An Active Transfer Learning (ATL) Framework for Smart Manufacturing with Limited Data: Case Study on Material Transfer in Composites Processing

Milad Ramezankhani
Materials and Manufacturing Research Institute
University of British Columbia
Kelowna, Canada
milad.ramezankhani@ubc.ca

Apurva Narayan
Department of Computer Science
University of British Columbia
Kelowna, Canada
apurva.narayan@ubc.ca

Rudolf Seethaler
Materials and Manufacturing Research Institute
University of British Columbia
Kelowna, Canada
rudolf.seethaler@ubc.ca

Abbas S. Milani*
Materials and Manufacturing Research Institute
University of British Columbia
Kelowna, Canada
abbas.milani@ubc.ca

2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS) | 978-1-7281-6207-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPS49255.2021.9468145

***Abstract*— Unprecedented advances in Machine Learning (ML), cloud computing and sensory technology promise to enable the manufacturing industry to respond rapidly to changes in marked needs while maintaining product quality and minimizing costs. Despite the unparalleled advantages that ML offers, critical limiting factors have prevented the exhaustive cultivation of ML in advanced manufacturing. Constant shifts in the process configuration and lack of sufficient fully-descriptive data restrict the performance of predictive ML models. This paper proposes to partly address these shortfalls with an active transfer learning (ATL) model that is applied to an aerospace composites manufacturing case study. The proposed ATL framework requires 1) developing an AI-based optimal experimental design using Active Learning (AL) to maximize the information gain from the limited number of allowable manufacturing trials, and 2) equipping the manufacturing process with a robust Transfer Learning (TL) model that is trained on limited available data and is immune to shifts in the process settings. The results suggest that uncertainty-based AL approaches can significantly decrease the dependency on large datasets for obtaining accurate process models. Furthermore, in comparison with traditional TL approaches, the proposed framework represents a practical solution to further reduce the necessity for generating expensive data in advanced manufacturing applications for developing reliable and transferable predictive models.**

Keywords— transfer learning, active learning, Industry 4.0, autoclave processing, advanced composites processing

I. INTRODUCTION

The emergence of new technologies such as cyber-physical systems (CPS), cloud computing and the internet of things has provided an unprecedented opportunity for manufacturing industries to fully exploit machine learning (ML) capabilities for obtaining valuable insights for prognostic and diagnostic decision-making tasks [1]. The current data streams in manufacturing can be divided into three main categories: 1) inventory and supply chain data, 2) equipment maintenance and monitoring data, and 3) product quality and design data [2]. Although under a CPS, massive datasets can be generated for the first and the second categories of data, the available *labeled* data (i.e., measured quality metrics) for product quality and design is often sparse and very limited. The main underlying factor for such data paucity is the notoriously high financial and temporal operational costs of manufacturing processes. For instance, the curing process of a composite part in the aerospace industry can take up to 10 hours to complete [3]. The absence of sufficient labeled data hinders the development of a robust and accurate predictive model for learning the underlying behavior of the data and extracting relationships between process settings (input space) and product quality (output variables). Even in the case of available historical data under some specific operational configuration, the trained ML model fails to accurately predict new data from a different probability distribution (e.g., new raw material). Such shifts and modifications in a manufacturing process can perturb the data distribution in such a way that it considerably differs from its original state [4]. This necessitates training a new model for each new process configuration from scratch, which is time-consuming and infeasible in many advanced manufacturing applications.

Such operational shifts, however, play a crucial role in smart manufacturing (Industry 4.0) as a response to the dynamic and rapidly-evolving needs of the customers and the market [5]. This entails re-designing the predictive models in such a way that they become immune to data distribution shifts and do not incur any performance decline under different conditions. Various methods have been introduced to tackle the issue of limited data in ML. Transfer Learning (TL) [6] and Active Learning (AL) [7] are the two widely studied solutions in the literature.

AL is an ML framework with the ultimate goal of learning an unknown function with the fewest possible number of data labeling steps (e.g., manufacturing trials). In this setting, the ML model itself is responsible for formulating the optimal design of experiments to minimize the required data to reach the expected performance level. Various AL strategies have been investigated [7]. Uncertainty sampling is one of the most popular approaches in which the estimated uncertainty is used for sample selection. This method is based on information theory and it targets maximizing the model entropy at each iteration [8]. Despite its effectiveness in reducing the need for big data, the application of AL in manufacturing has been lightly explored in the literature. Martínez-Arellano and Ratchev [9] implemented an online AL model for tool condition monitoring of Computer

This study was financially supported by the New Frontiers in Research Fund (NFRF) of Canada – Exploration Program.

Numerical Control (CNC) machines using an entropy-based acquisition function. In order to reduce the expert human costs for annotating large unlabeled images, Shim et al. [10] developed an AL-based CNN model for wafer map pattern classification in the semiconductor manufacturing process. It compared various AL acquisition functions for quantifying the prediction uncertainty. Yue et al. [11] implemented AL to minimize the required number of data for developing a predictive model for automatic shape control of a composite fuselage. Trained on finite element simulation data, the predictive model maps the actuators' configurations to the final shape of the composite fuselage.

In TL, however, the aim is to develop a reliable learning model with the limited available data by transferring the knowledge learned from other related domains. TL can mitigate the destructive effect of the distribution discrepancy between the training and test data on the model performance [1]. In the context of smart manufacturing, some studies implemented TL approaches in process optimization, fault diagnosis, and manufacturing maintenance. Ramezankhani et al. [4] used TL to predict the thermal history of composite parts manufactured in a two-hold autoclave curing process with limited data. The knowledge learned from training a model on abundant historical data with a different but similar process configuration (one-hold) was transferred for training the task of interest. Xu et al. [12] developed a fault diagnosis digital twin using TL by transferring the learned knowledge from simulation to a physical maintenance system. Sun et al. [13] used the abundant historical failure data to develop a deep TL model for estimating the cutting tools' useful life span. The TL-based CNN model proposed by Ferguson et al. [14] was used for detecting manufacturing defects by learning latent features from two large image datasets.

AL and TL have individually shown the capability to improve the performance of predictive models under limited data. Despite the considerable effectiveness, the performance improvement might not be sufficient in specific manufacturing applications, e.g., the aerospace industry with a very low tolerance for modeling error. Furthermore, the data collection procedure is expensive, time-consuming and intrusive. Hence, generating enough data is not always an option. In order to address these limitations, a novel hybrid ML architecture, referred to as Active Transfer Learning (ATL) is proposed in this paper. It integrates AL into TL architecture aiming to further reduce the dependency on expensive and scarce manufacturing data in developing smart and yet cost-effective predictive models.

II. Proposed Method

Fig. 1.a represents a schematic of the proposed ATL model. Initially, the source model is trained using sufficient (inexpensive) source data (Fig. 1.a, step 1). Next, using AL and under the constraints of a limited number of experiments (labeling), a limited but informative target (expensive) dataset is generated (Fig. 1.a, step 2). Finally, by integrating the knowledge gained from the source model through TL and the dataset generated by AL, a reliable model on the task of interest (target) is developed (Fig. 1.a, step 3). The details of the framework are elaborated in the sections below.

A. Active learning

AL is an ML model that learns to choose the data points to be labeled with the goal of maximizing the gained information from the task of interest with the minimum number of queries [10]. It begins with training a model from scratch using an initial set of data (e.g., historical data available in a factory), followed by using an acquisition function to decide which point to be labeled from a pool of unlabeled data. Next, an oracle annotates the data and the labeled data is then added to the training set. Then, the model is re-trained using the updated training set and new unlabeled data is chosen for labeling. This process repeats until it satisfies stopping criteria. AL usually yields a decline in the amount of required data for training a predictive model [7].

In AL, the acquisition function is used to decide what datapoint should be queried next (x^*). The function aims for maximizing the information gained at each iteration of the query process. The formal representation of AL's data query process can be shown as:

$$x^* = \text{argmax}_{x \in \mathcal{D}_{\text{Pool}}} a(x, \mathcal{M}) \quad (1)$$

where $\mathcal{M}$ is the learning model, $\mathcal{D}_{\text{Pool}}$ is the pool of unlabeled data, x is the labeled input data, and $a(x, \mathcal{M})$ denotes the acquisition function [10].

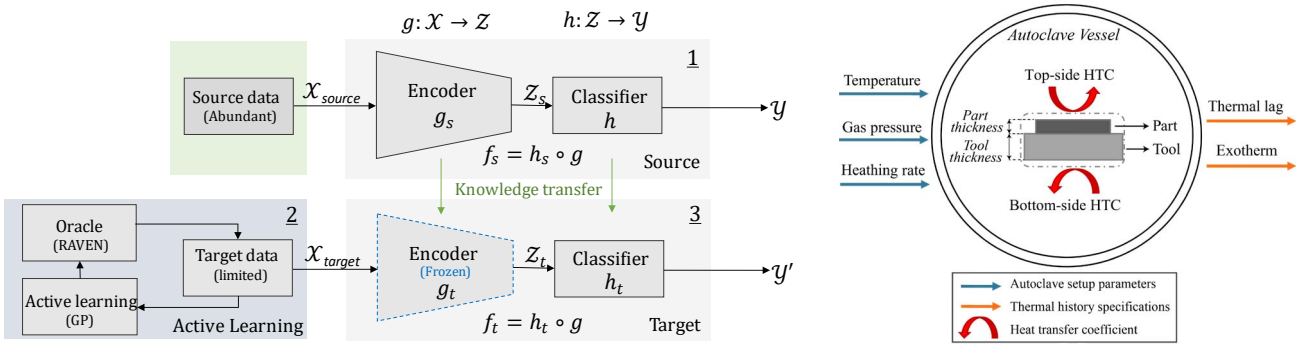


Fig. 1. (a) Schematic of the proposed ATL framework; (b) example of autoclave curing process in aerospace composites manufacturing (adapted from [4]).

In this paper, Gaussian Processes (GP) classification is used as the learning model in the AL settings. GP is a probabilistic inference method which provides prediction uncertainty. GP incorporates the knowledge from the training data with a Gaussian prior to estimate the posterior distribution over functions [15,16]. In particular, with the assumption of noisy data with Gaussian noise ($\mathcal{N}(0,\sigma_n^2)$), the prediction value y^* at test point x^* can be sampled from the joint posterior distribution $p(y^*|x^*,X,f)\sim\mathcal{N}(\bar{y}^*,v^*)$ where:

$$\bar{y}^* = K(x^*,X)\left[K(X,X)+\sigma_n^2 I\right]^{-1}y \tag{2}$$

$$v^* = K(x^*,x^*) - K(x^*,X)\left[K(X,X)+\sigma_n^2 I\right]^{-1}K(X,x^*). \tag{3}$$

Here, $\bar{y}^*$ and v^* are the prediction mean and variance at point x^*, K is the covariance matrix, and X and y denote the input and output of the training set, respectively.

In AL, the unknown function of interest (i.e., mapping between inputs and outputs) can be learned by minimizing the uncertainty about the posterior distribution [17]. The uncertainty can be expressed as entropy H, and in GP, it can be shown that the next point for labeling is the one that maximizes the *differential entropy score* [15]*:*

$$\Delta_j \triangleq H[p(y_j)] - H[p^{new}(y_j)] \tag{4}$$

where $H[p(y_j)]$ is the entropy at point $x_j \in \mathcal{R}$ and $H[p^{new}(y_j)]$ represents the entropy at the same point once the observation is included. Since the entropy of a Gaussian with variance v is $\frac{1}{2}\log(2\pi e v)$, it can be shown that the differential entropy score in GP is:

$$\Delta_j = \frac{1}{2}\log(1 + v_j/{\sigma_n}^2). \tag{5}$$

The monotonic relationship between Δ_j and v_j in (5) suggests that for maximizing the differential entropy score, the point with the *highest variance* must be selected [15].

In order to better investigate the effect of acquisition functions in the performance of AL, in this paper, two other acquisition functions (next to entropy) are also used to quantify the uncertainty of the model. In summary, the functions are as follows.

1) Differential entropy: For a GP model, (1) can be re-written as:

$$x^* = \text{argmax}_{x\in\mathcal{D}_{\text{Pool}}} v. \tag{6}$$

2) Distance from decision boundary: For non-probabilistic classification models, a common heuristic approach to evaluate the confidence of the model predictions is to measure the distance of unlabeled samples from the decision boundary. Then points with the closest distance to the decision boundary are selected as the most uncertain for annotating (by assuming a binary classification with class labels $\{-1,1\}$):

$$x^* = \text{argmin}_{x\in\mathcal{D}_{\text{Pool}}} |\bar{y}| \tag{7}$$

where $|\bar{y}|$ is the distance of estimates from the classification decision boundary [8].

3) Uncertainty: This custom acquisition function combines the heuristic confidence quantification and the uncertainty measures of probabilistic methods (i.e., GP) [8]:

$$x^* = \text{argmin}_{x\in\mathcal{D}_{\text{Pool}}} \frac{|\bar{y}|}{\sqrt{v+{\sigma_n}^2}}. \tag{8}$$

Unlike the two aforementioned methods, this approach considers not only the uncertainties that arise from the probabilistic prediction model but also the information about the labeled data and decision boundary-related uncertainties.

It is worth mentioning that unlike GP regression in which the assumption of Gaussian likelihood, prior and posterior makes the calculations tractable, GP classification requires approximate inference (as opposed to exact inference) as target values are discrete class labels, thus the assumption of Gaussian likelihood is not feasible [15]. In this paper, Laplace approximation is used to approximate the *non-Gaussian posterior* of the GP binary classification [15].

B. Transfer learning

By transferring useful knowledge from related domains, in order to avoid the need for expensive data generation efforts, TL provides a robust learning framework in the case of discrepancies between training and test data [1]. In the context of smart composites manufacturing example (Fig. 1.b), such discrepancies can resemble themselves by introducing a shift in the process configuration, e.g., implementing a new resin system in the raw material, inevitably modifies the underlying mapping between the inputs and outputs, resulting in poor performance of the learned model. TL, however, is able to efficiently adapt to new data distributions and maintain its high mapping accuracy [18]. In a TL framework, a task of interest (*target*) is learned by incorporating the knowledge learned from the *source* model [6]. For neural networks (NN), this can be accomplished e.g. by initializing the weights of the target model with the optimized weights of the source model [4].

Considering the input space $\mathcal{X}$ and label space $\mathcal{Y}$ for an ML model, the objective predictive function $f:\mathcal{X}\to\mathcal{Y}$ can be expressed as a composition of two functions, namely, $f = h\circ g$. The embedding function $g:\mathcal{X}\to\mathcal{Z}$ maps the input space to a latent low-dimensional feature space $\mathcal{Z}$, and $h:\mathcal{Z}\to\mathcal{Y}$ is a predictive function estimating the outputs $\mathcal{Y}$ from the feature space. In a supervised TL setting, the predictive functions of the source and target domains can be defined as $f_s = h_s\circ g_s$ and $f_t = h_t\circ g_t$, respectively (Fig. 1.a). Yosinski et al. [24] showed that in the NNs trained on related datasets, the first layers tend to learn similar low-level features and this is a general phenomenon in all networks regardless of their cost function and the nature of the input data. The last layers of the networks,

however, become more specific to the task they are being trained on (i.e., less suitable and informative for transfer). Inspired by the generality of the features across networks, in the proposed TL framework, the low-level features learned from the source (encoder g_s in Fig. 1.a) is directly transferred to the target network and is kept unchanged (frozen) during the initial fine-tuning steps. However, after transferring the weights of the second block of the source network (classifier h_s in Fig. 1.a) to the target network, they need to be fine-tuned as they are more specific to the source task and thus, need to be adapted to the target task accordingly. Finally, instead of transferring the weights of the output layer (fully task-specific layer), they are randomly initialized to be trained from scratch. Using *sequential unfreezing*, the layers of the target network are then fine-tuned sequentially [4,19]. First, the classifier is fine-tuned using limited data while keeping the target encoder frozen. This step is necessary for transforming the classifier layers to be more specific to the target task (i.e., reducing the magnitude of error gradients caused by non-optimal classifier weights during the backpropagation). Next, the whole network is unfrozen for further fine-tuning, so that all the layers contribute to the learning of the target task. By initializing the weights from a related trained network and fine-tuning them in a sequential manner, the size of the dataset needed for reaching a satisfactory model performance is reduced drastically [6].

III. EXPERIMENTS

In this section, the sample manufacturing case study of interest is described and the architecture of the proposed modeling framework is discussed.

A. Composites Autoclave Manufacturing Case Study

In order to validate the effectiveness of the proposed ATL method, a case study on the autoclave curing process of aerospace composite materials is carried out. Despite superior mechanical properties, manufacturing composite parts is a complex process with high levels of uncertainty [20]. A small deviation from the optimal process configuration can cause undesirable material properties in the cured part. In aerospace manufacturing, thermoset prepregs are often used as the raw material for fabricating the majority of the composite structures. The part is manufactured through a curing process, usually in an autoclave vessel, under the manufacturer's recommended cure cycle (MRCC), i.e., specific process settings that yield optimal product properties. Thermal lag and exotherm, as the thermal history specifications of the composite part during the curing process, are the primary quality metrics for evaluating whether the desirable mechanical, chemical and physical properties are achieved. Exotherm represents the maximum temperature of the part during the curing process and thermal lag measures the deviation between the temperature of the part and cure cycle [3]. Parts that exceed the thresholds for the exotherm and thermal lag are considered as a failed product. This study aims to develop a transferable and cost-effective ML classifier that can accurately predict whether the curing process configuration will provide acceptable parts, while different resin materials may be used. A two-hold cure cycle is investigated (Fig. 2). The curing process begins with a temperature ramp, followed by two isothermal holding steps separated by a second temperature ramp. The process ends with a cool-down step to 20°C with a 3.5°C/min

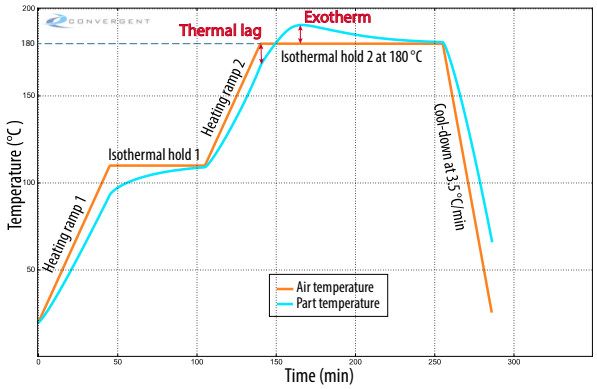


Fig. 2. Two-hold autoclave cure cycle. Thermal history specifications (thermal lag and exotherm) are shown with red arrows.

heating rate. Along with the aforementioned process specifications, the thermal and physical properties of the raw material and tooling also contribute to the process outcome. Furthermore, the thermal history of the part is also governed by the heat transfer coefficient (HTC) between the part and its surroundings (i.e., the air and tooling) which itself is determined by the autoclave settings such as the pressure and velocity of the gas flow [20]. Table I summarized the input variables used in the ATL classifier. It is worth mentioning that for this study, the second hold is assumed to be constant at 180°C.

The goal of the framework is to transfer the knowledge gained from the historical curing process data from one type of thermoset prepreg (source) and transfer it toward learning a model for a new material of interest (e.g., a thermoset prepreg with a new epoxy resin system). The data for the source material is assumed to be abundant, while it is assumed that only a limited number of data can be generated for the target material due to the expensive and time-consuming composites manufacturing processes. The AS4/8552 aerospace-grade composite material that is commonly used in the aerospace industry is selected as the source material. A different resin system, namely, AS4/8551, is chosen as the target.

The data for the case study was generated using RAVEN composites processing simulation software [21]. RAVEN simulates the curing process of composite parts which can be

TABLE I. DESCRIPTION OF CURING PROCESS INPUT VARIABLES

Input variables	Values	
Tool material	6061 Aluminum; AS4/8552 Composite; Invar 36; SAE 1020 Steel	
	Min	**Max**
Tool thickness (mm)	2.5	20
Part thickness (mm)	2.5	20
Heat rate – ramp 1 (°C/min)	1	5
Isothermal hold 1 (°C)	105	125
Heat rate – ramp 2 (°C/min)	1	5
Top-side HTC (W/m^2K)	10	125
Bottom-side HTC (W/m^2K)	10	125

used for evaluating the thermal and material properties throughout the process. The source dataset consists of 40,000 curing process simulations with 75% used for training (which is further divided into two sets: 75% training and 25% validation) and 25% kept for testing. For the target dataset, an initial set of 250 labeled data points were randomly generated. Setting aside 200 samples as the test set, the other 50 data points were used as the initial training set for AL. Separately, a dataset of 4000 simulations is generated for target material (as ground truth) to compare the performance of ML in the presence of abundant data (*upper-bound* performance) with ATL. For both source and target datasets, the class labels (pass/fail) were determined based on whether the thermal lag and exotherm are within the allowable envelope (i.e., exotherm < 5°C and thermal lag < 15°C per MRCC). A full factorial design of the input features was considered as the pool of unlabeled dataset from which samples were selected for the AL training process on target data.

B. *Experimental Design*

For the TL portion of the framework, two NNs (source and target) with similar architecture were implemented using the Keras Python library. The networks consist of an encoder with 1 input layer and 4 hidden layers, and a classifier with 3 hidden layers and 1 output layer. During the first phase of the sequential unfreezing process an Adam optimizer with a learning rate of 0.01 was chosen for a fast initial exploration of the parameters' space. For all subsequent learning phases, a learning rate of 0.001 was adopted. Scaled Exponential Linear Unit (SELU) [22] activation function was implemented in the hidden layers as it showed superior performance over other activation functions tested. The Sigmoid activation function was used in the output layer of the networks.

The AL performance was evaluated using three acquisition functions (section II.A) and a random querying procedure. At each iteration, the top 5 choices (based on the acquisition function output) were selected and passed to the oracle (RAVEN software) for labeling. The process ends when the dataset size reaches the threshold for the maximum allowable number of experiments (here, 200 was selected). To conduct a fair comparison and avoid any potential overfitting, all NNs were equipped with an early stopping regularizer with patience of 20 epochs. For the results in the following section, each ML experiment was performed with 5 repetitions and the average values were reported.

IV. Results and Discussion

The test set accuracy performance of the uncertainty estimation methods in AL (Uncertainty, maximizing differential entropy (STD), and Distance from decision boundary (DB)) is shown in Fig. 3. At each iteration of AL, the obtained labeled data was used to train an ML model (here, a NN with the same architecture of the TL models used) and its performance against the target dataset was reported. The accuracy of the upper-bound model and the performance of the random selection method are also presented. While dominant over the random selection, the three AL methods exhibit a similar performance comparing to each other. As the AL advances and the dataset size increases, the Uncertainty method exhibits slightly better performance. Thus, for developing the TL methods in the next section, the dataset obtained using the Uncertainty method was used.

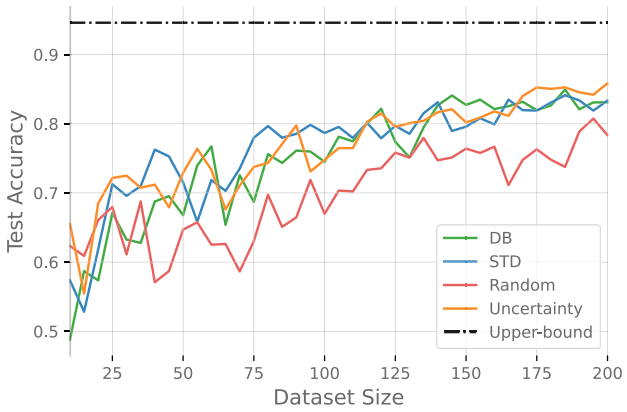
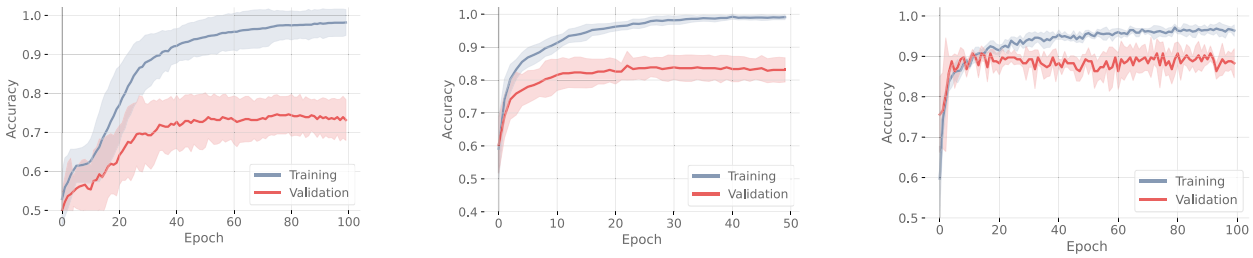


Fig. 3. Test accuracy as a function of dataset size for different AL acquisition functions, random, and upper-bound models.

TABLE II. Performance Comparison of the Developed

Model	Target Test Accuracy
Upper-bound	94.64 %
ATL	**91.9 %**
TL	87.88 %
AL	86.2 %
Random	80.08 %

In order to evaluate the effectiveness of the ATL framework (i.e., the TL model trained on the AL Uncertainty dataset), its performance is compared with different models; namely, a NN trained on random target data (Random), a NN trained on AL data (Pure AL; when using Uncertainty acquisition function; see Fig. 3), and a TL model trained on random target data (Pure TL). Note that in all cases, the test dataset size remained 200. The results are summarized in Table II. ATL outperforms Pure AL and Pure TL models, suggesting that the developed hybrid framework has enhanced the test accuracy more effectively than



(a) (b) (c)

Fig. 4. Model performance (training and validation accuracy values) during the training process for (a) random, (b) AL and (c) ATL.

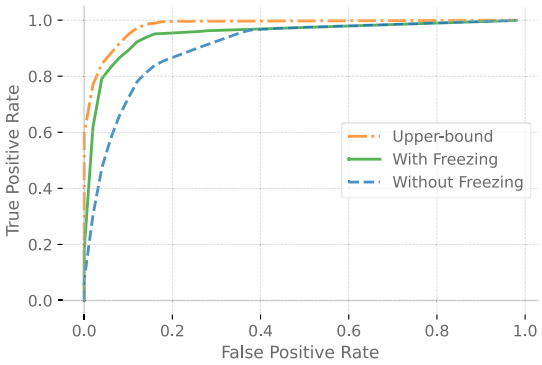


Fig. 5. ROC curves of TL models (with and without freezing) and upper-bound model.

AL or TL alone, in the current case study. In fact, the accuracy of the ATL (trained with 200 data) is very close to that of the upper-bound model (trained on 4000 data). *This comparable performance is achieved by using only 5% of the full data needed for a regular ML model to generate such high fidelity results.*

Fig. 4 compares the training curves of the Random, AL, and ATL models. Combining AL and TL dramatically improves the model accuracy of the validation set and reduces the gap between training and validation curves, making the model more robust against overfitting. Moreover, the performance uncertainty observed in the Random model is also drastically reduced in the ATL as the confidence intervals of the curves become very narrow.

Finally, to demonstrate the effect of sequential unfreezing, the receiver operating characteristic curves (ROC) of the ATL model with and without freezing are shown in Fig. 5. Without Freezing, the general knowledge stored in the weights of the learned encoder is destroyed during training instead of focussing on task-specific layers and randomly initialized output layer first (Area Under Curve (AUC)= 0.88). Sequential freezing blocks the effect of large initial errors and conserves the learned knowledge yielding better performance (AUC = 0.91).

V. Conclusion

In this paper, a hybrid ML framework composed of AL and TL is proposed to reduce the infeasible requirement of large data availability in advanced manufacturing for developing accurate learning models. It was demonstrated that uncertainty-based AL approaches outperform random data collection in terms of model accuracy. Furthermore, the TL model equipped with sequential unfreezing and trained on limited AL data can dramatically improve the model accuracy and come very close to the performance of the upper-bound model trained on abundant data, in the current case study.

Acknowledgment

The authors would like to thank the support and helpful comments from the colleagues at the Composites Research Network (CRN) and the University of British Columbia. This study was financially supported by the New Frontiers in Research Fund (NFRF) of Canada – Exploration Program.

References

[1] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. J Manuf Syst 2018;48:157–69. https://doi.org/10.1016/j.jmsy.2018.01.006.

[2] Kusiak A. Smart manufacturing. Int J Prod Res 2018;56:508–17. https://doi.org/10.1080/00207543.2017.1351644.

[3] Fernlund G, Mobuchon C, Zobeiry N. 2.3 Autoclave Processing. Compr Compos Mater II 2018;2:42–62. https://doi.org/10.1016/b978-0-12-803581-8.09899-4.

[4] Ramezankhani M, Crawford B, Narayan A, Voggenreiter H, Seethaler R, Milani AS. Making costly manufacturing smart with transfer learning under limited data: A case study on composites autoclave processing. J Manuf Syst 2021;59:345–54. https://doi.org/10.1016/j.jmsy.2021.02.015.

[5] Yang H, Kumara S, Bukkapatnam STS, Tsung F. The internet of things for smart manufacturing: A review. IISE Trans 2019;51:1190–216. https://doi.org/10.1080/24725854.2018.1555383.

[6] Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 2010;22:1345–59. https://doi.org/10.1109/TKDE.2009.191.

[7] Settles B. Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences; 2009.

[8] Kapoor A, Grauman K, Urtasun R, Darrell T. Active learning with gaussian processes for object categorization. 2007 IEEE 11th Int. Conf. Comput. Vis., IEEE; 2007, p. 1–8.

[9] Martinez Arellano G, Ratchev S. Towards an active learning approach to tool condition monitoring with Bayesian deep learning 2019.

[10] Shim J, Kang S, Cho S. Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification. IEEE Trans Semicond Manuf 2020;33:258–66.

[11] Yue X, Wen Y, Hunt JH, Shi J. Active Learning for Gaussian Process Considering Uncertainties with Application to Shape Control of Composite Fuselage. IEEE Trans Autom Sci Eng 2020.

[12] Xu Y, Sun Y, Liu X, Zheng Y. A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning. IEEE Access 2019;7:19990–9. https://doi.org/10.1109/ACCESS.2018.2890566.

[13] Sun C, Ma M, Zhao Z, Tian S, Yan R, Chen X. Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing. IEEE Trans Ind Informatics 2019;15:2416–25. https://doi.org/10.1109/TII.2018.2881543.

[14] Ferguson M, Ak R, Lee Y-TT, Law KH. Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning. Smart Sustain Manuf Syst 2018;2:20180033. https://doi.org/10.1520/ssms20180033.

[15] Rasmussen CE. Gaussian processes in machine learning. Summer Sch. Mach. Learn., Springer; 2003, p. 63–71.

[16] Ramezankhani M, Crawford B, Khayyam H, Naebe M, Seethaler R, Milani AS. A multi-objective Gaussian process approach for optimization and prediction of carbonization process in carbon fiber production under uncertainty. Adv Compos Hybrid Mater 2019. https://doi.org/10.1007/s42114-019-00107-6.

[17] Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression with a focus on exploration-exploitation scenarios. BioRxiv 2017:95190.

[18] Yang B, Lei Y, Jia F, Xing S. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mech Syst Signal Process 2019;122:692–706. https://doi.org/10.1016/j.ymssp.2018.12.051.

[19] Chronopoulou A, Baziotis C, Potamianos A. An Embarrassingly Simple Approach for Transfer Learning from Pretrained Language Models. Proc. 2019 Conf. North, Stroudsburg, PA, USA: Association for Computational Linguistics; 2019, p. 2089–95. https://doi.org/10.18653/v1/N19-1213.

[20] Astrom BT. Manufacturing of polymer composites. 1st Ed. London: Routledge; 2018.

[21] RAVEN simulation software 2016. http://www.convergent.ca/products/raven-simulation-software.

[22] Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. Adv. Neural Inf. Process. Syst., 2017, p. 971–80.