A multi-domain mixture density network for tool wear prediction under multiple machining conditions

Gyeongho Kim, Sang Min Yang, Sinwon Kim, Do Young Kim, Jae Gyeong Choi, Hyung Wook Park & Sunghoon Lim

To cite this article: Gyeongho Kim, Sang Min Yang, Sinwon Kim, Do Young Kim, Jae Gyeong Choi, Hyung Wook Park & Sunghoon Lim (06 Dec 2023): A multi-domain mixture density network for tool wear prediction under multiple machining conditions, International Journal of Production Research, DOI: 10.1080/00207543.2023.2289076

To link to this article: https://doi.org/10.1080/00207543.2023.2289076
A multi-domain mixture density network for tool wear prediction under multiple machining conditions

Gyeongho Kim\textsuperscript{a}, Sang Min Yang\textsuperscript{b}, Sinwon Kim\textsuperscript{b}, Do Young Kim\textsuperscript{c}, Jae Gyeong Choa, Hyung Wook Park\textsuperscript{b} and Sunghoon Lim\textsuperscript{a,d,e}

\textsuperscript{a}Department of Industrial Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea; \textsuperscript{b}Department of Mechanical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea; \textsuperscript{c}Department of Mechatronics Engineering, Chungnam National University, Daejeon, Republic of Korea; \textsuperscript{d}Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea; \textsuperscript{e}Industry Intelligentization Institute, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

\textbf{ABSTRACT}

Accurate tool wear prediction is an essential task in machining processes because it helps to schedule efficient tool maintenance and maximise the tool’s useful life, thus contributing to sustainable production via zero defect manufacturing (ZDM). However, there are limitations to existing methods; these cannot be used under multiple machining conditions, which is a common practice. This problem not only hinders accurate tool wear monitoring but also necessitates the use of multiple models, which increases operation and modelling costs. Therefore, the multi-domain learning problem should be addressed to enable tool wear prediction under various machining conditions. To this end, this work presents a novel method, a \textit{multi-domain mixture density network} (MD2N). In particular, a Bayesian learning-based feature extractor is proposed to learn domain-invariant representations. Additionally, an adversarial learning approach is developed to lead the predictive model in learning domain-invariant features. Lastly, a mixture density network-based predictor is used to generate probabilistic tool wear outputs. Experiments that use datasets from real-world milling processes under multiple conditions prove the proposed method’s promising efficacy, with the best mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) of 2.1748, 5.6422, and 0.0350, respectively, indicating the ability to learn multi-domain representations.

\textbf{1. Introduction}

The machining process is one of the most widely adopted techniques in manufacturing. Among various factors that affect the quality of manufactured products, the machine tool’s condition has the greatest impact. In particular, worn tools cause surface roughness and cracking via residual stress. Furthermore, tool wear often leads to tool breakage, which incurs additional costs for tool change and sudden stoppage, leading to decreased productivity. This problem is intensified for various commonly used materials that have desirable characteristics, such as a high strength-to-weight ratio and corrosion resistance. For instance, titanium alloy’s thermal properties cause the material to maintain a high temperature at the tool-material interface, which accelerates the tool wear. Therefore, accurately predicting ongoing tool wear during machining processes is essential not only for determining the proper time to change the tool, which can maximise available tool life, but also for preventing tool breakage, leading to improved overall process efficiency and sustainability. Hence, the development of a high-performing tool wear prediction system can contribute to digital manufacturing transformation (Konstantinidis et al. 2023) by lowering defects and increasing productivity, leading to zero defect manufacturing (ZDM).

Despite its effectiveness, developing an effective tool condition monitoring (TCM) method is a challenging task, because tool wear generation is highly complex to model, both analytically and numerically. Traditionally, mechanistic approaches that employ physical model-based methods and analytical solutions have been proposed. However, due to the difficulty of formulation
and computational cost, they cannot be easily applied and adapted in real-time TCM (Traini, Bruno, and Lombardi 2021). Furthermore, they cannot consider varying tool conditions during machining processes, which hampers flexible applications in diverse domains. On the other hand, data-driven approaches, which utilise sensor measurement data from machining processes to predict tool wear, have been recently proposed. In particular, machine learning (ML) and deep learning (DL)-based methods that exploit multivariate time-series data have shown superior performances in TCM for various machining processes (G. Kim et al. 2022; Shi et al. 2019; Traini, Bruno, and Lombardi 2021; Yang, Zhou, and Tsui 2016).

However, there are several limitations to the current state of data-driven tool wear prediction approaches. First, most existing methods can only be used in a single machining condition because each predictive model is trained on data from a particular machining process that corresponds to a single machining condition (C. Liu et al. 2021). However, in real-world machining practices, multiple machining conditions are widely used to produce the desired quality products. Various factors that determine the machining condition, such as cutting speed (i.e. a material removal rate), lubricant type, tool material, and machining type (e.g. rough, finishing), are adjusted to yield high product quality and process efficiency. In addition, recent machining technologies, which are used to restrain the progression of tool wear using coolants that lower material temperature, increase the diversity of machining conditions that affect tool wear. Despite the prevalence of machining process that uses multiple machining conditions, existing approaches have serious limitations, because multiple models should be used for prediction simultaneously or a model should be switched for use whenever the machining condition is changed (Z. Wang et al. 2021). Therefore, the lack of a capability to perform tool wear prediction under multiple machining conditions not only increases modelling cost but also operation cost due to the employment of as many independent models as the number of machining conditions, thus harming the sustainability of manufacturing processes.

In order to develop a unified model that can predict the degree of tool wear under multiple machining conditions to greatly improve efficiency in model development and deployment, several problems must be addressed. First, data from multiple machining conditions have different distributions because each data comes from a distinct domain. This heterogeneity increases the complexity of tool wear prediction and is the main reason why standard data-driven models perform poorly in multi-domain learning problems (Luo, Wen, and Tao 2017). Therefore, prediction models should learn multi-domain features to perform under multiple machining conditions. In addition, particularly in machining processes with multiple machining conditions, an inverse problem often exists where similar input data have different target values (e.g. degrees of tool wear), which hinders the application of models trained via the principle of a maximum likelihood estimation that produces a conditional average (C. Li and Lee 2019). For instance, in data-driven approaches where real-time sensor measurements are used as inputs to predict tool wear (i.e. output), similar input values may correspond to different output values depending on the machining condition. Due to these problems, tool wear prediction under multiple machining conditions has not been successfully performed in the current literature, and thus most of the existing methods have not proven efficient in realistic machining settings.

This work addresses the aforementioned problems to develop a unified data-driven tool wear prediction method, a multi-domain mixture density network (MD2N) that accounts for multiple machining conditions using various techniques. First, a Bayesian learning-based approach to multi-domain feature learning is developed. Using variational inference (VI), a convolution-based feature extractor learns domain-invariant representations. In addition, an adversarial learning approach is developed to lead the predictive model to learn domain-invariant features by optimising the objective function during training. Lastly, a mixture density network (MDN)-based predictor is used to generate probabilistic tool wear prediction outputs. In the experiments, real-world milling processes with various machining conditions are performed and used to validate the proposed method’s efficacy. Comparisons with existing data-driven methods and state-of-the-art (SOTA) methods in tool wear prediction prove the proposed method’s effectiveness.

The proposed method not only shows superior tool wear prediction performance but also greatly improves efficiency in modelling and deployment, because a single model can work versatility for multiple machining conditions. In particular, the ability to learn domain-invariant features enables the use of the proposed method under various machining conditions. Moreover, although the proposed method can handle data from multiple domains, it requires a short inference time, which enables an effective online prognostic application. The main difference between conventional approaches to multi-domain learning (MDL) in tool wear prediction and the proposed approach is illustrated in Figure 1. The contributions of this work to the literature are as follows.
• A novel tool wear prediction method (i.e. MD\textsuperscript{2}N) is proposed, which proves effective and efficient under various machining conditions.

• Bayesian and adversarial learning approaches are developed to learn domain-invariant features to achieve multi-domain learning for tool wear prediction under multiple machining conditions.

• A mixture density network-based model is proposed to generate probabilistic predictions, and various relevant techniques are presented.

• The superior prediction performance of the proposed MD\textsuperscript{2}N under various machining conditions indicates its practical applicability and the ability to learn multi-domain features and generate multimodal predictions.

2. Theoretical framework

This section outlines a theoretical framework and existing works in the literature pertaining to tool wear prediction, including conventional and the latest data-driven methods. Various ML and DL-based methods as well as their limitations are also discussed. In addition, the concept of multi-domain learning, which aligns with the objective of this work, which is to learn a model with domain-invariant characteristics, is presented, and the related works in the literature are reviewed in detail. Lastly, a specific probabilistic model architecture, named the mixture density network, is introduced, and its variety of usage in existing works is illustrated.

2.1. Data-driven tool wear prediction

Traditional approaches to tool wear prediction rely on analytical solutions and indirect methods given by a concrete formula between the degree of tool wear and the factors that affect it, such as cutting force, speed, and removal rate. This makes finding solutions difficult and also limits real-time TCM (G. Kim, Yang, et al. 2023; García-Ordás et al. 2018). Later, data-driven methods that utilise signals from computer numerical control (CNC) machines and attached sensors enable faster application and improved accuracy in tool wear prediction (Kuo and Kusiak 2019). Some hybrid approaches that combine analytical physics-based models with data-driven methods have also been proposed (Zhang et al. 2022). Conventional data-driven methods like state-space models rely on manual statistical feature extraction techniques coupled with ML-based predictors. Traini, Bruno, and Lombardi (2021) suggest a TCM framework for predictive maintenance using various time and frequency domain features extracted from sensor measurements for the milling process. Yang, Zhou, and Tsui (2016) employ a differential evolution algorithm with an extreme learning machine using statistical features and wavelet features for online tool wear estimation. Zamudio-Ramírez et al. (2020) propose using stray flux signals with a discrete wavelet transform and a fast Fourier transform. Zhu and Liu (2017) develop a TCM based on a hidden semi-Markov model using cutting force signals. However, conventional data-driven approaches that use manual feature extraction require additional computational time that hinders real-time TCM. In contrast, recent approaches that utilise predictive models with high expressive power, such as ML and DL, have shown superior performance in data-driven TCM (C. Liu, Zheng, and Xu 2021).

Recently presented approaches use deep neural network (DNN) architectures that can handle complex multivariate time-series data from machining processes. In particular, a convolutional neural network (CNN), a recurrent neural network (RNN), and their variants, which are suitable for handling sensor signals (e.g. force, audio, vibration), are widely adopted (G. Kim et al. 2021; J. Wang et al. 2019). Xu et al. (2020) present an integrated tool wear prediction model based on multiple DNNs, including a parallel CNN, a deep residual network, and a bi-directional long short-term memory (LSTM), using multisensory input signals. Shi et al. (2019) use a stacked sparse autoencoder to develop a tool wear prediction method based on vibration signals. Hahn and
Mechefske (2021) present a tool wear monitoring method based on a disentangled variational autoencoder (VAE) and a temporal CNN.

Among various DNN architectures, CNN-based methods that have model flexibility and adaptiveness to long time-series data are widely adopted in data-driven TCM. In fact, the convolutional operation is more efficient than other DNN operations, such as recurrent operations, in terms of computational complexity and speed (Kusiak 2020), thus is appropriate in online data-driven TCM. Ma et al. (2021) utilise CNN with a bidirectional gated recurrent unit (GRU) to predict tool wear during the milling process based on force signal. Guo et al. (2021) propose a multi-scale convolutional attention network-based tool life prediction method. Sun et al. (2020) propose employing CNN and LSTM, which use raw sensor signals to predict tool flank wear. This work also proposes a DL-based prediction method using CNN that is suitable for sensor signals. However, compared to most existing methods that can only work under a single machining condition, this work presents a domain-invariant model, which performs under multiple conditions and greatly improves prediction performance and efficiency. From the perspective of ZDM, the proposed method can help achieve sustainable and efficient production for digital manufacturing and production systems (Azamfirei, Psarommatis, and Lagrosen 2023; Psarommatis, May, and Azamfirei 2023). In particular, when coupled with existing ZDM strategies using intelligent quality management frameworks (Konstantinidis et al. 2023), the proposed method can improve product quality and cost by accurate tool wear prediction under multiple machining conditions. In addition, informed decisions based on predictions can greatly enhance process efficiency, leading to ZDM.

### 2.2. Multi-domain learning

Although the task of developing a unified model that can accurately predict tool wear under multiple machining conditions has not been previously well addressed, it is similar to the problem setting of multi-domain learning (MDL) in ML literature. Closely related to domain-invariant learning and domain generalisation, MDL aims to train predictive models that perform on data drawn from multiple domains. The existence of multi-domain leads to domain shift both in input and output spaces, which renders general model training difficult and impairs predictive performance (Zhao et al. 2023). To overcome this problem, the MDL-based method is developed to train models to learn feature representations that are invariant to data-generating distributions. Applying MDL not only provides the representation learning method for DNNs but also improves modelling efficiency, because it eliminates the need to train individual models for every domain. In the task of tool wear prediction, MDL can help develop models invariant to different domains due to changes in machining conditions. Therefore, this work employs an MDL-based approach to developing a tool wear prediction method that can well perform under multiple machining conditions.

Recent MDL approaches have relied on modifying the model architecture and optimisation process to make the feature representations of every domain similar. Rebuffi, Bilen, and Vedaldi (2017) propose using residual adapter modules, which contain domain-agnostic and domain-specific parameters of DNN. This approach is further improved by a universal parameterisation method (Rebuffi, Bilen, and Vedaldi 2018). Bilen and Vedaldi (2017) present a deep information-sharing method using the core parameters of DNN, except those used in normalisation techniques. Berriel et al. (2019) develop an efficient architectural strategy for MDL using budget-aware adapters that learn the most relevant features from novel domains. Y. Li and Vasconcelos (2019) suggest the use of covariance normalisation and an adaptive layer for each domain for MDL, which share the same idea as the MDL approaches mentioned above. Several approaches employ iterative pre-training and fine-tuning strategies for MDL, however, they do not guarantee domain-invariant feature learning (S. He et al. 2020). In addition, MDL approaches based on modifying model architectures limit the application of various DNN types, as they impose constraints on model selection.

Ganin and Lempitsky (2015) propose a seminal approach to MDL that uses a gradient reversal layer (GRL) and standard backpropagation to train models adversarially. During training, the domain classifier loss is maximised to obtain domain-invariant deep feature representations. Xiao et al. (2021) propose MDL using a Bayesian neural network by applying a partial Bayesian treatment to the last two layers with an explicit objective function that leads to learning invariant representations. Gao et al. (2022) also use a VAE to employ a Bayesian learning approach for conditional domain-invariant feature representation by aligning the class-conditional distributions of feature representations. This work develops an MDL-based tool wear prediction method that combines two key ideas; the Bayesian learning scheme and an adversarial learning-based approach (such as Ganin and Lempitsky 2015). To the best of our knowledge, except for similar but different approaches using meta-learning and transfer learning (J. He et al. 2022; Y. Li et al. 2019; Z. Wang et al. 2021; Zhao et al. 2023), no previous work
has been done on an MDL-based approach to tool wear prediction under multiple machining conditions.

2.3. Mixture density network

As mentioned previously, in addition to the heterogeneity of multi-domain data, one reason that standard DNN-based tool wear prediction methods cannot perform well under multiple machining conditions is that the output is forced to converge to the conditional average, while in reality, it could be multimodal (C. Li and Lee 2019). For instance, although the data samples are drawn from different domains, both might show similar patterns in input features while having a dissimilar degree of tool wear. This type of issue, which is called the inverse problem (Mao et al. 2022), brings detrimental effects to model training by simply leading to output average value rather than learning multi-domain features. Therefore, this is one of the reasons that prohibit the direct use of existing tool wear prediction methods in real-world practice under multiple machining conditions. To address this problem, the proposed method exploits a probabilistic model architecture called the mixture density network (MDN). In addition to the aforementioned advantages of the use of MDN in the proposed method, MDN also provides probabilistic prediction and high adaptability, which can be beneficial in other manufacturing domains.

MDN combines DNN with a mixture density model, which assumes that conditional likelihood \( p(y|x) \) consists of multiple plausible distributions. Due to its mixture assumption, MDN has high flexibility to model arbitrary distribution functions. In contrast to general deterministic DNNs, which can only output a single value, MDN can represent multimodal outputs and thus is deemed proper in the task of tool wear prediction under multiple machining conditions. In detail, MDN uses DNN to directly estimate the parameters of the mixture distributions from which the data are assumed to be drawn. For computational efficiency and theoretical guarantee, MDN often assumes a mixture of Gaussians, as expressed in Equation (1), to model the conditional density (H. Kim and Kim 2023). Therefore, the MDN outputs consist of three parts: (1) mixing coefficients of the mixture distribution, (2) mean values for every distribution, and (3) standard deviation values. As shown in Equation (1), the conditional probability distribution \( p(y|x) \) is modelled by Gaussian mixture components \( \mathcal{N}(y|\mu_k(x), \sigma_k(x)) \) with corresponding positive weights \( \pi_k(x) \). In practice, the training of MDN is not trivial due to some numerical instabilities (Choi et al. 2018). For this reason, this work develops practically useful techniques that facilitate the training procedure of MDN. In particular, several novel methods concerning the activation functions and the regularisation are presented in Section 3.

\[
p(y|x) = \sum_{k=1}^{K} \pi_k(x) \mathcal{N}(y|\mu_k(x), \sigma_k(x)).
\]

where \( \sum_{k=1}^{K} \pi_k(x) = 1 \) and \( \pi_k \geq 0 \ \forall \ k \).

MDN is widely used in ML and DL applications, where probabilistic predictive distributions should be generated to account for various hypotheses, which are often multimodal. C. Li and Lee (2019) use MDN on top of a CNN-based feature extractor to generate multiple hypotheses in a three-dimensional human pose estimation. Saunders, Camgoz, and Bowden (2021) use a transformer with MDN to produce various sign language poses. Ji, Ameri, and Cho (2021) employ MDN architecture to represent probabilistic effects on the prediction of non-conformance rates. MDN is also applied in other DL tasks, such as stereo matching (Tosi et al. 2021) and autonomous vehicle (Choi et al. 2018), due to its ability to model complex target distributions. However, despite its effectiveness in modelling multimodal and probabilistic outputs, MDN has not been commonly used in intelligent production domains. Therefore, another primary novelty of this work comes from the proper application and adaptation of MDN in a tool wear prediction task, which can be further developed in other manufacturing domains. In particular, the proposed method employs MDN to generate potential multimodal predictions using multi-domain features from various machining conditions.

3. Proposed method

3.1. Model architecture

This work proposes a novel tool wear prediction method named the multi-domain mixture density network (MD2N) that can be used under multiple machining conditions. To this end, multiple components are developed and combined to learn multi-domain features and generate multimodal predictions. In particular, a Bayesian learning-based feature extractor is proposed to learn domain-invariant representations. The MDN-based predictor is then used to generate a multimodal predictive distribution for tool wear. In addition, to achieve domain invariance, an adversarial learning-based auxiliary domain classifier with GRL is used to guide model training. An overall MD2N model architecture, which is illustrated in detail in the following sections, is shown in Figure 2.
3.2. Bayesian domain-invariant feature extractor

This work proposes a Bayesian domain-invariant feature extractor (BDIFE) that maps input data to representations that can be used for multiple domains. Based on the findings that Bayesian learning is effective in MDL (Gao et al. 2022; Xia et al. 2021), this work postulates that it will function as an ensemble of domain-specific representations (i.e. Bayesian model averaging), leading to probabilistically modelling multi-domain features. To this end, BDIFE employs Bayesian convolutional operations to extract domain-invariant features from the input data. A VI-based approximate inference (i.e. a reparameterisation technique) is used to realise Bayesian convolutions. Given input data \( x \in \mathbb{R}^{L \times C} \), where \( L \) and \( C \) are the length of the multivariate time-series and the number of channels (i.e. variables), respectively, BDIFE maps the input to the feature representation \( g \in \mathbb{R}^{L' \times C'} \) that is assumed to be domain-invariant by applying multiple Bayesian convolutions, batch normalisation, and pooling. Besides, convolutional operations are organised following a residual block structure (K. He et al. 2016).

Denoting the Bayesian parameters of BDIFE as \( w \), BDIFE \( f_w \), with a prior \( p(w) = \mathcal{N}(0, I) \), is optimised during training to generate domain-invariant representations. This work assumes that during training BDIFE would learn to generate diverse plausible representations for multiple domains via Bayesian inference. Bayesian learning provides a better generalisation ability to various domains because rather than learning deterministic values, Bayesian treatment allows model parameters to capture a diverse range of representations in a probabilistic manner (Ouyang et al. 2016). As mentioned above, since VI is adopted, a factorised Gaussian distribution is used as a variational posterior \( q(w|\theta) = \mathcal{N}(\mu, \rho^2) \), as expressed in Equation (2). Further, \( w \) is reparametrised as \( \theta = (\mu, \rho) \), leading to evidence lower bound (ELBO) objective of Equation (3). Therefore, during training, parameters \( \mu \) and \( \rho \) are optimised that determine BDIFE parameters \( w \), as shown in Equation (2). For a more detailed derivation of ELBO, refer to Kingma and Welling (2013).

\[
\begin{align*}
  w &= \mu + \rho \cdot \epsilon, \\
  \epsilon &\sim \mathcal{N}(0, I). \\
  L_VI &= KL(q(w|\theta)||p(w)) - E_q(w|\theta)\left[\log(p(y|x, w))\right] \\
  &\approx \sum_{i=1}^{n} \log(q(w^{(i)}|\theta)) - \log(p(w^{(i)}))) \\
  &- \log(p(y|x, w^{(i)}))).
\end{align*}
\]

On top of the extracted representations from the main BDIFE operations, including Bayesian convolutions, a squeeze-excitation (SE) block (Hu, Shen, and Sun 2018) is used to let the model adaptively select informative information by modelling dependencies between features. Hence, the final representation of BDIFE is denoted as \( g = f(x; w) \). A detailed architectural structure of BDIFE is illustrated in Figure 3.

3.3. MDN-based tool wear predictor

Using the extracted feature representations from BDIFE and SE block that are assumed to be domain-invariant due to the benefits of Bayesian learning, an MDN-based predictor is constructed to generate a multimodal predictive distribution of tool wear. The MDN outputs a Gaussian mixture distribution as expressed in Equation (1). In particular, the parameters that constitute the mixture distribution are produced by MDN \( f_\phi(x) = \{\pi_k, \mu_k, \sigma_k\}_{k=1}^K \) with the learnable parameters \( \phi \), where \( k \) represents the mixture component. The optimisation objective of the MDN-based predictor in MD\(^2\)N is expressed in
Despite MDN’s practical and theoretical effectiveness, its training requires careful treatment of the components. In particular, numerical instabilities often arise due to the likelihood function (Choi et al. 2018). In this work, several techniques are adopted to stabilise training. For an activation function inside the MDN-based predictor, a modified exponential linear unit is used, as expressed in Equation (5). In addition, Xavier initialisation and L2 regularisation are applied to every MDN parameter. To facilitate learning, two additional regularisation techniques are also developed, as shown in Equations (6) and (7). The use of Equation (6) regularises the negative entropy throughout MDN training, which prohibits the generated mixing coefficients from being overly imbalanced. Since the loss is to be minimised during model optimisation (i.e. training), the value of Equation (6) decreases as the distribution of mixing coefficients becomes more uniform (i.e. higher entropy). Equation (7) constrains MDN from producing too large standard deviation values for every mixture component, by directly regularising the magnitude of $\sigma_k(x_i)$.

\[
L_{\pi} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} (\pi_k(x_i) \log(\pi_k(x_i))).
\]  

\[
L_{\sigma} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} (\sigma_k(x_i))^2.
\]  

As mentioned above, during inference, the tool wear prediction output of MDN consists of a mixture distribution with each mixture component and its corresponding distribution parameters (i.e. mean and standard deviation). In this work, one of the most widely used heuristics in MDN literature to use the mixture component that has the highest mixing coefficient $k^* = \arg\max_k \pi_k$ is selected to yield the final tool wear prediction $N(\mu_{k^*}, \sigma_{k^*}^2)$.

### 3.4. Auxiliary domain classifier

As mentioned above, BDIFE is expected to achieve MDL via Bayesian learning and MDN. In addition, this work further enhances MDL ability by using adversarial learning-based GRL (Ganin and Lempitsky 2015). During training, an auxiliary domain classifier (ADC), which
predicts the domain class $t$ given extracted representations from BDIFE, is constructed and optimised. Importantly, ADC is optimised in an adversarial manner to maximise the training objective as shown in Equation (8), to learn domain-invariant representations. While cross-entropy loss (shown in Equation (8)) is used for ADC in this work, other types of loss functions for classification can also be employed. This idea of guiding model training adversarially for the sake of MDL is realised by GRL, which converts the sign of gradient for certain parts of BDIFE and MDN as expressed in Equation (10).

$$\Delta w = -\gamma \left( \frac{\partial L_{MDN}}{\partial w} + \lambda_1 \frac{\partial L_\pi}{\partial w} + \lambda_2 \frac{\partial L_\sigma}{\partial w} - \lambda_3 \frac{\partial L_{ADC}}{\partial w} \right)$$

$$\Delta \phi = -\gamma \left( \frac{\partial L_{MDN}}{\partial \phi} + \lambda_1 \frac{\partial L_\pi}{\partial \phi} + \lambda_2 \frac{\partial L_\sigma}{\partial \phi} \right)$$

$$\Delta \psi = -\gamma \left( \lambda_3 \frac{\partial L_{ADC}}{\partial \psi} \right)$$

where $\gamma$ is a learning rate and the KL divergence term is omitted for conciseness.

4. Experiments

4.1. Milling setup

In this work, the Ti-6Al-4V milling process is conducted under various machining conditions to obtain real-world data that can validate the efficacy of MD$^2$N. Detailed information on the milling experiments is provided in Table 1. The five-axis CNC machine (HTC-1000, HNK Co.) and a dynamometer as a cutting force sensor (Kistler, 9257B) are used. The machining distance of 100 mm is denoted as one pass in the experiments. The tool holder’s diameter (R220.291-0050-06.5A, Seco Tools Co.) is set at 50 mm, and the tool insert’s relief angle (RPHT1204M0T6-M13, Seco Tools Co.) is set at 11°. To further increase domain diversity (i.e. machining conditions), a cryogenic and minimum quantity lubricant (CryoMQL) is applied in addition to wet settings, where liquid nitrogen is sprayed on the cutting tool and the work material, leading to different trends in tool wear and sensor measurement. The milling setup is shown in Figure 4. The schematic diagram of the experimental setup is illustrated in Figure 5.

4.2. Data description

From the milling experiments, eight distinct datasets with different machining conditions are collected. Among a variety of sensor signals, including force, vibration, and audio, that can be used to monitor the milling process,

Table 1. Milling experiment conditions.

<table>
<thead>
<tr>
<th>Experiment number</th>
<th>Setting</th>
<th>Cutting speed (m/min)</th>
<th>Feed (mm/tooth)</th>
<th>Axial depth (mm)</th>
<th>Radial depth (mm)</th>
<th>Material removal rate (mm$^3$/ min)</th>
<th>Number of pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wet</td>
<td>60</td>
<td>0.6</td>
<td>1.0</td>
<td>18.0</td>
<td>5490.0</td>
<td>[1,5,10,15,20]</td>
</tr>
<tr>
<td>2</td>
<td>Wet</td>
<td>80</td>
<td>0.6</td>
<td>1.0</td>
<td>18.0</td>
<td>5500.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Wet</td>
<td>60</td>
<td>0.8</td>
<td>1.0</td>
<td>18.0</td>
<td>5500.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Wet</td>
<td>60</td>
<td>0.6</td>
<td>1.3</td>
<td>18.0</td>
<td>5363.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CryoMQL</td>
<td>60</td>
<td>0.6</td>
<td>1.0</td>
<td>18.0</td>
<td>5490.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CryoMQL</td>
<td>80</td>
<td>0.6</td>
<td>1.0</td>
<td>18.0</td>
<td>5500.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CryoMQL</td>
<td>60</td>
<td>0.8</td>
<td>1.0</td>
<td>18.0</td>
<td>5500.0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CryoMQL</td>
<td>60</td>
<td>0.6</td>
<td>1.3</td>
<td>18.0</td>
<td>5363.0</td>
<td></td>
</tr>
</tbody>
</table>
cutting force is used in this work for several reasons. First, cutting force during the milling process increases as the tool wear progresses, thus helping to predict tool wear in an online manner. Second, compared to vibration and audio signals, force signals can be used to monitor machining safety and stability and therefore have a higher real-world applicability. In addition, as force signals are measured with sensors attached to the work stage (as shown in Figure 5), it is less affected by external noise, unlike audio signals. Moreover, the force signals measured in three different axes can be used in other machining processes, hence having higher versatility. The input data used in this work consist of cutting forces sampled at a frequency of 1000 Hz, and the descriptive statistics are shown in Table 2. Figure 6 shows the exemplary sensor measurement input data.

### 4.3. Tool wear degree

For each pass of the milling process, the degree of tool wear is measured with a microscope (KEYENCE, VK-X200). The actual tool wear measurements from Experiments 2 and 6 are shown in Figure 7. To obtain a continuous degree of tool wear as true target values, the tool wear equation (shown in Equation (11)) is used. In particular, Usui’s tool wear model (Usui, Shirakashi, and Kitagawa 1984) is employed, where tool wear $VB$ is associated with machining time $T$. Based on the Levenberg–Marquardt (LM) method, the fitted equation is used to estimate the degree of tool wear between the machining distances at which the exact wear degree is measured. Detailed equations used in the LM method are provided in Equations (12), (13), and (14).

\[
VB = d(a + bT)^{-1}. \tag{11}
\]

\[
p_{k+1} = p_k - (J^T_r J_r + \mu_k \text{diag}(J^T_r J_r))^{-1} J^T_r r(p_k), \quad k \geq 0 \tag{12}
\]

\[
J_r(p) = \begin{bmatrix}
\frac{\partial r_1(p)}{\partial p_1} & \cdots & \frac{\partial r_1(p)}{\partial p_m} \\
\vdots & \ddots & \vdots \\
\frac{\partial r_n(p)}{\partial p_1} & \cdots & \frac{\partial r_n(p)}{\partial p_m}
\end{bmatrix} \tag{13}
\]

\[
r(p) = \begin{bmatrix}
r_1(p) \\
r_2(p) \\
\vdots \\
r_n(p)
\end{bmatrix} = \begin{bmatrix}
y_1 - f(x_1, p) \\
y_2 - f(x_2, p) \\
\vdots \\
y_n - f(x_n, p)
\end{bmatrix} \tag{14}
\]

### 4.4. Data preprocessing

Standardisation is applied to each variable to normalise the data. Compared to conventional data-driven approaches, raw time-series signals are used in this work to enable online prediction using temporal information inherent in the input data. In addition, a sliding window method is applied to transform raw data into a suitable format that can be used as inputs. To reduce training time and downsample the data, a sliding window of size 400 with a stride of 200 is used (Ma et al. 2021).
### Table 2. Descriptive statistics of datasets.

<table>
<thead>
<tr>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
<th>Dataset 5</th>
<th>Dataset 6</th>
<th>Dataset 7</th>
<th>Dataset 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
<td>Max</td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
</tr>
<tr>
<td>Variable</td>
<td>Fx</td>
<td>Fy</td>
<td>Fz</td>
<td>Fx</td>
<td>Fy</td>
<td>Fz</td>
<td>Fx</td>
</tr>
<tr>
<td></td>
<td>81.0290</td>
<td>45.0864</td>
<td>64.7218</td>
<td>78.1679</td>
<td>44.0454</td>
<td>103.4832</td>
<td>95.0169</td>
</tr>
</tbody>
</table>

**Figure 6.** Visualisation of sensor measurements.

### 4.5. Evaluation metrics

Since the target variable (i.e. tool wear degree) in this work is continuous, the appropriate evaluation metrics used in the regression analysis are employed to validate the experimental results. In particular, mean absolute error (MAE) and root mean squared error (RMSE), which measure the average magnitude of prediction error in linear and quadratic manners, respectively, are used. In addition, mean absolute percentage error (MAPE), which is one of the most popular regression metrics due to its robustness to outliers (Prestwich et al. 2014), is also used. The formulas for the metrics (i.e. MAE, RMSE, MAPE) are defined as expressed in Equation (15).

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|.$$  

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}.$$  

$$MAPE = \frac{100}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{|y_i|}.$$  

### 4.6. Implementation details

The experiments in this work were conducted with a TensorFlow library on a graphical processing unit. During model optimisation, Adadelta is used with a batch size of 256 and cosine decay as a learning rate scheduler with an initial learning rate of 0.01. In addition, early stopping is used to prevent model overfitting. In terms of computational time, training MD^2N for a single epoch takes around 40 seconds, whereas a prediction for a single sample at test time takes 0.0002 seconds on average. For a
fair evaluation, all data are divided into training, validation, and test sets, with a ratio of 70%, 10%, and 20%, respectively. Five trials with different random seeds are performed to obtain the error bars of the results.

5. Results

This work provides empirical results on tool wear prediction performance in three parts. First, experimental data from the wet setting with four different machining conditions are used. Second, data from the CryoMQL setting with the same number of machining conditions are used. In the last case, data from all eight machining conditions are used. In each part, training data and test data are composed of multi-domain data, thus the tool wear prediction results indicate whether the predictive model can accomplish MDL.

5.1. Evaluation of wet setting data

Using Datasets 1, 2, 3, and 4 collected in the wet setting, the tool wear prediction performance of the proposed MD$^2$N is validated. Using the same training set consisting of the four machining conditions, four different test sets are used to verify whether MD$^2$N has learned domain-specific representations for tool wear prediction. In addition, a single combined test set containing all machining conditions is also used to verify whether MD$^2$N has achieved MDL. As provided in Table 3, MD$^2$N shows good prediction performance in all four machining conditions even when trained on multi-domain data. In particular, even when trained on the dataset composed of multi-domain input features, the proposed MD$^2$N shows test MAPE from 0.04 to 0.07 for each input domain.

5.2. Evaluation of CryoMQL setting data

Tool wear prediction results for MD$^2$N using the training set collected in the CryoMQL setting with four different machining conditions are provided in Table 3. Similar to the results of the wet setting, the proposed MD$^2$N provides an accurate prediction performance on all four different machining conditions, showing that it has successfully learned domain-invariant representations, which can be used for multiple machining conditions simultaneously. The magnitude of the error for the CryoMQL setting has shown to be smaller than that of the wet setting because the ground truth tool wear degree is smaller due to the use of CryoMQL.
Table 3. Prediction performance for wet and CryoMQL setting data.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Test dataset</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAE</td>
</tr>
<tr>
<td>Wet</td>
<td>1</td>
<td>2.6772</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.4948)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.6201</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.4353)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.3429</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±1.1129)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.4466</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.3091)</td>
</tr>
<tr>
<td></td>
<td>All (1 + 2 + 3 + 4)</td>
<td>3.5198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.3430)</td>
</tr>
<tr>
<td>CryoMQL</td>
<td>5</td>
<td>1.2266</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.1676)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1.3638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.5127)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3.4514</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±2.1667)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.1237</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.4187)</td>
</tr>
<tr>
<td></td>
<td>All (5 + 6 + 7 + 8)</td>
<td>1.7204</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(±0.7231)</td>
</tr>
</tbody>
</table>

5.3. Evaluation of wet and CryoMQL setting data

Next, MD2Ni is trained and tested using a combined dataset containing all datasets from eight different machining conditions (including both wet and CryoMQL settings). As shown in Table 4, the proposed MD2Ni still performs well under eight machining conditions. Furthermore, compared to previous results in Table 3, although the number of machining conditions has doubled, the tool wear prediction performance does not change much and even increases in some cases (i.e. Datasets 1, 2, 3, and 7). In terms of MAPE, MD2Ni shows test values from 0.02 to 0.06, which indicates that it performs well on various test domains as well. Therefore, the proposed MD2Ni’s capability of learning multi-domain features to perform tool wear prediction under multiple machining conditions has been empirically verified.

The prediction results of MD2Ni trained in eight machining conditions, for every dataset, are visualised in Figure 8. For all test datasets, the proposed MD2Ni shows good prediction performance, well following the increasing trend of ongoing tool wear during the milling process. For some datasets, however, the prediction results have shown peaky points where the error values are larger than other time points. It might be caused by the increment of tool wear and the reduction of tool stiffness that generates abrupt vibrations. Moreover, as the machining process proceeds, the cutting tool might have adhered to the work material due to the additional friction force (e.g. ploughing force), which negatively affects the stable data acquisition.

Using the MD2Ni trained in eight datasets under wet and CryoMQL settings, the learned latent space is visualised with t-distributed stochastic neighbour embedding (t-SNE) (Van der Maaten and Hinton 2008). In particular, learned feature representations from the intermediate layer of MD2Ni, right after BDIFE $g_w$, are extracted. Representations from the layer that comes after BDIFE are used to visualise the latent space because its outputs are fed into both an MDN-based predictor and ADC. Five thousand test samples from each machining condition are used to generate feature representations from MD2Ni and are later transformed into t-SNE embeddings for visualisation. The latent space before training and its transformation during training is illustrated in Figure 9. The latent space of MD2Ni before training seems to have strong separability among different machining conditions, showing little overlapping
Figure 8. Tool wear prediction performance of MD^2N on all datasets under wet and CryoMQL settings. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3. (d) Dataset 4. (e) Dataset 5. (f) Dataset 6. (g) Dataset 7 and (h) Dataset 8.
regions between different datasets. However, as training proceeds, the latent embeddings gradually mingle together, decreasing the separability among machining conditions. In addition, after training, the embeddings become more spread out with some kind of structured patterns and less separation between machining conditions. Therefore, the visualisation of the MD²N’s latent space indicates its ability of MDL, with domain-invariant latent feature representations that enable tool wear prediction under multiple machining conditions.

5.4. Comparison with existing methods

Using the same experimental setting as in previous results, the performance of MD²N is compared with existing data-driven tool wear prediction models, including support vector regression (SVR), random forest (RF), LSTM, GRU, and CNN. As shown in Table 5, in all three cases: (1) four machining conditions in the wet setting, (2) four machining conditions in the CryoMQL setting, and (3) all eight machining conditions, the proposed MD²N yields a much higher prediction performance. The prediction results indicate that the proposed MD²N has more successfully learned multi-domain features for tool wear prediction under multiple machining conditions, compared to existing methods. In particular, the MD²N components, including BDIFE, MDN-based predictor, and ADC, have empirically shown effectiveness.

5.5. Comparison with state-of-the-art methods

In addition, the MD²N performance is compared to existing SOTA tool wear prediction methods (Hahn and Mechefske 2021; Sun et al. 2020; J. Wang et al. 2019). For the sake of reproducibility, existing methods with publicly available source codes are employed for performance comparison. Although the SOTA methods have shown high predictive performance in tool wear prediction tasks, they have not been used under multiple machining conditions. The experimental results indicate that the
Table 5. Performance comparison with existing models.

<table>
<thead>
<tr>
<th>Dataset Model</th>
<th>Wet</th>
<th>CryoMQL</th>
<th>Wet + CryoMQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>SVR</td>
<td>17.6989</td>
<td>22.2290</td>
<td>0.2897</td>
</tr>
<tr>
<td>(±0.0795)</td>
<td>(±0.0869)</td>
<td>(±0.0003)</td>
<td>(±0.0561)</td>
</tr>
<tr>
<td>RF</td>
<td>11.0783</td>
<td>15.3236</td>
<td>0.1947</td>
</tr>
<tr>
<td>(±0.0344)</td>
<td>(±0.0381)</td>
<td>(±0.0005)</td>
<td>(±0.0211)</td>
</tr>
<tr>
<td>LSTM</td>
<td>11.1720</td>
<td>13.7242</td>
<td>0.2730</td>
</tr>
<tr>
<td>(±1.1396)</td>
<td>(±1.2325)</td>
<td>(±0.1679)</td>
<td>(±1.4033)</td>
</tr>
<tr>
<td>GRU</td>
<td>9.2271</td>
<td>11.9802</td>
<td>0.1589</td>
</tr>
<tr>
<td>(±0.8675)</td>
<td>(±0.6444)</td>
<td>(±0.0133)</td>
<td>(±0.5707)</td>
</tr>
<tr>
<td>CNN</td>
<td>16.5959</td>
<td>18.8952</td>
<td>0.2875</td>
</tr>
<tr>
<td>(±0.0824)</td>
<td>(±0.0584)</td>
<td>(±0.0013)</td>
<td>(±0.0070)</td>
</tr>
<tr>
<td>MD²N(proposed)</td>
<td>3.5198</td>
<td>7.1057</td>
<td>0.0570</td>
</tr>
<tr>
<td>(±0.3430)</td>
<td>(±0.5456)</td>
<td>(±0.0061)</td>
<td>(±0.7231)</td>
</tr>
</tbody>
</table>

Table 6. Performance comparison with state-of-the-art methods.

<table>
<thead>
<tr>
<th>Dataset Model</th>
<th>Wet</th>
<th>CryoMQL</th>
<th>Wet + CryoMQL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>J. Wang et al. (2019)</td>
<td>17.4631</td>
<td>18.6138</td>
<td>0.2669</td>
</tr>
<tr>
<td>(±0.4771)</td>
<td>(±0.6608)</td>
<td>(±0.0550)</td>
<td>(±0.0352)</td>
</tr>
<tr>
<td>Sun et al. (2020)</td>
<td>10.4252</td>
<td>13.0826</td>
<td>0.1857</td>
</tr>
<tr>
<td>(±0.6919)</td>
<td>(±0.6727)</td>
<td>(±0.0173)</td>
<td>(±0.3657)</td>
</tr>
<tr>
<td>Hahn and Mechefske (2021)</td>
<td>9.5482</td>
<td>12.2764</td>
<td>0.1653</td>
</tr>
<tr>
<td>(±0.1381)</td>
<td>(±0.0980)</td>
<td>(±0.0040)</td>
<td>(±0.2353)</td>
</tr>
<tr>
<td>MD²N(proposed)</td>
<td>3.5198</td>
<td>7.1057</td>
<td>0.0570</td>
</tr>
<tr>
<td>(±0.3430)</td>
<td>(±0.5456)</td>
<td>(±0.0061)</td>
<td>(±0.7231)</td>
</tr>
</tbody>
</table>

proposed MD²N outperforms existing SOTA methods as provided in Table 6. In particular, the superior performance of the proposed method empirically proves that MD²N is capable of learning domain-invariant representations under multiple machining conditions.

6. Discussion

From the perspective of machining analysis, the precision of tool wear prediction by the proposed MD²N under multiple machining conditions, including wet and CryoMQL settings, is desirable. In particular, for difficult-to-cut materials (e.g. titanium), for which various machining conditions and types of lubricants are applied to meet the desired machining quality, the proposed method can greatly improve the efficiency of the machining process. The online prediction of tool wear using the proposed method can also contribute to the overall productivity of the machining process, as a single model can be used in real-time so that human interruption is minimised during production processes.

As mentioned above, titanium (i.e. Ti-6Al-4V) has poor thermal properties, such as low specific heat and low thermal conductivity, which are directly related to the increment of machining temperature. In the CryoMQL setting, liquid nitrogen (LN₂) and MQL help mitigate the heat generated during milling. In addition, their use leads to a reduction of friction and increased lubricant ability in the tool-chip and tool-work material interface, thus leading to a decrease in tool wear in the case of CryoMQL settings. This is shown in Figure 8, where the magnitude of tool wear degree is smaller in general for CryoMQL settings than wet settings. Specifically, in the case of Datasets 2 and 6, the difference in tool wear degree is larger than in other conditions because, as the machining speed increased, the generation of machining heat increased, which led to the increment of LN₂ effects. However, in the case of Datasets 7 and 8, the degree of tool wear is not considerably reduced. The increment of feed per tooth and axial depth is negatively induced by the large machining area, which is a high-contact area. Hence, increasing the cutting speed and decreasing the contact area of tool-work material is a way to maximise the effects of CryoMQL.

From the perspective of a data-driven tool wear prediction approach, the proposed method performs prediction solely based on real-time sensor measurements. Because a mechanistic model is not employed, the proposed method has higher flexibility and adaptability in real-world machining practices. However, as it is a purely data-driven method, when there is noise or abnormalities in sensors in real industrial applications, the quality of tool wear prediction could be affected (García-Ordás et al. 2018). For instance, when sensor precision degrades and measurement noise arises, the input data for the proposed method change. Therefore, the prediction results would differ. While the employed Bayesian learning-based approach to multi-domain feature learning can
mitigate the effects of sensor noise and abnormalities on the tool wear prediction performance to a certain degree, a severe amount of noise would potentially deteriorate the proposed method’s performance. For these reasons, using other factors of the machining environment that are less influenced by sensor abnormalities, which are contingent upon the state and geometric characteristics of the work material, can improve robustness to sensor noise and precision of prediction (Dreyfus et al. 2022). In addition, incorporating aleatoric and epistemic uncertainty for modelling and prediction can also help develop a robust tool wear prediction method under inherent sensor noise and abnormalities.

From the perspective of sustainable manufacturing and production, the proposed method can contribute to achieving ZDM in various ways. First, due to accurate online tool wear prediction performed by the proposed MD$^2$N, the tool costs, which account for a large part of manufacturing costs, can be reduced. In addition, the efficiency of tool replacement decisions can be improved. For instance, not only the tool can be maximally utilised but also the tool breakage that damages machinery can be prevented in advance. This results in the reduction of production defects and waste, thus contributing to ZDM. Furthermore, for industry decision-makers, the proposed MD$^2$N has advantages in reducing modelling and maintenance costs, because a single model is required for tool wear prediction under various machining conditions. In particular, due to the proposed method’s ability to learn domain-invariant features, the use of multiple models is not necessitated. In addition, when coupled with relevant digital technologies in modern manufacturing ecosystems, such as machine vision (Konstantinidis et al. 2023) and digital twins (Konstantinidis et al. 2022), the proposed method can be broadly applied to other types of production domains. Therefore, the aforementioned advantages of the proposed method lead to ZDM in the era of Industry 4.0 (Psarommatis, May, and Azamferei 2023).

7. Conclusion and future work

This work presents a novel tool wear prediction method (i.e., MD$^2$N) that can perform well under multiple machining conditions during milling process. Considering real-world machining practices, where various machining conditions are used interchangeably and that require learning domain-invariant feature representations, the proposed MD$^2$N can greatly improve efficiency and productivity. In particular, MD$^2$N employs a Bayesian learning-based feature extractor (i.e., BDIFE) to learn multi-domain representations for MDL. MDN-based predictor is then used to generate multimodal predictive distribution for ongoing tool wear, which might also handle potential inverse problems in multiple machining conditions. In addition, a GRL-based auxiliary classifier (i.e., ADC) is used in an adversarial learning manner to further enhance the MDL capability of MD$^2$N.

Extensive experiments using datasets from real-world milling processes conducted under multiple machining conditions prove the effectiveness of the proposed MD$^2$N. Using datasets collected under wet and CryoMQL settings consisting of eight different machining conditions, the proposed MD$^2$N’s efficacy in tool wear prediction under multiple machining conditions has been verified. Compared to existing data-driven tool wear prediction methods and SOTA methods in the literature, the proposed method shows superior prediction performance. The results indicate that MD$^2$N can effectively capture both domain-invariant and domain-specific representations during training. Furthermore, considering that MD$^2$N takes 0.0002 seconds (on average) for inference, a real-time prognostic application seems feasible.

One of the most prospective research areas in which the proposed method can be integrated into sustainable production domains is ZDM (Azamferei, Psarommatis, and Lagrosen 2023). Recently, the concept of ZDM, which aims to ensure product and process quality by minimising defects using advanced data-driven technologies, has arisen (Psarommatis et al. 2020). In particular, quality improvement based on digital technologies to improve the sustainability of production systems is one of the most widely studied topics both in academia and industry (Psarommatis et al. 2022). In this regard, the proposed method can contribute to achieving ZDM not only by reducing tool costs but also by improving product quality using accurate tool wear prediction results. In particular, the application of the proposed method in real-world machining processes would help determine proper tool change time and prevent unexpected process stoppage due to tool breakage, thus developing efficient tool maintenance strategies with lower defects (Psarommatis, May, and Azamferei 2023). Combining the proposed method with existing ZDM strategies in various aspects, such as detection, repair, and prevention (Psarommatis et al. 2020), would also bring improvement in sustainable manufacturing. For instance, using predictions from the proposed method and other virtual defect detection methods can enable more efficient operation for manufacturing systems. Furthermore, using a single model for multiple machining conditions could lower operation costs and improve productivity in practice. Especially for decision-making in the manufacturing industry, this would enhance production efficiency and sustainability. From the perspective of a prevention strategy for ZDM,
using a single model can even reduce the possibility of malfunction and maintenance costs.

It is worth noting that the objective of this work is different from that of the literature on tool wear prediction under a novel machining condition, in which domain adaptation and transfer learning techniques can be applied (Oh et al. 2022). In particular, a domain adaptation setting, where a few supervisions are provided for the target domain and a covariate shift occurs, is not entirely identical to the problem addressed in this work. However, the proposed method becomes the basis for it, since related tasks can be further improved when successful MDL is achieved.

Future works include the application of the proposed method to other predictive maintenance approaches (e.g., condition-based maintenance), quality prediction, and fault diagnosis, for industrial machinery (G. Kim, Choi, et al. 2023; Jun and Kim 2017; Konstantinidis et al. 2022). Using the proposed method with the tool recovery process, to determine when to halt the machining process for the effective reuse of the tools (Jun et al. 2012) remains one of the future works. Integrating the proposed method with other defect detection methods (e.g., virtual methods, vision-based methods) for improved inspection quality and reduced waste under the detection strategy of ZDM (Psarommatis et al. 2020), also remains future research. Another future work is to develop an uncertainty-aware tool wear prediction under multiple machining conditions utilizing the additional benefits that the MDN structure provides, which is the predictive uncertainty without the iterative sampling procedure (Choi et al. 2018). From the perspective of ZDM, developing approaches for automated quality inspection, process control, and maintenance (Azamfirei, Psarommatis, and Lagrosen 2023; Dreyfus et al. 2022), under various manufacturing conditions, remains a promising future research area. In addition, to achieve sustainable manufacturing, combining the proposed method with other advanced technologies, such as digital twins and robotics (Azamfirei, Psarommatis, and Lagrosen 2023; Konstantinidis, Mouroutsos, and Gasteratos 2021), would be one of the most prospective research directions for digital transformation in Industry 4.0.

Acknowledgments

The authors would like to express appreciation to the editors and referees for their helpful comments to improve the quality of our work. The authors thank Minjoo Ku at LG Electronics for helpful discussions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by [Ministry of Trade, Industry and Energy (MOTIE) of Korea] under Grant [number 20017932]; [National Research Foundation of Korea (NRF) funded by the Korea government (MSIT)] under Grant [number 2021R1F1A1046416]; [Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT), Artificial Intelligence Graduate School Program (UNIST)] under Grant [number 2020-0-01336]; [Science and Technology Commercialization Promotion Agency funded by the Korea government in 2023 (Ministry of Science and ICT)] under Grant [number RS-2023-00254286]; and [National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT] under Grant [number 2021H1D8A30652 0712].

Notes on contributors

Gyeongho Kim received his B.S. degree in industrial engineering from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2021. He is currently a Ph.D. Candidate with the Department of Industrial Engineering at UNIST. His research interests include semi-supervised learning, Bayesian learning, the application of machine learning in industries (industrial AI), and uncertainty in machine learning.

Sang Min Yang received a bachelor’s degree in mechanical engineering from Ulsan National Institute of Science and Technology (UNIST), Korea, in 2020. He is currently pursuing a combined master’s and doctoral degree in mechanical engineering at the same university. His research interests include machining processes and analyses.

Sinwon Kim received a bachelor’s degree in mechanical engineering from the University of Ulsan, Korea, in 2021. He is currently pursuing a combined master’s and doctoral degree at Ulsan National Institute of Science and Technology (UNIST) of Mechanical Engineering, Ulsan, Korea. His research interests include machining processes and analyses.

Do Young Kim received the B.S. and Ph.D. degrees from Ulsan National Institute of Science and Technology (UNIST). He is currently an Assistant Professor of Mechatronics Engineering at Chungnam National University. His research interests mainly include multiscale manufacturing processes.
Jae Gyeong Choi received her B.S. in design and human engineering from Ulsan National Institute of Science and Technology (UNIST), Republic of Korea, in 2019. Since 2020, she has been a Combined M.S./Ph.D student in industrial engineering at UNIST. Her research interests include machine learning/deep processing.

Hyung Wook Park is a Professor of Ulsan National Institute of Science and Technology (UNIST) from 2009, PhD in Mechanical Engineering at Georgia Institute of Technology, USA in 2008. His research interests include micro/macro machining, machining dynamics, robot arm machining, and multiphysics-based micro/meso-scale manufacturing (MP-M2) processes and systems.

Sunghoon Lim received his B.S. and M.S. in industrial engineering from KAIST, Republic of Korea, in 2005 and 2009, respectively, and his Ph.D. in industrial engineering from the Pennsylvania State University, University Park, PA, in 2018. He is currently an Associate Professor with the Department of Industrial Engineering, an Adjunct Associate Professor with the Graduate School of Artificial Intelligence, and a Head of the Industry Intelligentization Institute, Ulsan National Institute of Science and Technology (UNIST), Republic of Korea. His research interests include machine learning and deep learning, industrial artificial intelligence, and smart manufacturing.

Data availability statement
The data that support the findings of this work are available from the corresponding author, [SL], upon reasonable request.

References


Jun, Hong-Bae, Dong-Ho Lee, Jae-Gon Kim, and Dimitris Kiritsis. 2012. “Heuristic Algorithms for Minimising Total


Psarommatis, Foivos, Gökcan May, and Victor Azamfirei. 2023. “Envisioning Maintenance 5.0: Insights From a Systematic Literature Review of Industry 4.0 and a Proposed...


