Deep Learning-based Tool Wear Prediction under Multiple Machining Conditions

Gyeongho Kim[1], *Sang Min Yang*[2], *Sin Won Kim*[2], *Do Young Kim*[3], *Jae Gyeong Choi*[1], *Hyung Wook Park*[2], *Sunghoon Lim*[1,4,5]

[1] *Department of Industrial Engineering, Ulsan National Institute of Science and Technology,*
[2] *Department of Mechanical Engineering, Ulsan National Institute of Science and Technology,*
[3] *School of Mechanical Engineering, Kyungnam University,*
[4] *Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology,*
[5] *Industrial Intelligentization Institute, Ulsan National Institute of Science and Technology*

Contents

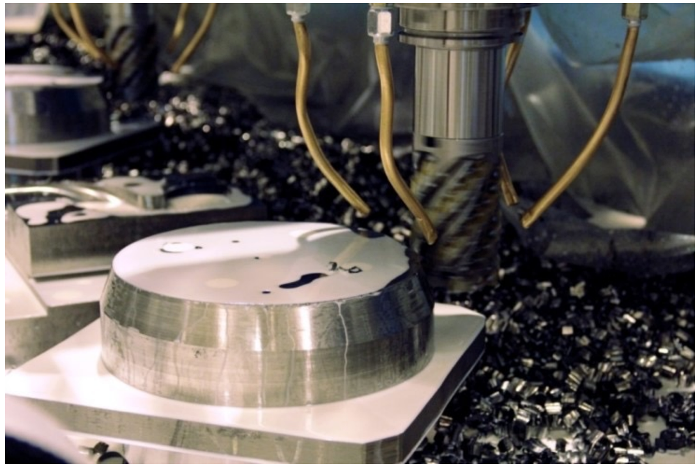
Abstract

Accurate prediction of tool wear is one of the most important tasks in the machining domain. It not only helps to manage efficient maintenance of machine tools but also maximize the useful life of tools. However, there are limitations to existing methods, including that these cannot be used under multiple machining conditions, which is common practice in the domain. To address the multi-domain learning problem in tool wear prediction under various machining conditions, this work presents a novel method. In particular, a Bayesian learning-based feature extractor is proposed to learn domain-invariant representations. In addition, an adversarial learning approach is developed to lead the prediction model to learn domain-invariant features. Lastly, a mixture density network-based module is used to yield final tool wear predictions. Experiments that use datasets from real-world machining processes under multiple conditions prove the effectiveness of the proposed method. Compared to existing methods to data-driven tool wear prediction, the proposed method shows superior prediction performance.

Keywords: Deep learning, machine tools, prediction methods, Bayesian approach

1. Introduction

- Manufacturing process
  - Additive manufacturing (i.e., 3D printing)
  - Casting
  - **Machining** (i.e., subtractive manufacturing)
    - High precision
    - Established theoretical analysis methods
    - Wide industrial applications (e.g., aerospace)

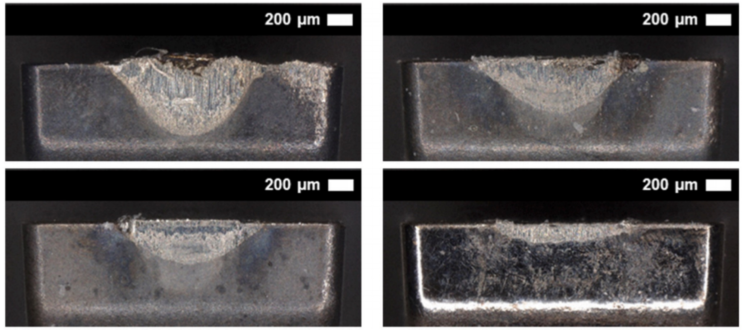


1. Introduction

- Milling process
  - High-level precision
  - End milling, slot milling, up/down milling, face milling
    - Rough machining, surface finish machining

  - Tool wear
    - Surface quality degradation
    - Rapid tool breakage
    - Tool maintenance (e.g., replacement)
      - Domain knowledge-based

➡ Impaired process and product quality



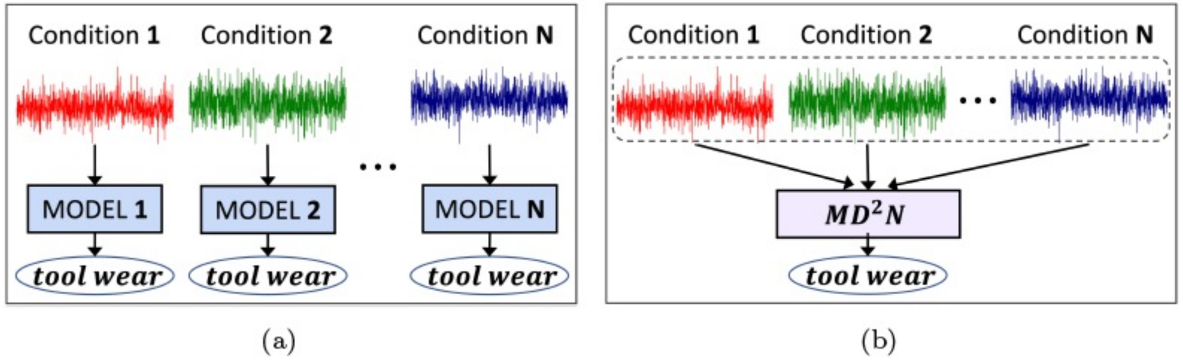
1. Introduction

- Tool wear prediction
  - Real-time estimation of ongoing tool wear → tool condition monitoring (TCM)
  - Advantages:
    - Efficient predictive maintenance strategies
    - Maximized available tool life and usage

  - Difficulties:
    - Complexity of machining and tool dynamics
      - 1) analytically, 2) computationally
    - Real-world machining practices
    - Existence of *multiple machining conditions*
      - Different configurations of: cutting speed, material, lubricant type, etc.
      - For data-driven approaches, several issues arise;
        - Multi-domain data
        - Use of multiple independent predictive models

1. Introduction

- Proposal:
  - A universal model for tool wear prediction under multiple machining conditions



  - Advantages:
    - Improved efficiency in model development and deployment

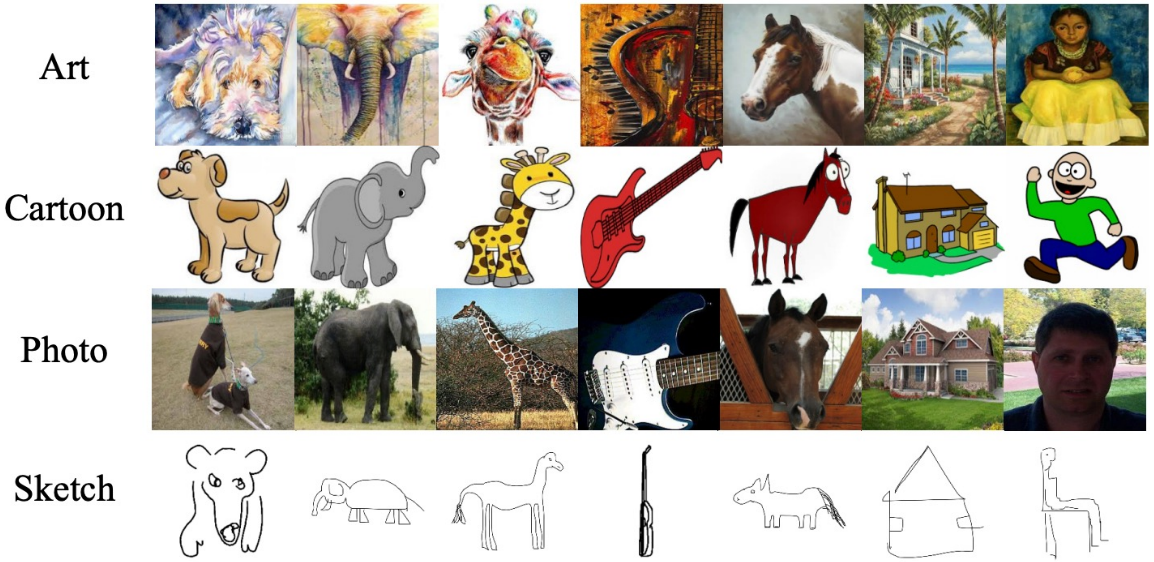
- Problems to be addressed:
  - Data from multiple machining conditions → multiple domains (data heterogeneity)
  - Inverse problem
    - Similar inputs corresponding to different output values (i.e., tool wear)

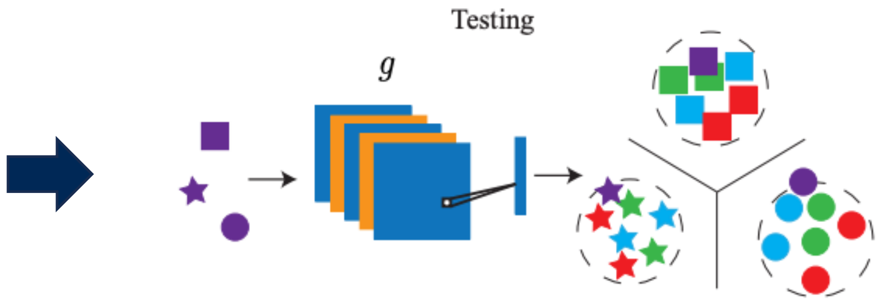
2. Preliminaries

- Data-driven tool wear prediction
  - Traditional approaches
    - Analytical solution-based, mechanistic approaches
    - Problems:
      - Requires high domain knowledge level, hinders online applications
  - Conventional data-driven approaches
    - Feature extraction + machine learning (ML)-based predictive models
  - Deep learning (DL)-based approaches
    - Multivariate time-series data as inputs → regression
    - Popular architecture types:
      - Convolutional neural network (CNN)
      - Recurrent neural network (RNN)
      - Transformer

2. Preliminaries

- Multi-domain learning (MDL)
  - Different machining conditions → different data domains
  - Objective of MDL:
    - Train predictive models to perform on data drawn from multiple domains
    - Learn domain-invariant feature representations





Yao, X., Bai, Y., Zhang, X., Zhang, Y., Sun, Q., Chen, R., ... & Yu, B. (2022). PCL: Proxy-based Contrastive Learning for Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7097-7107).
Motiian, S., Piccirilli, M., Adjeroh, D. A., & Doretto, G. (2017). Unified deep supervised domain adaptation and generalization. In Proceedings of the IEEE international conference on computer vision (pp. 5715-5725).

2. Preliminaries

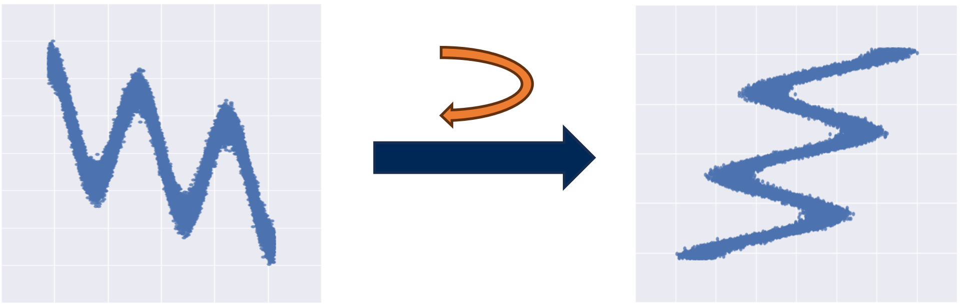
- Multi-domain learning (MDL)
  - Modifications on a model architecture
    - e.g., residual adapter
  - Training and optimization process
    - e.g., Iterations of pretraining → fine-tuning

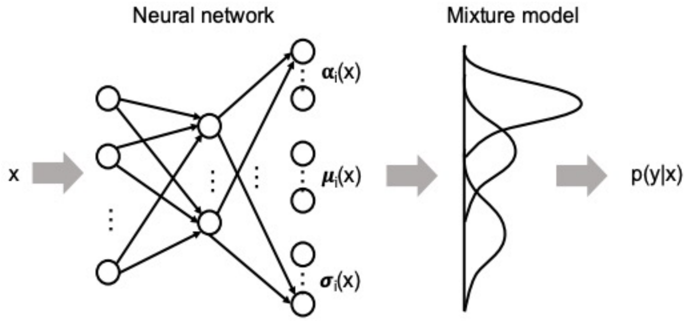
  - Adversarial learning
    - Make a model unable to discriminate between domains

  - Bayesian approach
    - Bayesian neural network (BNN)
      - Improved domain generalization ability
    - Uncertainty modeling

2. Preliminaries

- Mixture density network (MDN)
  - Inverse problem
    - A potential issue in tool wear prediction under multiple machining conditions
    - e.g., similar input signals → different tool wear degrees





  - Modeling multimodal outputs
    - Probabilistic outputs → mixture distribution

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

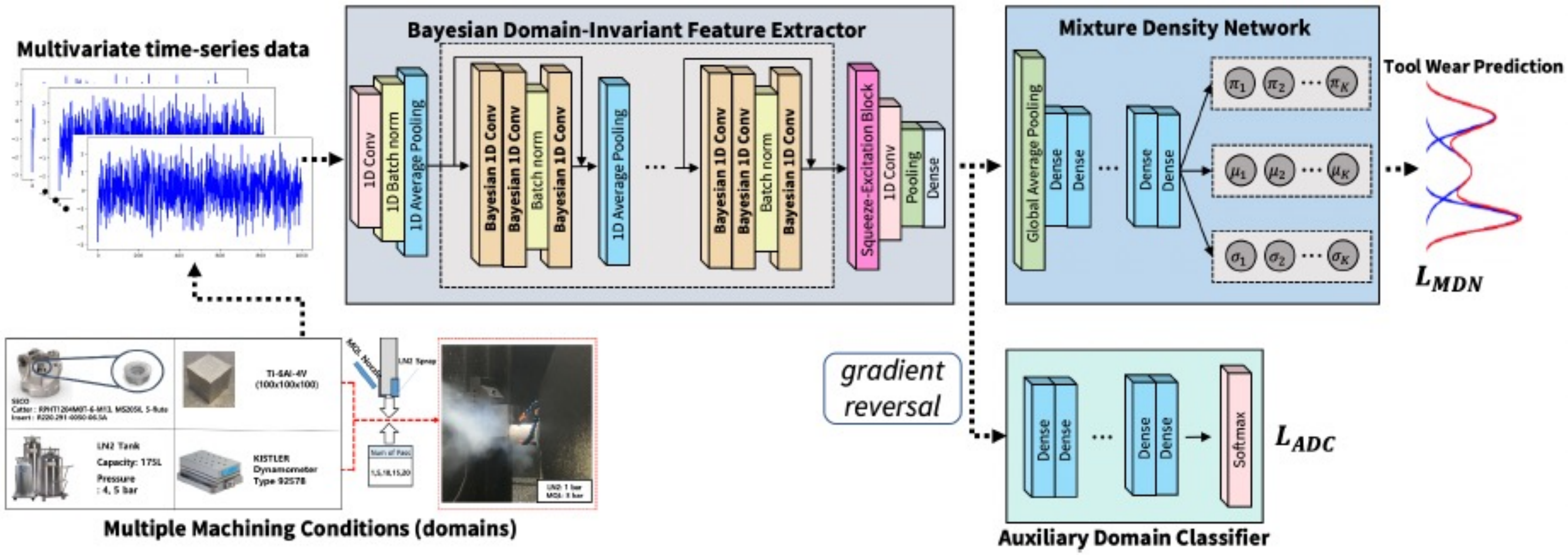
$$where \sum_{k=1}^{K} \pi_k(x) = 1 \ \ and \ \ \pi_k \geq 0 \, \forall \, k.$$

  - Diverse application areas
    - Pose estimation, autonomous vehicle

http://edwardlib.org/tutorials/mixture-density-network
Vossen, J., Feron, B., & Monti, A. (2018, June). Probabilistic forecasting of household electrical load using artificial neural networks. In 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS) (pp. 1-6). IEEE.

3. Proposed Method

- Multi-domain mixture density network (MD^2N)
  - Bayesian learning-based feature extractor
    - Learn domain-invariant representations
  - MDN-based predictor
    - Generate multimodal predictive distributions
  - Adversarial learning-based MDL



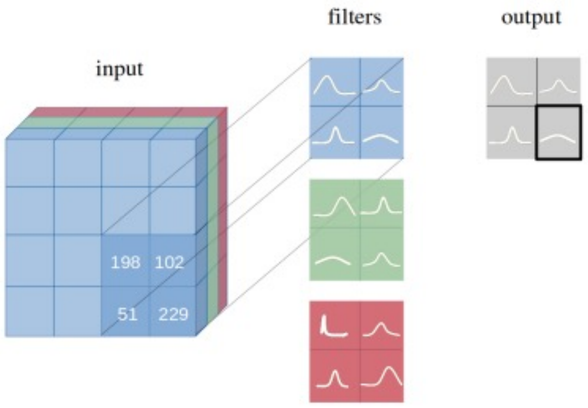
3. Proposed Method

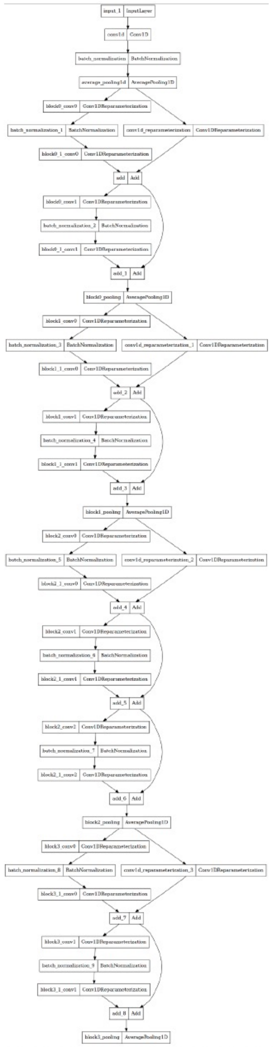
- Bayesian domain-invariant feature extractor (BDIFE)
  - Using BNN to learn domain-invariant representations
    - As an ensemble of domain-specific representations
  - Bayesian convolution
    - Variational inference (VI)-based
    - Reparametrization trick

$$w = \mu + \rho \cdot \epsilon.$$
$$\epsilon \sim \mathcal{N}(0, I).$$

    - Training objective

$$L_{VI} = KL(q(w|\theta)||p(w)) - E_{q(w|\theta)}[log(p(y|x,w))]$$
$$\approx \sum_{i=1}^{n} log(q(w^{(i)}|\theta)) - log(p(w^{(i)})) - log(p(y|x,w^{(i)})).$$





  - Squeeze-excitation (SE) block

Shridhar, K., Laumann, F., & Liwicki, M. (2019). A comprehensive guide to Bayesian convolutional neural network with variational inference. arXiv preprint arXiv:1901.02731.

3. Proposed Method

- MDN-based tool wear predictor
  - Using extracted features from BDIFE, perform tool wear prediction

  - Outputs:
    - $\{\pi_k, \mu_k, \sigma_k\}_{k=1}^{K}$ → constitutes a mixture distribution (i.e., a predictive distribution)

  - Additional techniques for MDN training
    - Activation function

$$h(x) = \begin{cases} x + 1, & if\ x > 0. \\ \alpha \cdot (exp(x) - 1), & if\ x \leq 0. \end{cases}$$

    - Regularization

$$L_\pi = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(-\pi_k(x_i)\ log(\pi_k(x_i))). \quad L_\sigma = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(\sigma_k(x_i))^2.$$

  - Inference:
    - Select the mixture component with the highest mixing coefficient $argmax_k \pi_k$

3. Proposed Method

- Auxiliary domain classifier
  - Adversarial learning approach to MDL
  - Gradient reversal layer (GRL)
    - Using features extracted from BDIFE, an auxiliary classifier trained in an adversarial manner

$$L_{ADC} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{d=1}^{D} t_d \, log(p(t|x_i)_d).$$

- Final training objective of MD^2N
  - Training with standard backpropagation (in an end-to-end fashion)

$$\begin{aligned} E_{w,\phi,\psi} = -\frac{1}{N}\sum_{i=1}^{N} log \sum_{k=1}^{K} \pi_k(x_i)\mathcal{N}(p(y|\mu_k(x_i),\sigma_k(x_i))) \\ +\lambda_1 L_\pi + \lambda_2 L_\sigma + KL(q(w|\theta)||p(w)) \\ -\lambda_3 \cdot -\frac{1}{N}\sum_{i=1}^{N}\sum_{d=1}^{D} t_d \, log(p(t|x_i)_d). \end{aligned}$$

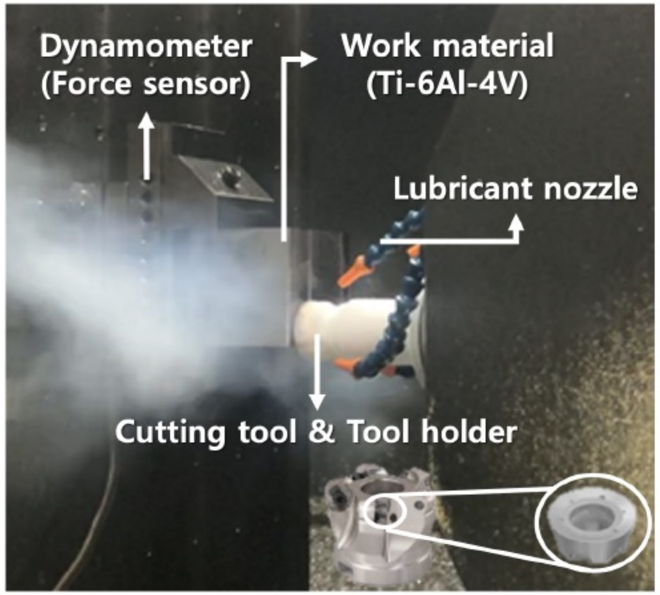
Ganin, Y., & Lempitsky, V. (2015, June). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180-1189). PMLR.

4. Experiments and Results

- Milling experiment
  - Setup
    - Work material: Ti-6Al-4V
    - 5-axis CNC machine
  - Multiple machining conditions:
    - 8 different machining conditions
    - Wet and CryoMQL setting

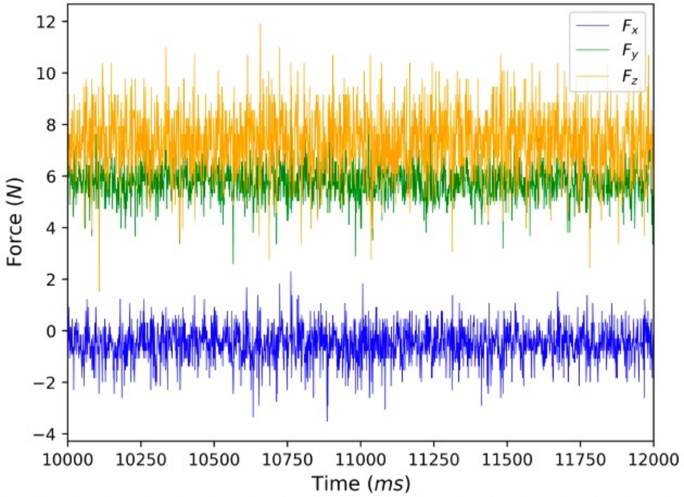
Experiment number	Setting	Cutting speed (m/min)	Feed $(mm/tooth)$	Axial depth (mm)	Radial depth (mm)	Material removal rate (mm^3/min)	Number of pass
1	Wet	60	0.6	1.0	18.0	5490.0	
2	Wet	80	0.6	1.0	18.0	5500.0	
3	Wet	60	0.8	1.0	18.0	5500.0	
4	Wet	60	0.6	1.3	18.0	5363.0	
5	CryoMQL	60	0.6	1.0	18.0	5490.0	[1,5,10,15,20]
6	CryoMQL	80	0.6	1.0	18.0	5500.0	
7	CryoMQL	60	0.8	1.0	18.0	5500.0	
8	CryoMQL	60	0.6	1.3	18.0	5363.0	

  - Data collection
    - Dynamometer (cutting force sensor)
    - 3-axes (x-,y-,z-)



4. Experiments and Results

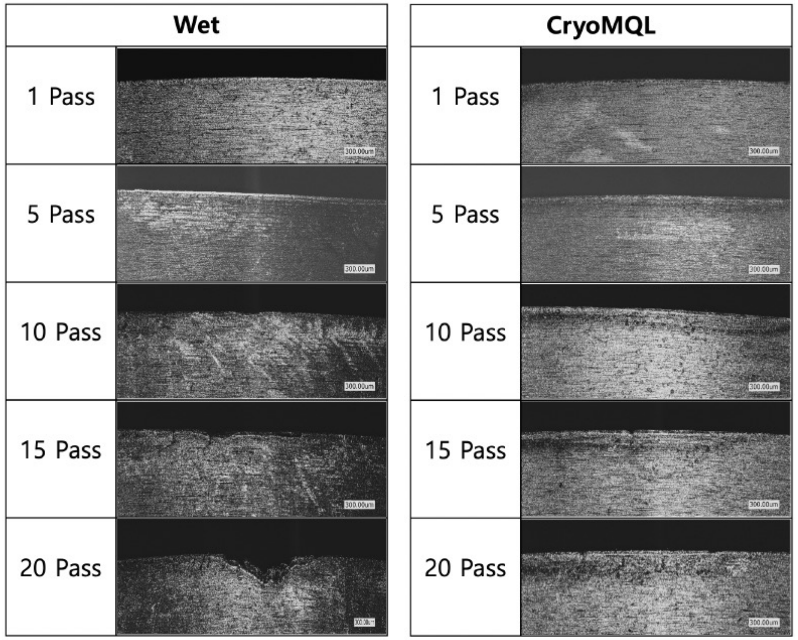
- Variable description
  - Multivariate time-series (acceleration data)



Number	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
Variable	F_x	F_y	F_z	F_x	F_y	F_z	F_x	F_y	F_z	F_x	F_y	F_z
Mean	81.0290	45.0864	64.7218	78.1679	44.0454	103.4832	95.0169	54.7226	106.5160	96.7278	53.5007	108.3236
SD	225.7679	112.8530	132.1324	232.5850	111.3457	153.2089	284.9614	144.2577	176.5907	283.6351	142.5736	168.3237
Min	-496.0630	-406.0360	-52.4902	-640.8690	-447.6930	-50.3540	-494.8430	-544.2810	-56.4575	-450.8970	-517.2730	-53.7109
Max	2102.0500	1121.2200	711.6700	2129.0600	1027.9800	823.3640	2738.4900	1344.4500	1087.6500	2629.3900	1323.5500	979.6140
Number	Dataset 5			Dataset 6			Dataset 7			Dataset 8		
Variable	F_x	F_y	F_z	F_x	F_y	F_z	F_x	F_y	F_z	F_x	F_y	F_z
Mean	80.0653	46.7010	-38.4820	82.0602	45.2524	1.4938	92.2392	55.6570	-52.6978	101.0826	53.7894	-83.4370
SD	228.4289	123.2950	147.1831	231.0209	117.2994	169.6341	264.6479	146.0505	159.0441	282.8738	146.6253	186.6293
Min	-519.8670	-456.0850	-232.8490	-637.0540	-456.6960	-217.5900	-439.9110	-550.2320	-257.8740	-729.2180	-618.2860	-358.5820
Max	2104.8000	1231.2300	689.0870	2107.2400	1193.0800	773.3150	2536.3200	1419.0700	847.1680	2671.8100	1533.9700	810.8520

4. Experiments and Results

- Tool wear measurement



(a) (b)

$$VB = d(a + bT^c)^{-1}.$$

$$\mathbf{p}_{k+1} = \mathbf{p}_k - (J_{\mathbf{r}}^T J_{\mathbf{r}} + \mu_k\, diag(J_{\mathbf{r}}^T J_{\mathbf{r}}))^{-1} J_{\mathbf{r}}^T \mathbf{r}(\mathbf{p}_k),\ \ k \geq 0$$

$$J_{\mathbf{r}}(\mathbf{p}) = \begin{bmatrix} \frac{\partial r_1(\mathbf{p})}{\partial p_1} & \cdots & \frac{\partial r_1(\mathbf{p})}{\partial p_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_n(\mathbf{p})}{\partial p_1} & \cdots & \frac{\partial r_n(\mathbf{p})}{\partial p_m} \end{bmatrix}$$

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

- Tool wear calculation
  - For ground-truth tool wear degrees (between measurements)

4. Experiments and Results

- Model training details
  - Data preprocessing via standardization
  - Sliding window method
  - Dataset split (train : valid : test = 70% : 10% : 20%)
    - Five independent trials

- Evaluation metrics
  - Regression measures

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

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

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

5. Results

- Experimental results
  - Training set: data from all machining conditions
  - Test set: different conditions

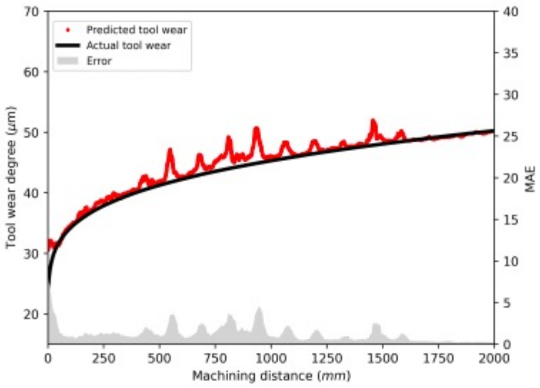
Test dataset	Metric MAE	RMSE	MAPE
1	1.7139 (±0.4186)	3.7887 (±0.3156)	0.0374 (±0.0032)
2	2.6190 (±0.5090)	5.2913 (±0.8235)	0.0452 (±0.0103)
3	4.1183 (±0.5812)	6.5092 (±0.7506)	0.0512 (±0.0025)
4	4.4241 (±0.6486)	9.9750 (±1.3586)	0.0559 (±0.0079)
5	1.4735 (±0.3564)	3.8300 (±0.6313)	0.0237 (±0.0050)
6	1.8446 (±0.5028)	5.6408 (±1.0257)	0.0428 (±0.0115)
7	3.3007 (±0.7133)	6.1715 (±1.4229)	0.0402 (±0.0063)
8	1.2603 (±0.1647)	3.8236 (±0.4435)	0.0199 (±0.0027)
All (1~8)	2.1748 (±0.2655)	5.6422 (±0.6205)	0.0350 (±0.0041)

5. Results

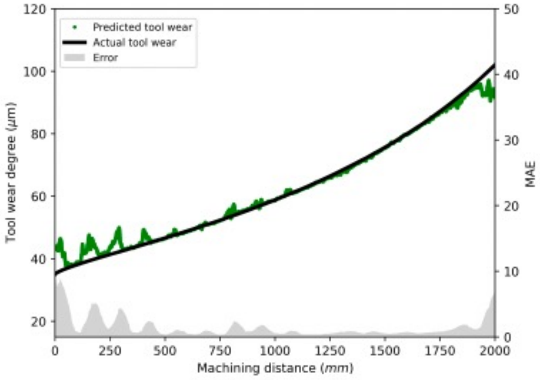
- Performance comparison with existing data-driven methods
  - Compared with:
    - Support vector machine (SVR), random forest (RF)
    - CNN
    - Long short-term memory (LSTM), gated recurrent unit (GRU)

Dataset	Wet			CryoMQL			Wet + CryoMQL		
Model	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
SVR	17.6989 (±0.0795)	22.2290 (±0.0869)	0.2897 (±0.0003)	11.0229 (±0.0561)	13.0161 (±0.0976)	0.1928 (±0.0010)	14.3869 (±0.0297)	16.8103 (±0.0310)	0.2570 (±0.0005)
RF	11.0783 (±0.0344)	15.3236 (±0.0381)	0.1947 (±0.0005)	7.1514 (±0.0211)	10.2964 (±0.0281)	0.1230 (±0.0004)	9.6556 (±0.0269)	13.2919 (±0.0313)	0.1681 (±0.0002)
LSTM	11.1720 (±1.1396)	13.7242 (±1.2325)	0.2730 (±0.1679)	6.4884 (±1.4033)	8.6648 (±1.1930)	0.1134 (±0.0276)	7.8924 (±1.1525)	10.4880 (±0.9437)	0.1368 (±0.0201)
GRU	9.2271 (±0.8675)	11.9802 (±0.6444)	0.1589 (±0.0133)	5.9730 (±0.5707)	8.2016 (±0.5442)	0.1028 (±0.0101)	7.7760 (±1.9583)	10.5061 (±1.9467)	0.1339 (±0.0335)
CNN	16.5959 (±0.0824)	18.8952 (±0.0584)	0.2875 (±0.0013)	9.9452 (±0.0307)	11.5631 (±0.0085)	0.1811 (±0.0011)	13.6391 (±0.0451)	16.3143 (±0.0499)	0.2187 (±0.0465)
MD^{2}N (proposed)	**3.5198** (±0.3430)	**7.1057** (±0.5456)	**0.0570** (±0.0061)	**1.7204** (±0.7231)	**3.6470** (±0.6684)	**0.0292** (±0.0111)	**2.1748** (±0.2655)	**5.6422** (±0.6205)	**0.0350** (±0.0041)

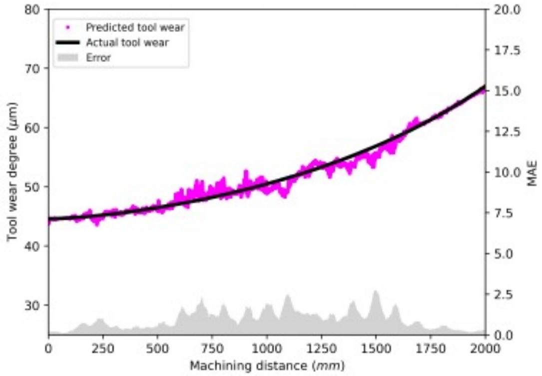
5. Results



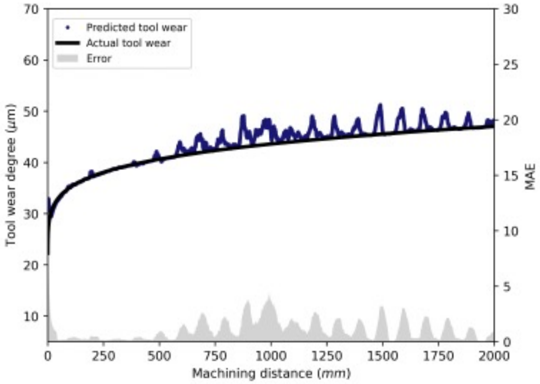
(a) Dataset 1.



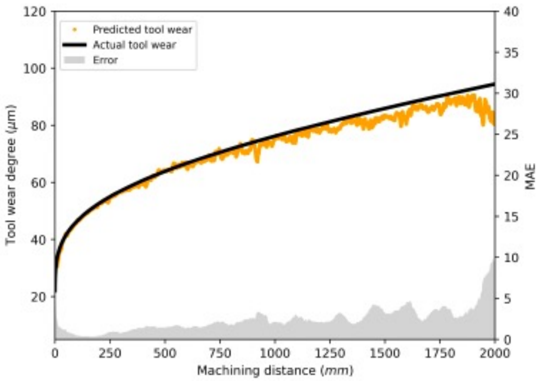
(b) Dataset 2.



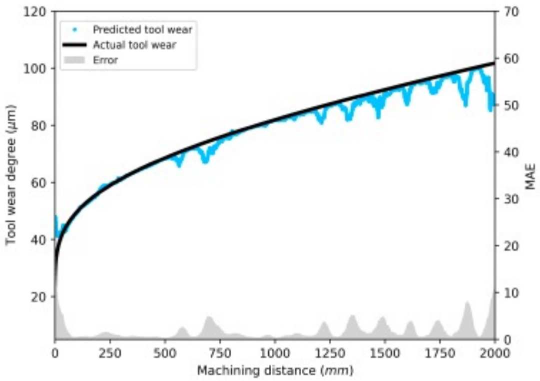
(e) Dataset 5.



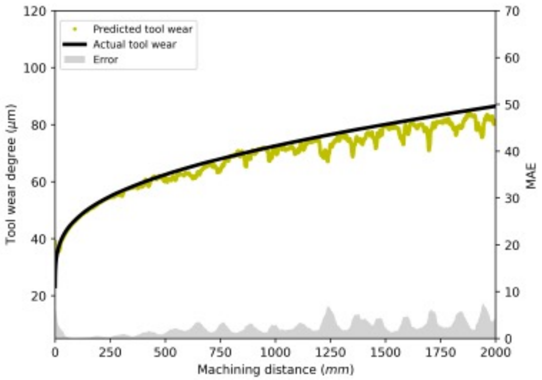
(f) Dataset 6.



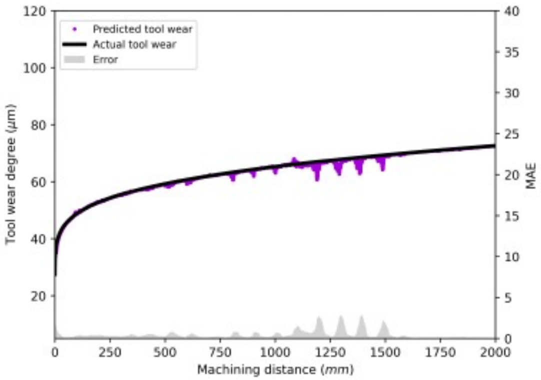
(c) Dataset 3.



(d) Dataset 4.



(g) Dataset 7.



(h) Dataset 8.

5. Conclusion and Discussion

- Summary:
  - A tool wear prediction method (i.e., MD^2N) that performs under multiple machining conditions

- Main points:
  - Multi-domain learning
    - Bayesian learning-based feature extraction
    - Adversarial learning approach with gradient reversal
  - Multimodal output
    - Mixture density network for probabilistic prediction
    - Solution to potential inverse problems

5. Conclusion and Discussion

- Future works
  - Increase data heterogeneity
    - More diverse machining conditions
    - Different work materials
  - Advanced modeling techniques
    - Recent approaches in related fields of MDL
      - e.g., domain generalization, transfer learning, etc.
  - Maintenance scheduling

Thank you for listening!

This work was supported by the Advanced Technology Center Plus (ATC+) Program (20017932, 50% Accident Prevention Focus to reduce accident rate Development of Risk Detection System for Road Facilities Based on Artificial Intelligence) funded by the Ministry of Trade, Industry and Energy (MOTIE) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1046416). This work was also partly supported by X-Corps Plus program of National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (No. 2021H1D8A306520712).

Presenter Information

Gyeongho Kim

Contact: kkh0608@unist.ac.kr