Journal of Manufacturing Systems 75 (2024) 288–298

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical paper

Multi-sensor fusion and machine learning-driven sequence-to-sequence translation for interpretable process signature prediction in machining☆☆,☆

Clayton Cooper [a], Jianjing Zhang [a], Ihab Ragai [b], Robert X. Gao [a,*]

[a] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH, USA
[b] Department of Mechanical Engineering, Penn State University, The Behrend College, Erie, PA, USA

ARTICLE INFO

Keywords:
Machining
Sensor fusion
Process signature
Interpretable machine learning
Acoustic signals

ABSTRACT

During machining, kinetic energy is imparted to a workpiece to remove material. The integrity of the machined surface, which depends on the energy transfer, affects the quality and performance of the product, therefore needs to be quantified. Prior studies have indicated the potential of using machining power, or the power consumption at the tool-chip interface, as a process signature for predicting machined surface integrity. However, direct measurement of machining power is constrained by the availability of special equipment and the associated cost. To address this gap, this paper presents a machine learning-based method for machining power prediction through multi-sensor fusion and sequence-to-sequence translation from acoustic and vibration signals, which represent portions of the in-situ kinetic energy dissipation, to the machining power signal as a process signature. Specifically, a neural network architecture is developed to separately translate the acoustic and vibration signals to corresponding machining power signals. The two predicted power signals are subsequently fused to arrive at a unified power signal prediction. To check for spurious decision logic, the sensor fusion model is interpreted using integrated gradients to reveal temporal regions of the input data which have the most influence on the machining power prediction accuracy of the fusion model. Systematic cutting experiments performed on a lathe using 1018 steel have shown that the developed sensor fusion method for process signature prediction can successfully map machine acoustics to power consumption with 5.6% error, tool vibration to power consumption with 8.2% error, and acoustics and vibration, jointly, to power with 2.5% error. Model parameter interpretation reveals that the vibration signal is more influential on the machining power prediction result than the acoustic signal, but that overall model accuracy is diminished if only the vibration signal is used.

Introduction

Commensurate with the rise of integrated sensing and industrial machine learning (ML), process signatures have emerged as valuable tools for manufacturing process monitoring and quality assurance [1,2,3]. Defined as the embodiment of all mechanical, thermal, and chemical energies imparted to the workpiece during a manufacturing process, a process signature comprehensively encodes the surface and subsurface conditions of a part caused by in-process energy transfer [4,5]. As such, process signatures have been shown to predict machined surface integrity characteristics and tool conditions with higher accuracy than process parameter-only models (e.g., speed, feed, depth of cut, etc.) or single process quantity-based approaches (e.g., acoustic, vibration, etc.). The benefit is earlier detection and correction of noncompliant product [6,7,8].

In the case of machining, recent studies have proposed machining power as a process signature since it represents, by definition, all energy imparted to the workpiece [9,10]. For example, Sealy *et al.* observed asymptotic relationships between the spindle specific energy and surface roughness, microhardness, and residual stress, indicating that the spindle specific energy embodies these product characteristics [11]. Zhu *et al.* identified an exponential relationship between specific cutting energy and undeformed chip thickness, which has a direct influence on the machined surface roughness [12]. Bustillo *et al.* used a radial basis function network to analyze average machining power readings and subsequently predict machined surface flatness deviations with 3 μm mean absolute error [13]. Cooper *et al.* modeled the relationship

☆☆ Peer-review under responsibility of the scientificcommittee of the NAMRI/SME.☆ 52nd SME North American Manufacturing Research Conference (NAMRC 52, 2024).

* Corresponding author.
E-mail address: robert.gao@case.edu (R.X. Gao).

https://doi.org/10.1016/j.jmsy.2024.04.010
Received 13 April 2024; Accepted 14 April 2024
Available online 13 May 2024

Nomenclature

c	Baseline specific cutting force
d	Depth of cut
$E_\star$	Energy
f	Cutting feed
g	Multi-sensor machining power prediction model
g_f	Decision fusion model, part of g
$h_\star$	Fully connected neural network
IG	Integrated gradient
$\mathscr{J}$	Jacobian
$k_\star$	Convolutional kernel
$\mathscr{L}$	Power prediction model loss
$L_\star$	Number of hidden layers
$l_\star$	1-D convolutional kernel length
MRR	Material removal rate
$P_\star$	Power
$\widetilde{P_\star}$	Predicted power
PP	Process parameters
$q_\star$	Multichannel 1-D convolutional neural network
S_{FC}	Specific cutting force
t	Time
t_c	Chip thickness
TC	Tool condition
v	Cutting speed
$w_\star$	Proportion of power
γ	Integrated gradient input interpolator
ε	Power fluctuations
ζ_0	Rake angle
θ	Model parameters
θ^*	Optimal model parameters
$\phi_\star$	Activation function
$\|\star\|$	Vector length
$\mathbf{1}_\star$	Ones vector of length $\star$

between energy consumption and surface roughness by passing time-frequency representations of machining power signals to a 2-D convolutional neural network (CNN) to predict machined surface roughness with 9% error [6].

Several studies have also used machining power to predict tool condition and remaining useful life, which have direct effects on the surface integrity [14,15]. Drouillet *et al.* used measured machining power as an input to an artificial neural network (ANN) to predict tool remaining useful life and achieved approximately 5% error [16]. Likewise, Corne *et al.* predicted tool wear in real time using the instantaneous machining power as input to an ANN, yielding as little as 1% error [17]. Wang *et al.* predicted tool wear by fusing recurrent neural network (RNN) representations of machining power signals with CNN-based representations of surface images, with about 12% error [18]. Moliner-Heredia *et al.* fit a polynomial model to determine tool change points as a function of machining power and observed reduced tool consumption and lower numbers of rejected parts compared to non-power-based approaches [19]. Recent power-based tool condition monitoring approaches and associated ML techniques are further reviewed in [20,21].

Unfortunately, acquisition of machining power data is oftentimes impractical and/or financially infeasible. High-quality power meters are cost prohibitive and their installation impedes production [22,23]. Considering these limitations, it is desired to infer the machining power process signature using readily available process quantities that are indicative of the power signal. Literature has explored three such quantities: acoustics, vibration, and thermal signals emanating from the tool-chip interface. The sum of these three sensing modes is considered to fully characterize the energy transfer into the workpiece and can thus be used to infer the machining power process signature. Among the three sensing modalities, acoustic-based and vibration-based monitoring have been widely used due to their availability and ease in installation.

Ragai *et al.* previously demonstrated that the principal component coefficients of acoustic signals acquired during turning processes were sufficient to achieve 99% cutting parameter classification accuracy using deep ML techniques, suggesting that machining acoustic data contains information unique to the cutting conditions [24]. Since cutting conditions have a predominant influence on machining power, it is hypothesized that the machining acoustics can also be used to predict the power signal while remaining responsive to in-machine conditions [25]. This hypothesis is supported by previous studies which used machining acoustics and ML to infer the tool and cutting conditions, which also determine the machining power [8,26,27,28]. Thus, the machining acoustic signature is considered to reflect the compound effect of process parameters and tool condition on the machining power consumption, and, as a result, represents an appropriate candidate for consideration for machining power process signature prediction.

In the field of vibration-based machining power prediction, several studies have developed mechanistic models to predict machining power as a function of tool vibration [29,30,31,32]. One limitation is that these approaches require the use of 3-D dynamometers, which may not be as easily accessible for the same reasons as a power meter, as discussed earlier. Alternatively, Kim *et al.* developed a transfer learning ML model for machining power prediction as a function of spindle position and cutting parameters, and demonstrated qualitatively good results under a limited data condition in the target domain [33]. Xu *et al.* applied improved case-based reasoning to leverage knowledge distilled from previous prediction queries and more accurately predicted average machining power based on the vibration signal, achieving a mean absolute percentage error (MAPE) of 8%. The method requires explicit knowledge of the tool wear condition [34]. These studies suggest that information on the machining power process signature is embedded in the tool vibration signal.

While information on the machining power is embedded in the acoustic and vibration signals, few published studies have directly predicted the time-varying machining power signals. Recent studies have elucidated the effectiveness of trend-seasonality decomposition for time series prediction [35], though this approach is yet to be seen in the manufacturing literature. Additionally, an acoustic to machining power prediction model has yet to be reported in literature. A consequence of this is that a joint acoustic-vibration sensor fusion model has not been realized, which is expected to have improved power prediction accuracy over single-sensor approaches [36]. Previous data-driven machining power prediction models have also yet to be interpreted to reveal their decision logic, making them black boxes with inherently opaque decision-making rules and potentially spurious behavior when processing never-before-seen data [37,38]. This study closes these research gaps by developing a multi-sensor fusion machining power signal prediction method using ML, comparing the machining power prediction accuracy of the fusion model against single sensing modality acoustic-only and vibration-only models, and observing the learned power prediction logic by assessing the sensitivity of the predicted power signal to the acoustic and vibration inputs. An overview of the proposed methodology is shown in Fig. 1.

Specifically, an ML architecture is designed to independently process raw acoustic and vibration signals and achieve sensor-specific predictions of the machining power signal as functions of time. Following from the current time series prediction literature, data from each sensor

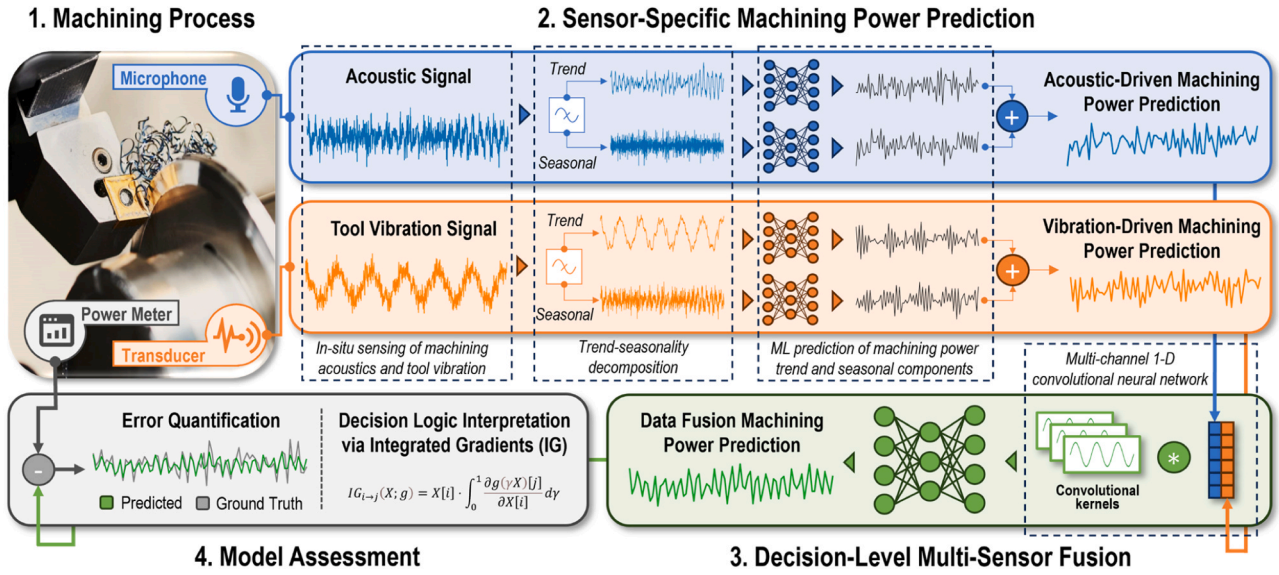

Fig. 1. Multi-sensor data fusion for machining power prediction and decision logic interpretation.

is decomposed into trend and seasonal components, which are transformed by separate ANNs into the corresponding trend and seasonal components of the power data, then recombined to yield a sensor-specific machining power signal estimate. A convolutional merging algorithm is then used to combine the predicted power signals from each sensor and arrive at a final unified machining power signal prediction, which considers the acoustic and vibration data simultaneously. Integrated gradients are calculated for the power signal predictions to identify which temporal regions of the acoustic and vibration signals have the greatest influence on the power signal predictions, as well as to identify any spurious logic. The contributions of this study are summarized as follows:

1) Developed a data fusion technique for machining acoustic and vibration signals, which extracts and integrates the signal-underlying information to predict the machining power process signature, using only raw in-process data as input.
2) Developed an interpretation method for the data fusion model using integrated gradients to reveal acoustic and vibration input patterns, which correspond to machining power signal features and elucidate the model's prediction logic for greater model transparency.
3) Quantitatively compared the performance of the fusion model with acoustic-only and vibration-only methods of machining power prediction.

The remainder of the paper is organized as follows: in Section 2 the background on the machining power process signature and multi-sensor fusion is provided. Section 3 presents the proposed power prediction model and model interpretation method based on integrated gradients. Section 4 describes the experiment performed to validate the sensor fusion method and model training. Section 5 presents the power signal prediction results using single-sensor and multi-sensor approaches and interprets the model to elucidate its prediction logic. Finally in Section 6, conclusions and future work are summarized.

Background of machining and multi-sensor fusion

The energy transferred into the workpiece during machining induces material removal as well as surface and subsurface transformations, which ultimately determine surface integrity. Thus, knowledge of the energy transfer in the form of machining power is paramount for the comprehensive characterization of post-machining product quality. In the absence of direct machining power measurement, prediction techniques must be developed to estimate the machining power signal using all available process observations. This section assesses the process physics embodied in the machining power and provides an overview of sensor fusion techniques to combine machining acoustic and tool vibration signals for prediction of the time-varying machining power.

Machining power as a process signature

Machining is a material removal process wherein a cutting edge imparts kinetic energy to a small region of the workpiece surface and subsurface as shown in Fig. 2. The source of this energy is the machine spindle, which rotates the tool during milling and rotates the workpiece during turning. The resulting shear deformation within the energy-affected region causes material displacement in the form of a chip, which grows until it breaks off or is manually removed. However, the input kinetic energy is simultaneously dissipated in several ways in addition to constituting the shear stress, namely acoustically, vibrationally, and thermally. Thus, the instantaneous energy balance can be written as follows:

$$\frac{dE_m}{dt} = \frac{dE_{out}}{dt} \tag{1}$$

$$P_m = P_\tau + P_a + P_v + P_h \tag{2}$$

where E_m and P_m are machining energy and power, respectively, E_{out} is energy dissipated into the workpiece and environment, P_τ is shear power, P_a is acoustic power, P_v is vibrational power, and P_h is thermal power. Since the sum of dissipated power cannot exceed P_s, each dissipation mode can be represented as dissipating a fraction of P_m:

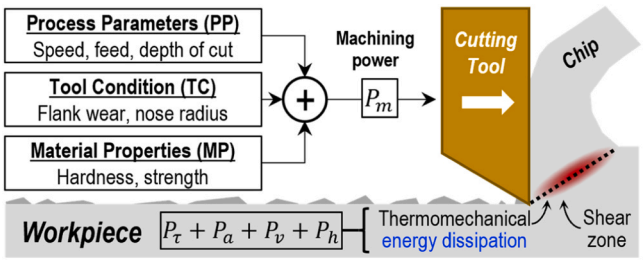

Fig. 2. Machining power determinants in machining energy dissipation.

$$P_m = w_\tau P_m + w_a P_m + w_v P_m + w_h P_m \tag{3}$$

where $w_\star$ is fraction of P_m dissipated by the dissipation mode listed in the subscript and $\sum w_\star = 1$.

Previous work has shown that the machining power demanded can also be modeled in terms of material removal:

$$P_m = MRR(PP) \cdot S_{FC}(PP, TC) + \varepsilon(t, TC, MP) \tag{4}$$

where MRR is the material removal rate (mm^3/min) as a function of process parameters (PP), S_{FC} is the specific cutting force (N/mm^2) as a function of PP and tool condition (TC), and ε denotes power fluctuations caused by regenerative vibration, etc., which are dependent on time (t), TC, and material properties (MP) [39]. In turning, MRR and S_{FC} are defined as:

$$MRR = vfd \tag{5}$$

$$S_{FC} = ck_c^{-c/t_c}\left(1 - \frac{\zeta_0}{100}\right) \tag{6}$$

where v is the cutting speed (m/min), f is the feed rate (mm/rev), d is depth of cut (mm), k_c is an experimental measurement of S_{FC} at 0° rake angle (N/mm^2), t_c is chip thickness (mm), and ζ_0 is the rake angle (deg) [40].

Combining (1)-(4), the following relationship between power dissipation and process physics is derived:

$$P_m = w_\tau P_s + w_a P_s + w_v P_s + w_h P_s = MRR(PP) \cdot S_{FC}(PP, TC) + \varepsilon(t, TC, MP) \tag{7}$$

Eq. (7) suggests that information on process parameters, tool conditions, and material property is encoded in the input power signal and its dissipation mechanisms. As a result, these power readings reflect the in-process physics caused by these physical properties. Previous work has shown that P_s in particular is an effective process signature for surface integrity quantification due to its embodiment of in-process disturbances not accounted for by parameter-only models [6]. However, few quantities in (7) are feasible to measure directly during machining and thus observation and inference of P_m are both difficult. Direct in-situ measurement of P_m, P_τ, P_h, and S_{FC} is expensive due to the cost of equipment and its installation, i.e., power meters, thermal cameras, and dynamometers. Additionally, there are not yet models to estimate the $w_\star$ terms in (7). MRR cannot be used to infer P_m without accurate readings of S_{FC}, which is infeasible for the reasons discussed above. Recent literature as reviewed herein suggests that acoustics and vibration are powerful predictors of machining power, although they are typically used individually [24,33]. This indicates that acoustic power, P_a, and vibrational power, P_v, are feasible means of estimating P_m.

Acoustic power and vibrational power are measured using microphones and force sensors, respectively, and the output of each of these sensors is directly proportional to P_m [24]. A larger machining power generally corresponds to greater spindle torque, higher spindle speed, increased tool wear, and/or more aggressive process parameters, which are prone to induce more prominent acoustic artifacts and more aggressive vibrations (vice versa for a lower machining power). These changes in the physical behavior of the system will be reflected in the amplitude of the microphone and vibration sensor readings, thus encoding P_m in the P_a and P_v signals.

Previous studies have successfully used vibration power as a predictor of machining power. However, these models do not account for all energy dissipation modes in machining and do not fully characterize the process, resulting in suboptimal machining power predictions [33,34]. A mapping from acoustic power to machining power has yet to be established despite evidence suggesting that the mapping is likely to be feasible [8,24,26,27,28]. This research gap has also limited the consideration of acoustic and vibration data simultaneously by way of sensor fusion, which is hypothesized to provide better machining power prediction capabilities than acoustics or vibration alone [36]. To fill this gap, the study presented herein considers previous vibration to machining power mappings, develops an acoustic-to-machining power prediction sub-model, and compares the machining power prediction accuracy of both single-sensor mappings with the multi-sensor fusion mapping. The decision logic of the multisensory model is then examined using integrated gradients to observe the learned mapping rules [41].

Fundamentals of multi-sensor fusion

The motivating idea of multi-sensor fusion is that decision making ability, e.g., power signal prediction, is improved when information from multiple sensing modalities is used. This is the case since using multiple sensors that represent varying modalities provides a more comprehensive characterization of the process or system being modeled [42]. Fusion can be achieved at the data, feature, and decision levels as shown in Fig. 3 [43].

In data fusion, multi-sensory data is concatenated to form a joint data representation with no loss of the original information. An example is seen in [44] where data from various sensors is fused by making each sensor's data a row of a matrix. However, this approach requires that each sensor's data have the same dimensionality, which may not always be possible without downsampling and data loss.

In contrast, feature fusion combines data representations rather than raw data directly. These features can be handcrafted, statistical, or latent representations attained via ML models. Most CNNs are in this category as described in [45], where convolutional layers are used to transform images of dogs into latent representations which are then merged and concatenated with age, weight, and gender data before being classified using an ANN. Despite this flexibility, feature fusion requires careful feature crafting, and it is difficult to determine if the selected features

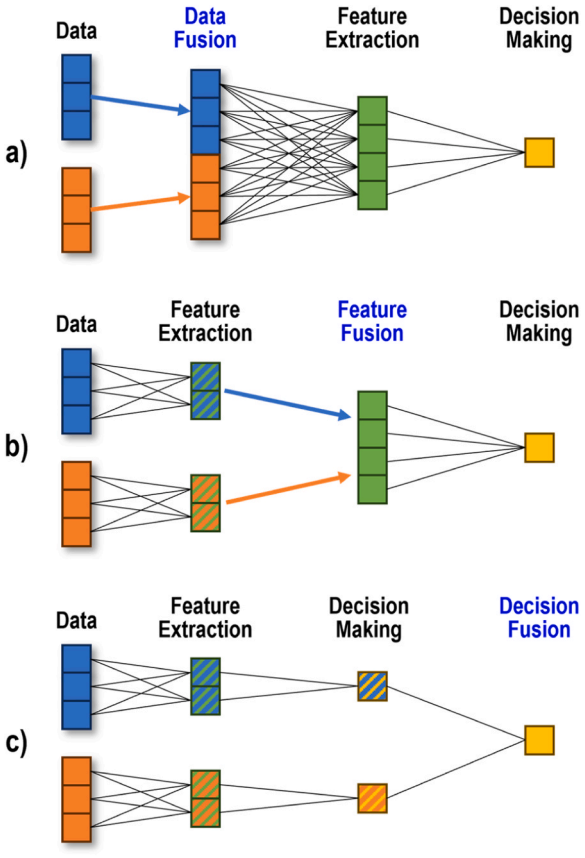

Fig. 3. Schematic representation of a) data fusion, b) feature fusion, and c) decision fusion.

will adequately characterize all possible model inputs.

Decision-level fusion considers the output of several predictive models at once and combines them into a single unified output. Ensemble-based models [46] fall into this category and have seen recent success in manufacturing process modeling [47,48,49,50]. A possible drawback of decision-level fusion models is that the decision logic of the fusion and of the constituent prediction models needs to be optimized concurrently. While some decision fusion train the constituent members in isolation, a preferable approach is to train all members relative to the accuracy of the full model since this will maximize the full model's predictive power rather than only that of the constituent members.

Proposed multi-sensor fusion model and interpretation

An ML model is proposed for decision-level multi-sensor fusion for machining power prediction. Acoustic and vibration signals are passed to the model and transformed into sensor-specific power signal predictions via ANNs inspired by the *DLinear* model architecture described in [35]. This model has been previously shown to outperform transformers, recurrent neural networks, and state-of-the-art variants in time series prediction tasks [51]. Additionally, *DLinear* retains transformers' advantageous ability to consider all input information at once without forgetting previously seen information and is substantially faster to train than sequential and transformer models since it can consider all input data simultaneously and has orders of magnitude fewer trainable parameters than do transformers. The sensor-specific predictions (decisions) are then fused via a convolution operation to arrive at a final power prediction. The model is fully differentiable from the power prediction output back to the sensor inputs, meaning that optimization via backpropagation optimizes all model parameters simultaneously and prediction accuracy of the full fusion model is maximized. This full differentiability also enables use of integrated gradients for model interpretation.

Fusion model

Assume discretely sampled acoustic signal $P_a[t]$, vibration signal $P_v[t]$, and machining power signal $P_m[t]$. Given the multi-sensor fusion model g,

$$g(P_a, P_v; \theta) = \widetilde{P}_m \tag{8}$$

where θ represents trainable model parameters and $\widetilde{P}_m$ is the estimated machining power signal corresponding to P_a and P_v, the objective of this study is to ascertain the parameters, θ^*, which minimize the mean absolute error (MAE) loss, $\mathscr{L}$, between each time step of predicted and actual power signals over a training dataset of size N:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathscr{L}(g; P_m) = \underset{\theta}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{n=1}^{N} \left| P_{m,n} - \widetilde{P}_{m,n} \right| \right)[t] \tag{9}$$

where n is the sample index and $T = \|P_m\|$ where $\|\cdot\|$denotes vector length. This is done using a decision-level fusion model optimized using backpropagation [52], as shown in Fig. 1.

Each sensor's signal is first decomposed into trend and seasonality components as follows:

$$P_{\star,trend} = \mathbf{1}_{l_t} * P_\star \tag{10}$$

$$P_{\star,seas} = P_\star - P_{\star,trend} \tag{11}$$

where $\mathbf{1}_{l_t}$ is a vector of all ones with length l_t, i.e., a uniform filter, $*$ is the convolution operator, and $\star$ denotes the sensing mode (a or v). The convolution is performed with mirrored padding such that $\left\|P_{\star,trend}\right\| = \left\|P_{\star,seas}\right\|$. Eq. (10) extracts the trend by filtering out high-frequency components from $P_\star$ using a moving average. When the trend is subtracted from the original signal, the remainder is, by definition, the high-frequency seasonal components.

The trend and seasonal components are then processed by separate neural networks, h, to predict the trend and seasonal components of the target power signal:

$$h_{\star,trend}\left(P_{\star,trend}; \theta_{\star,trend}, \phi_{\star,trend}\right) = \widetilde{P}^{\star}_{m,trend} \tag{12}$$

$$h_{\star,seas}\left(P_{\star,seas}; \theta_{\star,seas}, \phi_{\star,seas}\right) = \widetilde{P}^{\star}_{m,seas} \tag{13}$$

where $\theta_\star$ represents the parameters of each network and $\phi_\star$ is a nonlinear activation function applied to the output of each neuron. The final sensor-specific power prediction, $\widetilde{P}^{\star}_m$, is found as the sum of the trend and seasonal components:

$$\widetilde{P}^{\star}_m = \widetilde{P}^{\star}_{m,trend} + \widetilde{P}^{\star}_{m,seas} \tag{14}$$

Each $h_\star$ has $\|P_\star\|$input neurons, $\left\|\widetilde{P}_m\right\|$ output neurons, $L_\star$ hidden layers, and $\|P_\star\|$ neurons per hidden layer. This ensures that $\left\|\widetilde{P}^{a}_m\right\| = \left\|\widetilde{P}^{v}_m\right\| = \left\|\widetilde{P}_m\right\|$ and thus each h is trained to output a valid power prediction which aligns with the length of the ground truth signal, P_m.

After obtaining both sensor-specific power predictions, decision fusion model g_f is expressed as

$$g_f(\widetilde{P}^{a}_m, \widetilde{P}^{v}_m; \theta_f) = \widetilde{P}_m \tag{15}$$

This model first combines the incoming power signals using a multi-channel 1-D CNN, q_f, and then uses a final fully-connected layer to fine tune the fused result. First, J trainable 1-D kernels per sensor, $k_{a,j} \in \mathbb{R}^{l_a}$ and $k_{v,j} \in \mathbb{R}^{l_v}$, are synchronously convolved over each intermediate power signal prediction and averaged to merge the signals together:

$$q_f(\widetilde{P}^{a}_m, \widetilde{P}^{v}_m; k_a, k_v) = \frac{1}{J} \sum_{j=1}^{J} \left(k_{a,j} * \widetilde{P}^{a}_m + k_{v,j} * \widetilde{P}^{v}_m \right) \tag{16}$$

The convolutional result is then passed through a final fully connected ANN, h_f, for fine tuning:

$$h_f\left(q_f; \theta_f\right) = \widetilde{P}_m \tag{17}$$

The network has $\left\|q_f\right\|$ input neurons, $\left\|\widetilde{P}_m\right\|$ output neurons, L_m hidden layers, and $\left\|q_f\right\|$ neurons per hidden layer. A convolutional approach was chosen over a summation or fully-connected approach based on experimental findings, which suggested the 1-D convolutional merging yielded the highest machining power prediction accuracy.

Because each of the constituent models of g are fully differentiable, i.e., $h_{a,trend}$, $h_{a,seas}$, $h_{v,trend}$, $h_{v,seas}$, q_f, and h_f, all parameters of g can be optimized at once via backpropagation and (9) can be solved. The resulting model parameters are those which yield the most accurate machining power predictions using the fused acoustic and vibration data.

Model interpretation via integrated gradients

Model interpretation is performed using integrated gradients, which quantify the sensitivity of each of a model's outputs to the value of each of its inputs, i.e., the sensitivity of the power signal output to the acoustic and vibration signal inputs, respectively [41]. The purpose of quantifying the input-output sensitivity of the model proposed in this study is to check for signs of spurious decision logic. Nonspurious logic is indicated by well-distributed integrated gradient values over the acoustic and vibration inputs to the prediction model. In this case, the model is using all information at its disposal to predict the machining power rather than placing inordinate importance on only a few time steps, which is a potential indicator of training data memorization and compromised decision logic.

The integrated gradient, *IG*, is a direct measure of output variable sensitivity with respect to the inputs as quantified by the model gradients, which represent the rate of change of the outputs as a function of the inputs. The underlying intuition is that the influence of an input variable can be quantified by querying the model output with and without the variable being present and observing the difference in output between the two cases, with the assumption that the difference in model output is caused by the variable of interest being omitted. In the case of *IG*, this omission is calculated using integration of partial derivatives, which measure output sensitivity with respect to the inputs, with an interpolator, $\gamma \in [0,1]$ varying each input feature value from 0 to the ground truth value.

For time step t_i of input data $X = [P_a, P_v]$ processed by prediction model g and assessed at output time step t_j, the influence of the i^{th} input on the j^{th} output, $IG_{t_i \to t_j}$,is defined as:

$$IG_{t_i \to t_j}(X;g) = X[t_i] \cdot \int_0^1 \frac{\partial g(\gamma X)[t_j]}{\partial X[t_i]} d\gamma \tag{18}$$

Here, the partial derivative quantifies the instantaneous rate of change of the machining power value at the j^{th} output time step with respect to the acoustic or vibration signal value at the i^{th} input time step. This derivative is integrated using γ as an interpolator to quantify the average contribution of each input time step in the absence ($\gamma = 0$) and presence ($\gamma = 1$) of the input time step's information. Input time steps with little effect on the model output will have approximately equal gradient magnitudes at $\gamma = 0$ and $\gamma = 1$, which will cancel out during integration and yield low *IG* values.

Furthermore, the integration over γ overcomes the fact that the model gradient vanishes as $\gamma \to 1$ and thus assessment of the gradient at only $g(X)$is insufficient to capture the true feature importance. Small perturbations of X ($\gamma \approx 1$), are expected to have little effect on a well-trained $g(X)$ whereas larger perturbations ($\gamma \ll 1$) will be associated with larger gradient values as the network quickly and nonlinearly moves from $g(\boldsymbol{0}) = \boldsymbol{0}$ to $g(\gamma X) \approx \widetilde{P}_m$. This nonlinear gradient response is visualized in [53]. Thus, the integration over γ accumulates the gradients during interpolation to account for the large gradient values when $X \approx \boldsymbol{0}$ and fairly quantify the expected gradient of each ***i***.

The value of $IG_{t_i \to t_j}$ is the expected contribution of input time step *i* to the power signal output value at time step *j*. When multiplied by the original feature values, as shown in (18), the *IG* values yield an additive explanation of the model output:

$$\sum_{t_i} IG_{t_i \to t_j}(X;g) = g(X)[t_j] \tag{19}$$

Thus, input time steps with larger *IG* have a larger effect on the predicted power at time *j* relative to the other time steps.

Since *g* is a multi-input, multi-output (MIMO) function, (18) must be made into a multidimensional form to account for all dependencies between the input and output time steps. Integration over γ is also generally intractable, so the integral is discretized as a sum. The modified form is expressed as:

$$IG(X;g) = \frac{1}{R} \sum_{r=1}^{R} \mathscr{J}\left(g\left(\frac{r}{R}X\right)\right) \cdot X^T \tag{20}$$

where *IG* is a matrix whose entry at position (a,b) is the integrated gradient of the a^{th} power prediction time step with respect to the b^{th} input time step, $\mathscr{J}$ is the Jacobian, and X^T is a column vector of the model inputs. The vector of expected influences for each time step, ξ, is then the column-wise average of the absolute value of *IG*:

$$\xi(X;g) = \frac{\mathbf{1}_{\|g(X)\|} \cdot |IG(X;g)|}{\|g(X)\|} \tag{21}$$

where $\boldsymbol{1}_{\|g(X)\|}$ is a row vector of all ones with length $\|g(X)\|$. Each element of ξ describes how influential each input time step of the acoustic and vibration signals is, on average. Once calculated, ξ can be plotted, and patterns identified which reveal the time steps of the input signal that have the most influence over the power prediction.

Shapley additive explanation (SHAP) values [54] were also considered for feature influence quantification, however the combinatorial nature of SHAP value calculation made them infeasible for the developed fusion model. SHAP value computational complexity scales according to $O(2^N)$ where N is the total number of inputs [55]. The acoustic and vibration signals used for this study each have length 500, so $2^{1000} = 1.08 \times 10^{301}$ calculation iterations would be required to find the SHAP values. Previous work by the authors found that the average computation time per iteration is approximately 1 s on a single-threaded i7 CPU at 2.6 GHz, so computing all iterations would take an impossibly extensive amount of time [56]. Faster SHAP value approximation methods such as KernelSHAP were also considered, but the need to optimize several hyperparameters made them undesirable [55]. Contrastingly, integrated gradients have computational complexity $O(NM)$, where M is the number of outputs. Further, each integrated gradient iteration is computed in 0.01% the time of a SHAP value iteration, making them more feasible for this study. Integrated gradients also use only one tunable hyperparameter (R), making their computation more reliable. Additionally, previous work has shown that integrated gradients approximate Aumann-Shapley values, which quantify feature influence similarly to SHAP values [57,58]. Given these computational and theoretical benefits, integrated gradients are chosen over SHAP values for this work.

Experimental evaluation and model training

The proposed acoustic-vibration sensor fusion method for machining power prediction is validated using an experimentally acquired dataset of in-process signals gathered during turning at varying speeds and feeds.

Machining data collection

A CNC lathe equipped with TiN-coated CNMG432 TCN55 carbide

Table 1
Turning parameters for each cutting experiment.

Exp.	Speed (RPM)	DOC (mm)	Feed (mm/rev)
1	1200	0.635	0.254
2	1200	0.635	0.381
3	1200	0.635	0.508
4	1200	1.270	0.254
5	1200	1.270	0.381
6	1200	1.270	0.508
7	1200	1.905	0.254
8	1200	1.905	0.381
9	1200	1.905	0.508
10	1600	0.635	0.254
11	1600	0.635	0.381
12	1600	0.635	0.508
13	1600	1.270	0.254
14	1600	1.270	0.381
15	1600	1.270	0.508
16	1600	1.905	0.254
17	1600	1.905	0.381
18	1600	1.905	0.508
19	2000	0.635	0.254
20	2000	0.635	0.381
21	2000	0.635	0.508
22	2000	1.270	0.254
23	2000	1.270	0.381
24	2000	1.270	0.508
25	2000	1.905	0.254
26	2000	1.905	0.381
27	2000	1.905	0.508

turning inserts was used to perform 27 cutting experiments, each with a unique combination of speed, feed, and depth of cut (DOC) as listed in Table 1. The insert had a 12.700 mm inscribed circle, a thickness of 4.763 mm, a corner radius of 0.793 mm. The material being turned was AISI 1018 steel with initial nominal diameter of 50 mm and nominal length of 250 mm. Each cutting experiment had a fixed duration of 10 s and was repeated 3 times in a relatively quiet workshop and under chatter-free conditions [24].

The experimental setup and on-machine sensors are shown in Fig. 4. To record tool vibration, a high-sensitivity polyvinylidene fluoride (PVDF) vibration sensor (model LDT0–028 K) is affixed to the tool shank and outputs the dynamic strain of the tool as an analog voltage signal. The machining audio is simultaneously collected using a PmodMIC3 microphone module equipped with an analog-digital-converter (ADC) attached to the carriage and located approximately 100 mm away from the tool-chip interface. Both sensors communicate with a Discovery 2 data acquisition (DAQ) board attached to the carriage and are recorded by WaveForms software at 10 kHz [24]. The PVDF communicates over analog channel while the microphone uses a serial communication protocol (SPI).

Each of the three RMS current phases for the lathe are monitored and summed using a Simple Logger II-AL834 AC current logger with probes affixed directly to the machine's main circuit breaker. Since the machine has a fixed voltage as a lumped system and power is the product of current and voltage, the dynamic current embodies the power demanded by the machining operation plus the power consumption of the machine tool's auxiliary motors, pumps, computers, and lights. This latter power draw can be readily learned by the prediction model as a bias term. The sampling rate was set to 8 Hz, which is the maximum allowable by the logger. The current sample rate is deemed reasonable since machine current changes are attributed mainly to significant changes of the speed and/or torque of the spindle [24].

The DAQ was used to initialize the ADC, SPI, and current logger sampling simultaneously, yielding a common $t = 0$ point for all sensors, and the latency observed between the ADC and SPI data streams was on the order of 1 μs, indicating synchronized audio and vibration data. Further, the data timestamps were double-checked after acquisition to ensure signal alignment. Noise removal is performed following each 10 s acquisition period using a uniform filter of length 7. This filter size was determined experimentally by searching over odd-length filter sizes from 3–51 and selecting the length that yielded the best power signal prediction accuracy. After filtering, the acoustic and vibration signals are divided into 500-sample segments. Each segment is then used to predict the corresponding experiment's 80-sample power signal. A schematic overview of the experimental data acquisition is shown in Fig. 5.

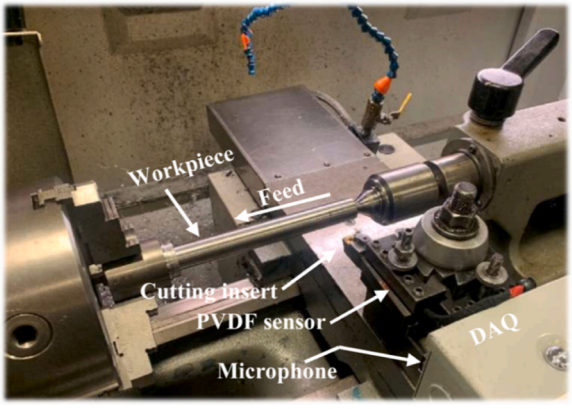

Fig. 4. Experimental setup.
Adapted from [24].

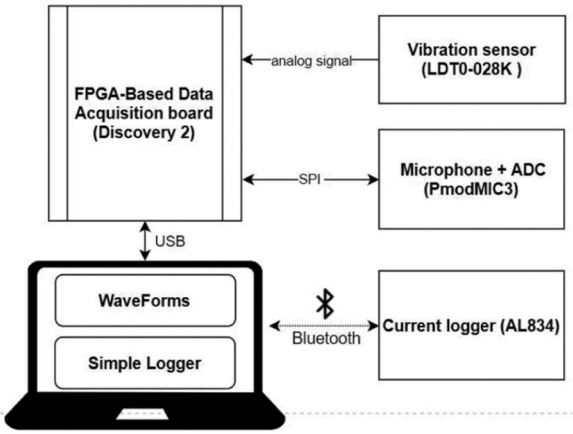

Fig. 5. Experimental data acquisition.
Adapted from [24].

Fusion model architecture and training

A stochastic grid search is performed to find good values for decomposition kernel length l_t, activation function $\phi_\star$, convolutional kernel lengths l_a and l_v, number of convolutional kernels J, hidden layer counts L_a, L_v, and L_m, model optimization algorithm, and training loss function. To reduce the search space, all $\phi_\star$ were identical, $l_a = l_v = l$, and $L_a = L_v = L_m = L$. The parameter search spaces were defined based on similar models in literature and, in the case of the activation functions [59,60,61] and optimizers [62,63,64], demonstrated performance over other approaches.

The search is performed by sampling 1000 randomly chosen parameter combinations and selecting the combination that produced the lowest $\mathscr{L}$ as defined in (9) as the model's parameter set for the study. The search spaces for each variable are defined in Table 2, with the optimal parameter values bolded. No signs of overfitting were observed during training with the selected model parameters [65]. A total of 75% of the experimental dataset is used for training, whereas 15% is used for validation and parameter tuning, and 10% is used for testing. Before acoustic and vibration signal segmentation, all subsets are shuffled and stratified based on their experiment number, as listed in Table 1. Training is performed using Z-score normalized versions of all signals. To contextualize the model's $\mathscr{L}$ from (9) on the testing dataset, testing error is reported in terms of MAPE, defined as:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{n=1}^{N}\left|\frac{P_{m,n} - \widetilde{P}_{m,n}}{P_{m,n}}\right|\right)[t] * 100\% \tag{22}$$

The expected integrated gradients, ξ, are then calculated for the acoustic and vibration inputs to the fusion model, with the number of computation steps, i.e., R in (20), set to 100 based on recommendations from literature [41].

Table 2
Model parameter search spaces; bold indicates values found via stochastic grid search which minimize model $\mathscr{L}$ (MSE = mean squared error).

Parameter	Search space
l_t	{3, 5, **7**}
ϕ	{ReLU, Leaky ReLU, **Mish**}
l	{**1**, 11, 21, 41}
J	{**1**, 3, 5, 7}
L	{1, **3**, 5}
Optimizer	{Adam, **Nadam**, RMSprop}
Loss function	{**MAE**, MSE}

Results and discussion

Multi-sensor fusion power prediction results

Exemplary power prediction results for the fusion model are shown for five evenly spaced MAPE percentiles in Fig. 6. The best result (100th MAPE percentile) exhibits only 0.8% error whereas the worst result (0th percentile) exhibits 6.2% error. The median (50th percentile) is 2.3%. As the percentile decreases, the model predictions remain centered about the ground truth, indicating good trend predictions, but increasingly exhibit sudden spikes, which deviate from the ground truth possibly indicating that the seasonal component could not be adequately learned for these power signals. Nonetheless, these results are initial indicators that the model performs well on the testing data and is not overfit to the training data.

The MAPE distribution for the testing dataset is shown in Fig. 7. The left y-axis is the absolute number of samples in each histogram bin whereas the right y-axis is the cumulative proportion of samples with a MAPE at or below the corresponding MAPE value on the x-axis. The plot shows that more than 80% of the testing dataset MAPE values are $\leq$ 3%, indicating a well-trained model capable of outputting highly accurate machining power predictions. The distribution also has a thin right tail, indicating a rapid drop in probability density as the MAPE exceeds 3%. The average MAPE of the testing data is 2.5%.

Fig. 8 presents violin plots of MAPE for RPM, depth of cut, and feed. The width of each violin is the kernel density estimate [66] of the test dataset MAPE's probability density function, marginalized to include only the selected process parameter, and assessed at the level shown in the x-axis. The mean MAPE of each distribution is plotted as a black dot. The RPM and feed distributions show that the model exhibits very little MAPE variance for the 2000 RPM and 0.508 mm/rev experiments, as evidenced by short and wide violins, but has a much larger and more uniformly distributed variability when predicting the power signal of the other two experimental conditions.

Meanwhile the variability of the depth of cut distributions are more even, with the 0.635 mm depth of cut exhibiting the highest maximum MAPE at 6.2%. The 1.905 mm depth of cut exhibits a noticeably multimodal distribution, as shown by the violin's double hourglass figure. The 0.381 mm feed and 2000 RPM conditions also exhibit multimodal MAPE distributions. Notably, average MAPE consistently decreases as the depth of cut increases, indicating better power signal predictability as the depth of cut is increased.

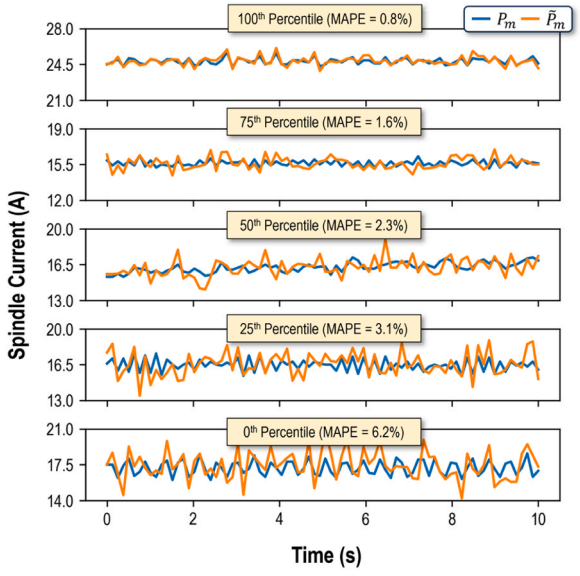

Fig. 6. Machining power prediction results by MAPE percentile; blue is experimentally observed ground truth from testing dataset, orange is fusion model prediction.

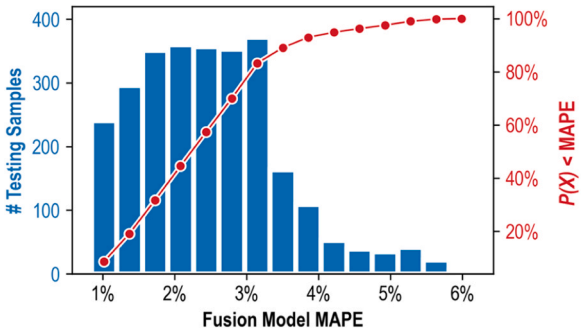

Fig. 7. MAPE distribution calculated on testing dataset; left y-axis denotes absolute number of testing samples; right y-axis denotes cumulative probability that a test dataset sample has a MAPE less than the corresponding MAPE on the x-axis; more than 80% of samples have less than 3% MAPE.

Comparison of multi-sensor and single-sensor prediction

Exemplary results of the acoustic-only and vibration-only power prediction models h_a and h_v are compared with the fusion model g and ground truth power signal, and the result is shown in Fig. 9. Each single-sensor model is retrained using the parameters in Table 2 to directly predict $\widetilde{P}_m$ rather than the intermediate representation $\widetilde{P}_m^{\star}$. As shown in the plots, the acoustic-only prediction model has a slightly lower MAPE than the vibration-based model, but both are much higher than the fusion model MAPE. The average acoustic-only, vibration-only, and fusion model MAPEs over the entire training dataset are 5.6% $\pm$ 0.5%, 8.2% $\pm$ 1.1%, and 2.5% $\pm$ 0.4% respectively, averaged over 5 model reinitializations.

The vibration-based predictions have a more jagged profile, indicating that the higher-frequency components of the power signal are more difficult to predict using vibrations and the proposed prediction model architecture. The same is somewhat true of the acoustic predictions as well, although the jaggedness is much less pronounced. Taken as a whole, these results indicate the advantages of sensor fusion for machining power prediction, as shown by the much-reduced MAPE of the fusion model, as well as elucidate the effectiveness of the proposed noninvasive acoustic-based method of indirect machining power prediction in machining.

Prediction model interpretation results

The average integrated gradient magnitude for the acoustic and vibration inputs to the fusion model, ξ, is shown in Fig. 10. The average was taken over 50 randomly sampled testing datapoints. The shaded region of the graph is $\pm$ 1 standard deviation. The plot demonstrates two notable phenomena: 1) the vibration signal is an order of magnitude more influential on the fusion model output than the acoustic signal, and 2) both sensors' *IG* values oscillate over the input signals.

In the former case, it is unclear why the vibration signal is so much more influential, considering the higher MAPE yielded by the vibration-only model. In terms of process physics, it may be possible that the tool vibration is more influential on machining power predictions since the tool is in direct contact with the workpiece and is thus more sensitive to workpiece condition, which determines power demand. Alternatively, the difference in influence may be a consequence of the convolutional merging layer taking advantage of hidden information in the vibration-specific latent power representation, $\widetilde{P}_m^v$ more so than the information contained in the acoustic-specific representation, $\widetilde{P}_m^a$. This would indicate that the vibration sensor-specific neural network h_v is better trained

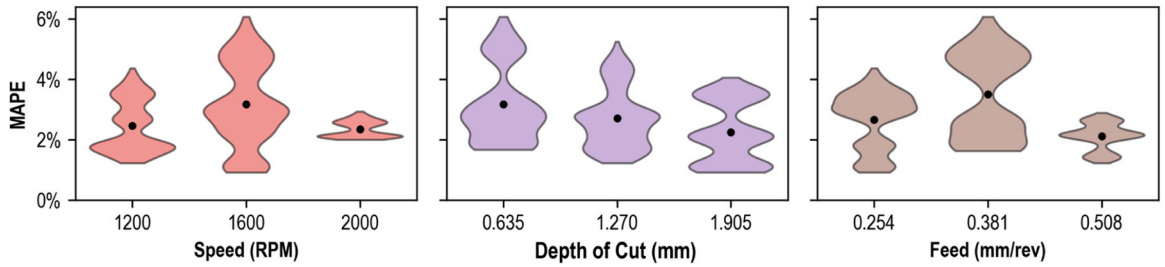

Fig. 8. Violin plots showing kernel density estimate of test dataset MAPE probability density function, marginalized to include only the chosen process parameter, and assessed at the level shown in the x-axis. Narrow violin segments indicate that few samples from the test dataset exhibited the corresponding MAPE on the y-axis and vice versa for wide violin segments. Each violin is clipped at its minimum and maximum MAPE values.

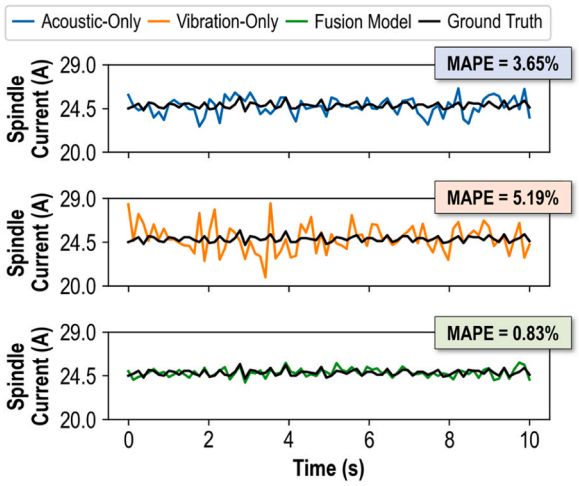

Fig. 9. 90th percentile machining power prediction results for acoustic-only, vibration-only, and sensor fusion models.

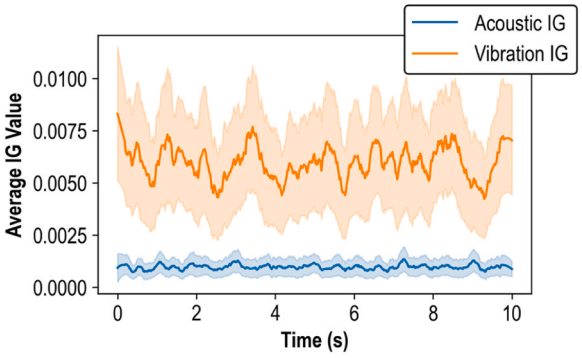

Fig. 10. Average integrated gradients for the acoustic and vibration signal inputs to the fusion model, calculated over 50 testing datapoints; the shaded region represents ± 1 standard deviation; the value of the integrated gradient at each time step is proportional to the average influence of the input signal at that time step on every time step of the machining power prediction.

to output latent representations while the acoustic network h_a is more adept at standalone power prediction. Further investigation of this hypothesis is recommended as future work.

Regarding the latter observation that the *IG* values are oscillatory, this may be an indicator that the model could inadvertently overlook input information. The average integrated gradient, as shown in Fig. 10, is a measure of input feature importance regardless of the feature value. In the case of time-series input, where each input feature corresponds to a time step, this means that temporal regions of low *IG* are deprioritized by the prediction model in favor of the high *IG* regions, regardless of the input signal's information distribution. Thus, if information that strongly predicts machining power happens to be in the low-*IG* areas, e.g., because of a phase shift, the model may produce a suboptimal prediction. Future work will quantify the detriment imposed by oscillatory *IG* values and develop *IG* smoothing techniques to avoid suboptimal process signature prediction.

Conclusion

To simultaneously investigate the potential of utilizing the acoustic and vibration signals as a predictor of machining process signatures and compare single-sensor signature prediction with multi-sensor prediction, this study has developed a multi-sensor fusion model for machining power prediction using both acoustic and vibration signals and compared its power predictions with acoustic-only and vibration-only models. The fusion model produces sensor-specific power predictions using ANNs then fuses them together convolutionally to yield a final unified prediction. The result exhibits 2.5% average error as compared to the experimentally observed ground truth. This contrasts with 5.6% and 8.2% expected power prediction error for acoustic-only and vibration-only models, respectively. Additionally, the fusion model's prediction logic has been elucidated using integrated gradients, revealing that the vibration signal has a pronounced influence on the machining power prediction result as compared to the acoustic signal and an oscillatory trend of input feature importance, both of which merit further investigation.

The study was performed with in-situ feasibility in mind. Power sensing equipment is both expensive to acquire and more complex to install whereas microphones and vibration sensors present themselves as easily accessible alternatives capable of power signal prediction, as demonstrated herein. It is foreseen that the developed multi-sensor fusion system will enable greater awareness of the machine state, by way of machining power signal monitoring, and product and tool state, by enabling predictive metrology using the surface and subsurface information represented by the machining power. The presented fusion approach for process signature prediction may also be adapted to other manufacturing processes such as milling, grinding, or additive manufacturing.

Regarding the use of the proposed model for real-world production, two possible pathways are envisioned:

1) Through transfer learning or other appropriate data-driven methods, the developed method is adapted to a different machining scenario where a different set of process parameters, cutting tool, material, and machine tool are used. The user will perform experimental data acquisition using microphones and vibration sensors in the same way as illustrated in this paper, without the need to purchase expensive power measuring device.
2) The work as presented in this paper is extended to develop a library of pretrained models with varying process parameters, cutting tools, materials, and machine tools. This library can then be queried as

needed by other users. Such an approach is being increasingly discussed for advancement of machine learning accessibility in academic, private, and governmental settings [67].

Beyond investigating the integrated gradient behavior identified earlier, future research will consider integration of thermal power dissipation in the fusion model to more comprehensively characterize the machining energy transfer and achieve more accurate machining power prediction results. Future studies are also anticipated to use a larger number of audio, vibration, and power samples covering a broader range of process parameters, e.g., 1000–5000 RPM cutting speeds, to more comprehensively assess the power prediction model's capability using process parameters seen in industrial applications. Ablation studies will also be performed to determine the necessity of the trend-seasonality decomposition and possibly increase the model's parsimony by removing constituent ANNs. Modeling the noise inherent to the sensor data will also be investigated to remove it from the trend and seasonality components and possibly eliminate the need for this decomposition altogether. Decision level fusion between the developed power prediction model and analytical machining power predictions will be investigated as well.

CRediT authorship contribution statement

Robert X Gao: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jianjing Zhang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Clayton Cooper:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ihab Ragai:** Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Science Foundation under grant CMMI-2040288. Clayton Cooper acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1937968. Robert Gao and Jianjing Zhang also acknowledge support from the NSF Engineering Research Center for Hybrid Autonomous Manufacturing: Moving from Evolution to Revolution (ERC-HAMMER) under award EEC-2133630.

References

[1] Karpuschewski B, et al. Process signatures–knowledge-based approach towards function-oriented manufacturing. Procedia CIRP 2022;vol. 108:624–9. https://doi.org/10.1016/j.procir.2022.01.001.
[2] Brinksmeier E, et al. Process signatures - the missing link to predict surface integrity in machining. Procedia CIRP 2018;vol. 71:3–10. https://doi.org/10.1016/j.procir.2018.05.006.
[3] Pimenov DY, Bustillo A, Wojciechowski S, Sharma VS, Gupta MK, Kuntoğlu M. Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review. J Intell Manuf 2022. https://doi.org/10.1007/s10845-022-01923-2.
[4] Brinksmeier E, Gläbe R, Klocke F, Lucca DA. Process signatures – an alternative approach to predicting functional workpiece properties. Procedia Eng 2011;vol. 19:44–52. https://doi.org/10.1016/j.proeng.2011.11.078.
[5] "Process Signature, *Collaborative Research Center 136*. German Research Foundation. (Accessed 19 September 2022). [Online]. Available: https://www.prozesssignaturen.de/en/about-us/glossary/p/process-signature.
[6] Cooper C, Zhang J, Guo YB, Gao RX. Surface roughness prediction through GAN-synthesized power signal as a process signature. J Manuf Syst 2023;vol. 62:660–9. https://doi.org/10.1016/j.jmsy.2023.05.016.
[7] "Process Quantities, *Collaborative Research Center 136*. German Research Foundation. (Accessed 19 September 2022). [Online]. Available: https://www.prozesssignaturen.de/en/about-us/glossary/p/process-quantities.
[8] Cooper C, Zhang J, Gao RX, Wang P, Ragai I. Anomaly detection in milling tools using acoustic signals and generative adversarial networks. Procedia Manuf 2020:372–8. https://doi.org/10.1016/j.promfg.2020.05.059.
[9] Liu ZY, Guo YB, Sealy MP, Liu ZQ. Energy consumption and process sustainability of hard milling with tool wear progression. J Mater Process Technol 2016;vol. 229:305–12. https://doi.org/10.1016/j.jmatprotec.2015.09.032.
[10] Yu. Pimenov D, Kumar Gupta M, da Silva LRR, Kiran M, Khanna N, Krolczyk GM. Application of measurement systems in tool condition monitoring of milling: a review of measurement science approach. Measurement 2022;vol. 199:111503. https://doi.org/10.1016/j.measurement.2022.111503.
[11] Sealy MP, Liu ZY, Guo YB, Liu ZQ. Energy based process signature for surface integrity in hard milling. J Mater Process Technol 2016;vol. 238:284–9. https://doi.org/10.1016/j.jmatprotec.2016.07.038.
[12] Zhu Z, et al. Specific cutting energy index (SCEI)-based process signature for high-performance milling of hardened steel. Int J Adv Manuf Technol 2019;vol. 103(1–4):1–13. https://doi.org/10.1007/s00170-019-03381-2.
[13] Bustillo A, Pimenov DYu, Mia M, Kapłonek W. Machine-learning for automatic prediction of flatness deviation considering the wear of the face mill teeth. J Intell Manuf 2021;vol. 32(3):895–912. https://doi.org/10.1007/s10845-020-01645-3.
[14] Groover MP. Fundamentals of modern manufacturing: materials, processes, and systems. 4th ed. Hoboken, NJ: J. Wiley & Sons,; 2010.
[15] Oberg E, Jones FD, Horton HL, Ryffel HH, McCauley CJ, Brengelman L. Machinery's handbook. 31st ed. South Norwalk: Industrial Press, Inc,; 2020.
[16] Drouillet C, Karandikar J, Nath C, Journeaux A-C, El Mansori M, Kurfess T. Tool life predictions in milling using spindle power with the neural network technique. J Manuf Process 2016;vol. 22:161–8. https://doi.org/10.1016/j.jmapro.2016.03.010.
[17] Corne R, Nath C, El Mansori M, Kurfess T. Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling. J Manuf Syst 2017;vol. 43:287–95. https://doi.org/10.1016/j.jmsy.2017.01.004.
[18] Wang P, Liu Z, Gao RX, Guo Y. Heterogeneous data-driven hybrid machine learning for tool condition prognosis. CIRP Ann 2019;vol. 68(1):455–8. https://doi.org/10.1016/j.cirp.2019.03.007.
[19] Moliner-Heredia R, Peñarrocha-Alós I, Abellán-Nebot JV. Model-based tool condition prognosis using power consumption and scarce surface roughness measurements. J Manuf Syst Oct. 2021;vol. 61:311–25. https://doi.org/10.1016/j.jmsy.2021.09.001.
[20] Iliyas Ahmad M, Yusof Y, Daud ME, Latiff K, Abdul Kadir AZ, Saif Y. Machine monitoring system: a decade in review. Int J Adv Manuf Technol 2020;vol. 108(11–12):3645–59. https://doi.org/10.1007/s00170-020-05620-3.
[21] Serin G, Sener B, Ozbayoglu AM, Unver HO. Review of tool condition monitoring in machining and opportunities for deep learning. Int J Adv Manuf Technol 2020;vol. 109(3–4):953–74. https://doi.org/10.1007/s00170-020-05449-w.
[22] Abate F, Carratù M, Liguori C, Paciello V. A low cost smart power meter for IoT. Measurement 2019;vol. 136:59–66. https://doi.org/10.1016/j.measurement.2018.12.069.
[23] Zhang P, Gao D, Lu Y, Ma Z, Wang X, Song X. Cutting tool wear monitoring based on a smart toolholder with embedded force and vibration sensors and an improved residual network. Measurement 2022;vol. 199:111520. https://doi.org/10.1016/j.measurement.2022.111520.
[24] Ragai I, Abdalla AS, Abdeltawab H, Qian F, Ma J. Toward smart manufacturing: Analysis and classification of cutting parameters and energy consumption patterns in turning processes. S027861252200067X J Manuf Syst 2022. https://doi.org/10.1016/j.jmsy.2022.04.016.
[25] Sealy MP, Liu ZY, Zhang D, Guo YB, Liu ZQ. Energy consumption and modeling in precision hard milling. J Clean Prod 2016;vol. 135:1591–601. https://doi.org/10.1016/j.jclepro.2015.10.094.
[26] Cooper C, et al. Convolutional neural network-based tool condition monitoring in vertical milling operations using acoustic signals. Procedia Manuf 2020:105–11. https://doi.org/10.1016/j.promfg.2020.07.004.
[27] Li Z, Liu R, Wu D. Data-driven smart manufacturing: tool wear monitoring with audio signals and machine learning. J Manuf Process 2019;vol. 48:66–76. https://doi.org/10.1016/j.jmapro.2019.10.020.
[28] Schmitz TL. Chatter recognition by a statistical evaluation of the synchronously sampled audio signal. J Sound Vib 2003;vol. 262(3):721–30. https://doi.org/10.1016/S0022-460X(03)00119-6.
[29] Luan X, Zhang S, Li G. Modified power prediction model based on infinitesimal cutting force during face milling process. Int J Precis Eng Manuf -Green Tech 2018;vol. 5(1):71–80. https://doi.org/10.1007/s40684-018-0008-7.
[30] Li B, Tian X, Zhang M. Modeling and multi-objective optimization method of machine tool energy consumption considering tool wear. Int J Precis Eng Manuf -Green Tech 2022;vol. 9(1):127–41. https://doi.org/10.1007/s40684-021-00320-z.
[31] Zhang X, Yu T, Dai Y, Qu S, Zhao J. Energy consumption considering tool wear and optimization of cutting parameters in micro milling process. Int J Mech Sci 2020;vol. 178:105628. https://doi.org/10.1016/j.ijmecsci.2020.105628.

[32] Zhang X, Yu T, Zhao J. An analytical approach on stochastic model for cutting force prediction in milling ceramic matrix composites. Int J Mech Sci 2020;vol. 168: 105314. https://doi.org/10.1016/j.ijmecsci.2019.105314.
[33] Kim Y-M, Shin S-J, Cho H-W. Predictive modeling for machining power based on multi-source transfer learning in metal cutting. Int J Precis Eng Manuf -Green Tech 2022;vol. 9(1):107–25. https://doi.org/10.1007/s40684-021-00327-6.
[34] Xu L, Huang C, Li C, Wang J, Liu H, Wang X. A novel intelligent reasoning system to estimate energy consumption and optimize cutting parameters toward sustainable machining. J Clean Prod 2020;vol. 261:121160. https://doi.org/10.1016/j.jclepro.2020.121160.
[35] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting. AAAI 2023;vol. 37(9):11121–8. https://doi.org/10.1609/aaai.v37i9.26317.
[36] Durrant-Whyte HF. Sensor Models and Multisensor Integration. Int J Robot Res 1988;vol. 7(6):97–113. https://doi.org/10.1177/027836498800700608.
[37] R. Yousefzadeh, Deep Learning Generalization and the Convex Hull of Training Sets." Jan. 24, 2021. Accessed: Jun. 13, 2021. [Online]. Available: http://arxiv.org/abs/2101.09849.
[38] P. Barbiero, G. Squillero, and A. Tonda, Modeling Generalization in Machine Learning: A Methodological and Computational Study." arXiv, Jun. 28, 2020. Accessed: Jun. 01, 2022. [Online]. Available: http://arxiv.org/abs/2006.15680.
[39] Wang X, Williams RE, Sealy MP, Rao PK, Guo Y. Stochastic modeling and analysis of spindle power during hard milling with a focus on tool wear. J Manuf Sci Eng 2018;vol. 140(11):111011. https://doi.org/10.1115/1.4040728.
[40] "Specific cutting force." Sandvik Coromant, 2023. Accessed: Oct. 23, 2023. [Online]. Available: https://www.sandvik.coromant.com/en-us/knowledge/materials/specific-cutting-force.
[41] M. Sundararajan, A. Taly, and Q. Yan, Axiomatic Attribution for Deep Networks." arXiv, Jun. 12, 2017. Accessed: Oct. 26, 2023. [Online]. Available: http://arxiv.org/abs/1703.01365.
[42] Castanedo F. A review of data fusion techniques. Sci World J 2013;vol. 2013:1–19. https://doi.org/10.1155/2013/704504.
[43] Wang J, Fu P, Zhang L, Gao RX, Zhao R. Multilevel information fusion for induction motor fault diagnosis. IEEE/ASME Trans Mechatron Oct. 2019;vol. 24(5):2139–50. https://doi.org/10.1109/TMECH.2019.2928967.
[44] Jing L, Wang T, Zhao M, Wang P. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. Sensors Feb. 2017;vol. 17(2):414. https://doi.org/10.3390/s17020414.
[45] Zhang T, et al. A feature fusion method with guided training for classification tasks. Comput Intell Neurosci Apr. 2021;vol. 2021:1–11. https://doi.org/10.1155/2021/6647220.
[46] Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: A review. Eng Appl Artif Intell Oct. 2022;vol. 115:105151. https://doi.org/10.1016/j.engappai.2022.105151.
[47] Cooper C, Zhang J, Gao RX. Error homogenization in physics-informed neural networks for modeling in manufacturing. J Manuf Syst 2023;vol. 71:298–308. https://doi.org/10.1016/j.jmsy.2023.09.013.
[48] Li Z, Liu X, Incecik A, Gupta MK, Królczyk GM, Gardoni P. A novel ensemble deep learning model for cutting tool wear monitoring using audio sensors. J Manuf Process Jul. 2022;vol. 79:233–49. https://doi.org/10.1016/j.jmapro.2022.04.066.
[49] Mongan PG, Hinchy EP, O'Dowd NP, McCarthy CT, Diaz-Elsayed N. An ensemble neural network for optimising a CNC milling process. J Manuf Syst Dec. 2023;vol. 71:377–89. https://doi.org/10.1016/j.jmsy.2023.09.012.
[50] Yang H, Li WD, Hu KX, Liang YC, Lv YQ. Deep ensemble learning with non-equivalent costs of fault severities for rolling bearing diagnostics. J Manuf Syst Oct. 2021;vol. 61:249–64. https://doi.org/10.1016/j.jmsy.2021.09.009.
[51] A. Vaswani *et al.*, Attention Is All You Need." arXiv, Aug. 01, 2023. Accessed: Feb. 07, 2024. [Online]. Available: http://arxiv.org/abs/1706.03762.
[52] Rojas R. The Backpropagation Algorithm. Neural Networks. Berlin, Heidelberg: Springer Berlin Heidelberg; 1996. p. 149–82. https://doi.org/10.1007/978-3-642-61068-4_7.
[53] Sturmfels P, Lundberg S, Lee S-I. Visualizing the Impact of Feature Attribution Baselines. Distill 2020. https://doi.org/10.23915/distill.00022.
[54] Shapley L. Notes on the N-Person Game – II: The Value of an N-Person Game. RAND Corporation; 1951. https://doi.org/10.7249/RM0670.
[55] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions, in Advances in Neural Information Processing Systems 30. California: Long Beach; 2017.
[56] Cooper C, Zhang J, Huang J, Bennett J, Cao J, Gao RX. Tensile strength prediction in directed energy deposition through physics-informed machine learning and Shapley additive explanations. J Mater Process Technol Jun. 2023;vol. 315: 117908. https://doi.org/10.1016/j.jmatprotec.2023.117908.
[57] M. Sundararajan and A. Najmi, The many Shapley values for model explanation." arXiv, Feb. 07, 2020. Accessed: Oct. 26, 2023. [Online]. Available: http://arxiv.org/abs/1908.08474.
[58] Aumann RJ, Shapley LS. Values of Non-Atomic Games. Princeton University Press; 1974. https://doi.org/10.1515/9781400867080.
[59] A.F. Agarap, Deep Learning using Rectified Linear Units (ReLU)." arXiv, Feb. 07, 2019. Accessed: Oct. 26, 2023. [Online]. Available: ⟨http://arxiv.org/abs/1803.08375⟩.
[60] A.L. Maas, A.Y. Hannun, and A.Y. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, 2013.
[61] D. Misra, Mish: A Self Regularized Non-Monotonic Activation Function." arXiv, Aug. 13, 2020. Accessed: Oct. 26, 2023. [Online]. Available: http://arxiv.org/abs/1908.08681.
[62] D.P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization." Jan. 29, 2017. Accessed: Jun. 14, 2021. [Online]. Available: http://arxiv.org/abs/1412.6980.
[63] T. Dozat, Incorporating Nesterov Momentum into Adam, in *Proceedings of the 33rd International Conference on Machine Learning*, San Juan, PR, 2016, p. 4.
[64] G. Hinton, N. Srivastava, and K. Swersky, Lecture 6.5 - rmsprop: Divide the gradient by a running average of its recent magnitude." 2012.
[65] S. Salman and X. Liu, Overfitting Mechanism and Avoidance in Deep Neural Networks." Jan. 19, 2019. Accessed: Sep. 14, 2021. [Online]. Available: http://arxiv.org/abs/1901.06566.
[66] Chen Y-C. A tutorial on kernel density estimation and recent advances. Biostat Epidemiol Jan. 2017;vol. 1(1):161–87. https://doi.org/10.1080/24709360.2017.1396742.
[67] Committee on Options for a National Plan for Smart Manufacturing. Options for a National Plan for Smart Manufacturing. Washington, D.C: National Academies Press; 2023. p. 27260. https://doi.org/10.17226/27260.