scientific reports

Check for updates

OPEN

Development of optimized ensemble machine learning-based prediction models for wire electrical discharge machining processes

Baneswar Sarker[1], Shankar Chakraborty[2]✉, Robert Čep[3] & Kanak Kalita[4,5]✉

This paper proposes development of optimized heterogeneous ensemble models for prediction of responses based on given sets of input parameters for wire electrical discharge machining (WEDM) processes, which have found immense applications in many of the present-day manufacturing industries because of their ability to generate complicated 2D and 3D profiles on hard-to-machine engineering materials. These ensembles are developed combining predictions of the three base models, i.e. random forest, support vector machine and ridge regression. These three base models are first framed utilizing the training datasets, providing predictions for all the responses under consideration. Based on these predictions, two optimization problems are formulated for each of the responses, while minimizing root mean squared error and mean absolute error, for subsequent development of two optimized ensembles whose predictions are the weighted sum of the predictions of the base models. The prediction performance of all the five models is ascertained through nine statistical metrics, after which a cumulative quality loss-based multi-response signal-to-noise (MRSN) ratio for each model is computed, for each of the responses, where a higher MRSN ratio indicates greater accuracy in prediction. This study is conducted using two experimental datasets of WEDM process. Overall, the optimized ensemble models having higher MRSN ratios than the base models are indicated to deliver better prediction accuracy.

Keywords Optimized heterogeneous ensemble, Wire electrical discharge machining, Response, Multi-response S/N ratio, Prediction performance

Wire electrical discharge machining (WEDM) is a very useful non-traditional machining process that removes material from different hard-to-cut work samples through an electro-thermal phenomenon of generation of sparks between the wire electrode (tool) and electrically conductive workpiece, on which conventional machining processes are difficult to apply[1]. They have found wide applications in mould, tool and die making, automotive, aerospace, medical, optical, dentistry and jewellery industries[2]. During WEDM process, the workpiece is immersed in a dielectric medium, like deionized or distilled water. The wire electrode, introduced into the machining zone in continuous motion, is separated from the workpiece by a small gap with the dielectric fluid in between[3]. The electrical energy generated in the form of continuous but distinct electric sparks is responsible for erosion of material through heating and subsequent melting of material from the workpiece[4]. In this process, minute amount of wire electrode is also removed, requiring introduction of new wire electrode in the machining zone, while the material removed is rinsed out by circulation of the dielectric fluid. Figure 1 demonstrates the working principle of a typical WEDM process.

The process of material removal due to electro-thermal action in WEDM is complex, governed by the influence of multiple factors (process parameters). There exist complex interrelationships between those

[1]Department of Industrial and Systems Engineering, Indian Institute of Technology, Kharagpur, India. [2]Department of Production Engineering, Jadavpur University, Kolkata, India. [3]Department of Machining, Assembly and Engineering Metrology, Faculty of Mechanical Engineering, VSB-Technical University of Ostrava, 70800 Ostrava, Czech Republic. [4]Department of Mechanical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi 600 062, India. [5]Jadara University Research Center, Jadara University, Irbid 21110, Jordan. ✉email: s_chakraborty00@yahoo.co.in; drkanakkalita@veltech.edu.in; kanakkalita02@gmail.com

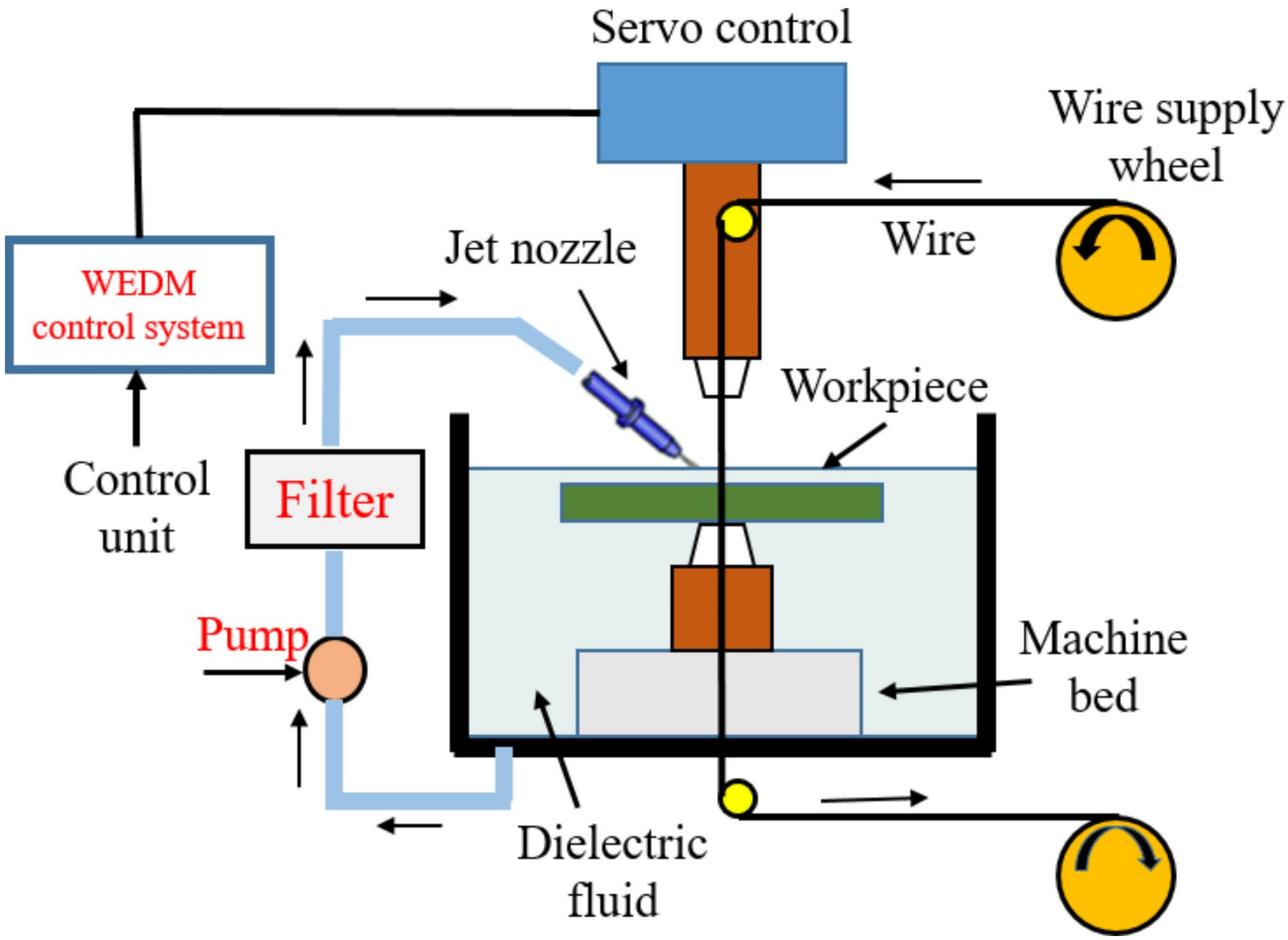

Fig. 1. Working principle of a WEDM process.

factors and outputs (responses), where the outputs have strong dependencies on the process parameters. Any slight change in the values of WEDM parameters, such as discharge current, voltage, pulse duration, wire electrode characteristics etc., can significantly influence the responses, like removal rate of workpiece material, surface conditions and other similar machining variables[2]. Application of WEDM processes thus enables the manufacturing industries to generate complex and intricate 2D and 3D shape profiles on many of the present-day engineering materials with high dimensional accuracy and close tolerance, thus requiring significant control over material removal and other machining indicators by adjusting the input parameters. As a result, prediction of the response values is ardently necessary to gauge the ideal or optimal setup of the process parameters, which would help the machine operators to explore the fullest machining potential of a WEDM process. It is of utmost importance to develop a model, which would include WEDM parameters and responses, and predict those responses, subsequently enabling to increase machining efficiency and process economy[5]. The modeling for prediction of responses can be achieved through application of suitable machine learning (ML) techniques.

Application of ML models/algorithms thereby form the basis for development of intelligent machining principles[6,7], which can subsequently be helpful in prediction and optimization of machining processes, like WEDM[6]. In this process, responses, like tool condition, material removal rate (MRR), surface conditions, geometrical errors etc. can be efficiently predicted based on different input parameters with the help of suitable ML models. Aided by these models, response monitoring of WEDM processes has become a less difficult task, incorporating diversified demands of the end users and dealing with the external factors influencing the process[7]. These models can take the help of an existing machining dataset to learn the hidden patterns, which can then aid in prediction of the response values for newly introduced data. The ML algorithms also show great promise for constant improvement while automating the associated decision-making functions through their efficient operation over large dimensions of the available datasets[8]. Since the response values of most of the machining processes, like WEDM, are quantitative and continuous in nature, a supervised (since the data is labeled) regression technique must be implemented. Since their inception, numerous researchers have adopted ML techniques for accurate prediction of the responses for various machining processes. However, those prediction results are required to be as accurate as possible so that proper monitoring of the responses can be undertaken. Modification in the existing ML approaches can improve ability of those models to provide more accurate prediction results, one such modification being the concept of ensemble methods[11–13].

Ensemble methods incorporate multiple ML algorithms to deliver enhanced prediction performance compared to that of any of the constituent models as a single entity[9]. Ensemble models can be broadly classified into two categories, e.g. homogeneous (aggregation of the same ML algorithm into a consolidated one) and heterogeneous (aggregation of different ML algorithms). Heterogeneous ensembles hold an advantage over homogeneous ensembles, as they can consider advantages of different ML algorithms in a consolidated manner, without multiplied effect of the drawbacks of the same ML algorithm as noticed in homogeneous ensembles. Since the responses of a machining process, like WEDM, are quantitative and continuous in nature, the applicable ensembles are regression models. The simplest regression heterogeneous ensemble model-based predictions are proposed while computing either the simple average or weighted average of the predictions of the base models.

This paper adopts a novel approach for developing optimized weighted average ensembles for prediction of WEDM responses based on two past experimental datasets. At first, three ML algorithms, i.e. random forest (RF), support vector machine (SVM) and ridge regression are implemented on the considered WEDM datasets for predicting the corresponding response values. The weight (importance) to be assigned to each of the base models is identified through minimization of two of the mostly reported performance metrics of regression models, i.e. root mean square error (RMSE) and mean absolute error (MAE) after solving the two developed optimization models. Hence, two ensemble models are developed, one with the optimized RMSE and other with the optimized MAE, by effectively integrating the three base models. The performance of the optimized ensemble models is subsequently compared against the base models using some well-reported statistical metrics, like mean squared error (MSE), correlation coefficient (R), mean absolute percentage error (MAPE), root mean square percentage error (RMSPE), root mean square logarithmic error (RMSLE), relative absolute error (RAE) and root relative squared error (RRSE). As mentioned earlier, these ensemble models are developed based on two WEDM experimental datasets to demonstrate validity and applicability of the results in all such cases.

Through the proposed methodology for predicting WEDM process responses, this paper contributes the followings:

(a) To the best of the authors' knowledge, it is the first attempt in implementing a parallel heterogeneous ensemble learning model for prediction of WEDM responses. This is further highlighted through a review of past research works of application of various prediction models for WEDM processes. Furthermore, such ensembles are also rare in all types of machining processes.
(b) Use of a mathematical programming-based optimization model for prediction error measurement while deriving weights of the base ML models is also a novel approach for the most accurate prediction of the WEDM responses. This is probably the first attempt in developing a mathematical programming-based optimization model for all types of the machining processes.
(c) Few studies have applied metaheuristic-based optimization of a single error measure while proposing an optimized heterogeneous ensemble learning model to predict the response values. This paper improves upon those studies by developing prediction models while minimizing two of the most commonly reported error measures, i.e. RMSE and MAE, followed by a comparison of those models against the base models.
(d) There is a lack of application of parallel heterogeneous ensemble learners for multiple machining responses. This is the first such endeavour in predicting multiple responses for any of the machining processes.
(e) Implementation of the proposed approach to multiple experimental WEDM datasets exhibits its diverse range of potential applications in experiments in the domain of both traditional and non-traditional machining processes.

This paper is structured as follows: "Literature review" provides a concise review on the applications of different ML algorithms as prediction models mainly for WEDM processes. Details of the considered ML algorithms, developed ensemble models and statistical metrics are presented in "Proposed ensemble model". The corresponding ensemble models are framed in "Development of ensemble models for WEDM processes" and "Conclusions" are drawn in Sect. 5 along with future research directions.

Literature review

In recent years, the experimental data-based prediction modeling of WEDM processes has been established as a focal point of research. Numerous works regarding prediction modeling of WEDM processes are available in the literature considering diverse responses and input parameters. Nain et al.[10] employed RF and M5P tree algorithms for development of prediction models for a WEDM process while accurately guessing values of MRR, surface roughness (SR) and dimensional deviation (DD), followed by sensitivity analysis and application of particle swarm optimization (PSO) technique for process optimization. Ulas et al.[11] applied four ML algorithms, i.e. extreme learning machine (ELM), weighted extreme learning machine (W-ELM), support vector regression (SVR) and quadratic support vector regression (Q-SVR) for prediction of SR values during WEDM of Al7075 alloy, and observed W-ELM as the best-performing model with the highest R^2, and lowest MAE and RMSE values. Lalwani et al.[12] employed a back-propagation artificial neural network (ANN) to develop prediction models for MRR, SR and kerf width (KW), and compared the model performance against the results derived using response surface methodology (RSM). Finally, the WEDM process was optimized using non-dominated sorting genetic algorithm (NSGA-II).

Naresh et al.[13] developed an adaptive neuro-fuzzy inference system (ANFIS), along with ANN models through back-propagation with three different algorithms, i.e. Elman regression, generalized regression and Levenberg-Marquardt, for prediction of MRR and SR values, and compared their prediction performance. Based on the considered statistical metrics, ANFIS had appeared as the best model having the highest accuracy for anticipation of WEDM responses. Dandge and Chakraborty[14]implemented two decision tree-based algorithms, i.e. classification and regression trees (CART) and chi-squared automatic interaction detection (CHAID) for prediction of four responses, i.e. machining rate, SR, wire wear ratio and DD, based on six WEDM parameters from an experimental dataset, followed by identification of the most important parameters influencing the responses. Shanmugasundar et al.[15] applied three ML algorithms, i.e. linear regression (LR), RF and AdaBoost to the datasets of two non-traditional machining processes, including WEDM. AdaBoost had emerged out as the most favourable and deployable model to predict MRR and KW, being insensitive to the number of regressors in the model.

Sharma et al.[16] modeled WEDM of titanium alloy with the help of two ML algorithms, i.e. Gaussian process regression (GPR) and SVM to predict the corresponding SR values. Their efficacy as prediction models was also validated based on test datasets. Three prediction models, i.e. multilayer perceptron, ensemble neural network (ENN) and optimization-based evolving product-unit neural network (EPUNN) were developed by Gurgenc and Altay[17] for SR prediction during WEDM of AZ91D magnesium alloy. The superiority of EPUNN model over the others was also established. Saha et al.[18] first employed four different ML algorithms, i.e. LR, GPR, SVM and regression trees as effective prediction models for cutting rate and SR. The GPR model having the maximum computational efficiency among the tested models was later considered for data-driven uncertainty quantification and sensitivity analysis of the said WEDM process. Verma and Singh[19]implemented SVR with four different kernel functions, i.e. linear, polynomial, radial basis and sigmoid to predict SR and slicing speed during WEDM-based wafer slicing. Based on the computed statistical measures, SVR with radial basis function had appeared to provide the best results in comparison to other kernel functions. Ishfaq et al.[20] attempted to predict cutting speed (CS) through application of three ML algorithms, like ANN, SVM and ELM. It was revealed that ANN model had performed comparatively better in predicting the considered response with respect to higher R^2, and smaller MAE and RMSE values. Saha et al.[21] conducted a study to investigate the influence of two clustering algorithms, i.e. grid partitioning (GP) and subtractive clustering (SC) on the performance of ANFIS-based prediction of MRR and SR values during WEDM of A286 superalloy. The predictive performance of ANFIS-GP model had appeared to be better than that of ANFIS-SC model.

In a recent study, Jithendra et al.[22] augmented an ANFIS model by encrypting snake optimizer (SO) for predicting SR, MRR and residual stress of Monel 400 alloy. The predictive performance of ANFIS-SO model was later contrasted against other hybrid approaches, like ANFIS- beetle antennae search, ANFIS-reptile search algorithm and ANFIS-COOT algorithm, showcasing its excellent performance with minimum computational effort. Treating pulse duration and peak current as the input parameters, and MRR, SR, DD and tolerance errors as the responses, Natarajan et al.[23] hybridized grey relational analysis (GRA) with ANFIS and ANN models. ANFIS-GRA had evolved out as an effective prediction model which could accurately evaluate the performance indicators of the WEDM process under consideration.

It can be unveiled from the above survey of the existing literature that most of the researchers have employed a single ML algorithm or some discretely chosen ML algorithms to develop the corresponding prediction models for WEDM processes. In multiple instances, ML algorithms are generally found to suffer from underfitting of data, i.e. they are not able to extract meaningful patterns from the training datasets, and thus, do not fit properly, making them unsuitable for prediction based on both the training and newer datasets. Conversely, ML algorithms may also find it difficult to generalize their predictive performance to newer datasets due to overfitting, i.e. fitting with the training dataset to such an extent that while prediction on the training data is highly accurate, the performance on newer datasets is dismal. Hence, the presence of bias and variance in the performance of those algorithms adversely affects their accuracy of prediction[24]. These issues can be resolved through modification of the existing ML algorithms, one such approach being the concept of ensemble methods. Ensemble methods minimize model variance, subsequently reducing overfitting, thus deriving more accurate prediction results[24].

Some of the research works have included homogeneous ensemble models, like RF or AdaBoost as the prediction model. Homogeneous models, however, have a disadvantage as the inadequacies of a single algorithm may be compounded, making them unacceptable for accurate prediction of the machining responses[25]. Conversely, heterogeneous models can derive the advantages of different algorithms and counter the drawbacks of the individual algorithms, making them more robust and generalized. So, it can be concluded that heterogeneous ensemble models enjoy considerable advantages, and are better suited for modeling and prediction of WEDM processes. As mentioned earlier, the simplest heterogeneous ensembles are simple average and weighted average approaches. For both the methods, there are multiple base regression models, which are trained on the same dataset. The difference lies in the prediction on a new dataset, as in the simple average method, the final prediction is the average of the predictions of the base regression models, while in the weighted average method, more weightage is allocated to the model with a better prediction performance on the new dataset and a weighted average is computed to be the prediction of the ensemble. Usually, these weights are computed based on the comparative prediction performance of the base models. As a result, those models, whose predictions are usually closer to the actual values, are assigned higher weights, making the predictions more accurate. This paper computes the weights of the base models through optimization ensuring a higher level of accuracy. Therefore, to attain greater prediction accuracy and minimum error, this paper proposes development of a heterogeneous weighted average ensemble methodology, where weights are obtained through mathematical programming, with minimized RMSE and MAE values.

Proposed ensemble model

As mentioned earlier, this paper proposes a methodology to develop novel optimized weighted average ensemble models for prediction of responses for two WEDM processes based on past experimental datasets. At first, three ML algorithms, i.e. RF, SVM, and ridge regression are implemented on two separate WEDM datasets as the base models. RF is a decision tree-based homogeneous ensemble technique that has found applications in some research studies regarding modeling of WEDM processes. On the other hand, SVM, a kernel-based algorithm, is one of the most commonly implemented ML algorithms, employed for development of prediction models for WEDM processes, as established through the literature review. Ridge regression is an extension of the classic multiple regression modeling, involving regularization to reduce overfitting errors. Even though multiple regression analysis is the most commonly applied modeling technique for WEDM processes, there is no literature available that studies application of ridge regression for the said purpose. Hence, it is fairly apparent that these three models are distinctly different from each other, not only on the basis of their theoretical and statistical framework, but also on the frequency of their applications in research articles, and hence, they are brought together for the development of a comprehensive ensemble model.

The weight to be assigned to each of the base models is estimated with the help of the corresponding optimization problem. To obtain the ensemble model with minimum RMSE or MAE value, an optimization problem is thus formulated for the proposed ensemble model with RMSE or MAE as the objective function, from which the predicted values of the ensemble model, and weights as well as RMSE and MAE values of the ensemble model are computed for each of the responses. Other statistical performance metrics are also computed from the predicted response values for each of the base models as well as RMSE-optimized and MAE-optimized ensemble models. Finally, among the considered prediction models, the best-performing model is identified based on the estimated multi-response signal-to-noise (MRSN) ratio. In case of an experiment involving multiple responses, as studied in this paper, these steps are reiterated for each of the responses. The proposed methodology is explained in Fig. 2 in the form of a flowchart. The details of the ML algorithms and optimization problem, along with the statistical metrics employed to compare prediction performance of the considered models and computation of MRSN ratios are presented here-in-under.

Ensemble for prediction of responses

As mentioned in the previous section, the ensemble approach in this paper includes RF, SVM and ridge regression as the base models. The theoretical fundamentals of these ML techniques are elucidated here through a brief description.

Random forest

RF is a decision tree-based ensemble ML technique, which serves as an aggregator of the predictions made by several decision trees, where generation of the trees occurs on the basis of the values of an independent set of random vectors[26]. It is a special case of bagging ensemble learner, as it implements bootstrap samples to create datasets for training of the base learning decision trees[27]. Training datasets are generated randomly from the original dataset, with the same number of observations or cases as that of the original dataset. Training is performed through sampling with replacement, which implies that while some observations in the original dataset may occur more than once in any given generated dataset, some observations may not occur at all. Each of the constituent decision trees can then be trained with each of the generated datasets. However, RF differs from bagging with respect to randomized selection of variables or features. During development of each of the trees constituting the model, a subset of variables is chosen followed by the typical approach of split selection within the selected feature subset[27]. Selection of the subset of variables has an advantage as it decreases similarities among the constituent decision trees developed by training with different generated datasets, while a specific number of variables (usually considered one-third of the total number of variables in case of regression) would balance reduction in variance and computational cost[28]. Aggregation of the predictions follows after a sufficient number of constituent trees are trained. For continuous numeric responses of a WEDM process, like MRR or SR, the aggregation approach involves a simple average of the predictions of the constituent decision trees.

Support vector machine

SVM is a kernel-based algorithm, generally used as a classification technique, and characterizes a decision boundary with the help of a subset of training data, which are called support vectors[26]. SVM functions through characterization of training data as spatial points by plotting the observations belonging to dissimilar categories or classes in order to maintain an established gap or decision boundary, dividing the points according to the categories. Thereafter, it makes predictions on a new dataset, by plotting the observations on the same space as above, where each observation is assigned a category based on the side of the boundary where that observation would be plotted[29]. The decision boundary is a hyperplane with its dimension equal to one less than the number of variables. In case of a dataset linearly separable by multiple hyperplanes, the one with the maximum possible gap on either side is selected as the decision boundary. This is obtained by moving a pair of hyperplanes away from the decision boundaries on either side, but parallel to each other until they coincide with a point on each side. The hyperplane with the maximum gap (called maximum margin) is finally selected.

However, in case of linearly inseparable data points, i.e. no linear decision boundary can completely separate the data points belonging to dissimilar categories, two approaches can be considered. One of them requires construction of soft margins, i.e. a hyperplane-based decision boundary with a considerable amount of margin which would not be able to classify a few points to either side, even if there exists a decision boundary with a lesser margin being able to completely separate the data points belonging to different categories. This would help in avoiding a case of overfitting which may arise in the latter case. For development of an SVM model with soft margins, there is a compromise between the margin width and classification error with the training dataset[26]. The other approach can be applicable for non-linearly separable data points and involves the kernel trick, i.e. conversion of the original data space to a high-dimensional space to enable a linear decision boundary to effectively separate the points[29].

As mentioned above, the responses of a WEDM process are quantitative and continuous in nature, requiring regression models for prediction of their values. SVM can be extended to a regression problem for the same where the central concept, based around Vapnik's ε-insensitivity approach, concerns utilization of data points

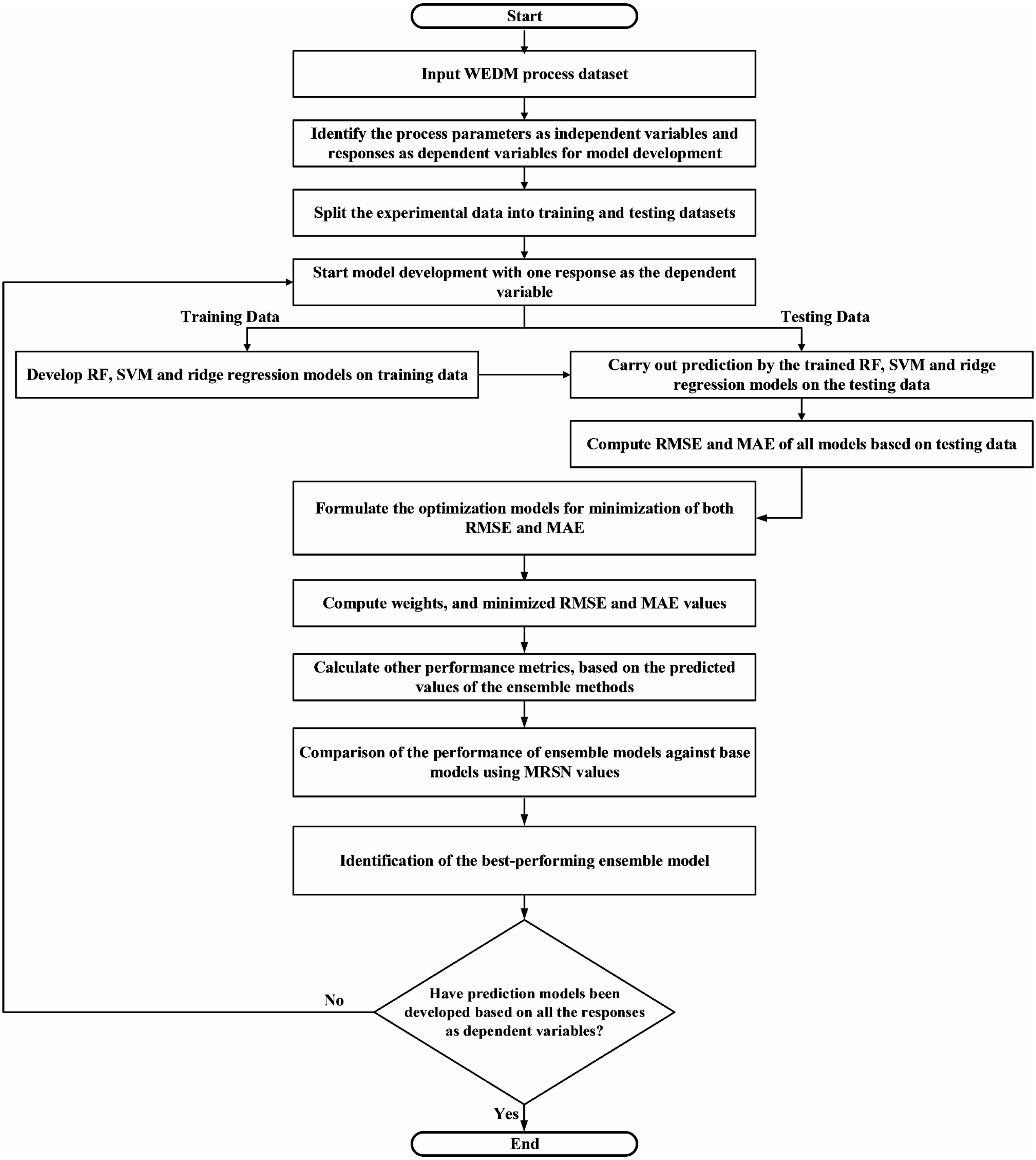

Fig. 2. Flowchart of the proposed methodology.

with absolute residual values less than some constant value ε, which is the width of a region symmetrically surrounding the fitted function represented as a hyperplane[30]. The ε-insensitive region around the function is also known as the ε-tube, which is designed to have a minimal width while including the maximum possible number of data points inside. The data points inside the tube, on either side of the fitted hyperplane, are ignored and not penalized while the ones outside the tube receive penalizations based on their separation from the fitted hyperplane. For regression model, the tube region around the hyperplane function with continuous values is obtained through a compromise between complexity of the model and error in prediction. The model development includes minimization of the convex ε-insensitive loss function and obtaining the tube with minimal width containing maximum possible number of data points. This necessitates multi-objective optimization of the loss function and tube geometry[30].

Ridge regression
Multiple linear regression (MLR) is one of the most basic prediction models for continuous response variables, and is easy to understand and interpret. However, an MLR model, developed based on the principle of least squares, underperforms with poorly-conditioned data in regards to the size of the model and prediction accuracy[31]. Such a regression model, developed to fit a training dataset, fails to maintain good prediction accuracy for a new dataset. The least square regression model also cannot deal with high dimensional data, i.e. data where the number of variables is greater than the number of observations. These drawbacks can be overcome through regularized regression models, like ridge regression. It improves estimation of the parameters while increasing bias up to an admissible level. Ridge regression introduces a biased estimator β_R, as shown in the following equation:

$$\beta_R = (X^T X + \lambda I)^{-1} X^T Y \tag{1}$$

where X is the design matrix, Y is the regressor (dependent variable), I is the identify matrix and λ is the biasing parameter having value greater than or equal to zero. If $\lambda = 0$, the above equation would be reduced to a least square estimation. The biasing parameter is responsible for regularization through controlling the size of the coefficients. The coefficients, which when unconstrained, may result in large variances, and are shrunk through penalization bringing them close to a value of zero[32]. If the value of the biasing parameter is high, bias in the estimator would then increase, while the variance would decrease. In a ridge regression model, the value of the biasing parameter is assigned so that the decrease in variance is equal to the square of the bias[33].

Optimization model for weight estimation

As aforementioned, the weight to be assigned to each of the constituent base models of the ensemble model is estimated after solving the corresponding optimization problems. Two ensemble models are obtained, one with minimum RMSE and another with minimum MAE values. They are developed with RMSE or MAE of the proposed ensemble as the objective function, which is required to be minimized. The objective function is subjected to two sets of constraint. One of them is the set of predicted response values based on the test data by the ensemble model, which is expressed as the weighted sum of the prediction of the constituent ML models. The other is the sum of the weights, which is always equal to one. The unknown values in the optimization problem are thus the response values predicted by the ensemble model, as well as weights of the constituent models, which are obtained after solving the said problem. The corresponding RMSE and MAE values are also obtained as solutions to the problems. As a result, predictions for the test data, and RMSE and MAE values based on the test data are computed for five different models, i.e. SVM, RF, ridge regression, RMSE-optimized ensemble model and MAE-optimized ensemble model. Equations (2)–(4) describe the optimization problem for RMSE or MAE-optimized ensemble model.

$$\text{Minimize}\, RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n}} \tag{2}$$

Or

$$\text{Minimize} MAE = \frac{\sum_{i=1}^{n} |y_i - \widehat{y}_i|}{n} \tag{3}$$

Subject to

$$\hat{y}_i = \sum_{j=1}^{3} w_j \times \hat{y}_{ij} \; for \; i = 1 \text{ to } n \text{ and } \sum_{j=1}^{3} w_j = 1 \tag{4}$$

whereis the *i*th observed value for the particular response, is the th predicted value for that response using the ensemble model, which is unknown and can be determined solving the optimization problem. Two sets of values are obtained, one by minimising the ensemble model RMSE and the other by the minimization of the ensemble model MAE. In each of the cases, is calculated as the weighted sum of the predictions of the base models, where w_j is the weight assigned to *j*th base model andis the predicted value for *i*th observation by *j*th base model. These weights are also unknown and can be computed through the developed optimization problems. Like the predicted response values of the ensemble models, two sets of weights are thus derived from this computation. The number of observations for which these ensemble model-based predictions can be obtained is *n*, which is equal to the number of observations in the test dataset. The two sets of values, computed for the RMSE-optimized and MAE-optimized ensembles, are thereby considered for comparison based on some statistical performance metrics. For further explanation, the codes written in LINDO for optimization of RMSE and MAE of the ensemble models for a particular WEDM response are provided in Appendix A.

Statistical performance metrics

As evident from above, two ensemble models are developed in this paper through minimization of RMSE and MAE values. The mathematical expressions for RMSE and MAE are provided as below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n}} \quad (5)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \widehat{y}_i|}{n} \quad (6)$$

Apart from the fact that these are the two most commonly reported metrics in similar studies, there is a fundamental difference in their concept. While $RMSE$ deals with squared error, MAE considers absolute error. Hence, the comparative results obtained through optimization of these metrics are likely to be different. While comparison can be made between the prediction models through the values of RMSE and MAE only, but, for a comprehensive study, it is also necessary to include other metrics to examine performance of the ensemble models and base models. For this purpose, values of MSE, R, $MAPE$, $RMSPE$, $RMSLE$, RAE and $RRSE$ are calculated here. Besides R, which is a larger-the-better metric, all the other metrics are of smaller-the-better type. Along with $RMSE$ and MAE, these seven metrics are adopted in this paper to compare performance of the considered models through calculation of MRSN ratio values, whose preliminaries are presented in the next sub-section. The expressions of the seven performance metrics are shown through Eqs. (7)–(13)[34]:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n} \quad (7)$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\widehat{y}_i - \bar{\widehat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2(\widehat{y}_i - \bar{\widehat{y}})^2}} \quad (8)$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right|}{n} \quad (9)$$

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{y_i - \widehat{y}_i}{y_i}\right)^2}{n}} \quad (10)$$

$$RMSLE = \sqrt{\frac{\sum_{i=1}^{n}(log(\widehat{y}_i + 1) - log(y_i + 1))^2}{n}} \quad (11)$$

$$RAE = \frac{\sum_{i=1}^{n}|\widehat{y}_i - y_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|} \quad (12)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (13)$$

Multi-response signal-to-noise ratio

The S/N ratio is a measure of quality characteristics for any set of observations obtained using Taguchi's orthogonal array-based experiments and the data derived from them. It was proposed as an acceptable indicator of performance, considering output characteristic of a process, and combining both the desirable and undesirable components. The desirable component is called 'signal' indicating the mean characteristic value, while the undesirable component, termed 'noise', indicates variability of the characteristic. The mathematical computation of S/N ratio is carried out on the basis of three situations, i.e. nominal-the-best, smaller-the-better and larger-the-better. In this paper, the metrics, based on which S/N ratios are calculated, are predominantly smaller-the-better, with one metric being larger-the-better type. In case of a multi-objective problem, where there are multiple contrasting parameters or objectives, a single overall S/N ratio is computed. This paper includes a mixture of smaller-the-better and larger-the-better metrics as output characteristics and hence, an overall S/N ratio, called MRSN ratio, is computed in the following steps[35].

Step 1: At first, quality losses are computed for both the larger-the-better and smaller-the-better quality characteristics, described for *k*th trial in it^{h} experiment and *j*th response, as shown below[36]:

$$L_{ij-smaller} = \frac{\sum_{k=1}^{n} z_{ijk}^2}{n} \quad (14)$$

$$L_{ij-larger} = \frac{\sum_{k=1}^{n} \frac{1}{z_{ijk}^2}}{n} \quad (15)$$

where and $L_{ij-larger}$ are the corresponding quality losses for smaller-the-better and larger-the-better quality characteristics, respectively for *n* experimental trials, whose response values are denoted as .

Step 2: This step involves scaling of the loss function by min-max normalization procedure, which is expressed using Eq. (16).

Parameter	Symbol	Unit	Level		
			− 1	0	1
Pulse-on time	T_{on}	µs	0.6	0.9	1.2
Pulse-off time	T_{off}	µs	16	22	28
Spark gap voltage	SV	V	20	30	40
Peak current	I_p	A	120	150	180
Wire tension	WT	g	850	1000	1200
Wire feed	WF	m/min	6	8	10

Table 1. WEDM parameters and their levels for example 1[37].

Response	Unit	Average	Standard deviation	Maximum	Minimum
CS	mm^2/min	75.73	18.04	104.45	35.48
SR	µm	3.35	0.30	3.89	2.75
SG	mm	0.0441	0.0060	0.058	0.033

Table 2. Summary of the WEDM responses for example 1.

$$S_{ij} = \frac{L_{ij} - L_j^{min}}{L_j^{max_j^{min}}} \tag{16}$$

where is the scaled loss function, and L_j^{min} and are the minimum and maximum loss function values for jth response, respectively.

Step 3: The total quality loss for th experiment is calculated as a weighted sum of all the scaled loss function values, as given in Eq. (17):

$$T_i = \sum_{j=1}^{p} \omega_j \times S_{ij} \tag{17}$$

where is the number of responses and ω_j is the weight assigned to *j*th response.

Step 4: In the final step, MRSN ratio for i^{th} experiment is computed using the following equation:

$$MRSN_i = -10log_{10}(T_i) \tag{18}$$

In this paper, ith experiment corresponds to ith model being studied, while jth response denotes jth model performance metric (out of p number of metrics). Here, the number of trials per experiment or model and per response or performance metric is equal to 1.

Development of ensemble models for WEDM processes

Example 1

Based on Box-Behnken design plan, Kumar et al.[37] randomly conducted 54 experiments (to avoid chance of any systematic error) on stir cast Al/SiCp metal matrix composites (MMCs) using a sprintcut CNC-WEDM setup (Electronica make), considering pulse-on time (T_{on}), pulse-off time (T_{off}), spark gap voltage (SV), peak current (I_p), wire tension (WT) and wire feed (WF) as the independent (process) variables, each set at three different levels (shown in Table 1), and CS, SR and spark gap (SG) as the dependent variables (responses). The dimensions of the MMC specimens were 5 × 5 × 15 mm. Brass wire having diameter 250 µm was used as the tool electrode and deionized water was utilized as the dielectric fluid. After each WEDM experiment, the average value of CS was estimated by measuring the total area machined and the corresponding machining time. On the other hand, a surface roughness measuring instrument (Mitutoyo make, Model S J 301) and a spark gap measuring device (TESAMASTER micrometer with least count = 0.001 mm) were employed to record SR and SG values, respectively. Furthermore, Table 2summarily presents the average, standard deviation, maximum and minimum values of each of the responses for the said WEDM process, based on the observations of 54 experiments, conducted by Kumar et al.[37].

From the experimental results, as shown in Table 2, the corresponding training and testing datasets are randomly considered based on an 80:20 split ratio, leaving 43 observations in training dataset and 11 observations in testing dataset. The split is randomly created using a function in R software v4.4.1, called 'createDataPartition', which maintains an approximate proportion between the training and testing datasets. All the three base models, i.e. RF, SVM and ridge regression, are developed on the training data, and later validated on the testing data. Those base models are framed using R software v4.4.1, applying the function train() (belongs to 'caret' package).

The train() function assesses the influence of the model tuning parameters on the prediction performance, followed by selection of the optimal model based on those parameters and gauging its performance employing the training dataset. It also checks performance of the model at different values of the tuning parameters and provides results for those tuning parameters at which the model delivers the lowest error metric (usually RMSE) or highest accuracy. At first, the base models are developed treating CS as the dependent variable. Development of the RF model involves consideration of the number of randomly chosen predictors or input variables at each split as the tuning parameter. The program calculates its value as 4, having the minimum RMSE. It is followed by the development of an SVM model with the linear kernel on the same training dataset, with the cost parameter (tuning parameter for SVM model) equal to (1) Development of the ridge regression model involves selection of λ value between 0.0001 and (2) Twenty different values of λ are tested (with a gap of 0.105258 in between each value of λ), and the one with the minimum RMSE is considered for the final model. In this case, the value of $\lambda = 1.052679$ is chosen and the corrseponding ridge regression model is developed accordingly. For the other responses, i.e. SR and SG, the RF model tuning parameter, i.e. (the number of input variables randomly chosen at each split) is set to 4, the cost parameter of the SVM model is equal to 1, while $\lambda = 0.0001$ is observed to provide the lowest RMSE for the ridge regression model.

All the three models developed on the training dataset are now implemented on the test dataset, where they provide predictions for each of the responses. These predictions as well as actual experimental observations are subsequently utilized for formulation of the corresponding optimization problem in the framework, as shown in Eqs. (2)–(4). As mentioned earlier, two optimization problems are developed for each of the responses, one minimizing RMSE and the other minimizing MAE. Therefore, a total of six optimization problems are framed in this example (two for each of the three responses) and their solutions are derived using LINDO software (v15). The outputs would provide the respective objective function values (RMSE and MAE) along with the predicted responses for RMSE-optimized and MAE-optimized ensemble models, and weight of each of the base models. The weights of the corresponding base models and predicted values for each of the responses (based on the test dataset) are exhibited in Tables 3 and 4, respectively. It can be observed that for MAE-optimized models, ridge regression, SVM and RF have the highest weightage for the responses CS, SR and SG, respectively. However, in case of RMSE-optimized models, SVM has the highest weightage for responses CS and SR, while RF has the maximum weight for SG response.

Following prediction of the responses using all the five models, performance of each of them is now ascertained using the corresponding statistical metrics. Thus, all the nine metrics, i.e. MAE, RMSE, MSE, R, MAPE, RMSPE, RMSLE, RAE and RRSE are computed for each model, for the three responses. These metrics, except R, are computed in R software v4.4.1, using the package 'MLmetrics'. Consequently, MRSN ratio is also calculated for each model, based on the computed values of all the metrics. These metrics are treated as multiple responses and each model is considered equivalent to an experimental setting. Hence, the number of trials per experiment in this paper would be equal to one, since there is only one value for a model-metric combination. The quality loss for each value of the metrics is calculated, followed by scaling of the quality loss, and computation of the total quality loss and MRSN ratio of each model for each response. Here, the weights assigned to the performance metrics are considered equal to each other, and hence, the total quality loss is computed as the average of all the scaled losses for a particular model. Tables 5, 6 and 7 show the performance metric values, total quality loss and MRSN ratio of all the five models, for CS, SR and SG respectively. Similarly, Figs. 3, 4 and 5 exhibit the scaled quality losses for CS, SR and SG respectively.

From Tables 5, 6 and 7, it is evident based on the cumulative MRSN ratio values that the accuracies of RMSE-optimized ensemble models are greater than the other models for all the responses. It is followed by MAE-optimized ensemble models, which have the second-largest MRSN ratio values. Figures 3, 4 and 5 reiterate these observations, as the two ensemble models have the lowest scaled quality losses. For responses CS and SR, RF models have the lowest MRSN ratio value, followed by ridge regression and SVM models. However, for response SG, ridge regression has the lowest MRSN value, followed by SVM and RF models. It is also apparent from the tables and illustrated figures that the two developed ensemble models distinctly outperform the individual base models. The response SR has the closest gap between the inferior ensemble model and best-performing individual base model, where MRSN ratio of MAE-optimized ensemble model is distinctly greater than that of SVM model by a margin of 2.998.

It can be noted that RMSE-optimized ensemble models perform best with respect to those metrics involving computation of the squared error. In basic regression, the aim is to minimize the loss function, i.e. sum of squared errors, through adjustment of the model parameters. Furthermore, while absolute errors provide equal weightage to all the errors, squaring an error magnifies the larger error more than the smaller error. Penalizing

Response	Ensemble model	Weight		
		RF	SVM	Ridge regression
CS	MAE-optimized	0.3856	0	0.6144
	RMSE-optimized	0.3884	0.6116	0
SR	MAE-optimized	0.0483	0.9372	0.0145
	RMSE-optimized	0.2336	0.7664	0
SG	MAE-optimized	0.8809	0.1191	0
	RMSE-optimized	0.5898	0.4102	0

Table 3. Weights of the base models for each ensemble model for example 1.

Response	Experimental observation	Prediction				
		RF	SVM	Ridge regression	MAE-optimized ensemble	RMSE-optimized ensemble
CS	78.29	76.1898	77.7557	78.3461	77.5147	77.1475
	48.73	55.0563	52.0667	53.9344	54.3669	53.2279
	60.68	77.3148	61.3703	62.1458	67.9945	67.5634
	42.56	52.4567	51.2898	52.5480	52.5128	51.7431
	60.37	56.2794	64.5442	65.2896	61.8155	61.3340
	96.78	92.7035	102.5207	101.1997	97.9238	98.7075
	47.92	51.1089	67.1700	67.7200	61.3152	60.9316
	104.45	86.7114	105.1465	103.6302	97.1067	97.9860
	96.29	87.7493	104.8567	101.6496	96.2900	98.2119
	72.48	78.2383	76.8362	76.7850	77.3454	77.3808
	72.38	78.2383	76.8362	76.7850	77.3454	77.3808
SR	3.21	3.3331	3.3961	3.3815	3.3928	3.3814
	3.08	3.1057	3.2320	3.2773	3.2265	3.2025
	2.96	3.3045	2.9413	3.0219	2.9600	3.0261
	2.95	3.0349	2.9365	3.0038	2.9423	2.9595
	3.02	3.0774	3.1821	3.1898	3.1771	3.1576
	3.62	3.5760	3.7490	3.6983	3.7399	3.7086
	3.08	2.9738	3.2819	3.2550	3.2666	3.2099
	3.89	3.5057	3.8488	3.7634	3.8310	3.7686
	3.76	3.4594	3.6793	3.6778	3.6687	3.6280
	3.36	3.3643	3.3540	3.3666	3.3547	3.3564
	3.32	3.3643	3.3540	3.3666	3.3547	3.3564
SG	0.048	0.0478	0.0444	0.0443	0.0474	0.0464
	0.041	0.0413	0.0392	0.0387	0.0410	0.0404
	0.043	0.0420	0.0440	0.0436	0.0422	0.0428
	0.042	0.0420	0.0387	0.0387	0.0416	0.0406
	0.036	0.0413	0.0387	0.0384	0.0410	0.0402
	0.054	0.0498	0.0491	0.0491	0.0497	0.0495
	0.033	0.0381	0.0367	0.0371	0.0380	0.0376
	0.047	0.0497	0.0472	0.0478	0.0494	0.0487
	0.047	0.0496	0.0465	0.0467	0.0492	0.0483
	0.04	0.0406	0.0443	0.0443	0.0410	0.0421
	0.041	0.0406	0.0443	0.0443	0.0410	0.0421

Table 4. Predictions based on base models and ensemble models for example 1.

Model	Performance metric									Total quality loss	MRSN ratio
	MSE	MAE	RMSE	*R*	MAPE	RMSPE	RMSLE	RAE	RRSE		
RF	83.427	7.655	9.134	0.915	0.114	0.137	0.127	0.445	0.450	0.9315	0.3081
SVM	56.523	5.503	7.518	0.967	0.094	0.145	0.124	0.320	0.370	0.3704	4.3133
Ridge regression	57.465	5.522	7.581	0.969	0.099	0.153	0.130	0.321	0.373	0.4508	3.4606
MAE-optimized ensemble	42.729	5.167	6.537	0.983	0.091	0.126	0.113	0.300	0.322	0.0961	10.1738
RMSE-optimized ensemble	38.336	5.082	6.192	0.982	0.087	0.119	0.106	0.295	0.305	0.0021	26.8590

Table 5. MRSN ratio-based prediction performance for response CS.

large errors by squaring them is necessary for improvement of performance of a model[38]. Keeping it in mind, multiple squared error performance metrics are thereby included in this paper.

Comparing values of the non-error-based metric included in the paper, i.e. *R*, which is the correlation between the actual and predicted values of a response, it is observed that RMSE-optimized ensemble models are consistently better than the individual base models. They are also observed to be better than MAE-optimized ensemble models in two of the three cases (for SR and SG), while for response CS, both ensemble models are almost equal in performance. So, RMSE-optimized ensemble models are proven to be the most versatile and accurate among all the considered prediction models.

Model	Performance metric									Total quality loss	MRSN ratio
	MSE	MAE	RMSE	*R*	MAPE	RMSPE	RMSLE	RAE	RRSE		
RF	0.036	0.138	0.190	0.824	0.041	0.056	0.043	0.516	0.605	1	0
SVM	0.014	0.093	0.117	0.951	0.029	0.037	0.028	0.348	0.373	0.0831	10.8051
Ridge regression	0.015	0.106	0.123	0.963	0.033	0.039	0.029	0.397	0.390	0.1873	7.2741
MAE-optimized ensemble	0.013	0.090	0.114	0.954	0.028	0.036	0.027	0.336	0.361	0.0417	13.8033
RMSE-optimized ensemble	0.011	0.093	0.107	0.964	0.028	0.033	0.025	0.346	0.340	0.0124	19.0478

Table 6. MRSN ratio-based prediction performance for response SR.

Model	Performance metric									Total quality loss	MRSN ratio
	MSE	MAE	RMSE	*R*	MAPE	RMSPE	RMSLE	RAE	RRSE		
RF	7.95×10^{-6}	0.002	0.003	0.893	0.050	0.073	0.003	0.461	0.503	0.3381	4.7098
SVM	9.31×10^{-6}	0.003	0.003	0.854	0.064	0.073	0.003	0.599	0.545	0.8636	0.6368
Ridge regression	9.71×10^{-6}	0.003	0.003	0.842	0.065	0.075	0.003	0.613	0.556	1	0
MAE-optimized ensemble	7.31×10^{-6}	0.002	0.003	0.907	0.049	0.070	0.003	0.445	0.483	0.1618	7.9109
RMSE-optimized ensemble	6.67×10^{-6}	0.002	0.003	0.925	0.051	0.066	0.002	0.474	0.461	0.0497	13.0388

Table 7. MRSN ratio-based prediction performance for response SG.

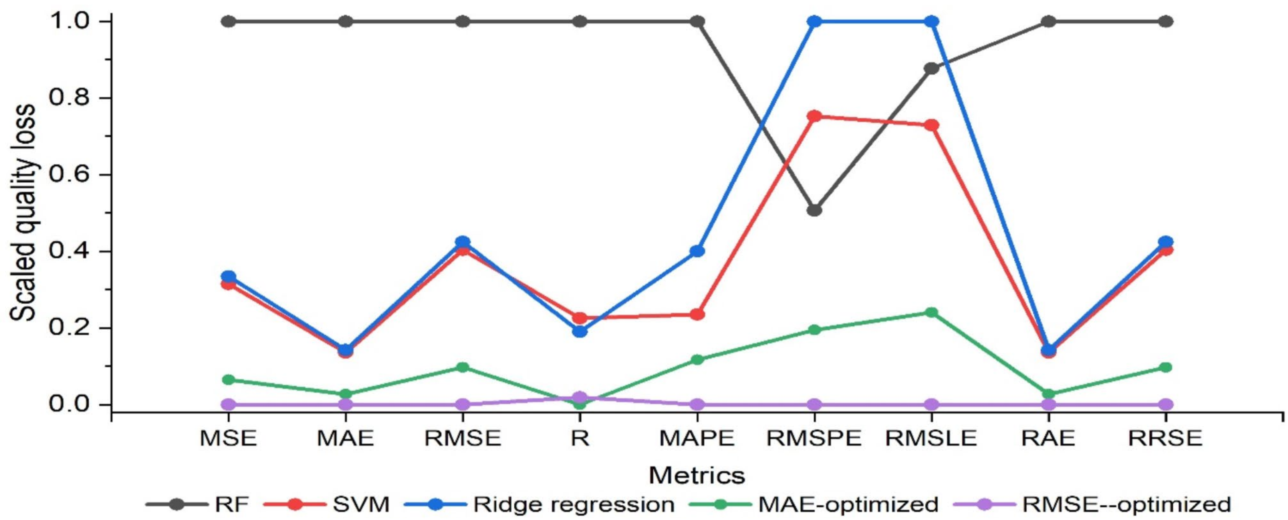

Fig. 3. Scaled quality losses for response CS.

Example 2

Kumar et al.[39] conducted 54 WEDM experiments on CP-Ti-G2 alloy (a bio-compatible material), with T_{on}, T_{off}, SV, I_p, WT and wire speed (WS) as the input variables, and MRR and SR as the responses. Each of the WEDM parameters was varied at three different levels, as shown in Table 8. While performing WEDM experiments, CuZn37 HH brass wire with diameter 250 μm was utilized as the tool electrode, while the dielectric pressure was maintained at 7 kg/cm^2. The workpiece dimension was considered as $10\times10\times26$ mm. Furthermore, Table 9provides a summary of each of the responses for the said WEDM process based on the research work of Kumar et al.[39]. From these experimental results, 45 observations are considered as training data, while 9 observations as testing data. As implemented in the previous example, the split is randomly created by 'createDataPartition' function, maintaining an approximate proportion between the training and testing datasets. The change in the number of observations in training and testing data, although derived randomly, also helps to demonstrate that the developed ensemble models are applicable to varying proportions of training and testing datasets.

Likewise, the previous example, three base models, i.e. RF, SVM and ridge regression, are developed on the training data, and subsequently validated on the testing data. It is followed by development of two weighted average ensemble models, obtained through minimization of RMSE and MAE values. The base models are first

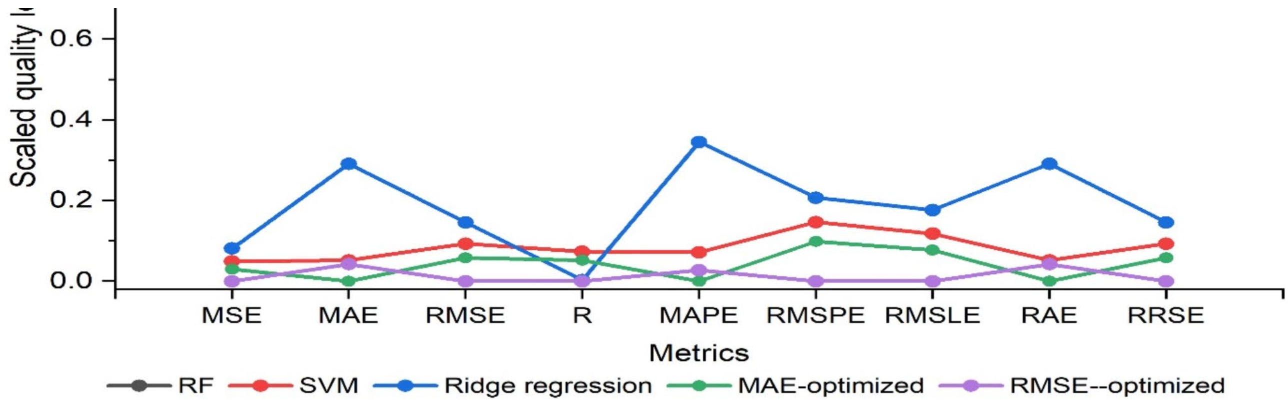

Fig. 4. Scaled quality losses for response SR.

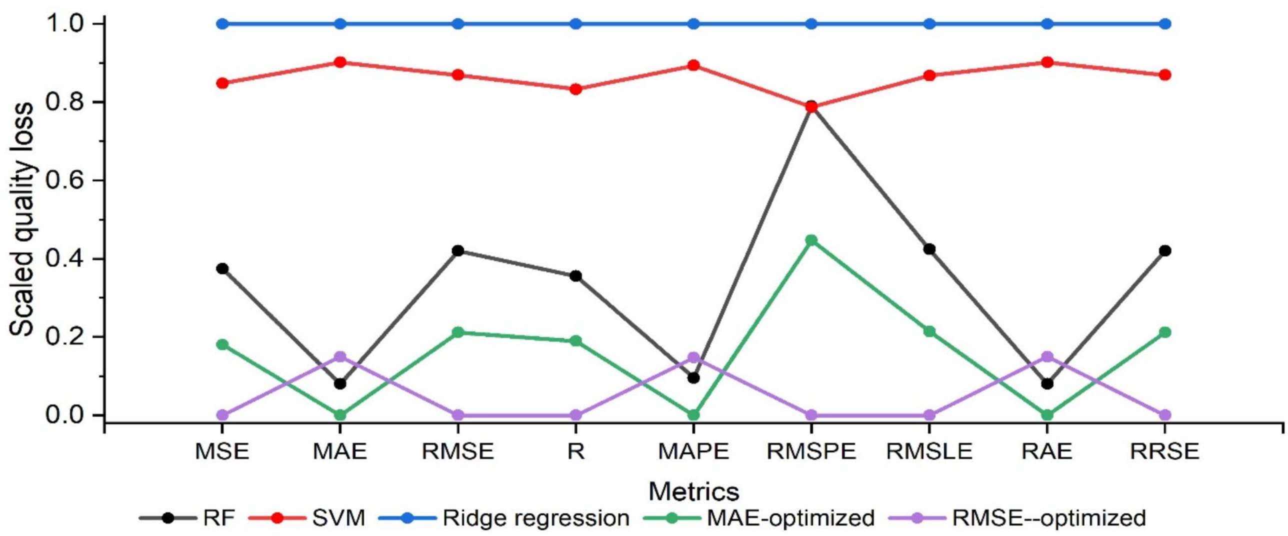

Fig. 5. Scaled quality losses for the response SG.

Parameter	Symbol	Unit	Level		
			−1	0	1
Pulse-on time	T_{on}	µs	0.7	0.9	1.1
Pulse-off time	T_{off}	µs	17	28	38
Spark gap voltage	SV	V	40	50	60
Peak current	I_p	A	120	160	200
Wire tension	WT	g	500	950	1400
Wire speed	WF	m/min	4	7	10

Table 8. WEDM parameters and their levels for example 2[39].

Response	Unit	Average	Standard deviation	Maximum	Minimum
SR	µm	2.67	0.24	3.28	2.15
MRR	mm^3/min	6.16	1.90	11.16	3.28

Table 9. Summary of WEDM responses for example 2.

Response	Ensemble model	Weight		
		RF	SVM	Ridge regression
SR	MAE-optimized	0.4018	0.5982	0
	RMSE-optimized	0.0749	0.9251	0
MRR	MAE-optimized	0.5419	0	0.4581
	RMSE-optimized	0.5747	0	0.4253

Table 10. Weights of the base models for each ensemble model for example 2.

Response	Experimental observation	Prediction				
		RF	SVM	Ridge regression	MAE-optimized ensemble	RMSE-optimized ensemble
SR	2.48	2.4812	2.5048	2.5183	2.4953	2.5030
	2.83	2.8332	2.7997	2.7903	2.8131	2.8022
	2.35	2.5175	2.5229	2.5524	2.5207	2.5225
	2.51	2.5211	2.5276	2.5391	2.5250	2.5271
	2.65	2.6590	2.6613	2.6713	2.6604	2.6611
	2.98	2.7868	2.9581	2.9583	2.8893	2.9453
	2.33	2.5169	2.5128	2.5314	2.5145	2.5131
	2.75	2.7577	2.7448	2.7526	2.7500	2.7458
	2.82	2.7656	2.8679	2.8697	2.8268	2.8602
MRR	4.77	4.9087	5.0339	4.8284	4.9237	4.8719
	7.03	7.5052	7.6014	7.5174	7.5413	7.5108
	4.27	3.8798	4.6207	4.8005	4.4336	4.3015
	4.44	4.5820	4.9247	4.9336	4.8134	4.7430
	5.61	5.1460	6.1110	6.1589	5.8053	5.6100
	8.28	8.0773	8.0660	8.1238	8.0890	8.0986
	4.41	3.8269	4.7025	4.9428	4.4907	4.3381
	7.61	8.0267	6.3265	6.7022	7.0185	7.4200
	7.53	8.0393	7.1057	7.2225	7.4558	7.6651

Table 11. Predictions of the base models and ensemble models for example 2.

developed with SR as the dependent variable. The number of randomly selected predictors (input variables) at each split, at which the corresponding RMSE value is observed to be minimum, is estimated as 4. The cost parameter of the linear kernel SVM model is set as 1. Among the 20 equidistant λ values from 0.0001 to 2, the value of $\lambda = 0.0001$, corresponding to minimum error, is chosen during development of the ridge regression model. On the other hand, during development of the base models for MRR, the number of input variables selected randomly at each split for the RF model is determined as 6. The cost parameter of the SVM model is observed to be equal to 1, while λ value considered for ridge regression is 0.0001.

These three base models, developed on the training dataset, are now validated using their predictions on the test dataset containing 9 observations for both the SR and MRR responses. Like the previous example, predictions of the base models and actual experimental observations form the basis for formulation of two optimization problems, one minimizing RMSE and the other minimizing RAE, for SR and MRR, resulting in a total of four such problems. For both the responses, solutions to the problems include the respective objective function values (RMSE and MAE), predicted response values for both the ensemble models and weights for the base models. The weights of the base models and predicted response values based on the base models as well as ensemble models are depicted in Tables 10 and 11, respectively.

It can be noticed that for response SR, SVM has the maximum weight for both the MAE-optimized and RMSE-optimized ensemble models. On the other hand, in case of MRR, RF has the maximum weight for both the ensemble models. Values of the nine metrics, i.e. MAE, RMSE, MSE, *R*, MAPE, RMSPE, RMSLE, RAE and RRSE, describing performance of the corresponding models based on their predictions, are computed for each of the five models. Finally, a cumulative univariate metric, in the form of MRSN ratio, is derived for each of the models for both the responses through estimation of quality loss, scaled quality loss and total quality loss values. Tables 12 and 13 depict those performance metrics, total quality loss and MRSN ratio of all the models, for both SR and MRR. The scaled quality losses for SR and MRR, are shown in Figs. 6 and 7, respectively.

From Tables 12 and 13, it can be revealed that the cumulative measures of errors for RMSE-optimized ensemble models are the maximum among the five models, for both SR and MRR. While for MRR, MAE-optimized weighted average ensemble model performs almost as accurately as RMSE-optimized model, and it is outperformed by SVM base model in case of SR. The MAE-optimized ensemble model is slightly better than SVM model in case of those measures not involving squared errors, while the opposite is observed for

Model	Performance metric									Total quality loss	MRSN ratio
	MSE	MAE	RMSE	*R*	MAPE	RMSPE	RMSLE	RAE	RRSE		
RF	0.012	0.070	0.107	0.924	0.028	0.042	0.030	0.367	0.497	1	0
SVM	0.008	0.057	0.087	0.960	0.023	0.037	0.025	0.298	0.403	0.0255	15.9320
Ridge regression	0.010	0.067	0.099	0.951	0.028	0.042	0.029	0.351	0.461	0.6834	1.6532
MAE-optimized ensemble	0.008	0.057	0.090	0.962	0.023	0.037	0.026	0.296	0.416	0.0645	11.9025
RMSE-optimized ensemble	0.008	0.057	0.087	0.962	0.023	0.037	0.025	0.298	0.403	0.0142	18.4683

Table 12. MRSN ratio-based prediction performance for response SR for example 2.

Model	Performance metric									Total quality loss	MRSN ratio
	MSE	MAE	RMSE	*R*	MAPE	RMSPE	RMSLE	RAE	RRSE		
RF	0.161	0.369	0.401	0.981	0.065	0.072	0.062	0.257	0.264	0.43916	3.57382
SVM	0.329	0.487	0.574	0.935	0.082	0.090	0.079	0.339	0.377	0.99968	0.00141
Ridge regression	0.255	0.447	0.505	0.959	0.079	0.090	0.074	0.311	0.332	0.7844	1.0543
MAE-optimized ensemble	0.047	0.166	0.218	0.991	0.027	0.036	0.030	0.116	0.143	0.0010	30.1094
RMSE-optimized ensemble	0.047	0.168	0.216	0.991	0.028	0.036	0.030	0.117	0.142	0.0010	30.1543

Table 13. MRSN ratio-based prediction performance for response MRR for example 2.

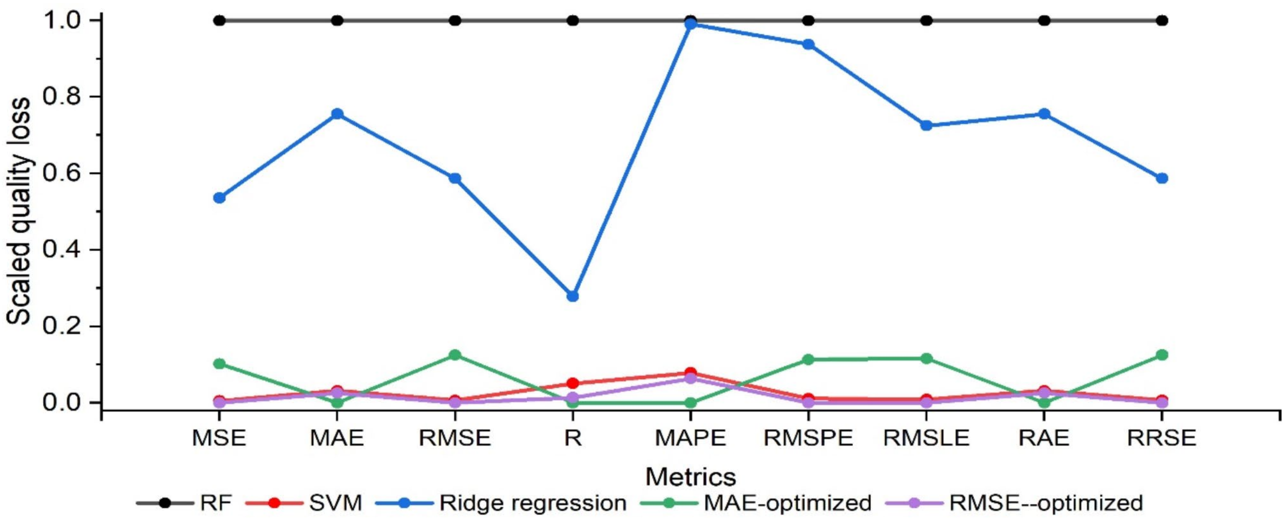

Fig. 6. Scaled quality losses for response SR for example 2.

squared error measures. Simultaneously, for response SR, RF has the lowest MRSN ratio, followed by ridge regression, while SVM shows the highest cumulative level of inaccuracy in prediction of MRR values, followed by ridge regression. The only metric that does not consider the prediction errors, *R*, shows that the predictions of response values by the two ensemble models are closely correlated to the actual values than the base models, for both the responses.

Conclusions

Ensemble models prove to be an effective means of predicting response values of WEDM processes, based on input parameters. A single ML-based prediction model may exhibit high bias or variability in its prediction. The combination of the predicted values by multiple ML models in ensemble models provides aggregated results, compensating for both bias and variance in prediction. In a heterogeneous ensemble, the base ML models employed are different from each other, integrating advantages of those models, and making the prediction framework more generalized and robust. In this paper, two heterogeneous ensemble models are developed for prediction of responses of WEDM processes based on two experimental datasets. It is followed by comparison of prediction performance of the base models and ensemble models for both the cases.

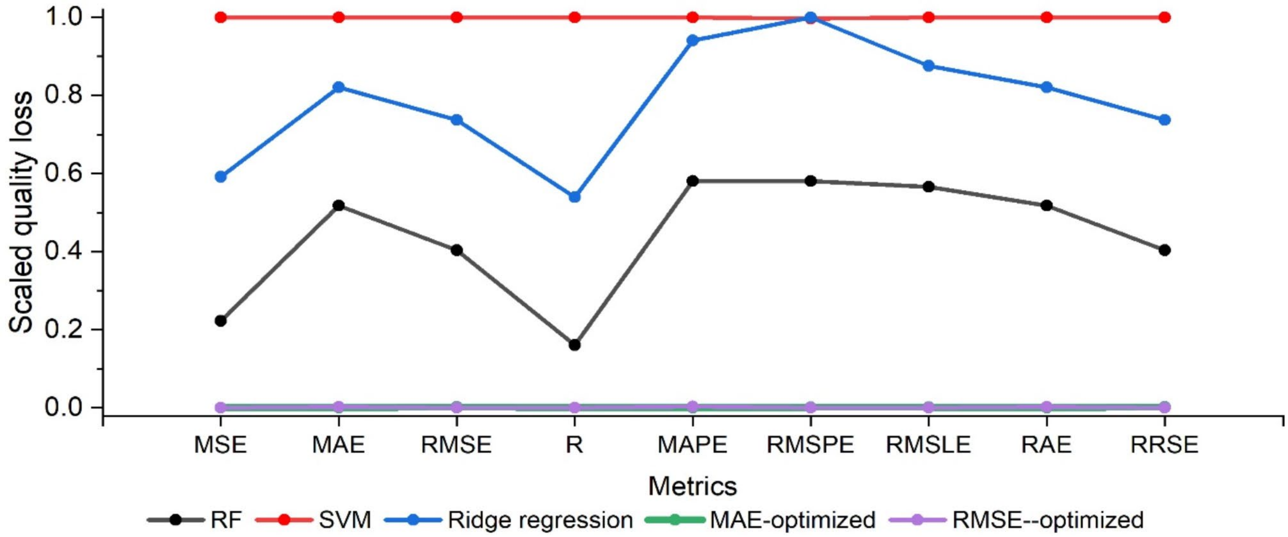

Fig. 7. Scaled quality losses for the response MRR for example 2.

After splitting the experimental data into corresponding training and testing sets, three base models, i.e. RF, SVM and ridge regression, are developed on the training datasets, providing predictions based on the input parameters in the testing datasets. From the predictions on the testing datasets, two optimization problems are developed where the objective functions involve minimization of RMSE and MAE values. Based on the derived solutions, two optimized weighted average ensemble models with minimum MAE and RMSE values are formulated, where the predictions of the two ensemble models are the weighted sum of predictions of the base models. Besides RMSE and MAE, other performance metrics, e.g. MSE, *R*, MAPE, RMSPE, RMSLE, RAE and RRSE are also calculated for all the models. A comprehensive MRSN ratio is finally computed based on quality losses for all the models, aiding in comparison between the models. Based on MRSN ratio values, it becomes evident that RMSE-optimized ensemble model has the best prediction performance for all the responses under consideration, while MAE-optimized ensemble, barring some exceptions, generally outperforms the base models.

The relationships between the responses and input parameters of a WEDM process are complex, where the interdependence can be linear or non-linear. Ensemble models, especially heterogeneous ensembles, are able to deal with such relationships, due to deployment of a combination of different models. As previously mentioned, heterogeneous ensembles assimilate benefits of different unique models, while negating their drawbacks, making the prediction more robust and generalized. These models usually have higher prediction accuracy, especially when the weights, allocated to the predictions of base models, are optimized. This combination of base model predictions balances the amount of bias and variance, thus making those models less susceptible to underfitting and/or overfitting. Consequently, the ensemble model-based predictions are stable, containing minimum noise. Based on these considerations, such ensemble models can be developed as an acceptable tool for modeling of WEDM processes and prediction of the responses of such machining processes. Future studies may include integration of more diverse ML algorithms into ensemble modeling, like ANN, ANFIS, K-nearest neighbours, AdaBoost, etc., as well as extension of such studies to other traditional and non-traditional machining processes.

Data availability
The data presented in this study are available through email upon request to the corresponding author.

Received: 23 July 2024; Accepted: 25 September 2024
Published online: 07 October 2024

References
1. Mukherjee, R., Chakraborty, S. & Samanta, S. Selection of wire electrical discharge machining process parameters using non-traditional optimization algorithms. *Appl. Soft Comput.* **12**, 2506–2516 (2012).
2. Ho, K. H., Newman, S. T., Rahimifard, S. & Allen, R. D. State of the art in wire electrical discharge machining (WEDM). *Int. J. Mach. Tools Manuf.* **44**, 1247–1259 (2004).
3. Puri, A. B. & Bhattacharyya, B. An analysis and optimisation of the geometrical inaccuracy due to wire lag phenomenon in WEDM. *Int. J. Mach. Tools Manuf.* **43**, 151–159 (2003).
4. Tsai, H. C., Yan, B. H. & Huang, F. Y. EDM performance of Cr/Cu-based composite electrodes. *Int. J. Mach. Tools Manuf.* **43**, 245–252 (2003).
5. Sarker, B. & Chakraborty, S. Parametric modelling of a wire electrical discharge machining process using path analysis approach. *Int. J. Model. Simul.* **42**, 902–919 (2022).
6. Zhang, F. et al. Data-driven and knowledge-guided prediction model of milling tool life grade. *Int. J. Comput. Integr. Manuf.* **37**, 669–684 (2023).

7. Wu, Q., Zhou, X. & Pan, X. Cutting tool wear monitoring in milling processes by integrating deep residual convolution network and gated recurrent unit with an attention mechanism. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **237**, 1171–1181 (2023).
8. Behera, B. C., Moharana, B. R., Rout, M. & Debnath, K. *Application of machine learning in the machining processes*, pp. 141–156 (Wiley, 2023).
9. Wuest, T., Weimer, D., Irgens, C. & Thoben, K. D. Machine learning in manufacturing: advantages, challenges, and applications. *Prod. Manuf. Res.* **4**, 23–45 (2016).
10. Bhattacharya, S. & Chakraborty, S. Prediction of responses in a cnc milling operation using random forest regressor. *Facta Univ. Ser. Mech. Eng.* **21**, 685 (2023).
11. Wang, Z., Yuan, Y., Zhang, S., Lin, Y. & Tan, J. A multi-state fusion informer integrating transfer learning for metal tube bending early wrinkling prediction. *Appl. Soft Comput.* **151**, 110991 (2024).
12. Xiang, Y. et al. Cross-sectional performance prediction of metal tubes bending with tangential variable boosting based on parameters-weight-adaptive CNN. *Expert Syst. Appl.* **237**, 121465 (2024).
13. Li, J. et al. Task incremental learning-driven Digital-Twin predictive modeling for customized metal forming product manufacturing process. *Robot. Comput. Integr. Manuf.* **85**, 102647 (2024).
14. Opitz, D. & Maclin, R. Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.* **11**, 169–198 (1999).
15. Nain, S. S., Garg, D. & Kumar, S. Performance evaluation of the WEDM process of aeronautics super alloy. *Mater. Manuf. Process.* **33**, 1793–1808 (2018).
16. Ulas, M., Aydur, O., Gurgenc, T. & Ozel, C. Surface roughness prediction of machined aluminum alloy with wire electrical discharge machining by different machine learning algorithms. *J. Mater. Res. Technol.* **9**, 12512–12524 (2020).
17. Lalwani, V., Sharma, P., Pruncu, C. I. & Unune, D. R. Response surface methodology and artificial neural network-based models for predicting performance of wire electrical discharge machining of Inconel 718 alloy. *J. Manuf. Mater. Process.* **4**, 44 (2020).
18. Naresh, C., Bose, P. S. C. & Rao, C. S. P. Artificial neural networks and adaptive neuro-fuzzy models for predicting WEDM machining responses of Nitinol alloy: comparative study. *SN Appl. Sci.* **2**, (2020).
19. Dandge, S. S. & Chakraborty, S. A data mining approach for analysis of a wire electrical discharge machining process, *Manag. Prod. Eng. Rev.* (2021).
20. Shanmugasundar, G. et al. A comparative study of linear, random forest and AdaBoost regressions for modeling non-traditional machining. *Processes (Basel)* **9**, 2021 (2015).
21. Sharma, V., Misra, J. P. & Singhal, S. Surface roughness modeling using machine learning approaches for wire electro-spark machining of titanium alloy. *Int. J. Struct. Integr.* **13**, 999–1012 (2022).
22. Gurgenc, T. & Altay, O. Surface roughness prediction of wire electric discharge machining (WEDM)-machined AZ91D magnesium alloy using multilayer perceptron, ensemble neural network, and evolving product-unit neural network. *Mater. Test.* **64**, 350–362 (2022).
23. Saha, S., Gupta, K. K., Maity, S. R. & Dey, S. Data-driven probabilistic performance of Wire EDM: A machine learning based approach. *Proc. Inst. Mech. Eng. Pt. B: J. Eng. Manuf.* **236**, 908–919 (2022).
24. Verma, A. S. & Singh, S. Experimental investigation and prediction modelling of slicing speed and surface roughness during wafer slicing using WEDM. *Eng. Res. Express* **4**, 035028 (2022).
25. Ishfaq, K., Sana, M. & Ashraf, W. M. Artificial intelligence–built analysis framework for the manufacturing sector: performance optimization of wire electric discharge machining system. *Int. J. Adv. Manuf. Technol.* **128**, 5025–5039 (2023).
26. Saha, S., Maity, S. R. & Dey, S. Prediction of WEDM performances using clustering techniques in ANFIS during machining of A286 superalloy. *J. Inst. Eng. (India) Ser. C* **104**, 315–326 (2023).
27. Jithendra, T., Sharief Basha, S., Das, R. & Gajjela, R. Modeling and optimization of WEDM of monel 400 alloy using ANFIS and snake optimizer: A comparative study. *Proc. Inst. Mech. Eng. Part C* **238**, 1573–1589 (2024).
28. Natarajan, M. et al. Optimization of wire spark erosion machining of Grade 9 titanium alloy (Grade 9) using a hybrid learning algorithm. *AIP Adv.* **14**, 1 (2024).
29. Kalule, R., Abderrahmane, H. A., Alameri, W. & Sassi, M. Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs. *Sci. Rep.* **13**, 9855 (2023).
30. Fang, Z., Wang, Y., Peng, L. & Hong, H. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Geogr. Inf. Syst.* **35**, 321–347 February (2021).
31. Tan, P. N., Steinbach, M. & Kumar, V. *Introduction to Data Mining* (Pearson Education, Inc., 2006).
32. Zhou, Z. H. *Ensemble Methods* (Chapman & Hall/CRC, 2011).
33. Altman, N. & Krzywinski, M. Ensemble methods: bagging and random forests. *Nat. Methods* **14**, 933–934 (2017).
34. Montesinos Lopez, O. A., Montesinos Lopez, A. & Crossa, J. *Multivariate Statistical Machine Learning Methods for Genomic Prediction* 1 edn (Springer Nature, 2022).
35. Khanna, R., Awad, M. & Khanna, A. Efficient Learning Machines, Elk, Grove: Apress, (2015).
36. Bhattacharya, S., Kalita, K., Čep, R. & Chakraborty, S. A comparative analysis on prediction performance of regression models during machining of composite materials. *Materials (Basel)* **14**, 6689 (2021).
37. Saleh, A. K. M. E., Arashi, M. & Kibria, B. M. G. *Theory of Ridge Regression Estimation with Applications* (Wiley, 2019).
38. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to Linear Regression Analysis* (Wiley, 2012).
39. Mitra, A. *Fundamentals of Quality Control and Improvement* 5 edn (Wiley, 2021).
40. Sapkota, G., Ghadai, R. K., Das, S., Das, P. P. & Chakraborty, S. A comparative study on multi-objective optimization of drilling of hybrid aluminium metal matrix composite. *Int. J. Interact. Des. Manuf. (IJIDeM)* **17**, 3177–3187 (2023).
41. Tong, L. I., Su, C. T. & Wang, C. H. The optimization of multi-response problems in the Taguchi method. *Int. J. Qual. Reliab. Manag.* **14**, 367–380 (1997).
42. Kumar, H., Manna, A. & Kumar, R. Modeling and desirability approach-based multi-response optimization of WEDM parameters in machining of aluminum metal matrix composite. *J Braz Soc. Mech. Sci. Eng* **40**, 1 (2018).
43. Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model. Dev.* **7**, 1247–1250 (2014).
44. Kumar, A., Sharma, R., Gupta, A. K. & Gujral, R. Investigation of biocompatible implant material through WEDM process using RSM modeling hybrid with the machine learning algorithm. *Sadhana* **46** (2021).

Acknowledgements

This article was supported by the project SP2024/087 from Ministry of Education Youth and Sport of Czech Republic and VSB-TUO.

Author contributions

Author ContributionsConceptualization: Baneswar Sarker, Shankar Chakraborty, Robert Čep, Kanak Kalita; Formal analysis: Baneswar Sarker; Investigation: Baneswar Sarker; Methodology: Baneswar Sarker, Shankar Chakraborty, Robert Čep, Kanak Kalita; Writing – original draft: Baneswar Sarker, Shankar Chakraborty; Writing – review & editing: Baneswar Sarker, Shankar Chakraborty, Robert Čep, Kanak Kalita; All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-74291-x.

Correspondence and requests for materials should be addressed to S.C. or K.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.