From Domain Adaptation to Federated Learning

Zahra Taghiyarrenani

School of Information Technology

Halmstad University Dissertations, 2024
From Domain Adaptation to Federated Learning

Zahra Taghiyarrenani
Abstract

Data-driven methods have been gaining increasing attention; however, along with the benefits they offer, they also present several challenges, particularly concerning data **availability**, **accessibility**, and **heterogeneity**, the three factors that have shaped the development of this thesis.

Data availability is the primary consideration when employing data-driven methodologies. Suppose we consider a system for which we aim to develop a Machine Learning (ML) model. Gathering labeled samples, particularly in the context of real-world problem-solving, consistently poses challenges. While collecting raw data may be feasible in certain situations, the process of labeling them is often difficult, leading to a shortage of labeled data. However, historical (outdated) data or labeled data may occasionally be available from different yet related systems.

A feasible approach would be to leverage data from different but related sources to assist in situations in which data is scarce. The challenge with this approach is that data collected from various sources may exhibit statistical differences even if they have the same features, i.e., data heterogeneity. Data heterogeneity impacts the performance of ML models. This issue arises because conventional machine learning algorithms assume what’s known as the IID (Independently and Identically Distributed) assumption; training and test data come from the same underlying distribution and are independent and identically sampled. The IID assumption may not hold when data comes from different sources and can result in a trained model performing less effectively when used in another system or context. In such situations, Domain Adaptation (DA) is a solution.

DA enhances the performance of ML models by minimizing the distribution distance between samples originating from diverse resources. Several factors come into play within the DA context, each necessitating distinct DA methods.

In this thesis, we conduct an investigation and propose DA methods while considering various factors, including the number of domains involved, the quantity of data available (both labeled and unlabeled) within these domains, the task at hand (classification or regression), and the nature of statistical heterogeneity among samples from different domains, such as covariate shift or concept shift.
It is crucial to emphasize that DA techniques work by assuming that we access the data from different resources. Data may be owned by different data owners, and data owners are willing to share their data. This data accessibility enables us to adapt data and optimize models accordingly. However, privacy concerns become a significant issue when addressing real-world problems, for example, where the data owners are from industry sectors. These privacy considerations necessitate the development of privacy-preserving techniques, such as Federated Learning (FL).

FL is a privacy-preserving machine learning technique that enables different data owners to collaborate without sharing raw data samples. Instead, they share their ML models or model updates. Through this collaborative process, a global machine learning model is constructed, which can generalize and perform well across all participating domains. This approach addresses privacy concerns by keeping individual data localized while benefiting from collective knowledge to improve the global model. Among the most widely accepted FL methods is Federated Averaging (FedAvg). In this method, all clients connect with a central server. The server then computes the global model by aggregating the local models from each client, typically by calculating their average.

Similar to DA, FL encounters issues when data from different domains exhibit statistical differences, i.e., heterogeneity, that can negatively affect the performance of the global model. A specialized branch known as Heterogeneous FL has emerged to tackle this situation. This thesis, alongside DA, considers the heterogeneous FL problem.

This thesis examines FL scenarios where all clients possess labeled data. We begin by conducting experimental investigations to illustrate the impact of various types of heterogeneity on the outcomes of FL. Afterward, we perform a theoretical analysis and establish an upper bound for the risk of the global model for each client. Accordingly, we see that minimizing heterogeneity between the clients minimizes this upper bound. Building upon this insight, we develop a method aimed at minimizing this heterogeneity to personalize the global model for the clients, thereby enhancing the performance of the federated system.

This thesis focuses on two practical applications that highlight the relevant challenges: Predictive Maintenance and Network Security. In predictive maintenance, the focus is on fault identification using both DA and FL. Additionally, the thesis investigates predicting the state of health of electric bus batteries using DA. Regarding network security applications, the thesis addresses network traffic classification and intrusion detection, employing DA.
To my husband and family.
Acknowledgements

I would like to express my sincere gratitude to my supervisors, Slawomir Nowaczyk, Sepideh Pashami, and Mohamed-Rafik Bouguelia, for their skilled guidance, engagement, and constructive suggestions. Undoubtedly, I could not have undertaken this journey without their support.

I would also like to thank Markus Sirnemark, Klas Thunberg, Niklas Tedenbrant, Yuantao Fan, and Mohammed Ghaith Altarabichi for always being ready to help me understand the problems from the application point of view to solve problems related to the EVE project.

I would like to extend my sincere thanks to all my friends and lab colleagues who provide a friendly and relaxed research environment. Thank you all, Sepideh, Denni, Slawomir, Rafik, Jens, Bjorn, Josef, Stefan, Eren, Martin, Peyman, Hamid, Tommy, Tomas, Mahmoud, Farzaneh, Amira, Hadi, Awaise, Alexander, Parisa, Kevin, Abdallah, Ghaith, Tiago, Ali, Pablo, Ece, Kunru, Anna, Atiyeh, Amirhossein and, of course, all other friends.

My deepest thanks go to my family. Words cannot express my gratitude for their continuing support and engagement, kindness and patience, hard work, and sacrifices. My mother, Mehri Kabiri, has been my best friend throughout my life. I could never even hope to express just how much appreciation I hold in my heart for you, my loving mom. My father, Ahmadreza Taghiyar, has been my teacher, my tutor, my protector, and my most trusted friend in my life. He is the strongest and the kindest I’ve ever met. My brother, Saeed Taghiyar, has always been an inspiration to me. Undoubtedly, he is the most intelligent, generous, and considerate person I have ever known. I am proud to have you as a brother. I would also mention my lovely sister-in-law, Mahnaz Hosseinian, and our little one, Chamaan Taghiyar.

And last but of course not least, I would like to give my special thanks to my loving husband, Hamed Farsi, for supporting me and providing me the engagement to reach my goals. His thoughtfulness, compassion, and kindness always make me feel fortunate and blessed that I have him in my life.
List of Papers

The following papers, referred to in the text by their Roman numerals, are included in this thesis.

PAPER I: Noise-robust representation for fault identification with limited data via data augmentation
Taghiyarrenani Z, Berenji A.

PAPER II: Facilitating Semi-Supervised Domain Adaptation through Few-shot and Self-supervised Learning
Taghiyarrenani Z, Nowaczyk S, Pashami S, Bouguelia MR.
submitted

PAPER III: Multi-Domain Adaptation for Regression under Conditional Distribution Shift
Taghiyarrenani Z, Nowaczyk S, Pashami S, Bouguelia MR.

PAPER IV: ITL-IDS: Incremental Transfer Learning for Intrusion Detection Systems
Mahdavi E, Fanian A, Mirzaei A, Taghiyarrenani Z.

PAPER V: Domain Adaptation with Maximum Margin Criterion with application to network traffic classification
Taghiyarrenani Z, Farsi H.
In International Workshops of Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2022 Sep 19 (pp. 159-169). Cham: Springer Nature Switzerland.

PAPER VI: Towards Geometry-Preserving Domain Adaptation for Fault Identification
Taghiyarrenani Z, Nowaczyk S, Pashami S, Bouguelia MR. In International Workshops of Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2022 Sep 19 (pp. 451-460). Cham: Springer Nature Switzerland.

PAPER VII: **Analysis of Statistical Data Heterogeneity in Federated Fault Identification**
Taghiyarrenani Z, Nowaczyk S, Pashami S.
In PHM Society Asia-Pacific Conference 2023 Sep 4 (Vol. 4, No. 1).

PAPER VIII: **Heterogeneous Federated Learning via Personalized Generative Networks**

The following papers also contributed to the domain adaptation topic and are included in the thesis to indicate potential future directions.

PAPER IX: **Why Industry 5.0 Needs XAI 2.0?**
The 1st World Conference on eXplainable Artificial Intelligence (XAI 2023)

PAPER X: **Towards Explainable Deep Domain Adaptation**
Szymon Bobek, Sławomir Nowaczyk, Sepideh Pashami, Zahra Taghiyarrenani, and Grzegorz J. Nalepa
Joint workshops on XAI methods, challenges and applications at the 26th European Conference on Artificial Intelligence

Papers that are not included in the thesis.

PAPER XI: **An Analysis of Vibrations and Currents for Broken Rotor Bar Detection in Three-phase Induction Motors**
Berenji A, Taghiyarrenani Z.
PAPER XII: **Fault identification with limited labeled data**
Berenji A, Taghiyarrenani Z, Rohani Bastami A.

PAPER XIII: **curr2vib: Modality Embedding Translation for Broken-Rotor Bar Detection**

PAPER XIV: **Data-Centric Perspective on Explainability Versus Performance Trade-Off.**
Berenji A, Nowaczyk S, Taghiyarrenani Z.
In International Symposium on Intelligent Data Analysis 2023 Apr 1 (pp. 42-54). Cham: Springer Nature Switzerland.
Contents

Abstract i
Acknowledgements v
List of Papers vii
List of Figures xiii

1 Introduction 1
1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
1.2 Challenges and research questions . . . . . . . . . . . . . . 3
1.3 Summary of papers and Authors’ contribution . . . . . . . . 4

2 Background 13
2.1 Domain Adaptation . . . . . . . . . . . . . . . . . . . . . . 13
2.2 Federated Learning . . . . . . . . . . . . . . . . . . . . . . 17

3 Summary of the methods proposed in the papers 23

4 Conclusion and future work directions 31

References 35

Paper I 39
Paper II 47
Paper III 65
Paper IV 79
Paper V 97
Paper VI 111
List of Figures

1.1 Contributions of this thesis and the appended papers. 5

2.1 Categorization of Domain Adaptation methods. The density of the gray boxes indicates the extent of the investigation conducted for that category. 17

2.2 Categorization of Domain Adaptation methods and the thesis contribution based on the availability of the samples. 21

3.1 Three steps of the proposed denoising method (paper I) 23
3.2 The architecture of the ACSSDA (paper II) 25
3.3 The figure on the left illustrates the class misalignment in the UDA setting. The figure on the right shows how ACSSDA deals with this problem in the SSDA setting. Each color represents an individual class, while the black displays unlabeled samples. The figures in the middle show the results of the alignment. A comparison of alignment results with actual sample labels is provided at the bottom. In contrast to unsupervised settings, semi-supervised settings prevent class misalignment (paper II). 25
3.4 An illustration of how DARC adapts the domains for regression tasks (paper III). 26
3.5 DARC (paper IV) 26
3.6 ITL-IDS (paper IV) 28
3.7 FedGenP (paper VIII) 30
3.8 Training procedure of generator proposed in FedGenP 30
1. Introduction

1.1 Introduction

Machine Learning (ML) algorithms are capable of learning patterns from data but need a crucial consideration: the assurance of generalizability. The generalizability of an ML model refers to its ability to make accurate predictions on new, unseen data beyond its training dataset. When a model generalizes well, it means that it has learned the underlying patterns and relationships in the training data, allowing it to make accurate predictions on test datasets. Generalization is often assessed by evaluating a model’s performance on a separate validation or test dataset that it hasn’t seen during training. However, the assessment of the generalization of an ML model is based on the IID (Independent and Identically Distributed) assumption \[1\]. The IID assumption states that the data used to train and test a model are independent and identically distributed. "Independent" means that the occurrence or value of one data point does not influence the occurrence or value of another data point. "Identically Distributed" means that all data points come from the same underlying probability distribution, which implies that the data’s statistical properties (e.g., mean, variance) are the same across the entire dataset. The IID assumption is necessary for generalization assessment because it ensures that the model learns relevant patterns and relationships from the training data that are likely to hold true in the test data. However, ensuring that the IID assumption holds can be challenging, especially in real-world scenarios.

Domain adaptation (DA) is a subfield of machine learning that deals with situations where the distribution of data in the target domain (a domain from which test data is drawn) is different from the distribution of data in the source domain (a domain where training data is drawn from) \[2\]. This situation can arise in various practical applications, including sentiment analysis \[3\], computer network security \[4; 5\], medical image analysis \[6\], predictive maintenance \[7; 8\] and so on. In this thesis, we particularly focus on predictive maintenance and computer network security applications.

In the context of predictive maintenance, an example of the practical use of DA is the maintenance of electric buses’ batteries. When a fleet obtains new buses to operate in significantly different environmental conditions, gather-
ing substantial training data specific to these new vehicles can be challenging. These new vehicles are considered to be from the target domain, and the goal is to develop a predictive maintenance model for them. However, historical data may exist from buses used in moderate climates, which is considered the source domain. In such situations, DA techniques adapt the historical data from the source domain to the data from new buses in the target domain in order to train an ML model for the target. These adaptation techniques consider the differences in environmental conditions, such as extreme temperatures, ensuring that the predictive maintenance model is effective for the new buses.

Similarly, as an example in the computer network security context, consider the situation where an organization expands operations into new regions, launches new applications, and adds new users. The data generated in these new environments can be different from what the organization has previously encountered. Additionally, gathering new data, especially data related to new cyberattacks, can be challenging and expensive, as simulating such attacks on a network requires substantial resources. In these circumstances, DA can be a practical approach that involves using the data available and adapting it to the new network environment. This adaptation helps in training an effective machine learning model by transferring knowledge about attacks from other environments to new ones [4].

Accordingly, DA is particularly valuable in real-world applications with a dynamic nature and diverse data that is continually changing, where it would be impractical to collect large amounts of domain-specific labeled data for training purposes. It allows existing data to be repurposed for new tasks or environments. In other words, DA can be used to address deviations from the IID assumption and improve the ML model’s generalization ability.

There is a fundamental requirement for DA in that data from various domains must be accessible. As an illustration, consider these domains as separate companies operating similar equipment. However, privacy concerns present a significant challenge, as these corporations may be unwilling to share their owned data. Consequently, leveraging the valuable insights contained within their data using conventional DA techniques can be challenging. In such circumstances, Federated Learning (FL) offers a potential solution.

FL is a machine learning approach designed to train models across multiple decentralized clients while keeping the data secured on the clients. It enables collaborative model training without sharing the actual data, making it particularly useful in scenarios where data privacy and security are essential.

A shared property between FL and DA relates to their focus on data originating from diverse domains, which results in data heterogeneity. DA, by having direct access to the data, aims to remove such heterogeneities and consequently improve the performance of the models. In the federated setting, a
specialized branch known as Non-IID FL or Heterogeneous FL addresses this challenge. Heterogeneous FL aims to moderate these divergences without directly accessing the underlying data, thus enhancing the performance of the final models while protecting privacy.

Thus, we begin this thesis by examining DA in different scenarios and then proceed to examine Heterogeneous FL from a DA perspective.

1.2 Challenges and research questions

Throughout our study of learning tasks across a variety of domains, we consider three key dimensions: data availability, data accessibility, and data heterogeneity. Data availability, whether labeled or unlabeled data, has been a continuous challenge within the machine-learning community. Addressing practical industrial problems in the real world emphasizes the importance of this challenge. Data accessibility, regardless of the volume of data available within each domain, is a critical factor to consider. Due to privacy concerns, many data owners refuse to share their data. Data heterogeneity, whether the data is shared or non-shared, may arise as a result of multiple sources of data. This heterogeneity can have notable consequences, potentially resulting in performance degradation for a model, whether it is learned through a centralized or federated setting. The following four research questions are outlined for this thesis based on the mentioned considerations.

**R1. How can we minimize the data requirement, both labeled and unlabeled data, for the adaptation process?** DA is primarily developed to tackle the challenge of insufficient labeled data in the target domain. Nonetheless, the availability of data, regardless of whether it is labeled or unlabeled, is a fundamental requirement for the adaptation process. Consequently, a significant challenge arises regarding the amount of data available and accessible for adaptation purposes.

**R2. How can we minimize misalignment while not having enough labeled data in a domain?** Given the insufficient target data, the right alignment of the diverse domains is a critical concern in DA. Class misalignment in adapting the wrong classes can lead to adverse consequences, including degradation in model performance rather than the intended enhancement, illustrating the difficulties involved in domain adaptation.

**R3. How do various forms of diversity influence the outcomes of the global model in Federated Learning, and how do these diverse characteristics relate to the principles of Domain Adaptation?** Up until this point, our assumption has been that we have access to data from different domains represented by different clients. However, taking into account that certain clients are unwilling to share their data, we move to a federated setting. In the con-
text of federated learning, we encounter heterogeneity in the data generated by different clients, and this heterogeneity can affect the performance of the final shared model, namely the global model. Consequently, we intend to address the effect of data heterogeneity in FL from a DA perspective.

R4. How can we effectively reduce statistical heterogeneity in a federated setting without direct access to the data? Our last research question concerns minimizing the distribution divergence among clients within a federated setting where we do not have direct access to the data from different clients.

Detailed information on which papers contribute to the answer to each research question can be found in Table 1.1.

1.3 Summary of papers and Authors’ contribution

Contrary to all other texts in this thesis, parts of this section are written in the “first person” format to highlight personal statements about the authorship of the papers.

Figure 1.1 visualizes the contributions of this thesis and the attached papers to the field. We examine these contributions from four perspectives: data availability within the target domain, practical applications, the number of domains, and the task addressed. Additionally, alongside the above factors, we consider data accessibility, defining the problem as either domain adaptation or federated learning. Roman numerals in the figure correspond to the respective papers associated with each of the aspects.

Below is a summary and the respective authors’ contributions for each paper. It is worth noting that the order of the papers is as follows: the first six papers relate to DA, while the last two are related to FL. The papers on DA are arranged based on their assumption of data availability in the target domain.

PAPER I: Noise-robust representation for fault identification with limited data via data augmentation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R3</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R4</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1.1: Details of papers contributing to the Research Questions.
Figure 1.1: Contributions of this thesis and the appended papers.
Taghiyarrenani Z, Berenji A.

Summary. The heterogeneity among the samples can be caused by the presence of noise introduced into the data. When considering industrial equipment, it is probable that varying degrees of noise may be added to the data, further contributing to the complexity of the dataset. In paper I [9], we addressed the presence of data with varying levels of noise, treating them as distinct domains. Then, we introduced a denoising method by leveraging the concept of adapting datasets. Using the fewest samples possible, we designed the denoising task to remove as much noise as possible.

Contribution. For this paper, I conducted background research and came up with the idea of using the concept of DA for denoising purposes. I designed the study and worked with Amirhossein on the experiments. Then, I analyzed the results, produced the tables and figures, and wrote the manuscript. Also, he was involved in writing the sections of the paper that were primarily concerned with the application rather than the method. Indeed, this work was a collaborative effort with Amirhossein, who holds expertise in mechanical engineering and provided an application-oriented perspective to this paper.

PAPER II: Facilitating Semi-Supervised Domain Adaptation through Few-shot and Self-supervised Learning
Taghiyarrenani Z, Nowaczyk S, Pashami S, Bouguelia MR.
(submitted)

Summary. Sometimes, collecting unlabeled data is straightforward, while gathering labeled data presents a challenge. In such scenarios, semi-supervised DA methods make use of the accessible labeled samples from different but related domains, referred to as source domains, to enhance the learning process within the desired domain, referred to as the target domain. In paper II, we introduced a novel semi-supervised DA method that leverages contrastive and adversarial techniques to facilitate adaptation between two distinct domains.

Contribution. For this paper, I conducted background research and came up with the idea presented in the paper. I designed
the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript with supervision from the other authors.

PAPER III: Multi-Domain Adaptation for Regression under Conditional Distribution Shift.

Summary. When differences between domains arise from variations not just in features but also in associated labels, it becomes crucial to utilize separate DA methods for regression and classification tasks. In paper III [10], we have presented a DA method for regression that reduces any arbitrary, either marginal or conditional, shift between domains. Furthermore, this paper extended our methodology to consider multi-domain scenarios where each domain contains a limited number of labeled data samples instead of restricting our focus solely to two source and target domains.

Contribution. For this paper, I conducted background research and came up with the idea presented in the paper. I designed the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript with supervision from the other authors.


Summary. Given the dynamic nature of the real world, it is evident that environments will evolve over time, and a model developed once may experience a decline in performance. This challenge is particularly critical in cases like intrusion detection systems. Therefore, methods designed for such environments should possess the capability to adapt and continue learning as new data arrives, ensuring ongoing effectiveness and relevance. In paper IV [5], we present an incremental DA method specifically for network intrusion detection systems. In addition, leveraging expert knowledge in machine learning enhances model performance and ensures that domain-specific insights contribute to more accurate predictions or decision-making. There-
fore, in this paper, we have also designed a solution to make use of expert knowledge to mitigate the negative transfer when performing DA.

**Contribution.** This paper is a collaborative effort with a colleague, Ehsan Mahdavi, who has expertise in Network Security applications. In this collaboration, I conducted background research on Transfer Learning, particularly on DA. The challenge I solved was developing and applying a DA method in network security, as this application has its own challenges. Specifically, my contribution to this paper involved suggesting an incremental DA method and outlining its application within the proposed framework. Furthermore, as an integral component of the framework, I incorporated expert knowledge to facilitate DA. Accordingly, I did experiments related to DA and also some experiments to compare the results with the state-of-art. The entire work was carried out under the supervision of two other authors. This work is a collaboration between Halmstad University in Sweden and Isfahan University of Technology in Iran.

**PAPER V:** Domain Adaptation with Maximum Margin Criterion with application to network traffic classification.

Taghiyarrenani Z, Farsi H. In International Workshops of Joint European Conference on Machine Learning and Knowledge Discovery in Databases 2022 Sep 19 (pp. 159-169). Cham: Springer Nature Switzerland.

**Summary.** It is possible that in certain scenarios, the difference between source and target can also be due to differences in the available classes. In this thesis, as detailed in paper V [11], a method for addressing situations where certain classes are missing in the source domain is developed.

**Contribution.** This paper is a collaborative effort with a colleague from industry, Hamed Farsi, who has expertise in Network Security applications. For this paper, I conducted background research and came up with the idea presented in the paper. I collaborated with Hamed to ensure that the method is valuable for the field. I designed the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript.

**PAPER VI:** Towards Geometry-Preserving Domain Adaptation for Fault Identi-

**Summary.** There are situations where some classes are absent in the target domain. In such cases, our objective is to enable our model to make predictions for these missing classes using only the information accessible from the source domain. In paper VI [12], we have designed a method that, by integrating DA and a geometry-preserving technique, adapts different domains when data for some classes are missing in the target domain during training, not test. We call this scenario as Limited DA.

**Contribution.** I conducted the background research and came up with the idea for the paper. I designed the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript with supervision from the other authors.


**Summary.** Data from various domains, though available, may remain inaccessible due to data owners’ privacy concerns, highlighting the value of federated learning. A widely recognized approach is FedAvg, which aggregates models from different domains (clients) and computes the average of local models to create a global model. In paper VII [13], we connect DA and FL principles, offering an empirical case study that demonstrates the impact of sample heterogeneity across different clients in FL.

**Contribution.** I conducted the background research and came up with the idea for the paper. I designed the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript with supervision from the other authors.

PAPER VIII: Heterogeneous Federated Learning via Personalized Generative Networks. Taghiyarrenani Z, Alabdallah A, Nowaczyk S, Pashami
Summary. In paper VIII[], we theoretically discuss the impact of sample heterogeneity among different domains (clients) in federated learning on the outcomes of the global model for each client. Accordingly, we propose a method to mitigate this heterogeneity and enhance the results within a federated setting.

Contribution. I conducted the background research and came up with the idea for the paper. I designed the study, conducted the experiments, analyzed the results, produced the tables and figures, and wrote the manuscript with supervision from the other authors.

The following papers also contributed to the domain adaptation topic and are included to indicate potential future directions.


Summary. This is a position paper about the importance of Explainable AI in Industry [14].

Contribution. I wrote the sections concerning the inevitability of data heterogeneity in various industries and the necessity for explainable Artificial Intelligence (XAI) in such contexts, particularly when applying DA. This paper is a collaboration between Halmstad University in Sweden, Jagiellonian University in Poland, and INESC TEC in Portugal.


Summary. This paper presents two complementary explanation mechanisms for aligning domains in DA: 1) explaining how the source and target distributions are aligned in the latent space. 2) descriptive explanations on how the decision boundary changes
in the adapted model with respect to the source model. **Contribution.** I prepared the datasets and also the codes for the DA part. I also contributed to the writing of the paper. However, the section related to explainability was mainly done by other authors. This paper is a collaboration between Halmstad University in Sweden and Jagiellonian University in Poland.

---

Papers that are not included in the thesis.


**Contribution.** This paper, [13], is a collaboration effort with Amirhossein, who has expertise in mechanical engineering.


**Contribution.** This paper [15] is a result of a master thesis, and I was involved as a supervisor.


**Contribution.** In this paper, [16] I was involved as a mentor.


**Contribution.** In this paper, [17] I was involved as a mentor.
2. Background

2.1 Domain Adaptation

Let’s define $\mathcal{X}$ as input space, $D$ as a probability distribution on $\mathcal{X}$, and $\hat{D}$ as an empirical probability distribution. We define the labeling function $f$ as $f : \mathcal{X} \rightarrow \mathcal{Y}$ where $\mathcal{Y}$ is output space.

A hypothesis $h$ is defined as $h : \mathcal{X} \rightarrow \mathcal{Y}$. $\mathcal{H}$ is a hypothesis space for $\mathcal{X}$. The risk of the hypothesis, $\varepsilon(h)$, captures the disagreement between the hypothesis and the labeling function as

$$
\varepsilon(h) = \mathbb{E}_{x \sim D}[|h(x) - f(x)|].
$$

Similarly, the empirical hypothesis risk is defined as

$$
\hat{\varepsilon}(h) = \mathbb{E}_{x \sim \hat{D}}[|h(x) - f(x)|].
$$

**Symmetric difference hypothesis space $\mathcal{H} \Delta \mathcal{H}$ [18]:** Given a hypothesis space $\mathcal{H}$, $\mathcal{H} \Delta \mathcal{H}$ is defined as

$$
\mathcal{H} \Delta \mathcal{H} = \left\{ h(x) \oplus h'(x) : h, h' \in \mathcal{H} \right\},
$$

where $\oplus$ represents the XOR operator. This means a hypothesis belongs to $\mathcal{H} \Delta \mathcal{H}$ if a given pair in $\mathcal{H}$, $h(x)$ and $h'(x)$ disagree.

$A_{\mathcal{H}}$ is a set of measurable subsets for some hypothesis $h \in \mathcal{H}$, so that,

$$
\{ x : x \in \mathcal{X}, h(x) = 1 \} \in A_{\mathcal{H}}, \forall h \in \mathcal{H}.
$$

Similarly, $A_{\mathcal{H} \Delta \mathcal{H}}$ is defined as

$$
\{ x : x \in \mathcal{X}, h(x) \neq h'(x) \} \in A_{\mathcal{H} \Delta \mathcal{H}}, \forall h, h' \in \mathcal{H} \Delta \mathcal{H}.
$$
**H-distance between two distributions**[18]: Given $D$ and $\hat{D}$ as two arbitrary distributions, $\mathcal{H}$-distance is defined as:

$$d_{\mathcal{H}}(D, \hat{D}) := 2\sup_{A \in \mathcal{A}_{\mathcal{H}}}|\Pr_D(A) - \Pr_{\hat{D}}(A)|,$$

similarly, $d_{\mathcal{H}\Delta\mathcal{H}}(D, \hat{D})$ is defined as distribution divergence induced by $\mathcal{H}\Delta\mathcal{H}$ as:

$$d_{\mathcal{H}\Delta\mathcal{H}}(D, \hat{D}) := 2\sup_{A \in \mathcal{A}_{\mathcal{H}\Delta\mathcal{H}}}|\Pr_D(A) - \Pr_{\hat{D}}(A)|.$$

The defined distribution divergence is defined over two arbitrary distributions $D$ and $\hat{D}$. Let’s define a representation function as $R : \mathcal{X} \rightarrow \mathcal{Z}$ and $\tilde{D}$ and $\tilde{\hat{D}}$ the corresponding distributions over $\mathcal{Z}$. Thus, one can calculate the distribution divergence over $\mathcal{Z}$ as:

$$d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}, \tilde{\hat{D}}) := 2\sup_{A \in \mathcal{A}_{\mathcal{H}\Delta\mathcal{H}}}|\Pr_{\tilde{D}}(A) - \Pr_{\tilde{\hat{D}}}(A)|.$$

First, we define a domain, $\tau$ as triple $\tau = \langle D, \mathcal{X}, f \rangle$; a distribution $D$ over input space $\mathcal{X}$ with the corresponding labeling function $f$.

**Generalization Bounds for Domain Adaptation**[19][18]: Lets assume two different source and target domain as $\tau_s = \langle D_s, \mathcal{X}, f \rangle$, $\tau_t = \langle D_t, \mathcal{X}, f \rangle$. Assuming $R : \mathcal{X} \rightarrow \mathcal{Z}$ as a shared representation function for source and target domains, $\tilde{D}_s$ and $\tilde{D}_t$ are corresponding distributions over $\mathcal{Z}$. The generalization bound on the target domain error is defined as follows, so that with the probability at least $1 - \delta$, for every $h \in \mathcal{H}$:

$$\varepsilon_{\tau_t}(h) \leq \hat{\varepsilon}_{\tau_t}(h) + \sqrt{\frac{4}{m} \left(d \log \frac{2em}{d} + \log \frac{4}{\delta}\right)} + d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_s, \tilde{D}_t) + \lambda \quad (2.1)$$

where $d$ is the VC-dimension of a set of hypothesis $\mathcal{H}$. $m$ is the number of source samples, and $\hat{\varepsilon}_{\tau_t}(h)$ is the empirical risk of the hypothesis trained on source data, and $e$ is the base of the natural algorithm. Considering $h^*$ as the optimal hypothesis so that $h^* = \arg\min_{h \in \mathcal{H}}(\varepsilon_{\tau_t}(h) + \varepsilon_{\tau_s}(h))$, then $\lambda$ is the optimal risk on two domains, $\lambda = \varepsilon_{\tau_s}(h^*) + \varepsilon_{\tau_t}(h^*)$.

Conventional machine learning methods consider training, and test samples are drawn from the same domain; the same distribution. Therefore, equation 2.2 will be reduced to the standard Vapnik-Chervonenkis theory, as follows:

14
\[
\epsilon_{\tau}(h) \leq \hat{\epsilon}_{\tau}(h) + \sqrt{\frac{4}{m} \left( d \log \frac{2em}{d} + \log \frac{4}{\delta} \right)}
\]  

(2.2)

Accordingly, having more than one domain, DA aims to minimize the distribution distance between the domains to solve the task \( f \) using the samples from all available domains.

Let’s assume a dataset \( S_{\tau} \) drawn from domain \( \tau \) is \( S_{\tau} = \{(x_{i}^{\tau}, y_{i}^{\tau})\}_{i=1}^{n} \), where \( n \) is the number the samples from domain \( \tau \).

The above definitions allow us to categorize and discuss DA methods from different perspectives. Generally, the term DA is used to describe the adaptation of two domains, which is the most commonly addressed scenario. In this case, we have two domains \( \text{source} \) and \( \text{target} \), respectively; the corresponding datasets are \( S_{s} = \{(x_{i}^{s}, y_{i}^{s})\}_{i=1}^{n} \) and \( S_{t} = \{(x_{i}^{t}, y_{i}^{t})\}_{i=1}^{m} \), where \( n \) and \( m \) are the numbers of source and target samples, respectively. The dataset \( S_{s} \) is assumed to be available; according to availability of \( \{y_{i}^{t}\}_{i=1}^{m} \) three different categories are defined:

- **Unsupervised-DA**: in this category, \( \{y_{i}^{t}\}_{i=1}^{m} \) is not available; In other words, only unlabeled target samples are available for training.

- **Semi-supervised-DA**: In this category, \( \{y_{i}^{t}\}_{i=1}^{p} \) is available where \( p \ll m \); In other words, very few target samples are labeled.

- **Supervised-DA**: In this category, \( \{y_{i}^{t}\}_{i=1}^{m} \) is available; however, it is usually assumed that \( m \ll n \), i.e., the number of target samples is less than the number of source samples.

Considering \( p(x, y) \) as a joint probability distribution over \( X \times Y \), \( p(x, y) = p(x)p(y|x) = p(y)p(x|y) \), where \( p(x) \) is a marginal probability distribution over the input space and \( p(y|x) \) and \( p(x|y) \) are conditional probability distributions. The difference between the joint probability distributions, \( p(x, y) \), of the domains is the central issue DA deals with. Considering this difference, two major categories are discussed in the literature: covariate shift and concept shift. The former refers to a situation where \( p_{s}(x) \neq p_{t}(x) \) but \( p_{s}(y|x) = p_{t}(y|x) \) whereas the latter means \( p_{s}(x) = p_{t}(x) \) but \( p_{s}(y|x) \neq p_{t}(y|x) \) or \( p_{s}(x|y) \neq p_{t}(x|y) \). If the \( p_{s}(x, y) = p_{t}(x, y) \), the problem will be reduced to the conventional machine learning.

Based on the number of domains, three different categories can be defined in cases where there are more than two domains: Multi-Source DA, Multi-Target DA, and Multi-DA. Their names describe the two first as having multiple sources or multiple targets, respectively. However, we define the Multi-DA
category in this way: there are multiple symmetric domains in terms of the availability of the samples. In other words, the available data for each domain in this category is insufficient to solve the task (to train domain-specific models). Thus, we adapt the samples from multiple domains so that each domain compensates for the sample shortage in the other domain.

The literature also introduces homogeneous and heterogeneous DA methods based on the differences between the domains' input feature spaces, $X$. In the case of identical input feature spaces of domains, homogeneous DA is the solution; otherwise, heterogeneous DA is needed. In this thesis, we only address homogeneous DA.

Additionally, the label set of the domains defines different categories. The categories in the standard DA benchmarks are the same in all domains. The authors in [20] call this setting Closed-set DA. They also define Open-set DA, which different domains include different categories. There are, however, different ways in which different categories can occur. Specifically, open-set DA occurs when both the source and target domains have some common categories and, at the same time, both domains have other unknown categories. It is possible that the unknown categories are indeed different from one another. As a result, the final target model should be able to predict the output of a sample as one of the known categories or an unknown category.

Another setting in [21] is defined as Partial DA. In a partial DA setting, the categories in the target domain are a subset of the categories in the source domain. As a result, there are some outlier categories in the source domain, which may have a negative impact on the results of DA. A partial DA model is trained to predict the output of samples as one of the available categories in the target. A new setting is defined in this thesis and is referred to as Limited DA. Similar to partial DA, available categories in the target domain in training time (of the adaptation model) are a subset of those in the source domain. The additional categories in the source domain are not outliers but will be encountered in the target domain during testing. In fact, we lack the samples from some categories in target during training, while we may encounter them during the test. Therefore, the objective is to construct a model that predicts the output of a target sample as a category within the source data.

**Contribution of the thesis in Domain Adaptation field** Figure 2.1 shows the mentioned categories. This thesis has addressed the categories that are indicated by darker boxes.

Concerning the availability of samples in the target domain and the subsequent categorization of different DA categories, Figure 2.2 illustrates the thesis’s contribution to DA by specifying the relevant papers for each category.
Figure 2.1: Categorization of Domain Adaptation methods. The density of the gray boxes indicates the extent of the investigation conducted for that category.

2.2 Federated Learning

Federated Learning (FL) is a machine learning approach that allows model training across multiple decentralized edge devices or data owners, generally called clients[22], [23]. The key idea behind FL is the ability to perform the training phase of machine learning models in a distributed fashion. This approach presents advantages like reducing network resource burden and enhancing the privacy of individual data sources by keeping data localized. As a result, FL is particularly well-suited for use cases involving sensitive data or those where centralized data aggregation is not feasible.

The scale of FL from the number of clients is categorized into cross-silo and cross-device. In the cross-silo scenario, the clients can be distinct institutions or companies. A cross-device setting involves an extremely large number of small clients, such as mobile devices [24].

Two terminologies, "client" and "server," are common in the federated setting. This thesis equates the "client" to the concept of a "domain" in the context of DA. Clients can be connected in a number of different topologies, with two
of the most well-known being "peer-to-peer" and "star-shaped." The peer-to-peer topology is known as fully decentralized. In a star-shaped connection, a central server is the client’s coordinator during model training. However, the server typically does not have access to the raw data.

Other key terminologies in FL include "local model," "global model," and "aggregation mechanism." The definitions of these terms are as follows:

**Local Model:** This refers to a machine learning model that a client individually trains using its own local data. Each client has its own local model that learns from its data.

**Global Model:** The global model is a unified model constructed by aggregating the local models to be generalized well across all clients.

**Aggregation Mechanism:** The aggregation mechanism is the method to construct the global model from the local models provided by clients.

Let’s use the scenario of predicting vehicles’ energy consumption in a federated setting. This example will help us explain the concepts in this section. Let’s consider a star-shaped system for the federated setting, where all clients are connected to a central server. Every federated system may have differences at different levels, where various components are involved.

There may be different types of clients. As an example, we refer to vehicles as clients. However, we include either buses or trucks in the system. In this case, the nature of the features between the clients may be different, which results in a heterogeneous feature set between clients [25].

The available memory or communication resources may differ in different vehicles, resulting in heterogeneous communication and asynchronous clients [26].

Alternatively, a federated system can be specifically designed for buses. Despite sharing a common nature and deploying similar sensors to collect data to have the same feature set, buses may operate in distinct locations with varying external conditions, such as diverse weather situations. This divergence can impact the statistical properties of the data. This scenario needs Non-IID or Statistical Heterogeneous FL [25; 27; 28].

This thesis focuses on the statistical heterogeneity between the clients.

There are different types of statistical heterogeneity, depending on the local labels and samples. As categorized by authors in [25], these differences fall into four categories: quantity skew, quality skew, feature skew, and label skew.

Quantity skew relates to an imbalanced dataset among clients, a challenge addressed by numerous heterogeneous federated learning approaches in the literature including [29-35]. For example, according to the example provided of a federated system of vehicles, certain types of faults are more likely to occur in certain types of vehicles. As a result, the number of samples per fault varies among clients.
Quality skew involves the presence of noise in either labels or samples. Feature skew occurs when the distribution of samples across the feature space differs. There are some vehicles that operate in cold climates and others that operate in warm climates. Such differences affect the statistical properties of the clients’ data. Lastly, label skew denotes differences in available label sets among clients or variations in label preferences. For example, what might be a normal engine temperature in a hot location could be considered an anomaly in a cold location. Label preference skew is notably problematic, introducing confusion to the global model [13].

To provide a formal definition, we continue with symbols as previously defined for DA. Nevertheless, in order to ensure the clarity of this section, we will repeat the initial definitions as follows:

Assume $X$ as input space, $D$ as a probability distribution on $X$ and labeling function $f$ as $f : X \rightarrow Y$ where $Y$ is output space. A hypothesis $h$ is defined as $h : X \rightarrow Y$. The risk of the hypothesis, $\varepsilon(h)$, captures the disagreement between the hypothesis and the labeling function as $\varepsilon(h) = \mathbb{E}_{x \sim D} |l(h(x), f(x))|$, where $l$ is a loss function. We consider every client as a domain, as in DA. Therefore, we define the set of clients as $T = \{\tau_k : k \in K\}$ with $K$ clients where $\tau_k = \langle D_k, X, f_k \rangle$.

Accordingly, all clients are assumed to share the same input space $X$. The clients may, however, have different data distributions $D_k$ over the input space. The labeling function $f_k$ may also be different between the clients. The case in which $f_i = f_j$ when $i \neq j$ means that all local data belong to the same global distribution. The setting with different client data distributions or labeling functions in the literature is called non-IID or statistical heterogeneous federated learning (HFL) [25].

The risk of every local model $h_k$ of client $\tau_k$ is $\varepsilon_{\tau_k}(h_k) = \mathbb{E}_{x \sim D_k} [l(h_k(x), f(x))]$.

The aim of federated learning is to learn a global model $h_g$. The risk of this model is calculated over all clients $T$ as $\varepsilon_T(h_g) = \mathbb{E}_{\tau_k \in T} [\varepsilon_{\tau_k}(h_g)]$.

This formula calculates the expectation risk of a single global model $h_g$ on all clients. The most well-known FL method in the literature is FedAvg [36]. In FedAvg, every client takes steps of gradient descent on its local model using its local data, and the server calculates the average of the local models as the global model. Intuitively, by averaging the local models, the model may be led toward the average optimum of the individual clients. However, in the presence of heterogeneity, this local optimum may not be beneficial for all clients equally, thereby influencing the overall performance of the global model [28]. In essence, greater heterogeneity among the clients’ data poses a
higher risk to the effectiveness of the global model. Non-IID or Heterogeneous FL techniques aim to overcome the challenge of data heterogeneity [25].

Among the approaches for heterogeneity is Personalization techniques [28]. This technique could be beneficial, especially for clients whose data drifts significantly. However, the extreme case of personalization is the local model without any knowledge transfer between the clients. Personalized FL aims to solve the challenge of client heterogeneity by establishing a tradeoff between personalization and generalization [28].

Contribution of the thesis in Federated Learning  This thesis initiates an empirical examination of the impact of data heterogeneity by simulating various types of heterogeneity using a dataset derived from vibration data of an inductive motor [13]. Building on the empirical findings, we subsequently present a theoretical analysis demonstrating that heterogeneity between each client and the rest influences the performance of the global model specific to that client.

Drawing on this theoretical foundation, we propose an approach inspired by the context of DA. Specifically, we proposed the construction of a generator for each client. Through adversarial training, we ensure that the generator produces data within the region of space where discrepancies among clients exist, enabling each client to augment its own dataset and thereby reduce heterogeneity with the rest of the clients. Consequently, our method aligns with the data augmentation category, aiming to mitigate heterogeneity in FL.
Figure 2.2: Categorization of Domain Adaptation methods and the thesis contribution based on the availability of the samples.
3. Summary of the methods proposed in the papers

Data may originate from multiple domains, resulting in discrepancies and thus adversely affecting the performance of conventional machine learning models.

**Paper I** presents a new denoising approach utilizing the concept of DA [9]. To explain, let’s take a dataset originating from a distribution. We assume that environmental noise alters the sample distribution, leading to a decline in a trained model’s performance in noisy environments. Therefore, training a model capable of adapting noisy and original data allows it to maintain performance in a noisy environment. However, for an effective model, it needs to endure not just one level of noise but multiple levels. Figure 3.1 illustrates the steps of a suggested denoising approach designed to eliminate various noise levels, even in the absence of sufficient labeled samples. We augment the samples with several noise levels to accomplish this objective, inspired by unsupervised contrastive learning techniques. By pairing original and noisy samples, we create paired sample sets. Subsequently, a Siamese neural network is trained using the contrastive loss function based on these paired samples. Training the network results in a feature extractor that maps the samples to a new space. In this space, the corrupted samples are aggregated with the original samples, resulting in denoising. Furthermore, because the new space is created using contrastive learning, it is class-distinguished and can be achieved through the use of a small number of labeled samples. From the DA perspective, we synthetically construct different domains for this paper by augmenting the data. Therefore, we have symmetric domains without any constraint on the number of samples per domain.

![Figure 3.1: Three steps of the proposed denoising method (paper I)](image)
In the remaining papers, however, we explicitly know the domains and propose DA methods in semi-supervised, supervised, and unsupervised DA settings in the papers II, III and IV, respectively. In addition, we propose solutions for Partial DA (PDA) and Limited DA (LDA) in papers V and VI, respectively.

**Paper II** addresses the adaptation problem for classification tasks in a two-domain setting where the source domain is fully labeled. We also have access to unlabeled target samples and a limited number of labeled target samples. In this paper, we propose a new representation-learning-based semi-supervised DA method called Adversarial Contrastive Semi-Supervised Domain Adaptation (ACSSDA). Through ACSSDA, a shared feature representation for both source and target domains is learned using very few target sample labels. ACSSDA merges two objectives, one for adaptation and the other one for distinguishing different classes. As shown in figure 3.2, ACSSDA trains a domain classifier $F$ with an adversarial training procedure to ensure that the resulting feature space is domain agnostic. Simultaneously, the contrastive loss inspired by self-supervised learning techniques discriminates different classes. In addition, figure 3.3 illustrates how ACSSDA prevents class misalignment and negative transfer; every available labeled target sample is paired with all source samples. The other target samples will gradually pair with the source samples as the training proceeds.

To conclude, ACSSDA is able to adapt domains and remove the probability shift between the domains for the classification task.

Numerous DA methods are available in the literature to address the covariate shift, cf. [37], [38], [39], [40] and etc. The objective of these methods, regardless of whether they are used for classification or regression, is to unite the marginal distributions of the source and target, denoted as $P_s(x)$ and $P_t(x)$, respectively. In other words, in this case, since neither the labels nor the outcome variables are considered for adaptation, the same adaptation methods can be applied for solving regression and classification problems.

On the other hand, for concept shift, DA should unify conditional distributions of source and target, i.e., $P_s(y|x)$ and $P_t(y|x)$. In the literature, several DA methods have been proposed to address the concept shift issue for the classification task [41], [42], [43]. A key principle of these methods is that samples within each class are analyzed separately, and marginal distributions within each class are minimized. Due to their assumption of the discrete nature of the labels (classification tasks), such methods cannot be directly applied to regression. Solving concept shift in regression tasks, therefore, is more challenging.

**Paper III** proposes a multi-DA regression method that aims to minimize arbitrary shifts between domains, whether marginal or conditional. We call this method Multi-Domain Adaptation for Regression under Conditional shift,
Figure 3.2: The architecture of the ACSSDA (paper II)

Figure 3.3: The figure on the left illustrates the class misalignment in the UDA setting. The figure on the right shows how ACSSDA deals with this problem in the SSDA setting. Each color represents an individual class, while the black displays unlabeled samples. The figures in the middle show the results of the alignment. A comparison of alignment results with actual sample labels is provided at the bottom. In contrast to unsupervised settings, semi-supervised settings prevent class misalignment (paper II).
**Figure 3.4:** An illustration of how DARC adapts the domains for regression tasks (paper III).

\[ l_{PSP} = |y_s^{ij} - d(x^l, x^l')| \]

\[ \frac{\partial l_{PSP}}{\partial \theta_f} \]

\[ y_s^{ij} = |y^l - y^{l'}| \]

**Figure 3.5:** DARC (paper IV)

**DARC.** As part of the DARC, we introduce a new loss function, called **Pairwise Similarity Preserver (PSP).** As shown in the figures [3.4 and 3.5] using the PSP loss function, DARC is capable of mapping the difference in labels between samples to the Euclidean distance between the samples in a new space; this results in aligning the domains while preserving the task-specific information within each domain. We evaluate DARC on a real-world dataset related to e-mobility to predict the state of health of heavy-duty vehicles’ batteries.

The remaining papers, again, are proposed for the classification problem. However, each considers a different situation.

**Paper IV** focuses on a specific application, network intrusion detection [5]. Although DA may be considered a Transfer Learning (TL) subcategory, we use TL and DA interchangeably in this study. The limitation of labeled samples is one of the challenges in this area. Meanwhile, new technologies and applications may introduce new vulnerabilities to computer networks.
To address this challenge, incremental learning is a practical approach. In other words, this paper addresses two issues: transferring knowledge between two instances of IDS and addressing the concept shift issue in each instance. Paper IV presents a new framework for intrusion detection systems called Incremental Transfer Learning for Intrusion Detection Systems (ITL-IDS) capable of learning in a network without any prior knowledge. An incremental clustering algorithm is used to detect cluster numbers and shapes without assuming anything about the attacks in advance. The outcomes of the clustering part transfer knowledge between other instances of ITL-IDS. With each iteration, transfer learning provides incremental knowledge to target environments. The proposed framework is designed so that we make use of the knowledge of domain experts to perform adaptation with the goal of minimizing class miss-alignment. Figure 3.6 shows the proposed framework.

**Paper V** considers a situation where some classes are missing in the source domain. For performing the adaptation, we utilize the Maximum Mean Discrepancy (MMD), which calculates the distance between two distributions. However, we aim to construct a transformation matrix that simultaneously adapts the shared classes and keeps the information about non-shared classes. To this end, we adopt the Maximum Margin Criterion (MMC), which maximizes the between-class and minimizes the within-class scatter. We apply MMC to both shared and non-shared classes. We achieve the desired transformation matrix by minimizing the summation of MMD and MMC.

**Paper VI**, as a work in progress work, designs a method for the Limited Domain Adaptation (LDA) setting. Some classes are missing in the target domain during adaptation in this setting. Therefore, we should only adapt the shared classes. Regarding the missing classes in the target domain, since there is no knowledge about them in the target, the only information that we can use is what we have from the source domain. Therefore, it is crucial to maintain their information and prevent them from distortion while adapting the source and target domains. By this intuition, paper VI proposes to keep the geometry of data while adapting the domains. To this end, this paper constructs a new shared feature representation for both source and target using a Siamese-shaped neural network. While it adversarially adapts the shared classes to each other, it retains the distance between the samples. In other words, the distance between samples in the original and constructed spaces will be kept equal. As a work in progress, this paper solves this problem for a synthetic 2-dimensional dataset and constructs a 2-dimensional new space. The distance measure used is Euclidean distance. However, other distance metrics, of course, can be utilized, for example, Geodesic distance.

Thus far, the emphasis has been on DA methods. However, in Papers VII and VIII, we consider the heterogeneity among samples in situations where di-
Source dataset

Target dataset

Source percentile subset

Target percentile subset

Incremental Clustering

Changes on clusters?

Yes

Active Learning

Incremental structure summary subset of source

Incremental structure summary subset of target

Structural matrix of source on incremental structure summary subset of source

Similarity matrix on incremental structure summary subsets

Dissimilarity matrix on incremental structure summary subsets

Structural matrix of target on incremental structure summary subset of target

Manifold alignment on incremental structure summary subsets

Source mapping function

Target mapping function

Incremental source samples

Incremental target samples

Machine Learning Algorithms

Labeling and Evaluation

Figure 3.6: ITL-IDS (paper IV)
rect access to the data is unavailable, i.e., in a Federated setting. Due to the significance of security and privacy in real-world applications, Federated Learning (FL) is experiencing increased adoption across various domains, including predictive maintenance. FL enables independent companies to build models while preserving data privacy collaboratively. However, as different companies operate in diverse environments, their working conditions may vary, leading to heterogeneity in their data distributions.

In paper VII, we focus on addressing the fault identification problem and simulate various scenarios of data heterogeneity. The presence of heterogeneity between the samples from different domains (clients) poses challenges for conventional FL algorithms, highlighting the crucial considerations that need to be taken into account when designing federated predictive maintenance solutions. Through experimental analysis, we demonstrate such challenges.

Finally, in paper VIII, as shown in figure 3.7, we theoretically explore the impact of client heterogeneity on each individual client. We propose a FL method, called FedGenP, that involves training client-specific generators on the server. These generators generate samples that help reduce the heterogeneity between each client and the rest of the clients. This generator indeed is adversarially trained using a Gradient Reversal Layer (GRL) as depicted in figure 3.8 for a system with two clients.
Figure 3.7: FedGenP (paper VIII)

Figure 3.8: Training procedure of generator proposed in FedGenP
4. Conclusion and future work directions

In this thesis, we have focused on the task of developing machine learning models in the presence of data originating from multiple domains rather than a single domain. Our study contains three key aspects: data availability, data accessibility, and data heterogeneity. Specifically, under the umbrella of data accessibility, we examined the spectrum from Domain Adaptation (DA), which involves full accessibility to data from different domains, to Federated Learning (FL), where data from domains is not accessible, shaping the title of the thesis: “From Domain Adaptation to Federated Learning.”

Our investigation began with DA and then moved to FL. The problem of FL is quite connected to the problem of multi-domain adaptation. In FL, not only the samples of different domains (clients) may be heterogeneous, but also they are not directly accessible.

Regarding DA, we examined the case from a variety of perspectives, including the volume of available labeled and unlabeled data, classification and regression tasks, and the nature of heterogeneity. We conducted our investigation in the context of two distinct applications: Predictive Maintenance and Network Intrusion Detection. The outcomes of our research in the domain adaptation field are described in papers I to VI. In all cases, regardless of the DA scenario at hand, it is crucial to prioritize the minimization of misalignment during the adaptation process.

In our study of federated learning, we have established both empirical and theoretical connections with domain adaptation. We demonstrated how heterogeneity among samples originating from different clients can impact the performance of a model trained within a federated setting. The outcomes of our research in the federated learning field are described in papers VII and VIII.

In the context of domain adaptation, one of the key requirements for developing or employing a domain adaptation method is understanding the distinctions between various domains. These distinctions may be due to statistical differences, especially when the feature sets are identical across domains. Differences may also refer to variations in feature sets, label sets, or even disparities in the level of difficulty when collecting data in different domains.
resulting in domains with distinct data quantities or imbalanced classes. Taking into account every type of difference, domain adaptation will intersect with other topics in the literature.

Concerning statistical differences, the key question is how the domains differ from one another. Answers to queries, such as the root causes of these disparities and the extent to which domains diverge, are helpful. By ensuring proper alignment between domains, this knowledge facilitates domain adaptation while maintaining model generalizability and preventing negative transfer.

In addition to implementing DA techniques, explaining the adaptation process is a guarantee for achieving the correct alignment of domains. This level of transparency holds particular significance when addressing real-world industry problems, as mentioned in paper IX. We demonstrated the validity of this research direction in the paper X. In relation to Feature Alignment and Decision Boundary Updates, we introduced two distinct explanation mechanisms for domain adaptation.

Regarding the feature alignment, the explanation provided serves as an explanation of how the distributions are aligned by the adaptation mechanism. Concerning Decision Boundary Updates, the explanations clarify the directions in which the decision boundaries of the source model (a model trained with source data, without the Domain Adaptation technique) are altered during the adaptation process, leading to the formation of the decision boundaries of the adaptation model (a model trained using domain adaptation technique using both source and target data). These two types of explanation can create opportunities to establish a more semantic connection between the two domains. This, in turn, opens up possibilities for transferring background knowledge from the source to the target domain.

On a per-paper basis, we address the extension of our research as follows.

An area for further exploration relates to the presented method for denoising in paper I. This method is only constructed by using labeled samples. However, we would like to leverage a large amount of unlabeled data that is usually available for many tasks. Unsupervised contrastive learning can be considered among the possible solutions, specifically those focusing on preventing collapses.

In paper II, we provided a framework in which self-supervised learning through data augmentation is possible. We aim to explore different data augmentation for the purpose of domain adaptation.

In a subsequent study, there is a potential to extend paper III, DARC, that solves the problem of DA for regression tasks. The input feature space of the source and target is assumed to be identical in this paper. Nevertheless, extending this method to work with source and target domains with different input feature spaces is interesting. This idea is inspired by a real-world exam-
ple: predicting the state of health of vehicles’ batteries when we want to use data sampled at high-frequency and low-frequency rates. We should transfer knowledge from high-frequency to low-frequency data to take advantage of both signals.

In order to extend our proposed method for limited domain adaptation, in paper V, we aim to apply the method to real-world datasets and evaluate it to find the robustness of this method and to analyze its robustness toward discrepancy or missing classes. In addition, in the limited domain adaptation setting, the challenge is to solve a problem for which the available target samples are only healthy, i.e., from none of the faulty classes we have the data. This case needs to deal with transferring knowledge about faulty classes from the source domain to the target that does not necessarily come from the labeled data. A possible solution to solve this problem is to consider physics-based models. In other words, while transferring knowledge from source to target for missing classes, the physics knowledge compensates for the missing data in the target domain.

The method proposed in paper VI solves the problem of network traffic classification. This paper finds a linear transformation to adapt source and target. This method can be extended to a kernelized one to support non-linear transformation. In addition, more experiments are needed to identify how this method is robust to the imbalanced classes and the difference between source and target, which are of interest.

In our studies on FL in paper VII and VIII, we have demonstrated the impact of heterogeneity on the performance of a global model across different clients. Considering the significant increase in interest in FL from both the research and industry sectors, there are many opportunities to explore. We particularly want to extend our proposed method to work under a non-fully supervised setting. Due to the challenges associated with obtaining labeled data, especially when addressing real-world problems, this area attracts our attention.
References


From Domain Adaptation to Federated Learning

Zahra Taghiyarrenani