Workshop on Cyber-Security and Functional Safety in Cyber-Physical (IWCFS 2019), August 26-29, 2019, Linz, Austria*, pages 108–118, Cham, 2019. Springer International Publishing.

[24] Mohit Kumar, Michael Rossbory, Bernhard A. Moser, and Bernhard Freudenthaler. Differentially private learning of distributed deep models. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20 Adjunct, page 193–200, New York, NY, USA, 2020. Association for Computing Machinery.

[25] Wenqi Li, Fausto Milletarì, Daguang Xu, Nicola Rieke, Jonny Hancox, Wentao Zhu, Maximilian Baust, Yan Cheng, Sébastien Ourselin, M Jorge Cardoso, et al. Privacy-preserving federated brain tumour segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 133–141. Springer, 2019.

[26] Yunhui Long, Vincent Bindschaedler, Lei Wang, Diyue Bu, Xiaofeng Wang, Haixu Tang, Carl A Gunter, and Kai Chen. Understanding membership inferences on well-generalized learning models. *arXiv preprint arXiv:1802.04889*, 2018.

[27] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282, 2017.

[28] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 691–706, 2019.

[29] Akshay Naresh Modi, Chiu Yuen Koo, Chuan Yu Foo, Clemens Mewald, Denis M. Baylor, Eric Breck, Heng-Tze Cheng, Jarek Wilkiewicz, Levent Koc, Lukasz Lew, Martin A. Zinkevich, Martin Wicke, Mustafa Ispir, Neoklis Polyzotis, Noah Fiedel, Salem Elie Haykal, Steven Whang, Sudip Roy, Sukriti Ramesh, Vihan Jain, Xin Zhang, and Zakaria Haque. Tfx: A tensorflow-based production-scale machine learning platform. In *KDD 2017*, 2017.

[30] Milad Nasr, Reza Shokri, and Amir Houmansadr. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 739–753, 2019.

[31] Ramin Nikzad-Langerodi, Werner Zellinger, Susanne Saminger-Platz, and Bernhard Moser. Domain-invariant regression under beer-lambert's law. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 581–586. IEEE, 2019.

[32] Helen Fay Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books.

[33] NVIDIA. Transfer learning toolkit, 2020. https://developer.nvidia.com/transfer-learning-toolkit.

[34] Sinno Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, Oct 2010. 00835.

[35] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[36] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1406–1415, 2019.

[37] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.

[38] Florian Seitner, Matej Nezveda, Margrit Gelautz, Georg Braun, Christian Kapeller, Werner Zellinger, and Bernhard Moser. Trifocal system for high-quality inter-camera mapping and virtual view synthesis. In *2015 International Conference on 3D Imaging (IC3D)*, pages 1–8. IEEE, 2015.

[39] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18, 2017.

[40] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016.

[41] Zhe Tong, Wei Li, Bo Zhang, and Meng Zhang. Bearing fault diagnosis based on domain adaptation using transferable features under different working conditions. *Shock and Vibration*, 2018, June 2018.

[42] Stacey Truex, Ling Liu, Mehmet Gursoy, Lei Yu, and Wenqi Wei. Demystifying membership inference attacks in machine learning as a service. *IEEE Transactions on Services Computing*, PP:1–1, 02 2019.

[43] Yang Wang, Quanquan Gu, and Donald Brown. Differentially private hypothesis transfer learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 811–826. Springer, 2018.

[44] Zhibo Wang, Mengkai Song, Zhifei Zhang, Quian Wang, and Hairong Qi. Beyond inferring class representatives: User-level privacy leakage from federated learning. In *IEEE International Conference on Computer Communications (INFOCOM)*, 2019.

[45] Werner Zellinger. Moment-based domain adaptation: Learning bounds and algorithms. *Doctoral Thesis (Johannes Kepler University Linz)*, April 2020.

[46] Werner Zellinger, Thomas Grubinger, Michael Zwick, Edwin Lughofer, Holger Schöner, Thomas Natschläger, and Susanne Saminger-Platz. Multi-source transfer learning of time series in cyclical manufacturing. *Journal of Intelligent Manufacturing*, 31(3):777–787, 2020.

[47] Werner Zellinger, Bernhard A Moser, Ayadi Chouikhi, Florian Seitner, Matej Nezveda, and Margrit Gelautz. Linear optimization approach for depth range adaption of stereoscopic videos. In *Electronic Imaging, Stereoscopic Displays and Applications XXVII*, pages 1–6. IS&T Electronic Imaging, 2016.