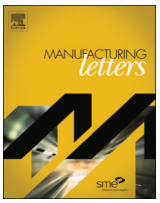
Available online at www.sciencedirect.com

Manufacturing Letters

Journal of Manufacturing Letters 41 (2024) 1515–1525



52nd SME North American Manufacturing Research Conference (NAMRC 52, 2024)

A stethoscope-guided interpretable deep learning framework for powder flow diagnosis in cold spray additive manufacturing

Jiho Lee[a], Semih Akin[b,*], Yuseop Sim[a], Hojun Lee[a], Eunseob Kim[a], Jungsoo Nam[c], Kyeongeun Song[c], Martin B.G. Jun[a,**]

[a] *School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Mechanical, Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
[c] *Smart Manufacturing System R&D Department, Korea Institute of Industrial Technology (KITECH), Cheonan 31056, Republic of Korea*

Abstract

Cold spray (CS) particle deposition, also known as cold spray additive manufacturing, has garnered great attention as an advanced additive manufacturing (AM) and surface deposition technology, facilitating rapid and scalable production of functional parts and surfaces in a solid-state manner. In CS, consistent and precise feeding of functional feedstock powders is crucial for achieving effective particle deposition. However, vibratory-based powder feeders often face challenges associated with powder delivery and powder segregation. This underscores the critical need for a precise diagnostic framework to effectively control powder flow during cold spraying. To this end, the present study proposes a powder flow monitoring framework for the CS process using a stethoscope sound-guided interpretable deep learning (IDL) model. Internal sound data from the vibrated powder feeder is collected through a stethoscope sensor to train a two-stage model. In the first stage, a convolutional autoencoder (CAE) is trained to build an unsupervised learning-based anomaly detector, which identifies classification thresholds based on the receiver operation characteristic curve. In the second stage, a convolutional neural network (CNN) model is trained as the powder flow diagnostic tool by considering process anomalies, namely i) no powder flow; ii) feeder clogging; and iii) no gas flow. The results reveal that the stethoscope sound-guided model achieves a classification accuracy of 95% on the test set, significantly outperforming benchmark utilizing typical external-sound recording microphones in diagnosing CS powder flow. Furthermore, the model is visualized and interpreted by employing t-distribution stochastic neighbor embedding and integrated gradients techniques to enhance reliability of CS powder flow diagnosis. This research highlights the effectiveness of the stethoscope sound-guided IDL model for *in-situ* powder flow monitoring and process diagnosis in the domain of cold spray additive manufacturing, contributing to effective particle deposition.

Keywords: stethoscope sound sensor, interpretable deep learning, multi-stage model, cold spray, powder flow monitoring, additive manufacturing

1. Introduction

Cold spray (CS) particle deposition is an emerging solid-state additive manufacturing technique that facilitates the high-throughput production of functional parts and surfaces. In CS process, micron-scale particles (i.e., typically metal particles in size from 5 to 50 μm) are fed into a converging-diverging nozzle through a powder feeder, where they undergo acceleration to supersonic velocities (>300 m/s) using compressed gases (e.g., air, nitrogen, helium) ([1]-[3]). Upon the impact of these particles onto a target substrate, material consolidation takes place, leading to the formation of a dense metal coating on the surface, all achieved in a solid-state manner [1]. Notably, owing to its unique features (vis., high deposition rate, minimal thermal input, corrosion resistance, material versatility), CS has garnered great attention in the application domains of additive manufac-

* Corresponding author. Tel.: +1-518-276-3244.
** Corresponding author. Tel.: +1-765-491-2793.
E-mail addresses: akins@rpi.edu (Semih Akin)., mbgjun@purdue.edu (Martin B.G. Jun).

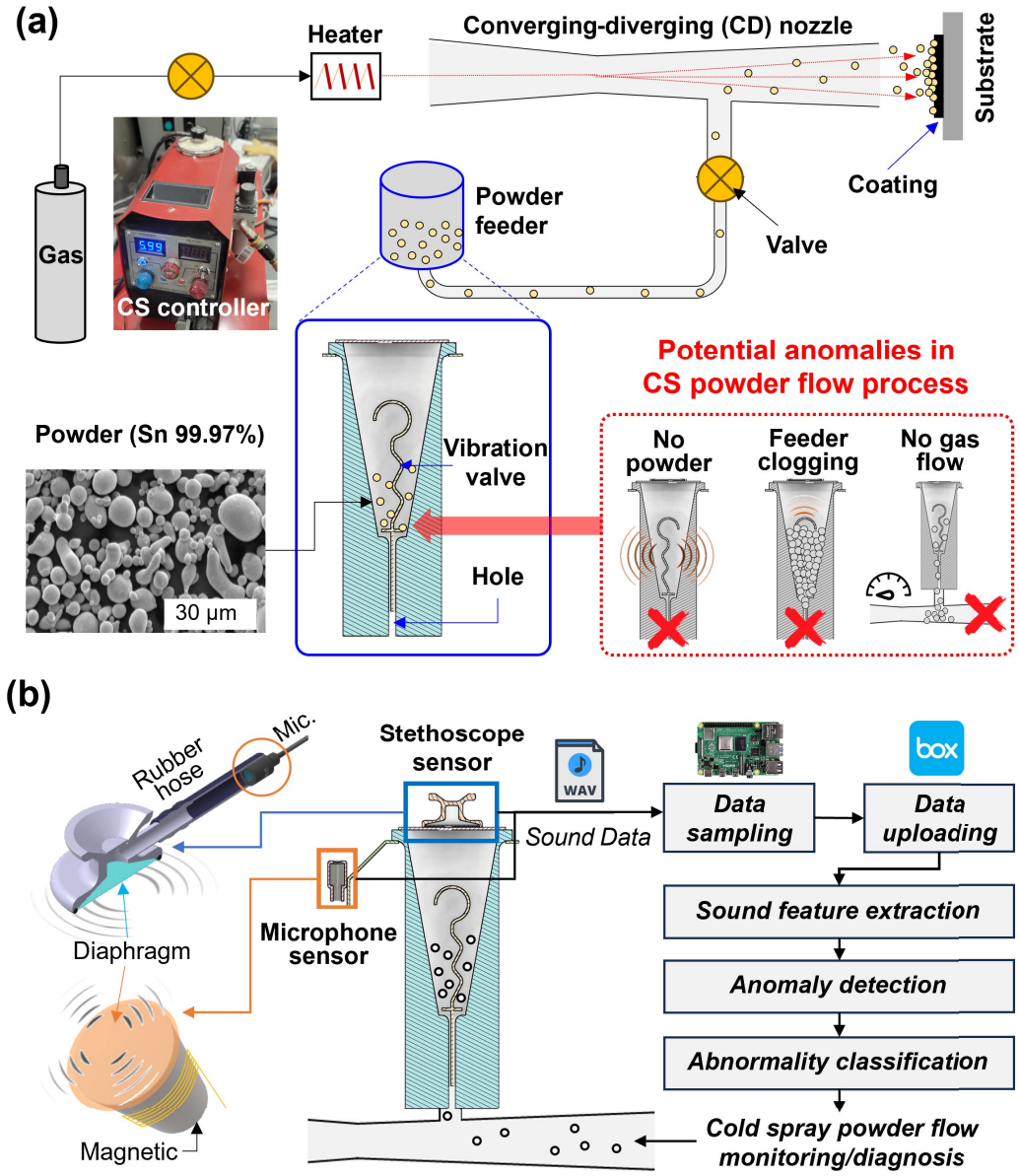


Figure 1. Schematic illustrations of CS powder flow monitoring framework: (a) LPCS system and its potential powder flow anomalies; (b) Sound sensing-based powder flow process monitoring framework.

turing, aerospace, electronics, defense, and energy industries ([4]-[7]).

There are mainly two types of CS systems, namely high-pressure cold spray (HPCS) and low-pressure cold spray (LPCS). The main difference between these setups arises from their operational pressure range (10-50 bar for HPCS and 6-10 bar for LPCS ([8]-[9])) and the location of the powder feed. In HPCS, the powders are fed through the upper stream of the gas flow (i.e., convergent section) using a dedicated high-pressure powder feeder, while LPCS involves downstream powder insertion achieved through the negative pressure (suction) created by the nozzle flow. Owing to its downstream powder injection, the LPCS eliminates the need for complex powder feeders, which offers a compact design, portability, application flexibility, and cost-effectiveness. These features collectively position LPCS as a suitable manufacturing approach for various practical applications, encompassing polymer metallization, antibacterial coating, on-the-site sustainable repairing, etc. ([10]-[13]).

Despite these advantages, LPCS often faces challenges associated with powder delivery and powder segregation due to the limitations of vibratory-based powder feeders widely used in LPCS setups. These shortcomings arise from inconsistent powder flow rate, sensitivity to powder characteristics, limited precision, and limited adaptability ([14]-[15]). Achieving consistent and precise feeding of functional powders is imperative for effective CS deposition. As such, these challenges underscore

the vital need for a precise process control and diagnostic mechanism to effectively manage and diagnose powder flow during CS in a non-invasive manner.

To this end, we introduce an interpretable deep learning (IDL) framework for powder flow monitoring for the CS process. In detail, a stethoscope sensor ([16]-[19]) is deployed to gather the internal sound data from the powder feeder for model training. Unlike traditional external microphones, the stethoscope sensor facilitates high-fidelity sound signal collection from the vibration of the feeder, while effectively surpassing ambient harmonic sound from high-frequency bands and noise from the working environments. Following the stethoscope sensor-based data collection, a convolutional autoencoder (CAE) is trained to build an unsupervised learning-based anomaly detector, identifying classification thresholds. A convolutional neural network (CNN) model is then trained as the powder flow diagnostic tool by considering three types of process abnormalities. The performance of the proposed two-stage model, especially the robustness for ambient noise, is compared with the model trained on the external microphone sound data to underscore the effectiveness of the statoscope sound-guided IDL approach. Finally, the model's decision-making is visualized and interpreted based on the IDL techniques to enhance the reliability of diagnostic results. The novelty of this work lies in the development of a stethoscope sound-guided IDL-based process control framework for powder flow monitoring and diagnosis, which ultimately contributes to effective and precise CS particle deposition.

2. Data Preparation

In a typical LPCS, as shown in Fig. 1(a), powders are fed into the supersonic nozzle (i.e., converging-diverging nozzle) through a vibrating-powder feeder. In detail, a vibration valve inside the hopper (reservoir) vibrates and applies a mechanical force to the powders, allowing the adjustment of the powder-flow rate. As the gas flows and rapidly expands through the nozzle, a significant pressure drop occurs in the divergent section. This phenomenon creates a negative pressure (suction) at downstream of the divergent section, facilitating the feeding of the powders into the nozzle for the spraying process. During this powder-feeding process, the vibration valve generates distinct sound signals to adjust the powder feed rate, which could be potentially used as a sensing metric for powder flow in the CS process. As such, we hypothesize that these sound signals can be collected and utilized to establish correlations between process (powder flow) stability and potential process anomalies. Furthermore, the gathered data can be utilized to train a deep learning model for *in-situ* powder flow monitoring and the detection and diagnosis of process anomalies.

In this regard, we considered three potential powder flow cases (anomalies) in the CS process, namely 1) no powder flow; 2) feeder clogging; and 3) no gas flow (see Fig. 1(a), below panel). To elaborate, the first case (no powder flow) can occur when the vibration valve is operational with compressed gas flow but there is no powder inside the hopper. High-fidelity detection of this case is particularly important to precisely terminate the spraying process to safeguard the CS equipment from potential damage. The second case (feeder clogging) arises when the feeder is filled with high-density powders above a threshold limit, restricting the vibration of the valve. In this scenario, the hopper was filled up to the top of the vibration valve at the beginning of the process, ensuring that the powder weight (inertia) is sufficient to clog the feeder. This overloading creates significant inertia on the vibration valve, impeding the powder feeding to the nozzle, thereby resulting in a distinct sound signal from the valve. This condition can eventually lead to the malfunction of the vibration valve due to excessive load. While this risk can be mitigated by filling the feeder with a certain amount of powder before spraying, detecting the over-loading remains valuable to prevent potential mechanical and electrical failures by detecting the compulsion of the valve. Lastly, irregularities in the driving gas flow can influence powder flow characteristics. Specifically, in the absence of gas flow, the valve may persist in vibrating, thereby causing the powders within the powder hopper to be pushed by gravitational effects. However, without the driving gas flow, there will be no suction to initiate the spraying process for powders. Consequently, this situation can eventually lead to agglomeration and segregation of the powders inside the tube, causing them to adhere to the tubing walls due to the cohesive forces (e.g., Van der Waals forces) [21]. Taken together, it is crucial to detect these anomalies during the CS in an unintrusive manner with high-fidelity precision.

To achieve this, in Section 2.1, we collect the vital sound data arised from the vibration valve through a stethoscope sensor (MDF Instruments USA LLC, Model MDF747-BO) attached to the cap of the powder feeder of the LPCS (BaltiCold Spray LTD, Model CSM 108.2). In these data collection experiments, quasi-spherical shaped tin (Sn) powders with a size range of 10-45 μm (see the SEM image in Fig. 1(a)). was sprayed using a constant gas (gauge) pressure of 0.7 MPa at room temperature without pre-heating the driving gas. The mean particle diameter (d_{50}) of the Sn powders is around 17 μm. In Section 2.2, the collected sound data is utilized to establish correlations for the aforementioned anomalies, with the ultimate goal of feature extraction for modeling purposes.

2.1. Sound data collection

Two types of sound sensors: (*i*) a typical external microphone; and (*ii*) an internal stethoscope, were installed on the powder feeder of the CS system as shown in Fig. 1(b). In the first case (external microphone), a USB microphone (Fifine Microphone, Model K053) serves as a typical type of sound sensor, capturing the auditory perceptive region with a frequency range spanning from 20Hz to 20kHz. This microphone converts diaphragm vibration in response to sound waves into electrical signals, which provides digital information on the acoustic environment. Meanwhile, the stethoscope sensor is specialized in detecting the internal sounds of machines through subtle vibrations on the attached surface and is characterized by its resistance to ambient environmental noise. This is particularly valu-

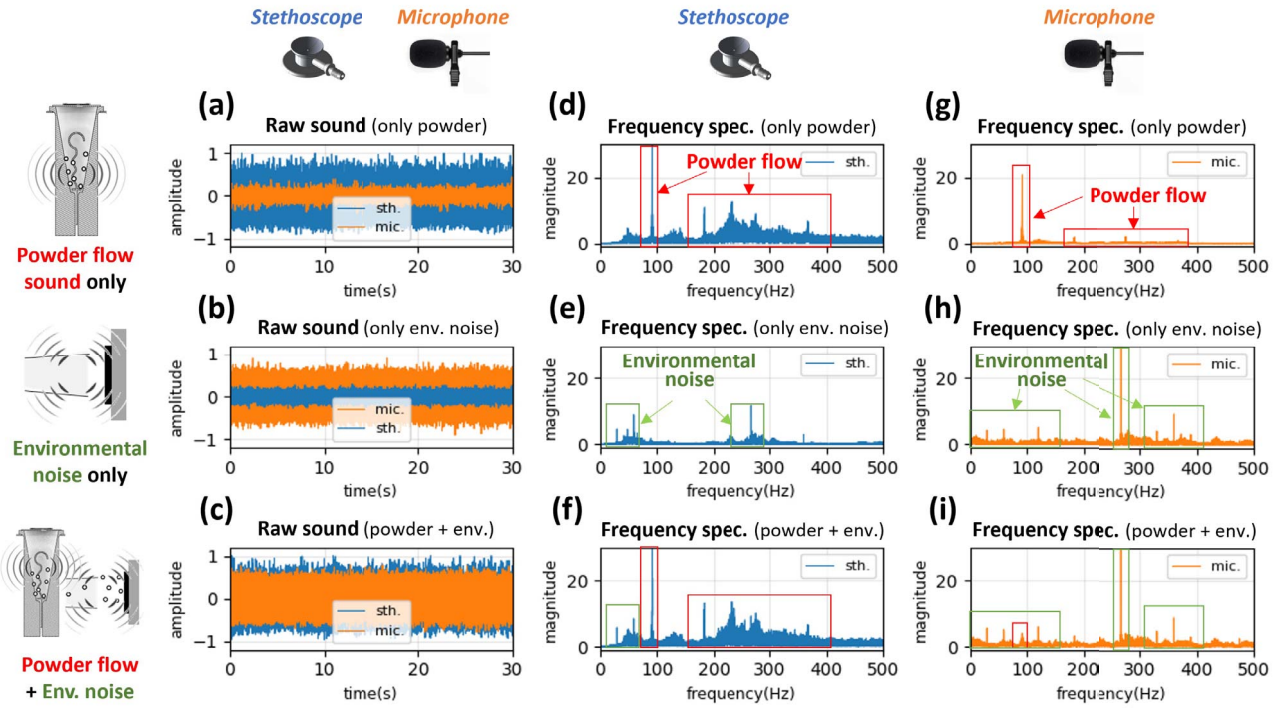


Figure 2. Sound signals by different sensor types and sound sources. (a-c) Raw sound signals for (a) powder flow, (b) environmental noise, and (c) a combination of both. (d-f) Frequency spectrums from the stethoscope sensor for (d) powder flow, (e) environmental noise, and (f) both combined. (g-i) Frequency spectrums from the microphone sensor for (g) powder flow, (h) environmental noise, and (i) both combined.

able for capturing sound signals in extremely noisy environments (> 80 dB), such as cold spraying. This sensor is designed to convert analog sound signals captured by a stethoscope head into digital data through a USB microphone (the same model as the first sensor) connected via a rubber hose. To minimize interference from external noise and concentrate on the internal sound for powder flow, the stethoscope sensor was deployed on the top surface of the powder feeder and secured with tape. Both sensors were linked to a single edge device (Raspberry Pi 4B), equipped with a data acquisition program developed in Python, to collect the synchronized sound data from both sensors. The sampling frequency was set to 48kHz, in line with the Nyquist–Shannon sampling theorem [22], to encompass the maximum audible frequency range, ≈ 24kHz.

To verify the sensors' ability to capture powder flow sounds amidst environmental noise from surrounding machinery, such as dust collectors and compressed air, sound data were initially collected under three distinct scenarios: (i) when only the powder feeder is operational (no external sound source), (ii) when only environmental noise sources (i.e., downdraft table) are active, and (iii) when both powder feeder and environmental noise sources are simultaneously active. Given that particles primarily collide on-the-fly rather than inside the powder hopper, collecting the sound signal from the vibrating valve of the hopper can be utilized to monitor the aforementioned anomalies. Additionally, environmental noise refers to the noise generated by the downdraft table (≈ 80 dB). According to the OSHA standards, CS equipment should be equipped with a dust collector or downdraft table to safely handle over-sprayed micron-scale metal particles ([23]-[24]).

Fig. 2(a-c) displays a 30-second length frame of raw sound wave signals recorded by stethoscope and microphone sensors under the above-mentioned three conditions. The microphone sensor recorded the environmental noise as more prominent compared to the isolated powder flow sound, which is consistent with human auditory perception. Remarkably, the stethoscope sensor captured the internal sound of powder flow, even over ambient noise. This distinction is further elucidated in the frequency spectrums (see Fig. 2(d-i)) derived via the Fast Fourier Transform (FFT). The spectra for both sensors during powder flow in Fig. 2(d, g)) reveal a prominent characteristic frequency peak at around 90Hz, which corresponds to the vibration frequency of the vibratory valve adjusted according to the power level of the powder feeder. The environmental noise spectra in Fig. 2(e, h) show characteristic peaks at around 270Hz, with fluctuating frequency bands below 120Hz and above 300Hz. Notably, the stethoscope's spectrum during the recording of both powder flow and environmental noise is significantly composed of powder flow frequencies (see Fig. 2(f)). On the other hand, the microphone's spectrum is more significantly affected by environmental noise (Fig. 2(i)). As such, the sound spectra results indicate that the stethoscope sensor provides more distinctive and permeable sound information and patterns for monitoring the powder flow state, even in the inherent noisy environment of CS process.

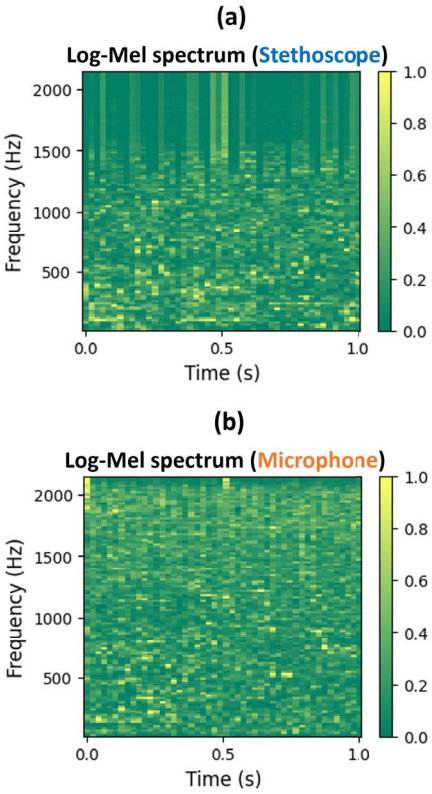


Figure 3. Log-Mel spectra extracted from 1-second frame of (a) stethoscope sound and (b) microphone sound

Subsequently, four sets of sound data were collected under varying powder flow conditions. The first dataset was obtained by recording the sound of normal powder flow state sprayed on the substrate (i.e., aluminum plate), which is not physically connected to the CS setup. The remaining datasets were collected during three different abnormal states: (*i*) no powder flow, where the powder feeder was empty, (*ii*) feeder clogging, characterized by a jammed feeder hole and a stuck shaking tip, and (*iii*) no gas flow, preventing the powder from being sprayed due to the absence of gas supply (see Fig. 1(a)). Each dataset comprises 30-second length of sound signals recorded by both stethoscope and microphone sensors to train the powder flow state diagnostic models.

2.2. *Sound feature extraction*

Feature extraction is an essential step for effective training of deep learning models, especially when working with a limited dataset. In this study, we have selected the Log-Mel spectrum as the feature for our sound data to train the diagnostic models. The Log-Mel spectrum is a powerful representation that captures the essence of sound signals in a format suitable for human auditory sensitivity. This is because the Mel-scale emphasizes finer details at lower frequencies, where the human ear is more sensitive, and wider bands at higher frequencies. Recent studies have presented the effectiveness of Log-Mel spectrum as features in training deep learning models for sound recognition ([25]-[26]). This characteristic also aligns well with sound monitoring applications, considering that the source frequencies produced by mechanical systems are relatively low compared to the high sampling frequency of sound sensors, such as 48 kHz in our case.

To extract the Log-Mel spectrum from the powder flow sound data, we first segmented each 30-second sound dataset into 1-second frames, shifting the start of each window by 0.1 seconds to create overlapping frames. Each sound frame was then converted into a spectrum via the Short-Time Fourier Transform (STFT) with a Hamming window. Subsequently, 40 Mel filter banks were applied to the STFT spectra to extract the Mel spectrum, focusing on a frequency range from 20 Hz to 4000 Hz, which is a suitable range for capturing relevant sound characteristics. The relationship between Mel and frequency (f) is expressed in Eq. (1):

$$Mel(f) = 2595 log_{10}(1 + \frac{f}{700}) \quad (1)$$

Each output of the filter bank was then converted to the decibel scale to obtain the Log-Mel spectrum, followed by a normalization process between 0 and 1. This normalization is crucial as it ensures that the model is not biased by variations in absolute signal amplitude. Fig. 3 shows examples of Log-Mel spectrums from stethoscope and microphone sound frames in normal powder flow states. The spectrum from the stethoscope sound emphasizes the low-frequency areas corresponding to the powder feeder vibration, while the microphone sound spectrum shows higher sound intensity across all frequency bands. These results indicate that the microphone is more susceptible to environmental noise from various sources than the stethoscope sensor, underscoring the high-fidelity performance of stethoscope-based CS monitoring. The resulting Log-Mel spectrum feature sets serve as the input features for the diagnostic models discussed in the subsequent chapter.

3. Model Construction

The powder flow diagnostic model in this study was constructed using a two-stage deep learning strategy as depicted in Fig. 4(a). Initially, the unsupervised model acts as an anomaly detector by learning from a dataset consisting only of normal conditions. It detects anomalies through the significant deviation in reconstruction error, indicating conditions diverging from the trained normal state. Following this, a supervised model, trained on datasets of three abnormalities, classifies the detected anomaly data into specific abnormal types, thereby diagnosing the powder flow state in the CS process. This hybrid method was chosen considering the practical difficulty of obtaining a comprehensive dataset that covers every case of anomalies. Recent studies have presented the effectiveness of unsupervised models in anomaly detection when sufficient abnormal data sampling is limited ([27]-[28]). While unsupervised models for anomaly detection and supervised models for abnormality classification have individually demonstrated their utility in existing studies, the strategic integration of these

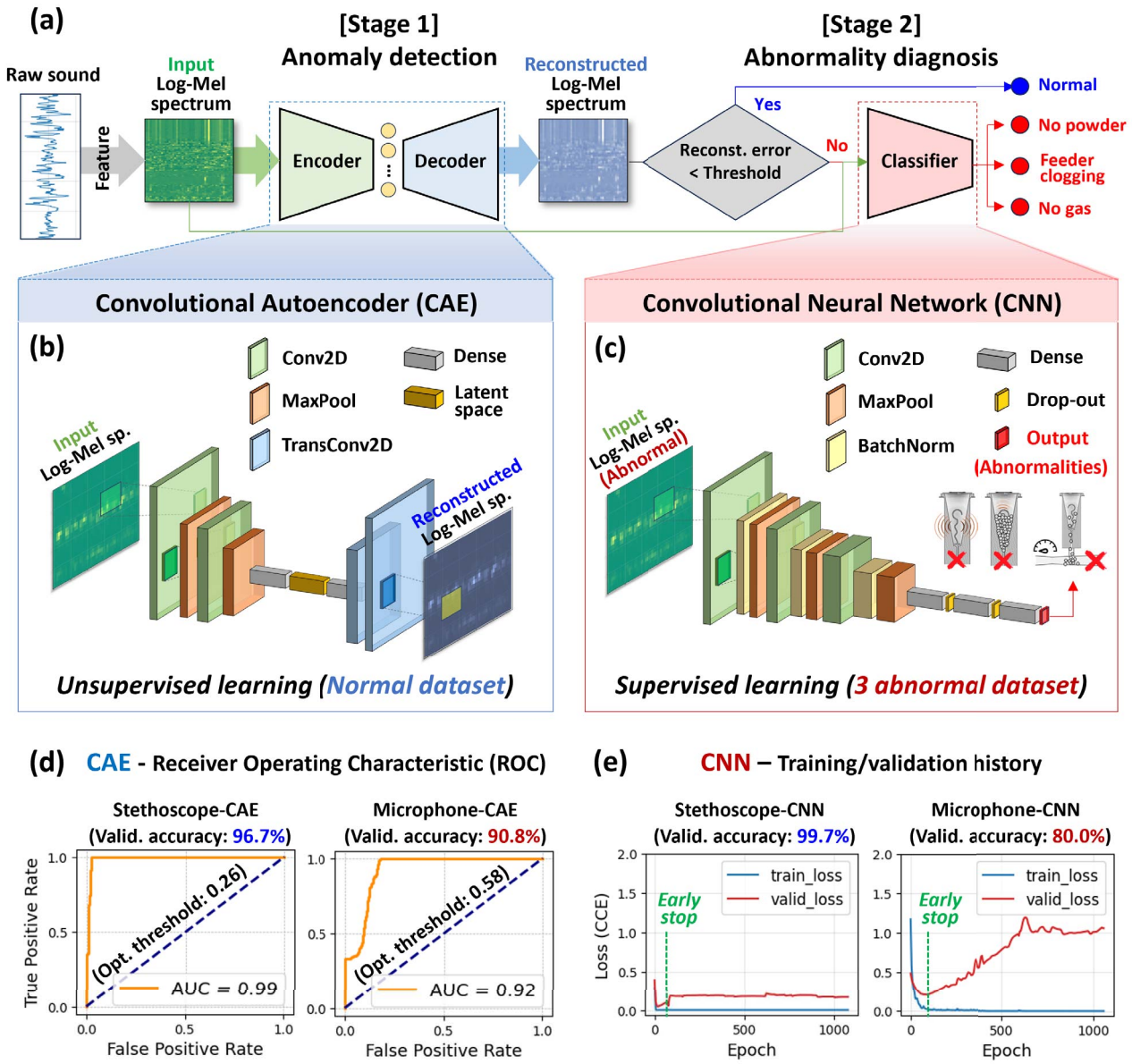


Figure 4. CS Powder flow diagnostic framework. (a) Schematic diagram of a two-stage powder flow diagnosis framework. (b) Architecture of a convolutional autoencoder (CAE)-based anomaly detector. (c) Architecture of a convolutional neural networks (CNN)-based abnormality classifier. (d) Receiver operating characteristic (ROC) curves of CAE models. (e) Training history of CNN models.

two models into a coherent framework for process monitoring presents a novel approach. To clarify the purpose of each stage in this paper, 'anomaly' denotes any variation from the normal condition, while 'abnormality' specifies the particular kind of abnormal state detected.

3.1. Anomaly detector construction

As depicted in Fig. 4(b), the anomaly detector within the powder flow diagnostic model was built by convolutional autoencoder (CAE). The specific type of architecture was chosen for its unsupervised learning capability with our sound data. An autoencoder is a neural network that encodes input data into a compressed representation and subsequently reconstructs the output to closely resemble the original input. Convolutional layers are utilized to better extract spatial features from 2-D input data—in this case, the Log-Mel spectra of sound signals. The convolutional layers apply multiple filters, also known as kernels, to the input data to create feature maps that highlight important patterns within the data, via the process described by Eq. (2):

$$F_{ij} = \sigma(b + \Sigma_m \Sigma_n I_{(i+m)(j+n)} \cdot K_{mn}) \quad (2)$$

where F_{ij} is the feature map value at position (i,j), σ is the activation function, b is the bias, I is the input, and K is the kernel of size $m \times n$. Max-pooling layers follow each convolutional layer

Table 1. Grid search results for stethoscope-CAE (s-CAE) and microphone-CAE (m-CAE) models.

Hyper-parameter	Search Level (Range)	Best Case	
		s-CAE	m-CAE
Dimension of latent space	10 (30-300)	60	120
Number of conv. layers	5 (1-5)	2	3
Size of kernels	3 (2-4)	(3x3)	(3x3)
Learning rate (Adam opt.)	3 (0.001-0.1)	0.05	0.001
Activation function	2 (ReLU, Sigmoid)	ReLU	ReLU
Avg. of MAE		0.2643	0.4219

Table 2. Grid search results for stethoscope-CNN (s-CNN) and microphone-CNN (m-CNN) models.

Hyper parameter	Search Level (Range)	Best Case	
		s-CNN	m-CNN
Number of conv. layers	5 (1-5)	3	4
Number of dense layers	5 (1-5)	3	3
Size of kernels	3 (2-4)	(3x3)	(3x3)
Learning rate (Adam opt.)	3 (0.001-0.1)	0.001	0.001
Activation function	2 (ReLU, Sigmoid)	ReLU	ReLU
Final validation accuracy		99.7%	80.0%

to downsample the feature map, retaining the prominent features. The decoder part of the CAE, in contrast to the encoder, uses transposed convolutional layers to upsample the feature maps, and this upsampling operation is denoted by Eq. (3):

$$F'_{ij} = \sigma(b' + \Sigma_m \Sigma_n O_{(i-m)(j-n)} \cdot K'_{mn}) \quad (3)$$

where F'_{ij} is the upsampled position (i,j), O is the output from the previous layer, and K' is the transposed convolution kernel. In short, the CAE model includes an encoder network with convolutional layers and a decoder network with transposed convolutional layers. The number of nodes in the latent space, which is the bottleneck of the CAE, is a critical hyperparameter that controls the reduced dimensionality of the input data.

In this study, two CAE models were constructed to compare performance based on the type of sensor used for sound collection: one trained on stethoscope-sound data and the other on microphone-sound data. For each model, 80% of normal dataset (24 seconds, resulting in 240 Log-Mel spectra) was used for training and the remaining 20% reserved for validation. A grid search was conducted to identify the best CAE model that minimizes reconstruction error of the Log-Mel spectra, using mean absolute error (MAE) between the original and reconstructed spectra as the evaluation metric. A total of 900 combinations of hyperparameters were tested for both the stethoscope-CAE (s-CAE) and microphone-CAE (m-CAE) models. Table 1 outlines the hyperparameters involved in the grid search and presents the optimal combination found for each CAE along with the best models' average MAE.

To determine the optimal threshold for anomaly detection based on the reconstruction error from the two above best models, the receiver operating characteristic (ROC) curve was employed. This curve plots the true positive rate (correctly predicting an actual anomaly) against the false positive rate (incorrectly predicting normal as an anomaly) at various threshold settings, offering a method to assess the balance between sensitivity and specificity. The ROC curve for each s-CAE and m-CAE model is shown in Fig. 4(d). The ROC analysis determined an optimal threshold for anomaly detection at 0.2619 for the s-CAE and 0.5834 for the m-CAE, with the lower threshold of the s-CAE reflecting its superior ability to reconstruct normal data as demonstrated in Table 1. Additionally, when validating anomaly detection across both normal and abnormal datasets, including the data used for training, the s-CAE model achieved a slightly higher accuracy of 96.7% compared to the m-CAE model at 90.8%.

3.2. Abnormality classi er construction

In the next phase of our diagnostic model, a convolutional neural network (CNN), a popular deep learning architecture, was employed as the abnormality classifier, as depicted in Fig. 4(c). Unlike the CAE used for anomaly detection, the CNN in this stage is designed to categorize specific abnormal conditions within the powder flow process. The architecture of the CNN incorporates batch normalization after each convolutional layer to accelerate training and improve performance by stabilizing learning. Additionally, dropout layers are strategically placed in the dense layers to prevent overfitting by randomly omitting subset of features during training. These elements are critical for enhancing the model's generalization capability, ensuring robust classification under various operational conditions.

Similar to the CAE model, grid search was conducted to find the optimal hyperparameter combinations for the best CNN models, the stethoscope-CNN (s-CNN) and microphone-CNN (m-CNN). For training, 80% of the data frames were randomly selected from each of the three types of abnormal datasets, and the remaining 20% were used as validation data to measure the tri-classification accuracy across 450 hyperparameter combinations. Table 2 displays the grid search results for the s-CNN and m-CNN models, along with the validation accuracy for each

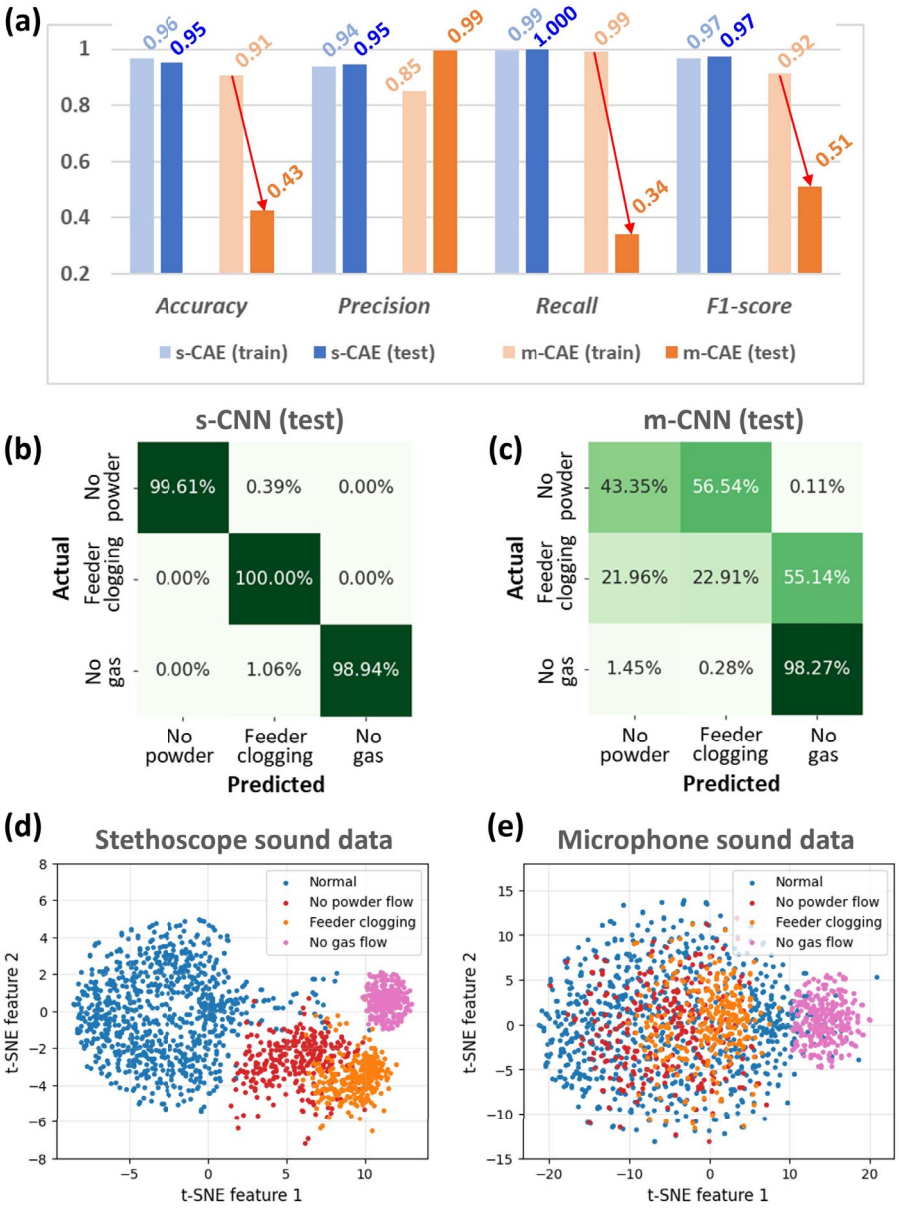


Figure 5. CS Powder flow diagnostic results and data visualization. (a) Anomaly detection robustness evaluation results. (b-c) Confusion matrices of abnormality classification by (b) stethoscope-CNN and (c) microphone-CNN. (d-e) Data visualization for (d) stethoscope sound and (d) microphone sound.

best model. Notably, just as the best m-CAE model from Table 1 required a relatively deeper network, the best m-CNN model also demanded more complexity compared to the s-CNN model. This could be interpreted as needing more trainable parameters to discern meaningful patterns from microphone-captured powder flow sounds masked by environmental noise, as observed in Figs. 2 and 3. Consequently, as seen in Fig. 4(e), there was a clear tendency for overfitting, with a significant drop in performance on validation data compared to the training data during the learning process.

4. Result and Discussion

4.1. Model robustness evaluation

As mentioned in the introduction, the robustness of models built on sound data—sensitive to noise—is a crucial factor determining the performance of process monitoring. Therefore, in this study, a separate dataset was collected to evaluate the robustness of powder flow diagnostic performance. The normal test data were sampled for one minute while 75 g of tin powder were sprayed without any issues. Similarly, data under abnormal conditions were also collected for one minute each. The datasets were segmented into 1-second frames and then trans-

formed into Log-Mel spectra, which served as input features for our powder flow diagnostic models.

In the first stage, the performance of anomaly detection was evaluated based on key binary classification metrics, namely accuracy, precision, recall, and F1-measure, which are derived from the classification outcomes: (*i*) True positive (TP) when the model accurately identifies anomaly data as such; (*ii*) True negative (TN) when normal data is correctly identified as normal; (*iii*) False positive (FP) when normal data is mistakenly identified as abnormal; and (*iv*) False negative (FN) when abnormal data is incorrectly labeled as normal. The calculations for these metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

Accuracy in Eq. (4) represents overall classification results and is typically the main performance index. Precision in Eq. (5) indicates the ratio of actual to predicted anomalies and is more reliable when normal data significantly outnumbers anomaly cases, which is not the case here. Recall in Eq. (6) focuses on accurately classifying abnormal data, crucial in process monitoring applications where missing an anomaly state can be critical. F1-score in Eq. (7) is the harmonic means of precision and recall, providing a balance between the two.

Fig. 5(a) demonstrates the results of evaluating the anomaly detection performance of the s-CAE and m-CAE models based on both training and test data. Notably, the performance of the m-CAE model on the test data significantly declined compared to its performance on training data, especially in terms of recall, a measure of accurately detecting anomalies, which dropped to 34%. This decrease is interpreted as being directly linked to the characteristics of the microphone sensor sound, which is highly influenced by environmental noise in the CS system. In contrast, the stethoscope sensor-based s-CAE model exhibited high anomaly detection performance nearly identical to that seen in training data evaluations, particularly notable in its recall on test data, which was confirmed to be 100%.

Fig. 5(b-c) presents confusion matrices evaluating the classification performance of the s-CNN and m-CNN models, respectively, on the test abnormality dataset. Similar to the first stage, the performance of the microphone sound-based model (m-CNN) is noticeably inferior, particularly in distinguishing between conditions of empty powder feeder (no powder flow) or feeder clogging. To analyze the reasons behind this performance disparity, the following sections will visualize the sound data and interpret the basis of the models' classifications.

4.2. Model visualization and interpretation

The CAE employed for unsupervised anomaly detection of our powder flow diagnostic model is also often utilized for effectively reducing the dimensionality of high-dimensional data. Particularly, when combined with the t-distributed stochastic neighbor embedding (t-SNE) technique, it demonstrates excellent performance in visualizing data in a two- or three-dimensional space that is perceptible to humans [29]. The t-SNE is adept at visualizing data by reducing its dimensionality to two or three dimensions while it maintains the local structure of data in this lower-dimensional representation. The algorithm first calculates pairwise similarities in the high-dimensional space, assigning higher probabilities to closely located points, based on Eq. (8):

$$P_{i|j} = \frac{exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\Sigma_{k \neq i} exp(-\|x_i - x_j\|^2/2\sigma_i^2)} \tag{8}$$

where $P_{i|j}$ is the conditional probability, representing the similarity of data point x_j to x_i, σ_i is the Gaussian variance determined by the perplexity setting in the t-SNE algorithm, reflecting the effective neighborhood size around data point x_i. It then seeks to replicate this similarity distribution in the reduced space by minimizing the Kullback-Leibler (KL) divergence between the distributions in both spaces. This optimization, typically achieved through gradient descent, ensures that similar objects are close together and dissimilar ones are far apart in the reduced space.

In this study, the features output from the latent space, by inputting the Log-Mel spectra into the CAE, were visualized on a 2-D plane by reducing their dimensionality using t-SNE. Fig. 5(d-e) displays the visualized stethoscope and microphone sound data based on the s-CAE and m-CAE models. Each point represented on the t-SNE feature plane corresponds to a Log-Mel spectrum sample from a 1-second sound frame, allowing for visual recognition of the similarities in sound according to the state of powder flow. Notably, in Fig. 5(d), distinct clusters can be easily observed for normal powder flow and the three types of abnormalities in the stethoscope sensor data; it supports the high performance of the s-CAE and s-CNN models shown in Fig. 5(a-b). Some normal data points being close to the 'no powder flow' state may indicate that the sound of a feeder nearly empty is similar to the sound of an entirely empty one. In contrast, Fig. 5(e)'s visualization aligns with the test results of the microphone sound-based models failing to detect and classify the powder flow anomalies. The similarity between the 'no powder flow' and 'feeder clogging' sound samples, almost merging into one cluster as opposed to the separately clustered 'no gas flow' sound, is interpreted as the basis for the abnormality classification results shown in Fig. 5(c).

For the CNN models employed to classify abnormalities, the application of the Integrated Gradients (IG) technique enables the interpretation of the classification basis. IG, gaining significant attention in the field of explainable artificial intelligence (XAI), is a deep learning model interpretation method commonly used in XAI [30]. It works by attributing the model's prediction to its input features based on Eq. (9):

$$IG_i = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(\, x_i' + \alpha \times (x_i - x_i')\,)}{\partial x_i} \, d\alpha \tag{9}$$

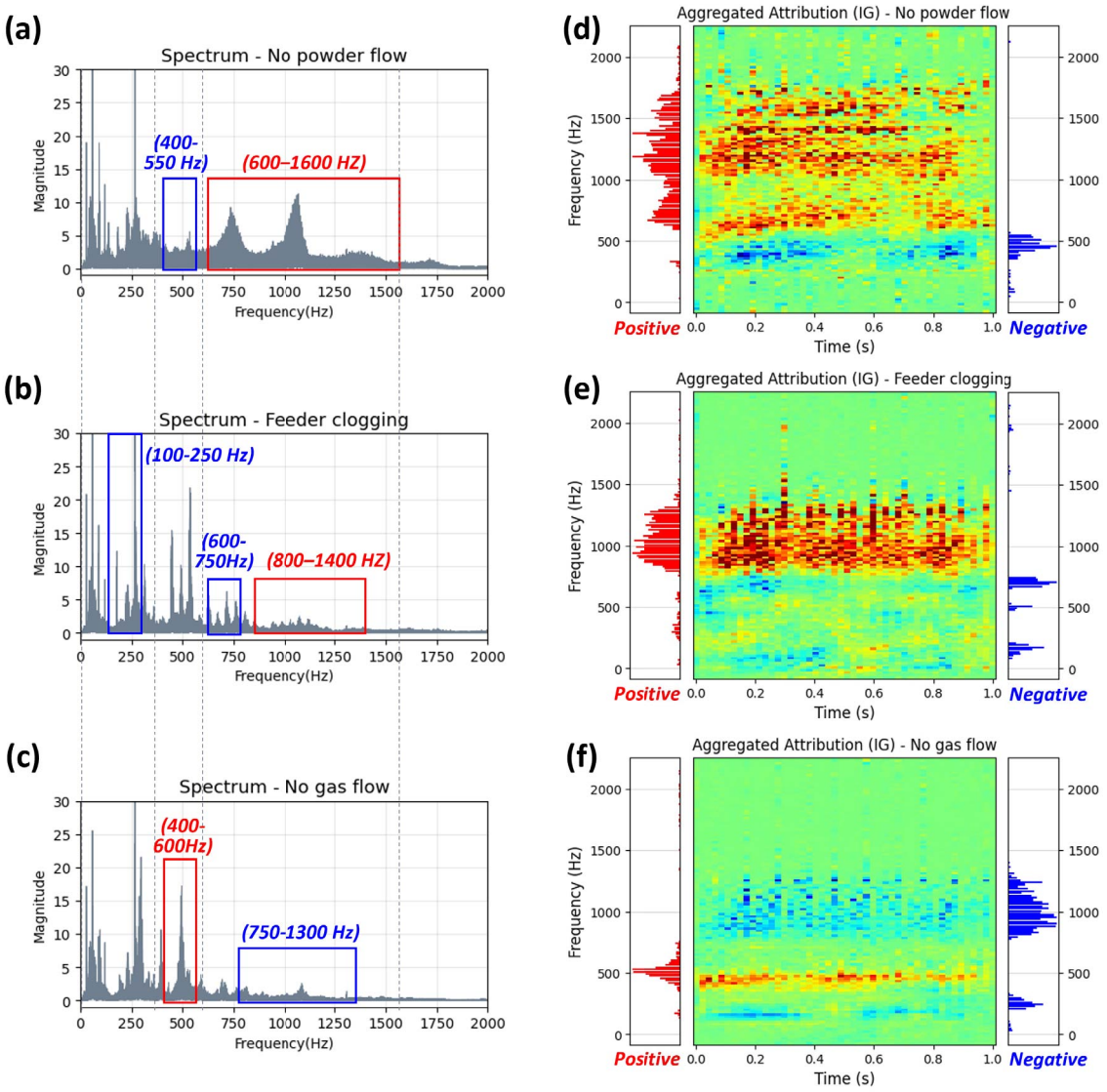


Figure 6. CS powder flow diagnostic model interpretation. (a-c) Frequency spectrums of sound where (a) no powder flow, (b) feeder is clogging, and (c) no gas flow. (d-f) Integrated Gradient (IG)-guided attribution for diagnosing (a) no powder flow, (b) feeder clogging, and (c) no gas flow.

where x_i is the input feature, x_i' is the baseline, F is the model, and $\frac{\partial F}{\partial x_i}$ is the gradient of the model's output with respect to x_i'. Consequently, the IG calculates how each feature contributes to the model's output, providing valuable insights into the decision-making process of the CNN models. In Eq. (9), the $\times$ symbol represents element-wise multiplication for the IG calculation. This evaluates the contribution of each input feature to the model's prediction by element-wise multiplying the gradient of the output with respect to each input by the input's deviation from a baseline. This approach details the impact of individual features on the model's decisions.

The IG was applied to the s-CNN model, which classifies three abnormal CS powder flow states with 99.5% accuracy, to identify which pixels in each Log-Mel spectrum contribute to the model's classification. Fig. 6(d-f) displays heatmaps visualizing the aggregated IG results, along with histograms of pixels that positively or negatively contribute according to frequency, for each type of abnormality. This can be interpreted in comparison with the frequency spectra of the corresponding sound signals shown in Fig. 6(a-c). For instance, the 'no powder flow' abnormality is distinguished by components significantly emphasized in the 600 Hz to 1600 Hz frequency range. This IG-based interpretation of the diagnostic model based on the frequency information can remarkably enhance the reliability of the CS powder flow diagnostics. Despite these advantages, a comprehensive analysis of sound signals throughout the entire CS process remains necessary to accurately identify the specific reasons behind the specific frequency bands observed during the tests. This suggests the potential for explainable sound monitoring that integrates domain knowledge of the CS systems with the interpretable framework.

5. Conclusion

In this work, a stethoscope sound-guided interpretable deep learning (IDL) framework was developed for powder flow monitoring and diagnosis in cold spray additive manufacturing. The sound signal data was collected from a stethoscope mounted on the powder feeder unit of the LPCS machine, allowing the capturing of distinctive sound signals from the vibratory valve inside the powder hopper. Utilizing IDL modeling, three powder flow anomalies, namely 1) no powder flow; 2) feeder clogging; and 3) no gas flow, were successfully diagnosed from the trained modeling. Moreover, the findings guided by stethoscope sound signals were compared with data obtained from the external microphone. This work yields the following conclusions:

- The use of stethoscope sensor sound, which is less affected by ambient noise compared to typical microphone sensors, can significantly enhance the robustness of monitoring models in the noisy CS process environments.
- The two-stage CS powder flow model demonstrates exceptional performance in sensitively detecting anomalies and classifying the types of abnormalities.
- Visualization of data and decision-making based on the trained IDL model can substantially improve the reliability of *in-situ* powder flow diagnostics in CS applications.

Future directions will focus on refining and enhancing these modeling endeavors for powder flowrate monitoring during the CS process, ultimately aiming to capture *in-situ* particle deposition efficiency (i.e., the ratio of the actual amount of powder deposited onto the substrate to the total amount of powder sprayed). In our pursuit, it is critical to evaluate the proposed approach under various powder size distribution and environmental conditions (e.g., humidity) to guide future research endeavors effectively. We envision that these modeling frameworks can significantly contribute to advancing the understanding and optimization of CS additive manufacturing processes. Furthermore, our approach can be extended to other powder-based additive manufacturing processes such as Directed Energy Deposition (DED), in which micron-scale powders are fed through a cladding nozzle at subsonic velocities for material consolidation.

References

[1] SUN, Wen, et al. Cold spray additive manufacturing (CSAM). In: Solid State Addit Manuf. CRC Press, 2024. p. 96-126.

[2] ZAVALAN, Florentina-Luiza; RONA, Aldo. A workflow for designing contoured axisymmetric nozzles for enhancing additively manufactured cold spray deposits. Addit Manuf, 2023, 62: 103379.

[3] WANG, Qian, et al. Unraveling microforging principle during in situ shot-peening-assisted cold spray additive manufacturing aluminum alloy through a multi-physics framework. Materials & Design, 2023, 236: 112451.

[4] ELLINGSEN, Marius; SMIRNOVA, Alevtina. Application of supersonic cold spray for solid-state battery manufacturing. In: Green Sustainable Process for Chemical and Environmental Engineering and Science. Elsevier, 2023. p. 295-317.

[5] MANGALARAPU, Tarun Babu, et al. Cold spraying of Al-aerospace alloys: Ease of coating deposition at high stagnation temperatures. Surface and Coatings Technology, 2023, 467: 129703.

[6] VILLAFUERTE, Julio. Current and future applications of cold spray technology. Metal finishing, 2010, 108.1: 37-39.

[7] YEOM, Hwasung; SRIDHARAN, Kumar. Cold spray technology in nuclear energy applications: A review of recent advances. Annals of Nuclear Energy, 2021, 150: 107835.

[8] HERNANDEZ-HERNANDEZ, Maricruz, et al. A correlational study of process parameters on properties of low-pressure cold sprayed copper coatings. Int J Adv Manuf Technol, 2023, 125.9-10: 4679-4691.

[9] ADAAN-NYIAK, Moses A.; TIAMIYU, Ahmed A. Recent advances on bonding mechanism in cold spray process: A review of single-particle impact methods. J Mater Res, 2023, 38.1: 69-95.

[10] RUZGAR, Duygu Gazioglu, et al. Multifunctional Cold Spray Hybrid Coatings on Flexible Polymers for Improved Surface Properties. Martin Byung-Guk, Multifunctional Cold Spray Hybrid Coatings on Flexible Polymers for Improved Surface Properties.

[11] GÄRTNER, Frank, et al. The cold spray process and its potential for industrial applications. J Therm Spray Technol, 2006, 15: 223-232.

[12] AKIN, Semih, et al. Cold spray-based rapid and scalable production of printed flexible electronics. Addit Manuf, 2022, 60: 103244.

[13] XU, Chenxi, et al. Nondestructive Characterization of Multiscale Defects in an Aluminum Alloy After Cold Spray Repair. J Nondestruct Eval, 2023, 42.3: 75.

[14] SUN, Wen, et al. Current implementation status of cold spray technology: A short review. J Therm Spray Technol, 2022, 31.4: 848-865.

[15] DYKHUIZEN, R. C., et al. Impact of high velocity cold spray particles. J Therm spray technol, 1999, 8: 559-564.

[16] YUN, Huitaek, et al. Autoencoder-based anomaly detection of industrial robot arm using stethoscope based internal sound sensor. J Intell Manuf, 2023, 34.3: 1427-1444.

[17] YUN, Huitaek, et al. Development of internal sound sensor using stethoscope and its applications for machine monitoring. Procedia Manuf, 2020, 48: 1072-1078.

[18] DONNAL, John, et al. "Stethoscopes" for nonintrusive monitoring. SAS. IEEE, 2017. p. 1-6.

[19] ABELLA, Manuel; FORMOLO, John; PENNEY, David G. Comparison of the acoustic properties of six popular stethoscopes. JASA, 1992, 91.4: 2224-2228.

[20] WIDENER, Christian A.; OZDEMIR, Ozan C.; CARTER, Michael. Structural repair using cold spray technology for enhanced sustainability of high value assets. Procedia Manuf, 2018, 21: 361-368.

[21] TSAI, Jung-Ting; JUN, Martin Byung-Guk; BAHR, David F. Measurement methods for quantifying powder flowability and velocity in cold spray systems. Adv Powder Technol, 2023, 34.1: 103910.

[22] POR, Emiel; VAN KOOTEN, Maaike; SARKOVIC, Vanja. Nyquist–Shannon sampling theorem. Leiden University, 2019, 1.1.

[23] CHAMPAGNE JR, Victor Kenneth; OZDEMIR, Ozan Cagatay; NARDI, Aaron (ed.). Practical cold spray. Cham, Switzerland: Springer International Publishing, 2021.

[24] AFSHARI, Aliakbar A., et al. Development of a thermal spray coating aerosol generator and inhalation exposure system. Toxicol Rep, 2022, 9: 126-135.

[25] LI, Juncheng, et al. A comparison of deep learning methods for environmental sound detection. In: ICASSP. IEEE, 2017. p. 126-130.

[26] MENG, Hao, et al. Speech emotion recognition from 3D log-mel spectrograms with deep learning network. IEEE Access, 2019, 7: 125868-125881.

[27] JEONG, Kyeong-Joong, et al. Two-Stage Deep Anomaly Detection With Heterogeneous Time Series Data. IEEE Access, 2022, 10: 13704-13714.

[28] LIU, Jianyu; YANG, Zhouwang; SONG, Yanzhi. A two-stage anomaly detection framework: Towards low omission rate in industrial vision applications. Adv Eng Inform, 2023, 55: 101822.

[29] VAN DER MAATEN, Laurens; HINTON, Geoffrey. Visualizing data using t-SNE. JMLR, 2008, 9.11.

[30] SUNDARARAJAN, Mukund; TALY, Ankur; YAN, Qiqi. Axiomatic attribution for deep networks. In: ICML. PMLR, 2017. p. 3319-3328.