



**HAL**  
open science

# Active transfer learning for data-driven manufacturing process modelling

Gengxiang Chen, Yingguang Li, Charyar Mehdi-Souzani, Xu Liu

► **To cite this version:**

Gengxiang Chen, Yingguang Li, Charyar Mehdi-Souzani, Xu Liu. Active transfer learning for data-driven manufacturing process modelling. *Procedia CIRP*, 2023, 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 118, pp.98-103. 10.1016/j.procir.2023.06.018 . hal-04174432

**HAL Id: hal-04174432**

**<https://hal.science/hal-04174432>**

Submitted on 2 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

16th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME '22, Italy

## Active transfer learning for data-driven manufacturing process modelling

Gengxiang Chen<sup>a,b</sup>, Yingguang Li<sup>a,\*</sup>, Charyar Mehdi-Souzani<sup>b</sup>, Xu Liu<sup>c</sup>

<sup>a</sup>Nanjing University of Aeronautics and Astronautics, College of Mechanical and Electrical Engineering, 210016, Nanjing, China

<sup>b</sup>Université Paris-Saclay, ENS Paris-Saclay, LURPA, Université Sorbonne Paris Nord, 91190 Gif-Sur-Yvette, France

<sup>c</sup>Nanjing Tech University, School of Mechanical and Power Engineering, 211816, Nanjing, China

\* Corresponding author. E-mail address: [liyingguang@nuaa.edu.cn](mailto:liyingguang@nuaa.edu.cn)

### Abstract

Manufacturing process modelling (MPM) aims to construct high-fidelity digital predictive models of the concerned properties of products, processes or manufacturing systems for the further optimisation and improvement of manufacturing activities. Data-driven modelling methods, including machine learning and deep learning, have drawn immense attention to MPM problems because of their powerful representative ability. However, the labelled data of concerning properties in the manufacturing process is often insufficient and sparse because of the expensive and time-consuming experiments or simulations. The scarcity of labelled data hinders the further development of data-driven models in MPM problems. This paper proposes an active transfer learning framework by integrating active generation of labelled data and the processing of relevant data to reduce the requirements of labelled data. Firstly, the initial active labelling module introduces the generation of a more representative and informative labelled dataset rather than a randomly generated one. Then, the transfer learning model can extract the general information from the relevant data to address the information scarcity for the target task. Besides, the iterative active labelling module can determine to query promising new labelled data according to the performance of the current model. The effectiveness of the proposed framework is verified in a tool wear prediction case. The experimental outcomes demonstrate that the three modules of the framework can reduce the labelled data requirements and enhance the performance of the data-drive model under limited labelled data.

© 2022 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering.

*Keywords:* Smart manufacturing; data driven; transfer learning; active learning

### 1. Introduction

Since formalised in the mid-2000s, the fourth industrial revolution has become the heart of the development of the more than ever globalised industry [1]. This revolution is above all the advent of information and its dissemination through communication at the heart of the so-called industry 4.0. The deep embedding of modern information technologies enables the manufacturing system to generate, transmit and analyse variable types of information, then improve and optimise the manufacturing activities reversely [2]. The integration of intelligent technologies and manufacturing technologies brings new changes to all levels of manufacturing problems, such as supply chain level, factory level, and process modelling level [1][3].

Manufacturing process modelling (MPM) aims to construct a high-fidelity digital predictive model of the concerned properties of products, processes or manufacturing systems, further enabling the optimisation and improvement of manufacturing activities [4]. Due to the complexity of the manufacturing process, traditional modelling methods pursue the first principle to focus on the primary mechanism knowledge while bringing inevitable assumptions, simplifications and approximations. Conversely, data-driven modelling methods can learn the relationship between complex influential factors and the concerned properties from the collected labelled data with low demand for domain knowledge. Therefore, data-driven MPM is drawn immense attention in different application scenarios such as additive manufacturing, milling, composites curing, etc. [5].

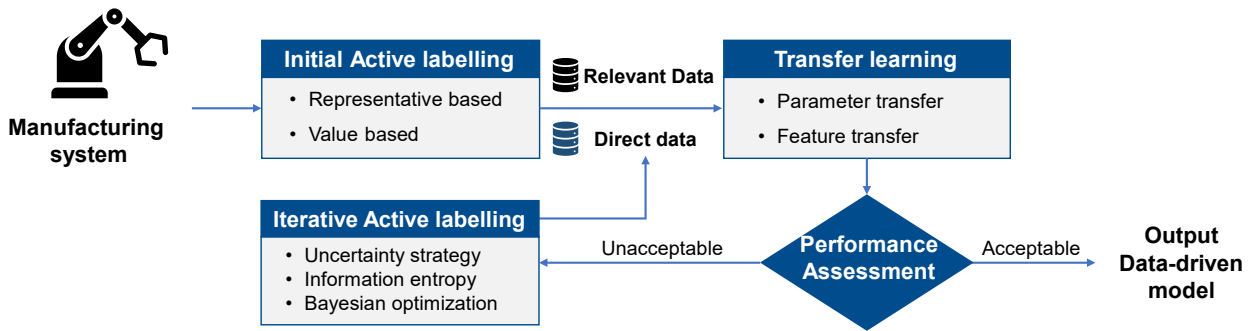


Fig. 1 Active transfer learning framework for data-driven manufacturing process modelling

Although data-driven modelling methods are powerful enough to represent the complex manufacturing process, the accuracy and the generalisability of the predictive model heavily rely on massive labelled data. However, the labelled data of concerning properties in the manufacturing process is often sparse and limited because of the expensive and time-consuming experiments or simulations. Since limited labelled data cannot provide sufficient underlying knowledge of the manufacturing process, even powerful modelling methods cannot compensate for the information scarcity.

When the labelled data is insufficient to train an accurate model, the most direct solution is to query new labelled data to improve the performance of the model, that is what defines the active learning technique [6]. Leco et al. proposed an active learning algorithm for robotic machining errors modelling that makes the online inspection decision based on the prediction confidence of the current model [7]. Hughes et al. presented a risk-based active learning solution for structural health classification, in which the promising class-label information is queried by the expected value for each incipient data point [8]. Similarly, Arellano et al. built a Bayesian Convolutional Neural Networks for online tool condition classification, that can determine whether the incoming data should be labelled. These applications demonstrate that active learning strategies can achieve a satisfactory model performance with a smaller training dataset [9].

Another solution for data scarcity is transfer learning, which aims to improve the predictive function of the target task by transferring the knowledge from related source data, so that the target task only requires a few target data. Transfer learning has been successfully applied in different manufacturing process modelling problems, such as composite thermal-chemical [10], tolerance estimation for additive manufacturing [11], manufacturing system maintenance [12], etc.

By combining active learning and transfer learning, active transfer learning (ATL) has also been studied recently. The transfer stage and active learning stage can be carried out in multiple sequence steps. For example, Ramezankhani et al. developed an ATL framework for composites curing, and decreased the dependency of the data-driven model on the large datasets [10]. However, existing ATL research only focuses on how to introduce external information to the existing insufficient labelled data without considering the generation of the labelled data. The initial labelled data of

most existing research are generated randomly or treated as default given [10].

In this research, an active transfer learning framework is proposed for both the active generation of labelled data and the processing of relevant data for manufacturing process modelling problems. The utility of the active generation here is two-fold, initial active labelling and subsequent iterative active labelling. The former provides the high-quality initial target labelled data, while the latter will determine the queried labelled data according to the performance assessment. The proposed framework is verified in the case study of tool wear prediction. The experimental results show that all the three modules in the framework, namely initial active labelling, iterative active labelling and transfer learning, can improve the performance of the model while reducing the requirements of labelled data.

## 2. The proposed framework

### 2.1. Structure

The proposed active transfer learning framework for manufacturing process modelling is shown in Fig . 1. Since it is difficult to collect sufficient labelled data, the valuable few labelled data deserve to be elaborately designed to carry more underlying knowledge. Therefore, the first step in the framework is initial active labelling to construct the initially labelled dataset. After that, transfer learning is applied to extract the valuable knowledge from the relevant data to the initially labelled dataset. According to the performance of transfer learning results, iterative active labelling will determine the querying of further labelled data to improve the performance of the model gradually.

Although all the three modules could reduce the requirements of labelled data, it is unnecessary and sometimes not feasible to include all these modules for a particular MPM problem. For example, iterative active labelling could be impractical for some tasks due to the complexity of the experimental setup. Besides, the initial label data of some tasks are predetermined via historical accumulation, which cannot be designed manually. Therefore, the solution of each MPM task should be tailored individually according to its characteristics.

### 2.2. Initial active labelling

Initial active labelling aims to find a subset  $T$  from the

potential unlabelled data pool  $N$  for labelling, where  $T \subseteq N$ . Therefore, initial active labelling becomes equivalent to selecting the initial unlabelled dataset. This problem here can be treated as a data sampling problem, which is a statistical analysis technique widely used in many fields [12]. The most intuitive solution is the representative sampling, which samples a coreset that represents the distribution of the total data [13]. From the perspective of probability, the objective of initial active labelling can be defined as:

$$\min_T d(p_T, p_N), \text{ subject to, } |T| = n_t \quad (1)$$

where  $n_t$  is the size of the selected subset.  $d(p_T, p_N)$  is the distribution distance between the total dataset and the selected dataset. Theoretically, the distance function can be Kullback-Leibler divergence or other statistical distance.

When the potential dataset is defined by the ranges of feature space, we can replace the representativeness of the probability density with the spatial representativeness. For example, Latin hypercube sampling (LHS) can generate representative samples with strict guarantees, especially when the input dimension is relatively low.

In practice, most potential datasets exist in the form of a finite amount of unlabelled data without predefined distribution information. Therefore, clustering becomes the most simple but reasonable strategy to select the representative initial unlabelled dataset [13]. Take k-means clustering as an example, suppose the potential dataset  $N$  is partitioned into  $n_T$  observation groups as  $\{N^1, \dots, N^{n_t}\}$ . The objective function is then defined as minimising the within-cluster sum of distances:

$$\min_T \sum_{i=1}^{n_t} \sum_{\mathbf{x} \in N^i} \|\mathbf{x} - \mu_i\|^2 \quad (2)$$

where  $T = \{\mu_1, \dots, \mu_{n_t}\}$  is the optimised representative unlabeled dataset, and  $\mu_i$  is the sample that represent the group  $N^i$ . Therefore, the dataset  $T$  can be actively labelled to construct the initial labelled dataset  $\mathcal{D}_T$ .

### 2.3. Transfer learning

For most MPM problems, there always exists auxiliary data which are less valuable but still can provide helpful information for the modelling of the target model. By transferring the knowledge from the auxiliary data, the data requirements of the target model can be reduced. Define the auxiliary data as the source data  $\mathcal{D}_S$ . From the probabilistic point of view, the transfer learning technique aims to adapt the distribution difference between the source data  $\mathcal{D}_S$  and the target data  $\mathcal{D}_T$ . The distribution adaptation solution should be designed according to the distribution difference assumptions of the problem [14]. Covariance shift and conditional shift are two widely studied scenarios for MPM problems.

#### 2.3.1 Covariance shift problems

Covariance shift means that the source data and target data have the different marginal distribution while sharing the same conditional distribution, that is  $p(\mathbf{y}_s | \mathbf{X}_s) = p(\mathbf{y}_t | \mathbf{X}_t)$ , whereas  $p(\mathbf{X}_s) \neq p(\mathbf{X}_t)$ .  $\mathbf{X}_s$  and  $\mathbf{y}_s$  are the features and labels

of the source data, namely  $\mathcal{D}_S = \{\mathbf{X}_s, \mathbf{y}_s\}$ .  $\mathbf{X}_t$  and  $\mathbf{y}_t$  are the features and labels of the target data, and  $\mathcal{D}_T = \{\mathbf{X}_t, \mathbf{y}_t\}$ . Bearing fault classification or structural health classification are classical covariance shift problems. The solution is to learn a new representation space to adapt the distribution difference of the feature space. For example, Maximum Mean Discrepancy (MMD) based methods try to find a mapping  $\phi$  to minimise the distance:

$$\text{Dist}(\mathbf{X}_s, \mathbf{X}_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{X}_{s_i}) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(\mathbf{X}_{t_i}) \right\|_{\mathcal{H}}^2 \quad (3)$$

where the mapping  $\phi$  can be a transformation matrix or a neural network. After distribution adaptation, the new features of the source task and the target task can be trained together to establish the target model.

#### 2.3.2 Conditional shift problems

Another widespread transfer learning configuration in MPM is conditional distribution shift regression problem, namely  $p(\mathbf{y}_s | \mathbf{X}_s) \neq p(\mathbf{y}_t | \mathbf{X}_t)$ , such as robot forward dynamics transfer, tool wear monitoring model transfer, etc. The conditional distribution difference can be adapted by minimising the Conditional Embedding Operator Discrepancy (CEOD) as [15]:

$$D_{CEOD}(\mathcal{D}_s, \mathcal{D}_t) = \|\hat{C}_{\mathbf{y}_s|\mathbf{X}_s} - \hat{C}_{\mathbf{y}_t|\mathbf{X}_t}\|_{HS}^2 \quad (4)$$

where  $\hat{C}_{\mathbf{y}|\mathbf{X}}$  is the conditional embedding operator.

In practice, we can pre-train the source model, reuse the parameters of the beginning layers, and fine-tune the last several layers by minimising the distribution difference.

#### 2.4. Iterative active labelling

Since the initial labelled data is insufficient to train a satisfactory model, active learning can query new labelled data according to the predefined acquisition function [6]. The general representation of the iterative active labelling is:

$$\max_{x \in N \setminus T} \Delta_{\mathcal{A}}(x | T) \quad (5)$$

where  $\Delta_{\mathcal{A}}(x | T)$  means the marginal gain of the sample  $x$  for the given subset  $T$  and data-driven algorithm  $\mathcal{A}$ .

The marginal gain can be the increment of the representativeness criteria or the information criteria. Representative-based query strategies focus on the exploration of the structure of the unlabelled dataset. Informative-based query strategies determine the promising samples based on the performance of the model on the current labelled dataset. The famous active learning methods under the two categories of strategies are summarised as follows:

Representative-based query strategies:

- Density-based approach
- Diversity-based approach
- Exploration graph structure

Informative-based query strategies:

- Uncertainty sampling
- Variance reduction
- Query by committee
- Support vector based approach

### 3. Case study

In this section, the proposed framework is applied in a tool wear prediction problem to verify the effectiveness of initial active labelling, iterative active labelling and transfer learning modules.

#### 3.1. Experimental settings

The healthy condition of cutting tools has a direct influence on the machining process stability and the final quality of the product. Due to the complexity of the cutting process, is it difficult to predict the tool wear accurately using mechanism models. Since the tool wear will result in inconsistent cutting width, further leading to the fluctuation in the cutting force, building data-driven models from the monitoring cutting force to the corresponding tool wear becomes a potential solution. There has been much research focusing on deep-learning-based tool wear prediction, while this research will investigate the potential of active labelling and transfer learning. The tool wear dataset from the PHM society conference data challenge consists of the monitoring signal and tool wear value for three blades of two cutting tools (C4 and C6). The experiments were conducted to answer the following questions:

- Whether the active initial labelled dataset performs better than the random initial labelled dataset?
- Whether transfer learning can improve the performance of the model?
- Whether the distribution of the target data will influence the performance of transfer learning?
- Whether active learning can improve the performance of the model effectively?

Tool C6B3 and C4B2 are selected as the source tool and the target tool, respectively. As shown in Fig. 2, the deep learning structures of the two tasks are the same. Four 1-D convolution modules are introduced to extract features from the input monitoring signal series. Each convolution module consists of a convolution layer, a batch normalisation layer and a max-pooling layer. After the pre-training model of the source tool, the parameters of the feature extraction layers are frozen and reused for the target tool. Three full connected layers that map the features to the output tool wear values are fine-tuned to adapt the distribution difference between the two datasets. The configurations of the neural network are listed in Table 1.

#### 3.2. Results of initial active labelling

We first investigate the influence of the distribution of the labelled data on the source tool. An initially labelled dataset, marked as the coreset, is first generated by K-means datasets

Table 1. The configurations of the neural network.

Item of the neural network	Column A (t)
Convolution modules	Cov(1,64,5, stride =5)+BN+ Relu
	Cov(64,64,5, stride =5)+BN+ Relu
	Cov(64,64,5, stride =5)+BN+ Relu
	Cov(64,64,3, stride =5)+BN+ Relu + flatten
Linear layer	384→64→16→1
Optimiser	Adam
Learning rate	1e-3
Epoch	100

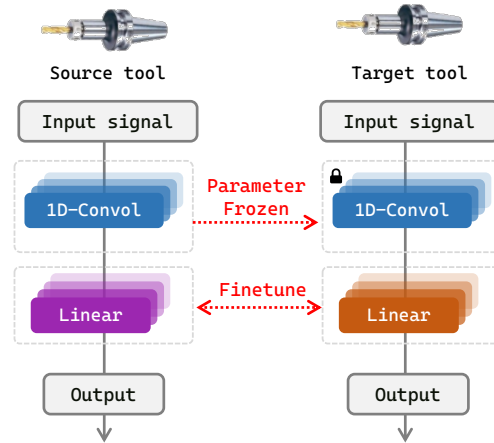


Fig. 2 The deep transfer learning structure

are generated for comparison. The rest labelled data are defined as test data to evaluate the Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

where  $y_i$  and  $\hat{y}_i$  are the real tool wear and the predicted result respectively.

As shown in Fig. 3a, the blue curve is the predicted tool wear result of coreset, while the multiple purple curves refer to the predicted results of 100 groups of random labelled dataset. It is clear that the coreset result shows a satisfactory consistency with the ground truth while the random results are particularly.

The statistical result of MSEs for all random datasets is shown in Fig. 3b. Most of MSEs lie between  $5 \mu m$  to  $10 \mu m$ , whereas the minimum is around  $3 \mu m$ , and the maximum can be larger than  $35 \mu m$ . This figure demonstrates that the distribution of labelled data has a significant influence on the performance of the data-driven model. The red line in Fig. 3b refers to the coreset result with  $MAE = 4.31 \mu m$ . Although the coreset cannot achieve the best result of random datasets, it can provide a relatively stable result better than 90% of random datasets.

#### 3.3. Results of iterative active labelling

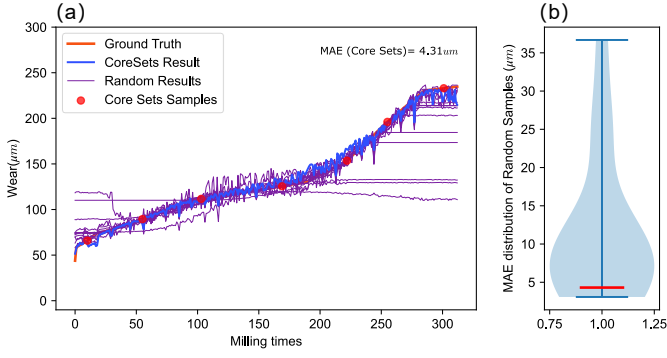


Fig. 3 The results of initial active labelling

The uncertainty based method is used as the iterative query strategy in this case study. For classification problems, the soft-max activation function of the output layer can be treated as probabilistic information of the corresponding layers. However, for regression problems, normal deep neural networks can only give a predicted variable as the output without any uncertainty information. Bayesian neural network or Gaussian Process Regression (GPR) are widely used probabilistic models which can provide uncertainty for iterative active labelling.

Then iterative active labelling procedure is shown in Fig. 4. We use the signal features extracted from the pre-trained convolution modules, represented by  $\psi(\mathbf{X})$ , as the input to the GPR. For a give signal feature  $x_*$ , the predicted tool wear value  $y_x$  can be represented as a conditional distribution:

$$p(y_* | \mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\mu_*, \Sigma_*) \quad (7)$$

The standard solution of GRP consists of the predicted mean and variance of the sample  $x_* \cdot g$

$$\mu_x = \mathbf{K}_x^T \mathbf{K}^{-1} \mathbf{y} \mathbf{K}_x \quad (8)$$

$$\Sigma_x = \mathbf{K}_{xx} - \mathbf{K}_x^T \mathbf{K}^{-1} \mathbf{K}_x \quad (9)$$

Then the model can query the sample with the maximum variance and add the sample to the new labelled dataset.

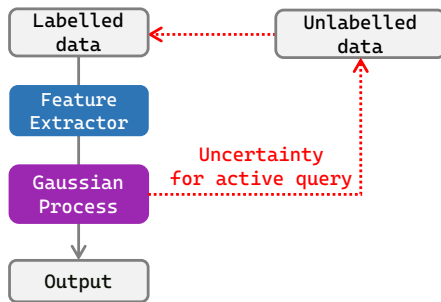


Fig. 5 The uncertainty-based active learning method in the case study

An implementation of the iterative active labelling for tool C6B3 is shown in Fig. 5. Starting with only two training samples, the initial MAE is very large (MAE = 55.45  $\mu\text{m}$  in Fig. 5a). Then the next sample could be determined by the uncertainty of the current prediction result. After adding the new samples, the maximum uncertainty would be reduced directly. The uncertainty function fluctuates wildly because of the noise of the input signal and the error from feature extraction. But the underlying trend of the uncertainty function does describe which samples in the feature space are more desirable. For example, Fig. 5b shows that exiting samples located in the region with fewer milling times, so the trained model shows great uncertainty for samples with larger milling time. Note that, the information 'milling time' is inaccessible during training. The query strategy determines the target based on the features of the monitoring signals. After querying five times, the distribution of the selected samples could cover the primary range of the feature space, and then the MAE is reduced significantly to only 6.13  $\mu\text{m}$  (Fig. 5e).

The results show that iterative active labelling can query desirable samples, which can potentially reduce the labelling efforts. An active learning solution to the tool wear prediction problem can work as an online monitoring system, predict the incoming data and determine whether it should be manually labelled.

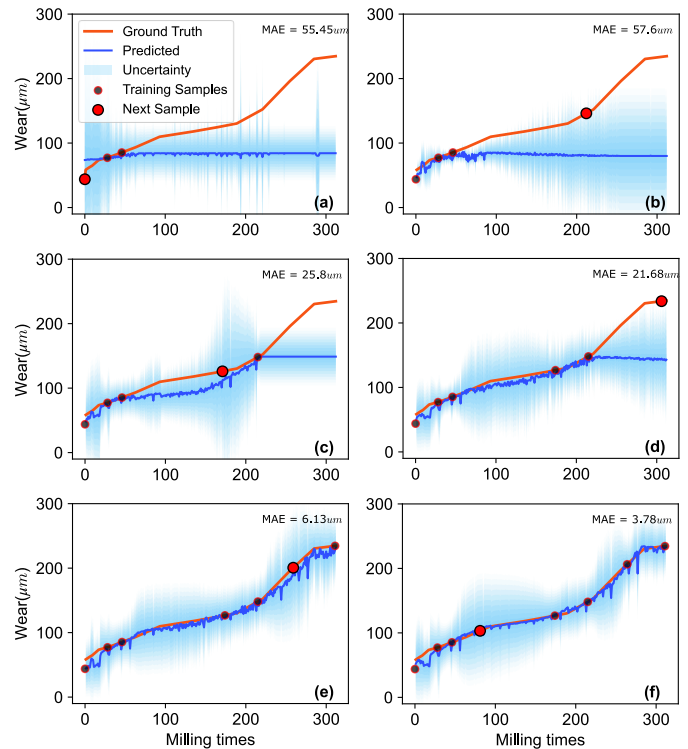


Fig. 4 The results of iterative active labelling

### 3.4. Results of transfer learning



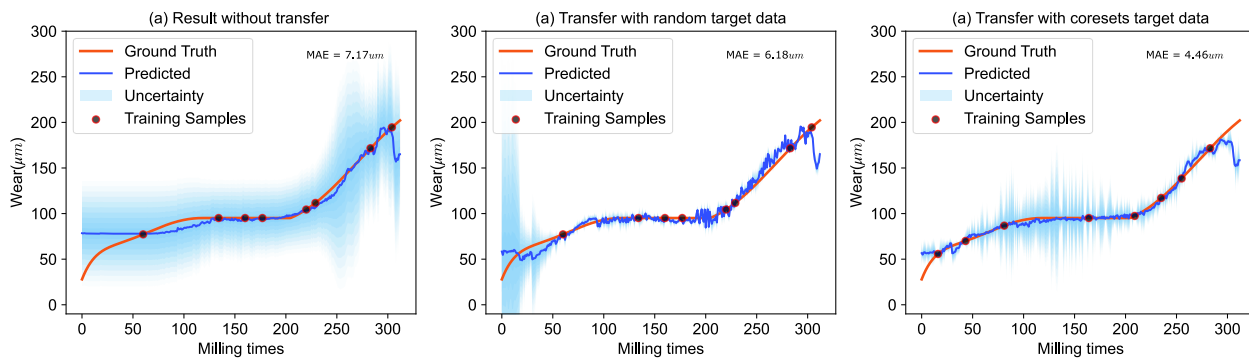


Fig. 6 The results of transfer learning

Fig. 6 reports the results of transfer learning from the tool C6B3 to the tool C4B2. The feature extraction layers were first trained on the source tool C6B3, and then reused in the target tool C4B2. Fig. 6a shows the prediction result of the model trained by a random group of target data without transfer learning. The uncertainty is relatively large because the insufficient training data cannot provide prior information for all the feature space. After transferring the knowledge from source data, Fig. 6b shows a much better prediction result with  $MAE = 6.18 \mu m$ . Besides, the initial active labelled coreset is defined as the target dataset to evaluate the influence of the target data distribution on the performance of transfer learning. The prediction results in Fig. 6c show that the coreset can improve the performance of transfer learning compared to the random target dataset.

#### 4. Conclusion

Although data-driven models have significant potential in modelling the complex manufacturing process, the scarcity of labelled data hinders its practicability and further development. An active transfer learning framework is proposed for the data scarcity scenarios by integrating the active generation of labelled data and the processing of relevant data. The contribution of this paper can be summarised as follow:

(1) The proposed framework can provide a general introduction for a wide range of data scarcity MPMs problems from three aspects: initial active labelling, iterative active labelling and transfer learning.

(2) A series of experiments are carried out on the tool wear prediction problem. The results on the tool wear demonstrate that the three modules of the framework can reduce the labelled data requirements and enhance the performance of the data-drive model under limited labelled data.

#### Acknowledgements

This work was supported by the National Key R&D Programs of China (Grant No. 2020YFA0713704).

#### References

- [1] Tan Jianrong, Liu Daxin, Liu Zhenyu, and Cheng Jin. Research on key technical approaches for the transition from digital manufacturing to intelligent manufacturing. *Strategic Study of Chinese Academy of Engineering*, 19(3):34–44, 2017.
- [2] Fei Tao, Nabil Anwer, Ang Liu, Lihui Wang, Andrew YC Nee, Liming Li, and Meng Zhang. Digital twin towards smart manufacturing and industry 4.0. *Journal of manufacturing systems*, 58:1–2, 2021.
- [3] Zuwei Zhu, Yassir Arezki, Na Cai, Charyar Mehdi-Souzani, Nabil Anwer, and Hichem Noura. Data fusion-based method for the assessment of minimum zone for aspheric optics. *Computer-Aided Design and Applications*, 18(2):309–327, 2020.
- [4] Andrew Kusiak. Smart manufacturing. *International Journal of Production Research*, 56(1-2):508–517, 2018.
- [5] Zuwei Zhu, Nabil Anwer, Qiang Huang, and Luc Mathieu. Machine learning in tolerancing for additive manufacturing. *CIRP Annals*, 67(1):157–160, 2018.
- [6] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Computing Surveys (CSUR)*, 54(9):1–40, 2021.
- [7] Mateo Leco, Thomas McLeay, and Visakan Kadirkamanathan. A two-step machining and active learning approach for right-first-time robotic counter-sinking through in-process error compensation and prediction of depth of cuts. *Robotics and Computer-Integrated Manufacturing*, 77:102345, 2022.
- [8] AJ Hughes, Lawrence A Bull, P Gardner, Robert James Barthorpe, Niko-laos Dervilis, and Keith Worden. On risk-based active learning for structural health monitoring. *Mechanical Systems and Signal Processing*, 167:108569, 2022.
- [9] Giovanna Martinez Arellano and Svetan Ratchev. Towards an active learning approach to tool condition monitoring with bayesian deep learning, 2019.
- [10] Milad Ramezankhani, Apurva Narayan, Rudolf Seethaler, and Abbas S Milani. An active transfer learning (atl) framework for smart manufacturing with limited data: case study on material transfer in composites processing. In *2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, pages 277–282. IEEE, 2021.
- [11] Ansi Zhang, Honglei Wang, Shaobo Li, Yuxin Cui, Zhonghao Liu, Guanci Yang, and Jianjun Hu. Transfer learning with deep recurrent neural networks for remaining useful life estimation. *Applied Sciences*, 8(12):2416, 2018.
- [12] Sif Eddine Sadaoui, Charyar Mehdi-Souzani, and Claire Lartigue. Multisensor data processing in dimensional metrology for collaborative measurement of a laser plane sensor combined to a touch probe. *Measurement*, 188:110395, 2022.
- [13] Punit Kumar and Atul Gupta. Active learning query strategies for classification, regression, and clustering: a survey. *Journal of Computer Science and Technology*, 35(4):913–945, 2020.
- [14] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.
- [15] Xu Liu, Yingguang Li, Qinglu Meng, and Gengxiang Chen. Deep transfer learning for conditional shift in regression. *Knowledge-Based Systems*, 227:107216, 2021.