

Deep Learning-Based Real-Time Defect Detection and Kerf Geometry Prediction in Abrasive Water Jet Machining (AWJM) using Acoustic and Vision Sensor Data

Sharanya K, Sharvari G, Shreya B. P., and Sheethal V. Kamath B.

Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India

Abstract - Abrasive Water Jet Machining (AWJM) is widely used to cut materials that are difficult to machine using conventional tools. Although the process avoids the heat-affected zones associated with many thermal methods, the quality of a cut can still vary with machining conditions. Common concerns include kerf taper, striation marks, burr formation, incomplete penetration, and dimensional variation. These issues are often identified during post-process inspection, when an opportunity to respond during cutting may already have passed.

This paper proposes a multimodal deep-learning framework for monitoring AWJM by combining acoustic-emission (AE) signals with camera images. The two sensing methods provide different kinds of information: AE signals capture changes in process activity over time, while images provide spatial evidence about the kerf and visible surface conditions. In the proposed architecture, an acoustic branch based on a one-dimensional convolutional neural network (1D-CNN) or a temporal convolutional network (TCN) extracts signal features. A separate convolutional neural network (CNN) processes images. The resulting feature vectors are combined and passed to two output branches: one for defect classification and another for kerf-geometry regression.

The intended regression targets are top kerf width, bottom kerf width, and kerf taper angle. The paper also defines an evaluation plan using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean squared error (RMSE), the coefficient of determination (R^2), and inference latency. The work is currently a proposed framework: the team has not yet collected an AWJM dataset or validated the model, so no experimental performance values are claimed. The framework establishes a basis for future testing of sensor fusion for in-process AWJM quality monitoring.

Keywords— Abrasive Water Jet Machining, Acoustic Emission, Machine Vision, Deep Learning, Multimodal Sensor Fusion, Defect Detection, Kerf Geometry Prediction, Real-Time Monitoring.

1. INTRODUCTION

1.1 Background

Manufacturing components from hard-to-machine materials or producing complex profiles can be challenging with conventional cutting tools. Non-traditional machining processes address some of these challenges by removing material through mechanisms other than direct tool-edge cutting. Abrasive Water Jet Machining (AWJM) is one such process. A high-pressure water stream carries abrasive particles through a nozzle, and the resulting jet removes material through erosion. AWJM has been investigated for a range of engineering materials, including composites and titanium alloys [1], [7], [9].

The quality of an AWJM cut depends on several process variables, including water pressure, abrasive mass flow rate, traverse speed, standoff distance, and nozzle dimensions. Changes in these variables can affect the kerf profile, surface finish, cutting depth, and dimensional accuracy [2], [4], [9]. One important quality measure is kerf taper, which describes the difference between the width of the cut at the top and bottom of the workpiece. Other concerns include striation, burr formation, incomplete penetration, and visible surface damage.

Researchers have used data-driven models to estimate AWJM quality characteristics. For example, neural networks have been studied for surface-roughness prediction in Kevlar 49 composite cutting and for kerf-related prediction in waterjet machining [1], [4]. Kerf modelling for layered structures and hybrid Taguchi-ANN modelling for Ti-6Al-4V have also been investigated [2], [9]. These studies demonstrate the relevance of predictive modelling, but the sensor information used during the process is also important if the aim is to monitor quality while machining is underway.

1.2 Motivation

A conventional inspection workflow often measures the finished part after cutting. This can identify a quality problem, but it may not provide information early enough to respond during the same operation. In-process sensing offers a possible way to observe changes as they occur. Earlier work has examined acoustic emission for online monitoring of AWJ penetration depth and for assessing cutting capability [3], [6]. Machine vision has also been explored for burr measurement and in-situ monitoring of waterjet processes [5], [10].

AE and vision are complementary rather than interchangeable. Acoustic measurements may reveal changes in the process response, while images can show the location and shape of visible features. Combining the two could give a more informative representation of the cutting operation than either sensor alone. This paper therefore proposes a system that processes the two data streams separately, combines their learned features, and uses the combined representation for defect classification and kerf prediction.

1.3 Research Gap and Objective

The cited literature covers AWJM process modelling, acoustic monitoring, and vision-based inspection [1]–[6], [9], [10]. These research directions provide a foundation for data-driven quality assessment. The present study focuses on the opportunity to integrate acoustic and visual information in a single learning framework and to use that shared representation for two related tasks: identifying a machining condition and estimating kerf geometry.

The main objective is to specify a technically testable multimodal framework for AWJM monitoring. The proposed system will (i) align AE windows with corresponding camera observations, (ii) learn acoustic and visual features using separate networks, (iii) fuse those features, and (iv) predict a defect class together with top kerf width, bottom kerf width, and kerf taper angle. These are proposed objectives; their effectiveness must be established through future experiments.

1.4 Proposed Contributions

- A sensing concept that combines acoustic-emission measurements and machine-vision images for AWJM monitoring.
- Separate feature-extraction branches suited to temporal signals and spatial images.
- Feature-level fusion for joint defect classification and kerf-geometry regression.
- An evaluation plan that includes predictive quality, comparison against single-modality baselines, and real-time feasibility.

The contribution at this stage is a proposed architecture and validation plan, not a claim of experimentally demonstrated performance.

2. LITERATURE REVIEW

2.1 Data-Driven Modelling of AWJM

Several studies have applied neural-network methods to model AWJM quality measures. Madara et al. examined surface-roughness modelling for AWJ cutting of Kevlar 49 composite [1], while Pillai et al. investigated neural-network prediction of kerf width and surface roughness [4]. Singh et al. studied kerf formation in a double-layered structure [2]. More recently, a hybrid Taguchi-ANN approach was reported for multi-input, multi-output modelling of kerf characteristics in Ti-6Al-4V [9]. Collectively, these works show that data-driven models can represent relationships between machining conditions and quality responses.

2.2 Acoustic-Emission Monitoring

Acoustic emission has been considered as a source of information about the behaviour of an abrasive waterjet. Mohan, Momber, and Kovacevic studied online monitoring of AWJ penetration depth using AE [3]. Copertaro and Annoni later investigated airborne acoustic emission as a means of monitoring cutting capability [6]. These studies support the use of acoustic measurements as a process-monitoring signal. However, an AE-only approach does not directly provide the same spatial detail as an image of the kerf or machined surface.

2.3 Vision-Based Monitoring

Machine vision provides spatial information that can be used to inspect visible machining features. Ravi et al. studied vision-based burr-length measurement in abrasive waterjet trepanning [5], and Datta et al. investigated vision-based in-situ process monitoring for productivity improvement [10]. The results of this research area motivate the inclusion of image data in a monitoring system. In practice, image quality may be affected by water droplets, splash, reflections, illumination changes, or partial obstruction, so the acquisition arrangement and preprocessing steps require careful evaluation.

2.4 Position of the Proposed Study

The present paper brings these strands together at the framework level. Instead of relying only on process parameters, only acoustic measurements, or only images, it proposes learning from synchronized AE and vision inputs. Its distinguishing research question is whether their fused representation can support both defect classification and kerf-geometry prediction. This question remains open until a dataset is collected and controlled comparisons are completed.

3. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The methodology below describes a proposed experimental workflow. Sensor specifications, machine settings, sample counts, model hyperparameters, and final performance values have not yet been established. They must be selected and reported after the experimental setup is defined.

3.1 Sensing Arrangement

The proposed setup uses an AE sensor to capture process-generated acoustic activity and an industrial camera to observe the cutting region or machined surface. The workpiece material and thickness should be selected according to equipment availability and the research objective. Stainless steel or Ti-6Al-4V are possible candidates, but neither is assumed to have been tested in the current study.

The two sensors should be mounted so that they collect useful data without interfering with machining. The camera should provide a sufficiently clear view of the relevant region, while the AE sensor should be coupled or positioned according to the sensor type and the manufacturer's recommendations. The final mounting method, bandwidth, sampling rate, camera resolution, frame rate, illumination, and trigger mechanism must be documented once selected.

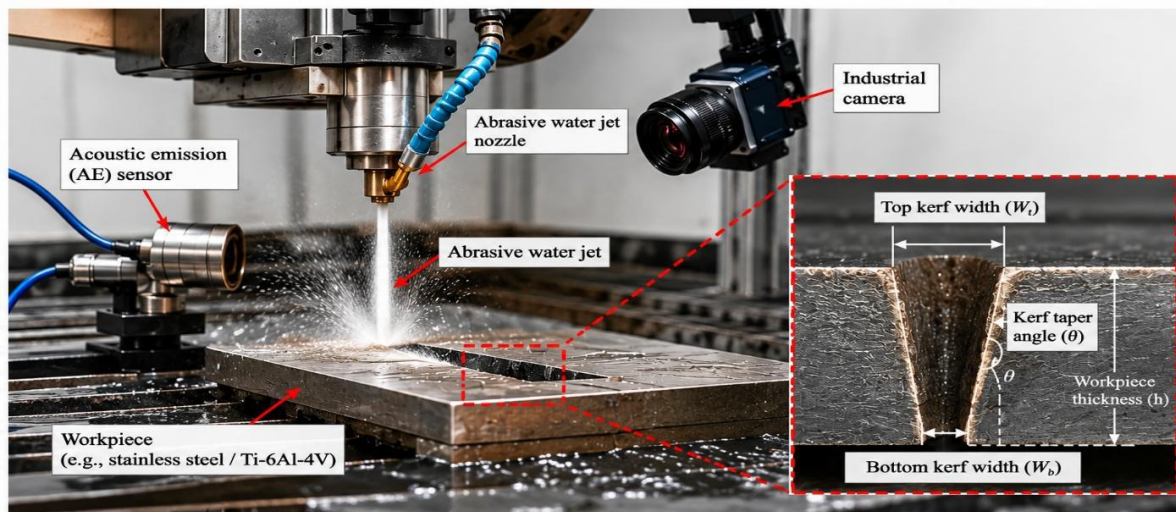


Fig. 1. Proposed AWJM sensing arrangement showing the cutting zone, acoustic-emission sensor, industrial camera, workpiece, and measurable kerf geometry.

Fig. 1. Proposed AWJM sensing arrangement and measurable kerf geometry.

3.2 Data Acquisition and Synchronization

A future experimental record should connect each cutting operation with its process settings, AE signal, relevant camera frames, and post-process kerf measurements. Synchronization is important because the acoustic segment and image must describe the same machining event. A common trigger, timestamping system, or another documented alignment method should be used. The selected method should be checked for timing error before model training.

For each operation, the dataset should record water pressure, abrasive mass flow rate, traverse speed, standoff distance, nozzle/orifice dimensions, material, and workpiece thickness, as applicable. The target measurements should include top kerf width (W_t), bottom kerf width (W_b), workpiece thickness (h), kerf taper angle, and a defect label assigned using a documented inspection procedure. The number of cutting trials should be determined from available resources and the range of conditions to be represented; it should not be reported until the trials have actually been performed.

3.3 Acoustic Signal Processing

AE signals can contain information about material removal and changes in the cutting process, but they may also contain machine vibration and environmental noise. The proposed pipeline therefore begins with signal inspection and conditioning. Filtering or

wavelet-based denoising may be considered if the acquired data show that these operations are needed. The continuous signal is then divided into time windows that can be aligned with the corresponding camera observations.

Normalization can reduce scale differences between samples. Both the time-domain signal and time–frequency representations may be evaluated. The Fast Fourier Transform (FFT) provides a frequency-domain representation:

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp(-j2\pi kn/N)$$

Here, $x[n]$ is the discrete signal, N is the number of samples, $X[k]$ is the frequency-domain representation, and k is the frequency index. A spectrogram may also be used to retain information about how frequency content changes over time. A 1D-CNN is one candidate for learning local patterns from the signal, while a TCN may be considered when temporal dependencies across a longer window are important. The final input representation should be selected based on pilot data rather than assumed in advance.

3.4 Vision Preprocessing

The camera branch receives images of the cutting region or machined surface. Preprocessing may include selecting a region of interest, resizing images, normalizing brightness and contrast, and reducing noise where necessary. Edge detection and morphological operations may be used as supporting image-processing steps if they help identify kerf boundaries. Their usefulness should be checked against the actual image conditions, particularly where splash, mist, reflection, or low contrast is present.

A CNN can then learn visual features associated with kerf boundaries, surface texture, striation patterns, burrs, incomplete cutting, or visible material damage. Data augmentation—such as modest cropping, scaling, or illumination variation—may be used after labelled images have been collected. Augmentation should preserve the meaning of the defect label and should not create unrealistic samples.

3.5 Multimodal Feature Fusion

The proposed architecture has two feature-extraction branches. The acoustic branch maps the processed signal X_A to a feature vector F_A , and the vision branch maps image input X_V to a feature vector F_V :

$$F_A = f_A(X_A) \quad F_V = f_V(X_V)$$

The feature vectors are concatenated to create the fused representation:

$$F_{\text{fusion}} = [F_A ; F_V]$$

The fused representation is passed to shared fully connected layers and then to two output heads. The classification head estimates probabilities for the defined defect classes. The regression head estimates the three kerf-related targets. This feature-level fusion design is a starting point; attention-based fusion or other alternatives could be compared in future work if the dataset is sufficiently large.

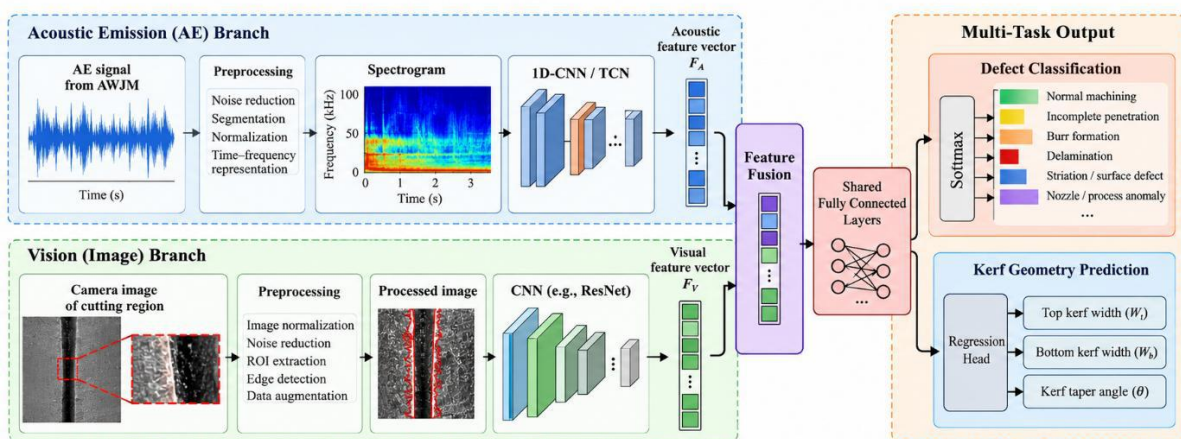


Fig. 2. Proposed multimodal deep-learning architecture showing independent acoustic and vision feature extraction, feature-level fusion, and simultaneous defect classification and kerf-geometry regression.

Fig. 2. Proposed multimodal deep-learning architecture for defect classification and kerf-geometry regression.

3.6 Defect Classification

The classification task is intended to identify the machining condition. Candidate labels include normal cutting, incomplete penetration, burr formation, delamination or material damage, striation or surface defects, and nozzle/process anomalies. These are proposed classes only. The final label set should be based on defects that can be identified consistently in the collected samples, with clear labelling criteria and, where possible, independent verification.

For C classes, categorical cross-entropy can be expressed as:

$$L_{\text{classification}} = -\sum_{c=1 \text{ to } C} y_c \log(\hat{y}_c)$$

Here, y_c is the true class indicator and $\hat{y}_c$ is the predicted probability for class c. The classifier produces a probability distribution across the defined classes; the final predicted label can be selected from that distribution.

3.7 Kerf-Geometry Regression

The regression head is intended to estimate top kerf width, bottom kerf width, and kerf taper angle. If the top and bottom widths and workpiece thickness are measured consistently, the taper angle can be calculated as:

$$\theta = \tan^{-1}((W_t - W_b) / (2h))$$

In this expression, W_t is the top kerf width, W_b is the bottom kerf width, and h is workpiece thickness. The predicted target vector is:

$$\hat{Y}_{\text{kerf}} = [\hat{W}_t, \hat{W}_b, \theta]$$

The units used for width and thickness must be consistent. The measurement procedure should also specify how kerf boundaries are determined and how repeated measurements are handled. This is necessary for interpreting the regression error in physical units.

3.8 Joint Training Objective

The two output heads can be trained jointly using a weighted multi-task objective:

$$L_{\text{total}} = \lambda_1 L_{\text{classification}} + \lambda_2 L_{\text{regression}}$$

The coefficients λ_1 and λ_2 control the relative contribution of the classification and regression losses. A possible regression loss is mean squared error:

$$L_{\text{regression}} = (1/N) \sum_{i=1 \text{ to } N} (y_i - \hat{y}_i)^2$$

The loss weights and other hyperparameters should be selected using the training and validation data. They should not be presented as established values until tuning has been completed.

3.9 Training and Data-Leakage Control

Once the dataset is available, it should be divided into training, validation, and independent test subsets. Samples from the same cutting run or workpiece should not be randomly spread across these subsets if doing so would allow near-duplicate observations to appear in both training and testing. Group-based splitting can reduce this risk. Class imbalance should be measured, and suitable class weighting or sampling strategies may be considered where needed.

An optimizer such as Adam may be evaluated. The learning rate, batch size, number of epochs, stopping criteria, and model configuration should be selected experimentally. Early stopping can help limit overfitting, while cross-validation may be useful if the dataset is small. All preprocessing steps that learn parameters from the data must be fitted using the training subset only.

3.10 Baseline Comparisons

To test whether fusion adds value, the multimodal model should be compared with at least a vision-only model and an acoustic-only model. Conventional methods such as Random Forest, Support Vector Machine (SVM), or XGBoost may also be evaluated where their inputs and tasks are appropriate. All models should use the same independent test set and comparable evaluation conditions. Until these comparisons are performed, it is not possible to conclude that the multimodal model is more accurate or more robust.

4. EVALUATION PLAN

No experimental results are reported because the AWJM dataset and trained model are not yet available. This section defines the measures that should be used when validation is carried out.

4.1 Classification Metrics

Classification should be evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. For binary classification, accuracy can be written as:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Precision and recall are:

$$\text{Precision} = TP / (TP + FP) \quad \text{Recall} = TP / (TP + FN)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

For a multi-class task, the paper should state whether metrics are reported per class or averaged using macro, micro, or weighted averaging. A confusion matrix can show which defect categories are most often confused.

4.2 Regression Metrics

The kerf predictions should be evaluated separately for top width, bottom width, and taper angle. The proposed measures are mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2):

$$\text{MAE} = (1/N) \sum_{i=1}^N |y_i - \hat{y}_i|$$

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

$$R^2 = 1 - [\sum_{i=1}^N (y_i - \hat{y}_i)^2] / [\sum_{i=1}^N (y_i - \bar{y})^2]$$

Here, y_i is the measured value, $\hat{y}_i$ is the prediction, and $\bar{y}$ is the mean measured value. MAE and RMSE should be reported in the relevant physical units for each target.

4.3 Sensor-Fusion Ablation

The central evaluation question is whether AE and vision provide complementary information. The acoustic-only, vision-only, and fused models should therefore be evaluated on the same held-out samples. Any claimed improvement should be supported by the actual metric values and, where feasible, repeated runs or uncertainty estimates. Failure cases should also be examined—for example, image degradation caused by water splash or AE contamination caused by machine vibration.

4.4 Real-Time Feasibility

A system intended for in-process monitoring must be evaluated for more than predictive quality. Relevant measurements include end-to-end inference latency, acoustic-window processing time, image throughput, memory use, and processor or GPU utilization. The hardware and software environment should be documented. Real-time capability should only be claimed if the measured latency and sampling requirements support it.

4.5 Process-Parameter Sensitivity

Future experiments should cover a suitable range of water pressure, abrasive flow rate, traverse speed, and standoff distance. These variables may influence material removal, sensor responses, and kerf geometry. A sensitivity analysis can help determine which operating conditions and sensor representations are associated with prediction changes. The experimental design should avoid confounding the effect of one parameter with another where possible.

5. PRACTICAL IMPLICATIONS

5.1 In-Process Quality Monitoring

If validated, the proposed system could support quality assessment while AWJM is in progress rather than relying entirely on post-process inspection. AE may provide a continuous indication of process behaviour, while camera images may help localize visible features. The practical value will depend on whether the sensors can operate reliably in the wet, noisy machining environment and whether the model generalizes to new cutting runs.

5.2 Potential Process Control

A future extension could connect the monitoring system to the AWJM controller. For example, a validated warning about unacceptable quality might prompt an operator to inspect the cut or consider changing traverse speed, abrasive flow rate, or pressure. Automatic adjustment should not be implemented solely on the basis of unvalidated predictions; it would require controlled testing and safeguards.

5.3 Waste and Manufacturing Efficiency

Earlier identification of a developing quality problem may reduce avoidable rework or scrap, but such benefits cannot be assumed. They should be measured in controlled trials by comparing inspection outcomes, rework rates, material use, and process time with and without the monitoring system. Until then, waste reduction remains a potential benefit rather than a demonstrated result.

5.4 Potential Applications

The framework may be relevant to manufacturing settings that use AWJM for composites, titanium alloys, stainless steel, and other engineering materials. Transfer to a new material or machine should be assessed rather than assumed, because material properties, machine dynamics, and sensor placement can change the data distribution.

6. LIMITATIONS AND FUTURE WORK

6.1 Current Limitations

The main limitation is that the framework has not yet been validated using an experimental AWJM dataset. Consequently, no accuracy, regression error, or latency result can be reported. Sensor selection, acquisition settings, defect definitions, dataset size, and model hyperparameters also remain to be finalized. In addition, AE data can be affected by environmental noise and machine vibration, while images can be degraded by splash, mist, reflection, and poor illumination. Accurate synchronization and careful separation of training and test data will be essential.

6.2 Future Work

The first step is to define and document the experimental arrangement, select sensors, establish a repeatable kerf-measurement protocol, and collect synchronized signals and images across multiple machining conditions. The dataset should then be labelled and split without leakage between related cutting runs. The acoustic-only, vision-only, and multimodal models can be trained and compared using the evaluation measures defined above.

Further work could investigate attention-based fusion, Transformer architectures, self-supervised learning for unlabelled signals, edge-device deployment, and multi-material generalization. Once prediction reliability has been established, integration with a digital twin or a safeguarded adaptive-control system may also be explored. These directions are possibilities for future study, not capabilities demonstrated by the present paper.

7. CONCLUSION

This paper has presented a proposed multimodal deep-learning framework for defect detection and kerf-geometry prediction in Abrasive Water Jet Machining. The approach combines acoustic-emission signals, which describe changes in process activity over time, with camera images that provide spatial information about the cutting region and visible surface conditions. Separate feature-extraction branches are followed by feature-level fusion and two prediction tasks: defect classification and regression of top kerf width, bottom kerf width, and kerf taper angle.

The paper also defines a methodology for data acquisition, synchronization, signal and image preprocessing, joint training, baseline comparison, and evaluation of predictive and real-time performance. Previous research on AWJM modelling, acoustic monitoring, and vision-based inspection provides motivation for this integrated approach [1]–[6], [9], [10]. However, the proposed architecture has not yet been experimentally tested. Its accuracy, robustness, and usefulness for real-time monitoring must be established using measured data and controlled comparisons. The present contribution is therefore a testable framework and evaluation plan, rather than a report of completed experimental results.

REFERENCES

- [1] S. R. Madara, S. R. Pillai, M. C. Pon Selvan, and J. Van Heirle, "Modelling of surface roughness in abrasive waterjet cutting of Kevlar 49 composite using artificial neural network," *Materials Today: Proceedings*, vol. 46, pp. 1–8, 2021.
- [2] N. P. Singh, D. S. Srinivasu, and N. R. Ramesh Babu, "Modelling of abrasive waterjet kerf in a double-layered structure," *Journal of Manufacturing Processes*, vol. 69, pp. 514–531, 2021.
- [3] R. Mohan, A. W. Momber, and R. Kovacevic, "On-line monitoring of depth of AWJ penetration using acoustic emission technique," in *Jet Cutting Technology*, pp. 649–664, 1994.
- [4] S. R. Pillai, S. R. Madara, and C. Pon Selvan, "Predication of Kerf Width and Surface Roughness in Waterjet Cutting using Neural Networks," *Journal of Physics: Conference Series*, vol. 1276, Art. no. 012011, 2019.
- [5] R. R. Ravi, C. Datta, and D. S. Srinivasu, "Machine vision based burr length measurement in abrasive waterjet trepanning," *Manufacturing Letters*, vol. 35, pp. 417–422, 2023.
- [6] E. Copertaro and M. Annoni, "Airborne acoustic emission of an abrasive waterjet cutting system as means for monitoring the jet cutting capability," *The International Journal of Advanced Manufacturing Technology*, vol. 123, pp. 2655–2667, 2022.
- [7] S. R. Shekolkar, H. A. Chavan, and D. D. Deshmukh, "Recent advances in abrasive water jet machining for structural materials: A comprehensive review," *Research on Engineering Structures & Materials*, vol. 12, no. 3, pp. 1395–1417, 2026.

- [8] A. Gaonkar and R. S. P. Gaonkar, "Predictive analysis and performance measures optimisation in sustainable abrasive jet machining," Materials Research Express, vol. 13, Art. no. 065501, 2026.
- [9] Y. I. Mogul, J. D. Quadros, G. Subramanian, P. Suhas, R. Hanumantharaya, I. Mogul, and A. Begum, "Hybrid Taguchi-ANN based multi-input multi-output modelling of kerf characteristics in AWJM of Ti6Al4V with experimental validation," Scientific Reports, vol. 16, Art. no. 30200, 2026.
- [10] C. Datta, R. R. Ravi, and D. S. Srinivasu, "Vision-based in-situ process monitoring for increased productivity in abrasive waterjet trepanning," Procedia Computer Science, vol. 253, pp. 2086–2095, 2025.