# Applying Multimodal Data to Meta Learning for Time-Series Analysis in CPS

Philipp Ruf, Christoph Reich
Institute for Data Science, Cloud Computing and Security *(IDACUS)*
*Hochschule Furtwangen University (HFU)*
Furtwangen, Germany
email:{Philipp.Ruf, Christoph.Reich}@hs-furtwangen.de

Djaffar Ould-Abdeslam
*IRIMAS*
*Université de Haute-Alsace (UHA)*
Mulhouse, France
email: djaffar.ould-abdeslam@uha.fr

***Abstract*—Up until now, it has been shown that big parts of the so called Industry 4.0 are impacted by Machine Learning (ML) in some way or another. In many shopfloor situations, there are different sensors involved and all data is eventually structured, accumulated and prepared for application in various ML-based scenarios, e.g., predictive maintenance of a machine, quality monitoring of manufactured workpieces or handling domain-specific aspect of the respective fabricator or product. As the physical environment of Cyber Physical System (CPS) can change rapidly, the overall Data Acquisition (DAQ) process and ML training is impacted, too. This work focuses on datasets which consist of small amounts of tabular information and how to utilize them in image-based Neural Networks (NN) with respect to meta learning and multimodal transformations. Therefore, the conceptual utilization of an DAQ system in industrial environments is discussed regarding a variety of techniques for preprocessing and generating visual material from multimodal data. The outcome of such operations is a new dataset which is then applied in model training. Therefore, the presented approach is three-fold. In first analysing the concept of predicting the similarity of structured and numerical data in different datasets, indicators of the feasibility when applying the methodology in related but more sophisticated learning scenarios can be gained. Although ongoing time-series data is differing from simple multi-class data in terms of a chronologically dimension, basic classification concepts are applied to it and evaluated. In order to extend the similarity prediction with a temporal component, the discussed methods are extended by multimodal transformations and an subsequent utilization in Siamese Neural Networks (SNN). By discussing the concept of applying visual representations of structured time-series data in a meta-learning context, known trends and historic information can be utilized for generating real-world test-samples and predicting their validity on inference.**

***Keywords*—*Data Acquisition; Time-Series Analysis; Multimodal Data; Meta Learning; Cyber Physical Systems.***

## I. Introduction

In recent years, the application of Machine Learning (ML) techniques has increased in many parts of the manufacturing domain. Since industrial-grade Internet of Things (IoT) setups, which are also known as Cyber-physical System (CPS), are fusing more and more with ML-based solutions, the term Artificial Intelligence of Things (AIoT) [1] has also been introduced as descriptive expression. The complexity of such a system is increasing with the utilization of additional sensors which are distributed in the environment or placed inside the manufacturing machines. Therefore, concepts regarding a fully integrated and distributed ML-based Continuous Integration/ Continuous Delivery (CI/CD) pipelines [2] can enhance the overall project structure and management. There are many different quality properties when it comes to CPS-based data [3], [4], especially when it is applied in an subsequent ML model training or inference phase.

In reality, there are situations in which only a few data points are available for the utilization in the training a model. Current proposals which tackle such restrictions are summarized as Few-Shot Learning (FSL) [5], where one differs between transfer learning and a variety of meta-learning techniques, e.g., metric-, optimization- and model-based approaches. Metric learning is commonly assessing the similarity or dissimilarity of two samples, based on a calculation which corresponds to their respective distances [6]. Thereby, the distances between mismatches is maximized while the length of an edge to a positive element, e.g., a matching sample, is minimized, enabling analysis through clusters. One metric-based FSL technique makes use of so called Siamese Neural Network (SNN)s, which utilizes a pair of identical neural networks which are sharing the same weights for processing the respective element of a sample pair, eventually determine their distances. This enables real-world applications, as for example calculating the similarity of hand-written signatures or the structure of human faces [7], where only a few data points of each class are used during model training. Although there are some established strategies for producing effective Convolutional Neural Network (CNN)-based ML models from only a few structured data points, the majority of published work targets image-based implementations. When there are multiple modalities of a specific happening, the transformation and fusion of multimodal data [8], e.g., generating another representation of a modality or combining them, is often part of the solution. Such approaches are usually of generative nature, based on a pre-defined grammar or is utilizing dictionaries for translating between unimodal signal structures. When transforming and fusing structured modalities within a unstructured modality as in a visual representation, established image-based frameworks can be utilized anyhow. Another aspect of this work is the transformation of different datasets samples into visual data, e.g., creating image representations of the *Iris* [9] dataset, the *"Mill Data Set"* [10] as well as of the *Sunspots* [11] dataset.

In addition to applying such new and synthetically generated datasets for training a model in SNN manner, additional experiments are considering ML approaches regarding the original, numeric data. In visualizing and comparing results of the respective approaches across the three varying datasets, selected aspects of utilizing meta-learning approaches for classifying multimodal time-series data are discussed.

A brief overview of requirements in the AIoT domain, as well as procedures regarding multimodal data transformation and applicable meta-learning methods is given in Section II. By discussing the overall methodology and utilized datasets in Section III, a deeper contextual understanding of the experiments, which are carried out and discussed in Section IV, can be gained. The work on hand is concluded in Section V.

## II. Related Work

In the last years, the potential of end-user IoT hardware has evolved significantly, even allowing for small-scale CPS setups. In [12], the ML-focused Data Acquisition (DAQ) system dAta collectoR sysTem witH distribUted sensoRs (ARTHUR) was proposed, suggesting Raspberry-Pi hardware as worker nodes in combination with a distributed edge-cloud environment. In [13], a comprehensive overview of the AIoT domain is given, clarifying enabling technologies and architectural elements of distributed and ML-dependent operations. By considering Artificial Intelligence (AI) tools for utilization in the IoT domain, a overview of potential use-cases and open challenges is discussed. Although the ARTHUR system is applicable in the work on hand, no holistic view of comprehensive hardware considerations or use-case specifics is given.

In [8], the different approaches in the domain of multimodal ML are extensively discussed. When having multiple modalities of a specific event or happening, as for example acoustic emission, vibration data, temperature and a visual observations, there are situations in which a *multimodal fusion* can be appropriate. In using such an approach, there is an increased robustness of predictions due to handling missing values by design, as well as exposing complementary information which may be missed when processing unimodal samples by themselves. Although multimodal transformations are used on various datasets throughout the work on hand, the focus is on applying the resulting representation with respect to a time-series classification. Obviously, the feature-dimension, e.g., modalities, of available samples is also impacting the choice of an target representation. In [14], a simple two-dimensional plot of numeric values was applied to a CNN for further classification. Naturally, this transformation must be well-defined, which is why multiple preprocessing steps like cropping, rotating or framing had to be carried out. The approach of generating visual material differs from the proposal on hand, although a well-defined target transformation is necessary for stable predictions. The augmentation of image data was discussed in [15], where a variety of visual manipulation techniques were described with respect to their utilization in deep learning approaches. Although approaches like kernel filters, random erasing or color space transmissions could be part of the multimodal transformation or simply be utilized for increasing the available test- and training data, no additional augmentation was implemented in the work on hand. In [16], a solution to a temporal Common Representation Learning (CRL) problem regarding image and time-series data was introduced. The main idea is that the additional result of an image classification is enhancing the time-series classification task based on a triplet loss calculation, while the actual inference of the model is exclusively concerned with time-series data. In both synthetic, e.g., time-series sinus value with noise and Gramian Angular Summation Field (GASF) image representation, and real-world handwriting recognition datasets, the cross-modal triplet selection enhanced the inference even though only the main-modality was present. In [17], numeric data was used for generating image filters which were applied to a base-image in order to classify tabular data. Therein, a specifically formed matrix was converted into a convolutional kernel which was applied in the CNN, significantly altering the base-image in a recognizable and class-dependent manner.

In general, deep metric learning consists of informative input samples, structure of the network model, as well as a metric loss function [6]. In the context of chosing a metric loss function, there are various approaches for finding relations between samples [18] and relations between the respective sample's features. Usually approaches are based on distance metrics [6], which are implemented in order to assess sample distances in a triplet, e.g., an anchor element and a positive, as well as a negative element. The aim is to learn a metric which represents negative element further away from the anchor, than the positive element. In his studies regarding the likeness of different human races, Mahalanobis proposed a measure procedure in 1930, where the *Mahalanobis Distance* was applied for the purpose of craniometry, e.g., determining the proportions of the human skull. Since then, his method was increasingly applied and extended in a variety of domains [19], ranging from archaeology to medical diagnosis and remote sensing while solving a multitude of problems like classification, numerical taxonomy or statistical pattern recognition. In [18], the Sparse Compositional Metric Learning (SCML) approach was introduced, where the focus was on learning the Mahalanobis distances which are parameterized by a positive semidefinite square matrix. The metrics are learned as a sparse combination of rank-one basis elements, enabling local, global and multitask metric learning.

In [20], a approach for learning to determine the identity of a masked person was proposed, where a triplet loss function is applied for learning meta-features by considering positive and negative examples within a SNN while using only a few image samples during the training. In [21], a unsupervised meta learning method was proposed in the context of multivariate time-series, e.g., data that contains multiple time-dependent features. A time- and memory efficient system was proposed for learning the universal embeddings in time-series datasets, using an encoder which employs dilated causal convolutions, as well as triplets where elements can be of varying length. Although there is the commonality of triplet creation in time-series data, the approach on hand focuses on the classification of potential future samples. In [22], a framework for image-based feature extraction regarding visual time-series data was proposed with respect to plant phenotyping, e.g., observations of biologic material over time. Therein, a novel transfer learning approach of finding feature representations from image time-series data was proposed. In applying an pretrained *ImageNet* architecture as basic feature extractor, triplet based learning is applied for reducing dimensionality. A siamese-like architecture was proposed, where a series of five temporal subsequent images are processed by five SNN, each consisting of a feature extractor and a ranking module. In [23], two ML approaches, e.g., using a CNN, as well as an SNN, were presented for analyzing supernova phenomenon time-series data with respect to the classification of spectral light-curves, e.g., deciding if the combination of green, red near-infrared and infrared is a type 'I.a supernovae' or not. The CNN approach is initially processing a matrix with four rows, each of which is representing a ongoing color time-series, transforming it into a one-dimensional representation and subsequently analyzing and classifying the sample combination. In the SNN approach, an additional anchor element, which is a subset of the actual anchor's time-series data, is applied in order to tackle the sparsity of available data. Although the work on hand is utilizing SNN technology for classification, the primal focus is on using multimodal transformations in a temporal context.

## III. Environment and Methodology of the Experiments

The generic and distributed DAQ system ARTHUR [12] can be applied for rapid prototyping of productive CPS environments with respect to streaming data, it's analysis and utilizing AI operations, while relying on low-cost end-user IoT hardware and open source technologies. The DAQ showcases are demonstrated using a *Redis* streaming system for transmitting data originating from different sensors, which are mounted on Raspberry Pi embedded devices, towards a cloud infrastrucure. Every shopfloor, e.g., a collection of spatial close worker nodes, is managed by a coordinating node and occurring data can be preprocessed, e.g., cleansed, taken into consideration for aggregation or even be utilized in other quality assuring procedures like applying ML models for predicting live insights into manufacturing processes. In Figure 1, the context of applying deep integrated systems like ARTHUR is depicted in a high-level manner with respect to the overall methodology of the approach on hand. Every worker node is equipped with a so called Digital Twin (DT), which is the intermediate between the phsyical and digital world, e.g., a set of logic for actuating the physical as well as the digital environment according to sensed information. By implementing the multimodal transformation of information which was gathered regarding one or multiple shopfloors, devices or sensors, in a cloud environment, virtual resources can be utilized. With subsequently processing the generated dataset in multiple meta learning approaches, a way of training models for a multitude of situations is given. In deploying the resulting model at a respective worker or coordination node, added value can be created for multiple use-cases, as for example predictive maintenance, the quality of a respective work piece, an assessment of a manufacturing machines tools, and many others.

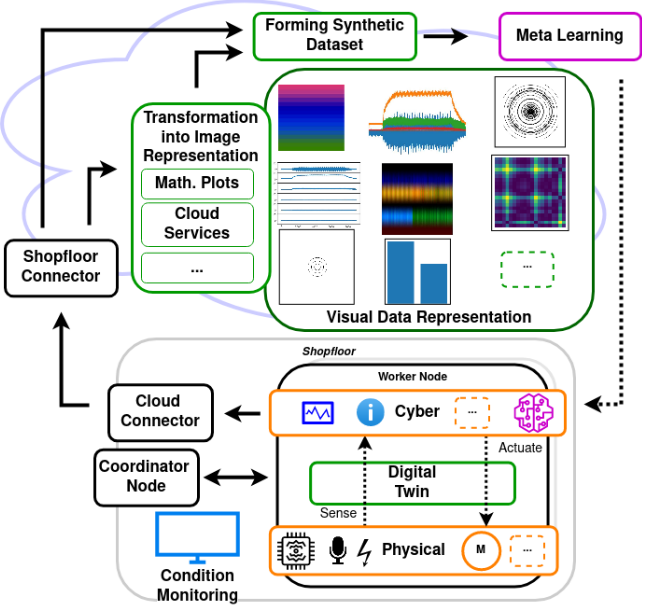

Figure 1. A High-level depiction of the overall pipeline of transforming structured shopfloor data into visual material and its utilization within cloud-based machine learning environments.

### A. Utilized Datasets

In the following, three relative simple but fundamentally differing and well-known datasets are described with respect to their utilization for meta-learning in Section IV. While choosing them, a restriction was that each of them should differ in the primarily domain of application or utilization, e.g., data for classification of samples, for finding trends and for classification in the context of time-series data.

*1) Iris:* The *Iris* [9] dataset consists of 150 samples of three different flower species, differing in petal and sepal length and width. Specifics of how the respective species are distinguishing themselves from each others can be obtained from Figure 2-a, which contains all available values. In this balanced dataset, no kind of preprocessing other than translating the respective species labels into a numerical representation has been done for the experiments. Although there are many possibilities of applying the Iris dataset, there is no relation to an additional dimension for expressing variations in time.

*2) NASA Milling:* The well known *NASA Ames & UC Berkeley "Mill Data Set"* [10] contains multimodal data of a milling machine's runs under various operating conditions and is content of several works. In [24], a summary of best

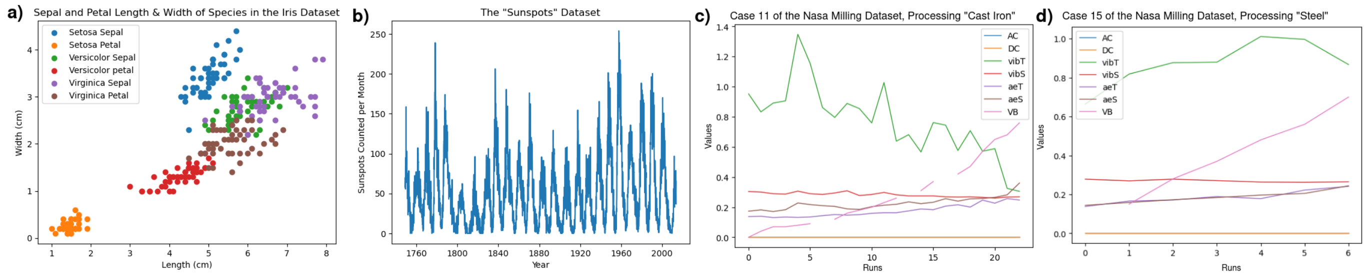

Figure 2. Depiction of datasets which were utilized in the experiments: a) Values in the Iris dataset; b) Plot of the Sunspots dataset; c) & d) Measured tool-wear values (VB) across two cases in the NASA Milling dataset, expressing the mean values of processing two different materials throughout multiple runs.

practices for applying ML in Computer Numeric Control (CNC) machining was given, while considering the dataset from [10]. Therein, each of the recorded runs was visually inspected with the aim of selecting an approximate region of *stable cutting* and additionally extract *sub-cuts* using a sliding-window of 1024 data points, which were labeled according to [25]. In [26], the concept of autoencoders is demonstrated for predicting the tool wear over time using self-supervised ML techniques and anomaly detection with respect to the Milling [10] dataset. The dataset contains data of 167 distinct recordings of occurred vibration, acoustic emission and consumed current of an spindle's individual cuts on different working material types. There are 16 cases, varying in amount of cuts from six to 23, where different parameters are applied, e.g., material type, feed rate of the cutting tool and the depth of an cut. For each of the 167 cuts, 9000 sampling points were collected at 250Hz and persisted within an structured MATLAB array. After each of the recordings, a manual assessment, e.g., an numeric representation *VB*, was carried out with respect to the tool's flank wear, manually measuring the tool's unwanted contact with already finished parts of an workpiece using an microscope. According to [25], the flank wear status can be interpreted as healthy when $VB < .2mm$, worn or degraded when $.2mm < VB < .7mm$ and failed when $VB > .7mm$ is exceeded. Given that definition of an appropriate working condition, the obvious approach is a three-class classification problem. When applying this interpretation to the whole dataset, a representation of the VB's distribution can be further taken into consideration regarding the most appropriate ML strategy. On one hand, a ML model which utilizes a binary classification with respect to a VB threshold of $.2mm$ is presumably directly applicable. On the other hand, a more granular view of the tool wear could be applied in business models where non-premium customers are satisfied with products which were manufactured with a certain degree of tolerance.

In Figure 2-c, the median values across the 11th case in the dataset are plotted. For the sake of visibility, the power consumption, e.g., information regarding AC and DC during the runs, was set to zero as it would conceal graphs of the vibration and acoustic emission. Therein, it is clearly visible that the VB value is increasing over the 23 runs of case 11 when processing cast iron. In comparison with the processing of a steel-based workpiece, a different course of sensed information is recognizable, as shown in Figure 2-d.

There was a certain kind of preprocessing necessary for further utilization in the experiments. First, the milling data [10] was extracted from the matlab structure and visually inspected. Although the majority of runs are free from sensing errors, some obviously inaccurate recordings can be determined by considering the plotted information. Those specific runs had been manually excluded from the experiment. By associating all available runs to the respective measured tool wear status, e.g., healthy, degraded or failed, three classes can be distinguished. Since there are samples for which no such value has been measured, a median value is calculated between the surrounding runs where the flank wear was determined.

*3) Sunspots:* The *Sunspots* [11] dataset consists of monthly observations regarding the number of counted sunspots, e.g., activities at the surface of the sun. Although there is only one target value, e.g., the number counted within a respective month as depicted in Figure 2-b, there are many observations ranging from January 1749 until September 2013, resulting in 3177 ongoing time-depended data points. Although there are well-proven and established preprocessing techniques for time-series data like normalizing values, solely the original data was considered in the experiments.

### B. Visual Representation of Structured Data

Although the data available in CPS environments is usually structured, e.g., numerical values, a transformation of these modalities into visual material is almost always possible. When effectively applying a CNN as feature extractor regarding visual material, the focus is primarily on textures, e.g., a distinction between intact grass and burned grass will be more successful than learning to predict the number of grass stalks within a picture. In Figure 1, examples of data transformation are contained and some of them are exemplary discussed in the following, although there are virtually no boundaries to creativity.

1) *Feature-wise Color Pixels*: Each numeric feature of a sample can be represented by a RGB color representation, e.g, three features may be normalized to values from zero to 255. On the other hand, each feature may

be represented by a gray-scale pixel. When consistently concatenate such pixel representations, time-series can be expressed, which is also true for all following transformations.

2) *Geometric Shapes*: There is a multitude of two-dimensional shapes which can be applied as feature representation, e.g., cubes, circles or triangles, where a second dimension may be expressed by the color, stroke-width or filling of the shape with a color map. Dependent on amount of features per sample pentagons, heptagons and higher-dimensional shapes can be build or multiple basic shapes may be projected on top of each other.
3) *Visual Data Analysis Approaches*: Dependent on the dataset, different approaches like pie charts, bars, lines or cycle plots can be applied as representation of multi-variate data. When for example generating a polygon radar plot, the feature's value is corresponding to a vertex within the plot, preserving their relative magnitude.
4) *Time-Series Plots*: When there is a temporal component to a dataset, the values can simply be represented by a line- or scatter plot, where multiple modalities are specifically colored or styled. Dependent on the respective problem, a grid, axis, labels and legends can be either an obstacle or support when analyzed by the CNN.
5) *Gramian Angular Fields*: This method effectively interprets a time-series as an polar coordinate system, which is then transformed into an Gramian Angular Field representation. As there are many textures within the resulting image, it is assumed to be an appropriate transformation for utilization within an CNN.

### C. Utilizing Triplets with Time Series Data

In the following, a brief overview of the applied distance learning and sampling strategies is given.

*a) Learning with SCML:* As the SCML methodology is by now a established approach, it will be utilized in experiments where only numerical data is considered. This well-performing meta-learning technique will be used for creating (baseline) models, which will be assessed and compared to performances of subsequent experiments.

*b) Learning with Siamese Neural Networks:* One assumption of the work in hand is, that when the amount of existing image representations is way to low for traditional ML approaches, it may be sufficient for SNN-based approaches anyhow. Commonly, SNN architectures are created with respect to the comparison of two visual inputs. Throughout the three datasets, the capability of processing visual representations of numeric data in SNNs is investigated.

*c) Triplet Sampling for Time-Series Data:* Although there are many real-world applications of predicting the similarity of two data points, no standard exists regarding the crafting of triplets with respect to time-series data. Another aspect to consider is that the problem formulation is moved from predicting a class affiliation by the relation of anchor, positive and negative elements, towards an assessment of their respective appropriateness. When considering classification with respect to a regression problem, an *approximate regression* may be conducted by classification, e.g., a situation in which well-defined classes are utilized as a representation of an associated value. The choice of such a strategy may also be impacted by a multitude of aspects, as one might for example differentiate if possible data values are recurrent, exponential, linear or seemingly random. For example, there can be a static or dynamic 'sliding window', where triplet elements are positive when the window has proceeded the anchor within a certain threshold, negative respectively when the threshold was exceeded, e.g., learning to predict if a specific time-frame is associated with a preceding one. Another strategy may be a distinction and classification by splits for days, weeks, months or certain events. One might also copy the positive sample as the negative but overwriting a specific part with random or conditional values. A completely different approach was proposed in [21], where samples can be of different length and the positive element is a random subset of the anchor time-series, while the negative element is outside the anchor time frame, another modality respectively when multivariate time-series data is available. In [22], six triplets are defined for a pre-defined "time course", where direclty neighboring elements, e.g., t+1, are treated as positives and non-neighboring elements as negative element. Specifics of the further applied approach for utilizing time-series data in distance learning by forming triplets is depicted in Figure 3. The values involved in this example are depicted as circle representation and associated with eight consecutive events, e.g., *t0 - t7* of the Sunspots [11] dataset, e.g., sample nr. 1000 to 1007 which are observations between April and November of 1832. Throughout this example, it is recognizable that the *Frame Length* is constant over all triplet elements and amounts to four. In order to form a positive triplet element, the initial frame which is expressed by the anchor element, e.g., *t0 - t3*, is shifted by the amount of *Positive Offset* to the succeeding position, e.g., *t1 - t4*. A negative or non-conforming element is formed by selecting elements within a time-window of length *Negative Length*, which is positioned *Negative Offset* elements in the future regarding when the positive element has ended, and overwriting the end of the positive time-frame with it.

## IV. Experiment Description

In the following experiments, the datasets which were described in Section III-A are applied to different meta-learning approaches. In addition to processing the raw data for learning a metric with SCML, synthetic datasets are formed by generating different data representations and train on them in an SNN. The experiments aim at investigating different possibilities of multimodal data transformations with respect to meta learning in an temporal context.

### A. Learning Classification Metrics

Although the Iris [9] dataset has no temporal component, it was applied during experiments as additional indicator of the respective methodologies appropriateness. When forming triplets with elements of the three classes and implement a

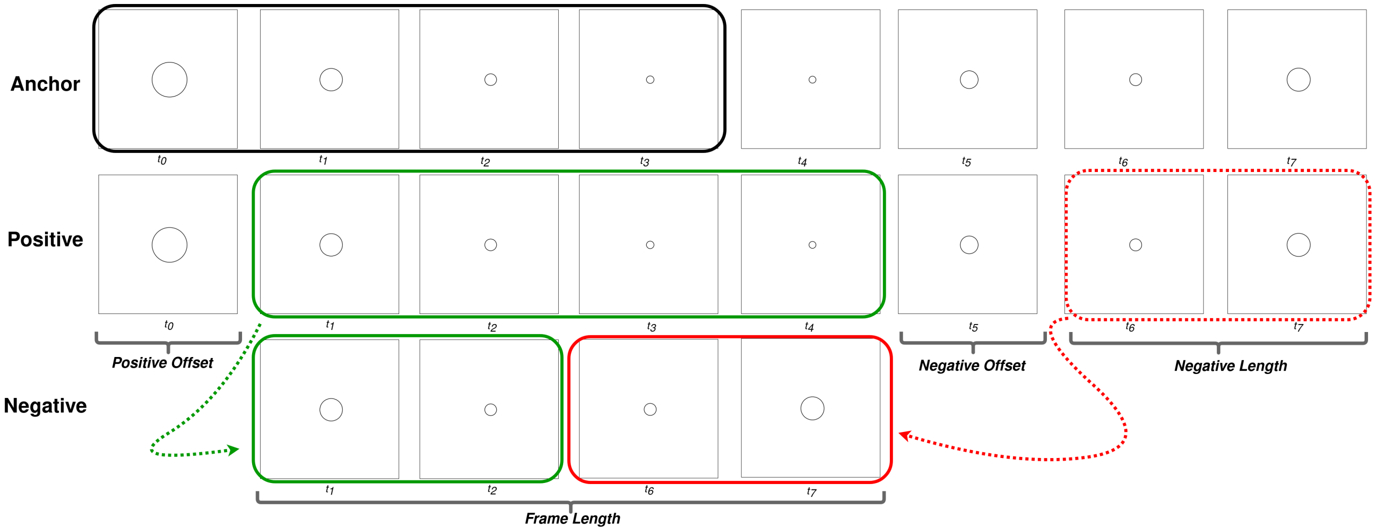

Figure 3. An example of the configurable parameters when forming triplets in ongoing unimodal time-series data, using circle representations.

random selection of the negative class, the performance of a SCML models is exceeding 90% accuracy with even a very smal amount of triplets, as depicted in Figure 4-a. The SCML algorithm was additionally applied to a simplified version of the NASA milling time-series dataset. Based on the flank wear status metric as defined in [25], three classes were considered, creating conditions similar to the Iris dataset. For synthetically decreasing the available information in training, solely the the minimum, maximum and median of values associated with a single run, e.g., power consumption, acoustic emission and vibration values of the complete time-series, were considered, alongside the processed material. The results are depicted in Figure 4-b and are supporting the assumption that measuring of distances between data points which represent different states can be applied in such time-series prediction scenarios, too.

### *B. Classification in Time-Series Forecasting*

As there is no classification in the Sunspots [11] time-series dataset, triplets have been generated by a configurable function, as already described in Section III-C. The impact of all applicable parameters was investigated in multiple benchmarks, where the SCML model training indicated that different offset sizes have not a huge impact within this dataset. Therefore, the subsequent experiments results are depicted in Figure 4-c, where different total- and negative frame lengths were tested and results suggest that the negative frame length is the most impacting parameter. This was confirmed during experiments where the *Positive Offset* parameter was set to the value of *Negative Length*, where comparable results were achieved.

### *C. Classifying Image Representations in CNN*

In order to assess the capabilities of SNN regarding a visual representation of structured data, multimodal transformations were carried out, beginning with the *Iris* [9] dataset. Beginnig with the official tensorflow tutorial examples on CNNs and making minor adjustments in the dense layer and loss function, a grayscaled *Filled Pie* representation was passing the 90% threshold within six epochs. The aim of the actual, subsequent, experiment was to reach this encouraging result using the same transformations in an SNN.

### *D. Classification of Time-Series in SNN*

With having promising results from the previous experiments which were based on structured data, the following image-based approaches were conducted. The SNN architecture begins with implementing a batch normalization of the inputs, followed by a two-dimensional convolutional layer with a stride of 2x2, 16 filters and 'tanh' activation function. Afterwards, a two-dimensional average pooling layer with a pool-size of 2x2 is applied. This combination of convolutions and average pooling resumes to 32, 64, 128 and finally 256 filters, before it is flattened, normalized and applied to a dense layer with 'l2' kernel regularizer, 'tanh' activation function and ten units. This SNN structure is effective utilized as feature extractor of the information present in the image representations and trained with categorical crossentropy, RMSProp optimization with a learning rate of 0.001 and an euclidean distance function. In training multiple models with different amounts of triplets, viable results are emerging, as depicted in Figure 4-d.

*1) Milling:* For demonstrating the utilization of tabular data in image-based ML methods, the numeric values of the mill dataset [10] were transformed into various visual representations, divided into test- and train sets and subsequently fed to an SNN in order to train a model for categorizing the flank wear of an work piece. A simple RGB plot transformation was applied regarding the various runs modalities, e.g., acoustic emission, power consumption and vibration. The results of training on different amounts of triplets in this three-class

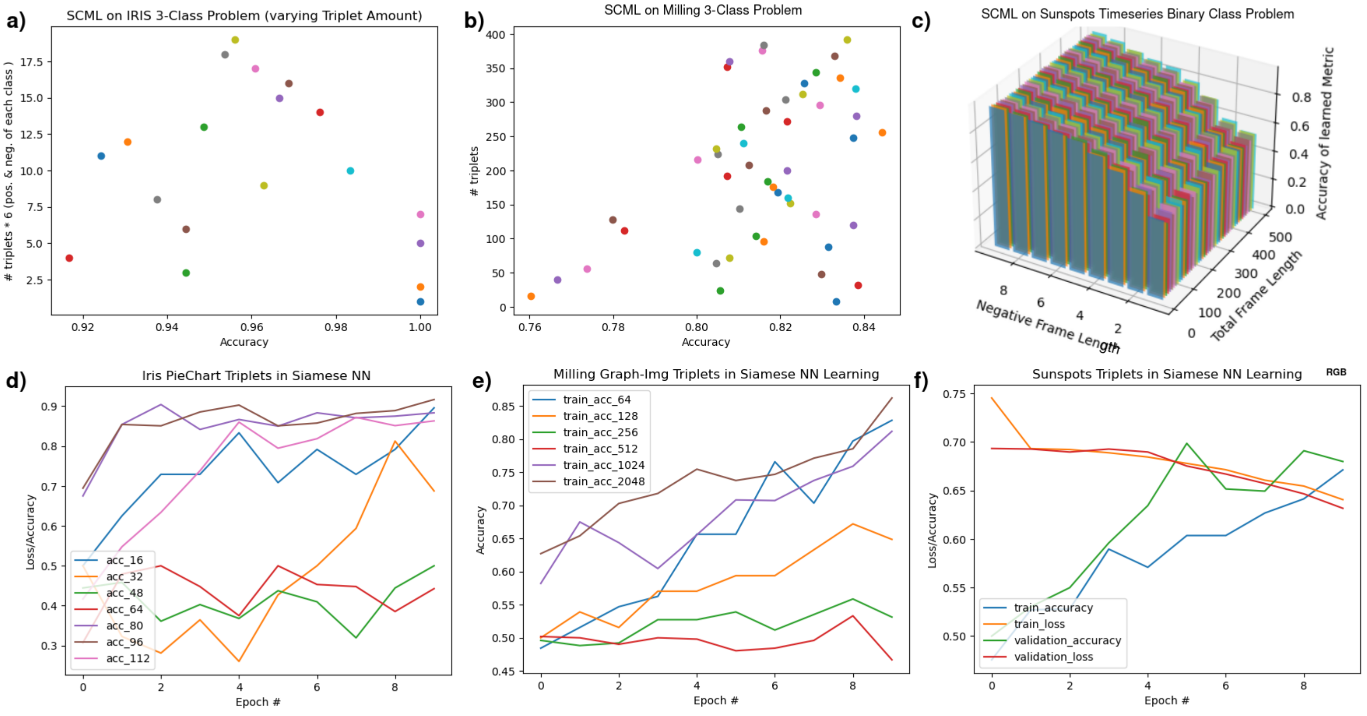

Figure 4. Results of the Experiments: **a)** *SCML* on different amounts of triplets for classifying samples in the Iris dataset; **b)** *SCML* on minimum, maximum and median values in the NASA Milling dataset; **c)** Impact of the 'Negative Length' parameter training on different amounts of triplets with positive and negative offset of 1; **d)** *SNN* training on grayscale *Pie-Chart* transformation in the Iris dataset; **e)** Training a *SNN* with different amounts of triplets in the NASA Milling dataset consisting of the particular signals *RGB plots*; **f)** Training a *SNN* with all possible Sunspots sample triplets with Frame-Length of 23, offsets of 1 and Negative-Frame-Length of 4, transformed as *Gramian Angular Field* RGB image.

classification problem are depicted in Figure 4-e. Therein, an accuracy similar to when utilizing SCML is recognizable.

*2) Sunspots:* Since there is only one value in samples of this time-series data, it cannot be put into relation with another modality. Therefore, the *Gramian angular field* transformation was applied to the Sunspots [11] dataset. Since the total frame length parameter of a triplet element is apparently not a impacting factor, a window-length of 23 was chosen. In setting the offsets both to one and configuring the negative frame length to four, results similar to when training a SCML model on numeric data were expected. In Figure 4-f, the results of this experiment are shown.

### *E. Discussion of the Obtained Results*

All in all, the experiments with SCML achieved good accuracy results, even on a small amount of training triplets. Although the chosen transformations have a heavy impact on the feature extraction of SNN approaches, the results indicate a comparative accuracy. In comparing different learning approaches or variations in their configuration, the significance of model candidates can be determined. There are also situations, where depending on the problem on hand and the amount of available data, such examinations require the concurrent long-term utilization of multiple ML models on production data.

*1) Structured Data:* The experiments with the Iris [9] and Milling [10] dataset have shown, that a classification based on the numeric values is realizable using an approach of distance-based SCML. In Figures 4 a) and d), it can be found that the random selection of triplets is causing doubtable results, e.g., accuracy scores of 1.0, where triplets are heavily biased. As there is a decent score for SCML computations right away, the SNN approach begins to perform on $triplets \geq 80$ and stabilizes after four epochs. When considering Figures 4 b) and e), an related aspect is the flank wear assessment, which is causing an unbalanced dataset interpretation as there are naturally less samples for runs with a failed tool than for degraded or healthy ones. Another point is that the image representation in e) contains more information in terms of signals when compared to the SCML experiment from b), but is missing the type of processed material.

*2) Similarity in Time-Series Classification:* When considering Figures 4 c) and f), the experimentation with the Sunspots [11] dataset suggests that with a growing *negative frame lenght* parameter, accuracy is non-stop increasing. The chosen total and negative frame length of 23 and for is scoring approximately 75% accuracy with SCML approach as contrasted with nearly 70% after eight epochs of training on the SNN, while the loss is steadily decreasing. Although the used dataset may allow for an arbitrary elevation of this parameter, the respective use case data must allow for generating 'realistic' negative samples, as well as different strategies of selecting samples for the actual inference of the model. When increasing the positive offset, the assumptions about future values, which are usually not available in a productive environment, must be formalized. Therefore, historic data may be analyzed for finding trends, outliers or other significant observations, which can be combined with, or appended to, current values. When there is only a small amount of possibilities or there are unlimited resources, the model can be inferred with a multitude of samples which consist partly of random values and determining the most likely synthetic element.

## V. Conclusion and Future Work

In this work, an overview of the utilization of multimodal data in meta-learning strategies was given with respect to time-series analysis in the context of CPS operations. Therefore, different approaches of distance-learning were investigated and applied in experiments using SCML and SNN. In implementing a novel approach of compiling temporal triplets, a classification of future time-series data seems possible. Although the results are capable of improvement, it was shown that strategies for predicting specific situations in CPS environments are possible for even small datasets using meta-learning approaches.

As this paper is concerned with basic experimentation, a better suited SNN architecture may be found and applied to additional datasets, using a broader variety of multimodal transformations and preprocessing approaches. In addition, problem-specific significance tests could be implemented for determining the feasibility of model candidates. There are many promising applications of DT technology, as for example the management and representation of concurrent modules in a ML pipeline or model-specific preprocessing operations on inference. The utilized datasets could in general be extended with augmented [15] versions, e.g., adding samples which initially were copies of the originals but are subject to random noise, blurring, colorizing, rotating and other image manipulation techniques. Such a methodology could then be assessed with respect to increasing the samples of poorly-represented classes and impacts on prediction accuracy. Regarding the Milling [10] dataset, a more ganular flank-wear classification, as for example in .2mm steps, could contribute to more stable predictions. Another factor cold be to additionally fuse the information of the applied material with the respective signal plots, increasing the specifics of samples and potentially the model's accuracy, too. When forming a series of temporal subsequent same-class samples, effects of the structured time-series data could also be represented as video stream and further be interpreted by ML methods for classifying short visual sequences, challenging aspects like the compilation of triplets, real-time inference or multimodal fusion.

## Acknowledgement

This work as outcome of the project DQ-Meister*in with the project number P2021-01-013 has received funding from the Carl Zeiss Stiftung.

## References

[1] E. Raj, D. Buffoni, M. Westerlund, and K. Ahola, "Edge mlops: An automation framework for aiot applications," in *2021 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 2021, pp. 191–200.

[2] P. Ruf, M. Madan, C. Reich, and D. Ould-Abdeslam, "Demystifying mlops and presenting a recipe for the selection of open-source tools," *Applied Sciences, MDPI*, vol. 11, no. 19, p. 8861, 2021.

[3] N. Zubair, A. Niranjan, K. Hebbar, and Y. Simmhan, "Characterizing iot data and its quality for use," 06 2019.

[4] V. Jane *et al.*, "Survey on iot data preprocessing," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 238–244, 2021.

[5] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," *arXiv preprint arXiv:2203.04291*, 2022.

[6] M. Kaya and H. Ş. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, p. 1066, 2019.

[7] X. Chen and K. He, "Exploring simple siamese representation learning," *CoRR*, vol. abs/2011.10566, 2020. [Online]. Available: https://arxiv.org/abs/2011.10566

[8] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[9] R. A. Fisher, "Iris," UCI Machine Learning Repository, 1988, DOI: https://doi.org/10.24432/C56C76.

[10] A. Agogino and K. Goebel, "Best lab, uc berkeley. "milling data set ", nasa prognostics data repository," 2007, nASA Ames Research Center, Moffett Field, CA.

[11] J. Rogel-Salazar, "Sunspots - Monthly Activity since 1749," http://doi.org/10.6084/m9.figshare.6728255.v1, 2018.

[12] N. Schneider, P. Ruf, M. Lermer, and C. Reich, "Arthur: Machine learning data acquisition system with distributed data sensors," in *Proceedings of the 13th International Conference on Cloud Computing and Services Science - Volume 1: CLOSER*, INSTICC. SciTePress, 2023, pp. 155–163.

[13] B. Chander, S. Pal, D. De, and R. Buyya, "Artificial intelligence-based internet of things for industry 5.0," in *Artificial Intelligence-based Internet of Things Systems*. Springer, 2022, pp. 3–45.

[14] A. Sharma, E. Vans, D. Shigemizu, K. Boroevich, and T. Tsunoda, "Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture," *Scientific Reports*, vol. 9, 08 2019.

[15] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[16] F. Ott, D. Rugamer, L. Heublein, B. Bischl, and C. Mutschler, "Cross-modal common representation learning with triplet loss functions," 2022, oSF Preprints.

[17] L. Buturović and D. Miljković, "A novel method for classification of tabular data using convolutional neural networks," *bioRxiv*, 05 2020.

[18] Y. Shi, A. Bellet, and F. Sha, "Sparse compositional metric learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.

[19] G. J. McLachlan, "Mahalanobis distance," *Resonance*, vol. 4, no. 6, pp. 20–26, 1999.

[20] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Self-restrained triplet loss for accurate masked face recognition," *arXiv preprint arXiv:2103.01716*, 2021.

[21] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," *Advances in neural information processing systems*, vol. 32, 2019.

[22] P. A. Marin Zapata, S. Roth, D. Schmutzler, T. Wolf, E. Manesso, and D.-A. Clevert, "Self-supervised feature extraction from image time series in plant phenotyping using triplet networks," *Bioinformatics*, vol. 37, no. 6, pp. 861–867, 2021.

[23] A. Brunel, J. Pasquet, J. Pasquet, N. Rodriguez, F. Comby, D. Fouchez, and M. Chaumont, "A CNN adapted to time series for the classification of supernovae," *arXiv preprint arXiv:1901.00461*, 2019.

[24] T. von Hahn and C. K. Mechefske, "Machine learning in cnc machining: Best practices," *Machines*, vol. 10, no. 12, p. 1233, 2022.

[25] Y. Cheng, H. Zhu, K. Hu, J. Wu, X. Shao, and Y. Wang, "Multisensory data-driven health degradation monitoring of machining tools by generalized multiclass support vector machine," *IEEE Access*, vol. 7, pp. 47 102–47 113, 2019.

[26] T. V. Hahn and C. K. Mechefske, "Self-supervised learning for tool wear monitoring with a disentangled-variational-autoencoder," *International Journal of Hydromechatronics*, vol. 4, no. 1, pp. 69–98, 2021.