Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2015
August 2-5, 2015, Boston, Massachusetts, USA

DETC2015-47957

ENSEMBLE NEURAL NETWORK MODEL FOR PREDICTING THE ENERGY CONSUMPTION OF A MILLING MACHINE

Ronay Ak*
Systems Integration Division
National Institute of Standards and Technology
Gaithersburg, 20899, Maryland
E-mail: ronay.ak@nist.gov

Moneer M. Helu
Systems Integration Division
National Institute of Standards and Technology
Gaithersburg, 20899, Maryland
E-mail: moneer.helu@nist.gov

Sudarsan Rachuri
Systems Integration Division
National Institute of Standards and Technology
Gaithersburg, 20899, Maryland
E-mail: Sudarsan.rachuri@nist.gov

ABSTRACT

Accurate prediction of the energy consumption is critical for energy-efficient production systems. However, the majority of existing prediction models aim at providing only point predictions and can be affected by uncertainties in the model parameters and input data. In this paper, a prediction model that generates prediction intervals (PIs) for estimating energy consumption of a milling machine is proposed. PIs are used to provide information on the confidence in the prediction by accounting for the uncertainty in both the model parameters and the noise in the input variables. An ensemble model of neural networks (NNs) is used to estimate PIs. A k-nearest-neighbors (k-nn) approach is applied to identify similar patterns between training and testing sets to increase the accuracy of the results by using local information from the closest patterns of the training sets. Finally, a case study that uses a dataset obtained by machining 18 parts through face-milling, contouring, slotting and pocketing, spiraling, and drilling operations is presented. Of these six operations, the case study focuses on face milling to demonstrate the effectiveness of the proposed energy prediction model.

* Corresponding author. Tel: +1-301-975-8655.
Address: Systems Integration Division, NIST, 20899, MD, USA.

Keywords: Energy prediction, Ensemble, Manufacturing, Milling, Neural networks, Prediction intervals.

INTRODUCTION

The ability to measure environmental impacts, such as energy and material efficiency, of different industrial processes has become critically important for manufacturing companies as they attempt to balance economic growth with environmental impacts. On the economic side are traditional metrics including cost, quality, reliability, and productivity; i.e., the basis for measuring the performance of individual processes and systems. On the environmental side are measures for CO_2 emission, energy and water consumption. These types of measures historically have not been collected at the level of an individual manufacturing process; however, the availability of much more extensive data at the machine level is making it possible to assess these factors at that level. Machine tool controllers [1, 2] and standards (e.g., MTConnect [3]) allow manufacturers to collect and use increasingly larger amounts of data.

This large amount of data being collected from manufacturing processes and systems [4, 5] enables manufacturers to gain value from the data, i.e., transform information into insights and then into action. In doing so, data mining including advanced statistical techniques has proven to be successful for discovering useful knowledge from this large data for data-driven decision making. As part of data mining, predictive analytics have been used for a wide range of manufacturing problems such as fault diagnosis and preventive maintenance [6], product demand forecasting [7], surface roughness prediction [8]. Surface roughness prediction is a widely used index of product quality, and energy prediction for different machining operations based on measured data [1, 9, 10]. For example, since there is a lack of adequate and acceptable physics-based (e.g., mechanistic) models for metal cutting processes, empirical models (e.g., statistical regression, artificial neural networks, fuzzy set theory) are generally used in metal cutting processes [11].

Accuracy plays a crucial role in making reliable decisions based on the predicted results. Thus, accurate prediction of the required energy consumption for machine tool operations can enable manufacturers to improve operating efficiency and cost, respond to new regulations and business drivers (e.g., the Smart Grid [12]), and develop useful process monitoring approaches relying on the predicted results. Moreover, one should consider in a smart-grid-future all agents can be assumed to be prosumers; i.e., a manufacturing company can be both a power generator and a power consumer. Thus, data-driven predictive analytics will play a more determinant role in estimating the generated and required energy within a certain time period for design, planning, and operation decisions.

A prediction model must be capable of providing a quantification of the uncertainty associated with the prediction for informed decision-making. To this aim, in the present work, a novel framework is proposed to estimate prediction intervals (PIs) rather than point predictions. PIs are used to provide information on the confidence in the predictions accounting for both the uncertainty in the model parameters and noise in the input variables. The framework proposed is based on an ensemble of Neural Network (NN) models. Specifically, an NN ensemble is a learning paradigm where a certain number of NNs are combined to estimate the desired output for the target of interest (see Fig. 1) [13]. To train each individual NN; i.e., to determine the optimal values of the weights of a multi-layer perceptron neural network (MLP NN), a multi-objective genetic algorithm (MOGA) framework (non-dominated sorting genetic algorithm–II, NSGA-II) [14] has been used to estimate the bounds defining the PIs. The performance of the MOGA-NN model has been tested via different case studies and demonstrated high accurate prediction results [15, 16].

In general terms, it is well known that an ensemble of different predictors can generate predictions that are more accurate than those obtained by individual predictors [13]. Typically, an NN ensemble is constructed in two steps: i) training a number of individual NNs and ii) combining the predictions yielded from these NNs. The aim of assembling a number of NNs into an ensemble is to obtain high training accuracy and to improve the generalization ability, which is defined as the performance of the trained network on unseen data patterns.

In short, the aim of this work is to develop a reliable method to predict the total energy consumption of a milling machine tool performing cutting operations including face milling, contouring, pocketing, slotting, spiraling and drilling. The objective is to understand the effects of the machining parameters on energy consumption during the production phase. Since this is ongoing research work, this paper provides only the preliminary prediction results for a face milling operation, whereas the overall objective is to estimate the total energy required for cutting and non-cutting operations for a Computer Numerical Control (CNC) milling machine.

MULTI-LAYER PERCEPTRON NEURAL NETWORKS AND INTERVAL PREDICTION

NNs are universal approximators, often used as regression models, capable of learning non-linear patterns from historical data [17]. MLP NNs are a common and popular type of feed-forward NNs [17, 18]. A MLP NN consists of processing units called neurons or nodes that are interconnected by weights. The nodes are ordered into one input layer, one or more hidden layers, and one output layer. Each layer receives input signals generated by the previous layer, produces output signals through an activation function (e.g., a sigmoid transfer or activation function), and distributes the output to the neurons in the subsequent layer. Herein, a three-layered (i.e., input-hidden-output) MLP is used.

The prediction accuracy of an NN can depend on several factors, such as the network topology, the level of variability and uncertainty in the input data, the number of data samples used for training, the learning algorithm, and the set of initial parameters. Zhang et al. [18], Rojas [19], and Bishop [20] provide a good overview of the theoretical basics of NN modeling.

By definition, given an input-output process $\boldsymbol{x}$, $y(\boldsymbol{x})$, a PI is a statistical estimator composed by lower and upper bounds, $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$, within which a future target $y(\boldsymbol{x})$ is expected to lie with a certain probability $1-\alpha$, denoting its nominal coverage rate [21, 22]:

$$Pr(L(\boldsymbol{x}) < y(\boldsymbol{x}) < U(\boldsymbol{x})) = 1-\alpha \quad (1)$$

There are two indicators used to evaluate the quality of the PIs: the prediction interval coverage probability (PICP) and the prediction interval width (PIW). This work uses PICP as the empirical coverage rate to analyze the actual performance of the predictive modeling. It is estimated as the proportion of true output values lying within the estimated PIs [22]:

$$PICP = \frac{1}{n_p}\sum_{i=1}^{n_p} c_i \quad (2)$$

where n_p is the number of samples in the training or testing sets, and $c_i = 1$, if $y_i \in [L(\boldsymbol{x}_i), U(\boldsymbol{x}_i)]$ and otherwise $c_i = 0$. Note that Eq. (2) is an empirical version of PICP, which yields an estimate of PICP according to the frequentist interpretation of probability theory.

The prediction interval width (PIW) measures the extension of the estimated interval as the difference between the estimated lower and upper bound values. Herein, the normalized mean prediction interval width (NMPIW) is used as the second measure indicating the quality of the constructed prediction intervals. The mathematical definition of PIW and NMPIW measures are as follows [22]:

$$PIW = (U(x_i) - L(x_i)) \tag{3}$$

$$NMPIW = \frac{1}{n_p} \sum_{i=1}^{n_p} \frac{(U(x_i) - L(x_i))}{y_{max} - y_{min}} \tag{4}$$

where y_{min} and y_{max} represent the true minimum and maximum values of the target (i.e., the bounds of the range in which the true values fall) in the training set, respectively. The denominator of Eq. (4) corresponds to the range of the training set y. The regression function $y = f(\boldsymbol{x})$ is constructed by supervised learning, which requires a training dataset, $\{(\boldsymbol{x}^i, y^i) | i = 1, 2, \cdots, D\}$ where D is the number of the samples in the dataset, and the input vector $\boldsymbol{x}$ and the corresponding output vector y are given observations [23]. NMPIW is a dimensionless measure representing the average width of PIs as a percentage of the underlying target range [22]. Normalization against the range of the target allows one to objectively compare PIs of different targets.

The main requirement on the quality of the estimated PIs is a high coverage probability (CP) that the true values will be within the predicted intervals; on the other hand, to give useful practical information, the intervals need to have small widths. The two requirements conflict since a small interval will induce a low probability that the true value be within the interval itself, whereas wide intervals may be required to obtain high coverage probability.

To obtain optimal PIs based on interval size and coverage, we follow the method presented by Ak et al. [15, 16] using NSGA-II. NSGA-II is a powerful multi-objective evolutionary algorithm (MOEA) [14]. The algorithm finds the optimal values of the parameters of the NN that maximize PICP and minimize NMPIW simultaneously in the sense of Pareto optimality. The implementation steps of training a MLP NN by NSGA-II can be found in [15, 16]. With respect to the characteristics of the PI estimation problem, we can write a formal expression to better define the problem in a multi-objective framework:

Objectives: find the optimal solution set of weights $\hat{\boldsymbol{w}}$ such to:

$$\begin{array}{ll} \text{Maximize} & PICP(\boldsymbol{w}) \\ \text{Minimize} & NMPIW(\boldsymbol{w}) \\ \text{Subject to} & 0 \leq PICP(\boldsymbol{w}) \leq 1 \\ & NMPIW(\boldsymbol{w}) \geq 0. \end{array} \tag{5}$$

where $\boldsymbol{w}$ is the vector of the NN weights (parameters) to be optimized during the training process. To simplify implementation, maximizing PICP was accomplished by minimizing 1-PICP.

Note that the proposed MOGA-NN method aims to obtain optimal PIs without the prior knowledge, statistical inference or distribution assumption of forecasting errors required in most traditional approaches.

Khosravi et al. [22] proposed a “Lower and Upper Bound Estimation Method (LUBE)” that obtains NN-based PIs by considering both CP and PIW in the PI construction phase. They performed the training by using a single-objective PI-based cost function called Coverage Width Criterion (CWC), which combines two separate objectives: PICP and NMPIW. The adapted CWC objective function used in the present work is defined as [21, 22]:

$$CWC = NMPIW(1 + e^{-\eta(PICP - \mu)}) \tag{6}$$

where η and μ are user-specified constants. The role of η was to magnify any small difference between μ and PICP. The value μ gives the nominal confidence level. Then η and μ determine the penalty paid by the PIs with low coverage probability.

In our framework for NN ensemble construction, we use the CWC scoring function proposed by Khosravi et al. [22] to rank the individual NNs on the validation set (see Section 3). We stress that we do not use CWC (Eq. 6) during the training of the individual networks. We instead use the multi-objective formulation of the PI estimation problem in terms of Eq. (5). In other words, in this work, we use a multi-objective framework rather than a single objective one. CWC is a single-objective function used to combine both objectives (PICP and NMPIW) in one single quality measure for optimization [21, 22]. As such, it yields only one single solution not a Pareto front of optimal solutions as we obtain in this work. CWC is just a scoring function a posteriori used in our case study; any other scoring function could also be used to rank the individual NNs on the validation set.

PI ESTIMATION VIA AN ENSEMBLE OF MOGA-BASED NNS

Herein, a method to construct PIs using an ensemble of NNs is proposed. The method uses the same training data set and the same number of hidden neurons for each individual NN, but initial parameters of NNs are chosen randomly, thus each NN has different set of initial weights. The random initializations are applied to increase the diversity of NN models. After we train each NN an overall Pareto front, which can also be called as combined Pareto front, is obtained applying non-dominated sorting to the Pareto fronts yielded from each trained network. The implementation procedure for the proposed method is summarized as follows:

Step 1: Divide the input data set into training (D_{tr}) and validation (D_{vald}).

Step 2: Set the number of hidden neurons and other initial parameters.

Step 3: Train m MOGA-based NNs using D_{tr}, where each network uses the full training set and differs only in its random initial weight settings. Thus, obtain m optimal Pareto fronts.

Step 4: Obtain an overall best Pareto front from the *m* optimal Pareto fronts (see Fig. 2).

Step 5: Perform validation using D_{vald} with the solutions on the overall best Pareto front.

Step 6: Obtain the non-dominated solutions on the validation front obtained in Step 5.

Step 7: Calculate the CWC_{vald} value of the non-dominated solutions and sort (rank) them with respect to their CWC_{vald} values.

Step 8: Select the n_{best} NNs giving the smallest CWC_{vald} and discard the others.

Step 9: For each testing sample *i* in the testing dataset D_{test}, determine the *k*-nn in the training dataset of each selected n_{best} NNs.

Combination of the outputs

Step 10: Combine the lower and upper bounds of the selected *k*-nn by mean and median calculations, respectively.

Step 11: Perform testing with the selected n_{best} NNs on the testing set. Then, compare the estimated PI results of the selected n_{best} NNs with the combined PI results.

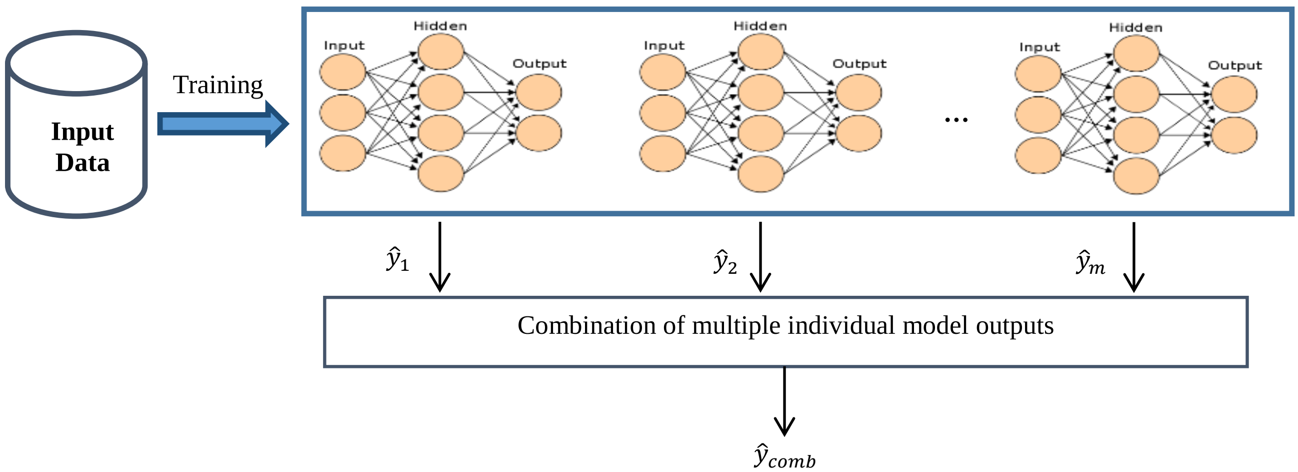

Figure 1. A BASIC SCHEME OF NN ENSEMBLE.

CASE STUDY

The case study focuses on predicting the energy consumed by a milling machine tool. The data set used in this case study was generated by Park et al. [10]. They describes an intelligent machine monitoring framework that consists of two basic models: (1) a data-management and extraction agent, and (2) a data-driven machine-learning and knowledge-extraction agent. The former agent consists of the MTConnect agent [3, 10] and a data post-processor. The MTConnect agent retrieves raw sensor data from a machine tool and systematically converts and organizes the data into semantically meaningful input features and response output.

The raw data includes the timestamp, power demand, feed rate, spindle speed, and numerical control (NC) code block (an NC code block corresponds to a specific cutting operation). The raw data is processed to obtain derived data, such as the average feed rate, average spindle speed, and cumulative energy consumption for each NC code block. The data set also includes simulated data, such as the depth of cut and cutting strategy (i.e., climb versus conventional milling). This data is determined by comparing two sequential tool positions reported by the MTConnect agent relative to the actual dimensions and location of the workpiece.

The data set was generated from a total of 18 parts machined with 196 face milling, 108 contouring, 54 slotting and pocketing, 16 spiraling and 32 drilling experiments. Of these six operations, the case study focuses on face milling to demonstrate the effectiveness of the proposed energy prediction model. We refer the reader to Bhinge et al. [1], Helu et al. [2], and Park et al. [10] for detailed information about the experimental design and data processing techniques used to generate the data set.

Predicting the energy consumed during machining requires that we first understand the power demanded by the machine tool. Diaz et al. [24] defined three components of the power demand of a machine tool: cutting (i.e., process), variable, and constant power components. The cutting power demand is the power needed to remove material and generate a chip. The variable power component is due to machine tool components that have constant power demands but that may not always be active (e.g., drive motors). The constant power component represents the demand of peripheral equipment that are always active when the machine is powered on (e.g., computer panel) [25]. This case study focuses on the entire power demand of the machine tool during face milling operations.

We predict the interval $[y_L, y_U]$ for energy consumption y (output) corresponding to the machine operational parameters $\boldsymbol{x}$ (i.e., the input feature vector) by constraining the prediction function $y = f(\boldsymbol{x})$ via each individual NN. Thus the input-output data set is defined as $\{(\boldsymbol{x}^i, y^i) | i = 1, 2, \cdots, D\}$ where D denotes the number of points (in this case, the number of NC code blocks) in the training dataset, and $\boldsymbol{x}^i$ denotes the input feature vector for the i^{th} machine operation. The optimum input feature vector is determined using the holdout cross-validation technique [10, 23]. This feature vector $\boldsymbol{x} = \{x_1, x_2, \dots, x_5\}$ for face milling is defined as follows:

$x_1 \in \mathbb{R}$ Feed rate: the velocity at which the tool is fed

$x_2 \in \mathbb{R}$ Spindle speed: rotational speed of the tool

$x_3 \in \mathbb{R}$ Depth of cut: the actual depth of material that the tool is removing

$x_4 \in \{1,2,3,4\}$ Active tool cutting direction: 1 is for x-axis, 2 for y-axis, 3 for z-axis, and 4 for x-y axes

$x_5 \in \{1,2,3\}$ Cutting strategy: 1 is for conventional, 2 for climbing, and 3 for both.

Each component of the optimum input feature vector represents a characteristic that helps define the energy consumed by the machine tool. For detailed information about the selection of the input features we refer the reader to Park et al. [10].

For the testing parts, three parts with identical geometry are machined but using different sets of spindle speeds (see Table 1). The machining operations used to produce the test part involve the cutting operations (face milling, pocketing, drilling), the non-cutting operations (air-cut in x-y direction air-cut in z direction), and auxiliary operations (such as rapid tool motion) [10].

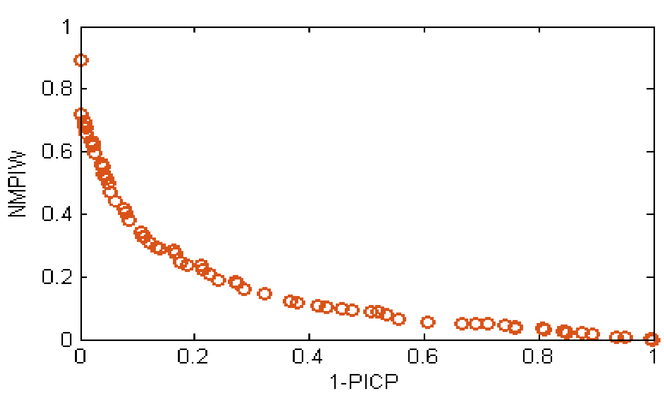

Figure 2. THE OVERALL BEST PARETO FRONT OBTAINED FROM 5 OPTIMAL PARETO FRONTS AFTER TRAINING THE NETWORK.

For the case study, of the total of 1466 pairs of input machine parameters and the corresponding energy output (J) for the face milling operation data, 1173 have been used for training each individual NN and the rest for validation. We have used 99 single NC code blocks in the test dataset. For all datasets, the inputs were normalized to be in the value range [0, 1].

Each individual NN in the ensemble is trained independently to minimize the prediction error with respect to the target. We have used the same architecture (i.e. the number of hidden neurons) for each individual NN. Thus, the number of hidden neurons has been set to five after a trial-and-error process. The μ and η values (see Section 2) were set to 90% and 50, respectively, in our experiment. Table 2 contains the parameters of the NSGA-II for training the NN. "MaxGen" indicates the maximum number of generations used as a termination condition and N_c indicates the total number of individuals per population. P_{c_int} indicates the initial crossover probability and is fixed during each iteration of the training process. P_{m_int} is the initial mutation probability and it decreases at each iteration (generation) by the formula:

$$P_{m\_int} \times e^{\left(-\frac{gen}{MaxGen}\right)} \tag{7}$$

Table 1. SPINDLE SPEED CHOSEN FOR THE BLIND TESTS.

	Used spindle speeds (in RPM)
Training parts 1-18	{1500; 3000; 4500}
Test part 1	{1500; 3000; 4500}
Test part 2	{1700; 2800; 4300}
Test part 3	{2125; 2400; 3750}

Table 2. PARAMETERS USED IN THE EXPERIMENTS.

Parameter	**Numerical value**
MaxGen	300
N_c	50
P_{m_int}	0.06
P_c	0.8
D_{tr}	1173
D_{vald}	293
D_{test}	99
n_{best}	10
m	5
η	50
μ	0.9

The final results for testing part 3 obtained by proposed ensemble of NNs integrated with a k-nn approach are shown in Table 3. Those are the PICP and NMPIW results of the ensemble of NNs calculated via median and mean aggregation. In this case study, the k value has been set to 3 and 5, respectively. In the Table 3, the "Best of n_{best}" column demonstrates the PICP and NMPIW results corresponding to one whose CWC score is the smallest among the selected 10 best individual NNs to construct ensembles. In other words, this is the best prediction result yielded by one of the selected 10 best NNs. From the results, it is seen that the proposed ensemble method gives higher PICP with smaller NMPIW value than those we obtained from the individual NN training. Figure 3 shows the combined PIs results for the testing set of "test part 3" estimated by the proposed ensemble method via median aggregation with k value of 5. Note that test parts 2 and 3 use different spindle speeds from those of the training sets. Thus, one can obtain higher prediction results for "test part 1" since it uses the same spindle speed as the training sets.

Table 3. COMBINED PI RESULTS ON THE TESTING SET ACCORDING TO TWO AGGREGATION (MEAN AND MEDIAN) CALCULATIONS.

	***k* = 3**		***k* = 5**		
	Mean	Med	Mean	Med	Best of n_{best}
PICP (%)	81.82	92.93	74.75	93.94	90.91
NMPIW (%)	43.87	45.26	44.05	42.96	90.40

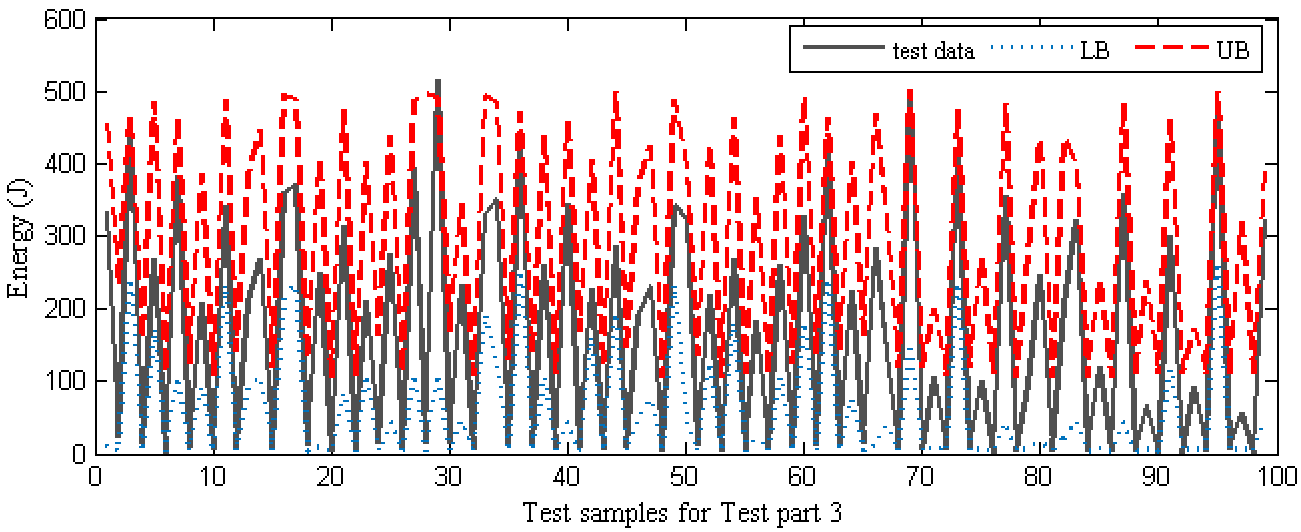

Figure 3. ESTIMATED PIS FOR TEST PART 3 ENERGY PREDICTION (DASHED LINE FOR UPPER BOUND AND DOTTED LINE FOR LOWER BOUND), AND MEASURED ENERGY CONSUMPTION DATA IN THE TESTING SET (SOLID LINE).

CONCLUSION

In this paper, the use of an NN ensemble model to predict the energy consumption of a machine tool is demonstrated. We perform prediction with NNs accounting for the uncertainty associated both to the input data and the regression model itself. Thus, we have estimated the prediction results in the form of intervals. The results obtained by the proposed ensemble method are promising in terms of the quality of the predicted PIs. We can conclude that the proposed ensemble modeling framework yields a reliable estimation of the PIs characterized by a high coverage probability and a small interval size.

While we have demonstrated the preliminary case study results only for face milling operations, we are currently working on estimating the energy consumption of other cutting and non-cutting operations. Then, the estimated energy consumption pertinent to each machine operation can be aggregated to calculate the required total energy consumption for machining a part. This type of data-driven prediction method can support better design, planning, and operation decisions. It can help improve machine tool efficiency and reduce cost through optimal process parameter selection, enable advanced process monitoring, or allow manufacturers to respond to new regulations and business drivers.

The prediction model can be improved as further experimental datasets are collected. In doing so, different topology (i.e., number of hidden neurons) might be used for each individual NN. Moreover, using disjoint subsets of the training set for each individual NN might be also considered.

ACKNOWLEDGMENTS AND DISCLAIMER

The authors would like to acknowledge the support of Jinkyoo Park from Stanford University, USA and Raunak Bhinge from University of California Berkeley, USA in collecting, preparing and providing machine operation data. Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by NIST; nor does it imply that the products identified are necessarily the best available for the purpose.

REFERENCES

[1] Bhinge, R., Biswas, N., Dornfeld, D., Park, J., Law, K. H., Helu, M., and Rachuri, S., 2014, "An Intelligent Machine Monitoring System for Energy Prediction Using a Gaussian Process Regression," *Proc. IEEE International Conference on Big Data*, Washington DC, USA, pp. 978–986.

[2] Helu, M., Robinson, S., Bhinge, R., Timo, B., and Dornfeld, D., 2014, "Development of a Machine Tool Platform to Support Data Mining and Statistical Modeling for Machining Processes," *Proc. MTTRF 2014 Annual Meeting*, San Francisco, USA.

[3] Vijayaraghavan, A., Sobel, W., Fox, A., Dornfeld, D., and Warndorf, P., 2008, "Improving Machine Tool Interoperability Using Standardized Interface Protocols: MT Connect," 2008, *Proc. ISFA*, Atlanta, GA, USA.

[4] Lavelle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N., 2011. Big Data, Analytics and the Path from Insights to Value, Technical Report, MIT Sloan Management Review, **52** (2).

[5] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H., 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity, Technical Report, McKinsey Global Institute, June. See also URL http://bigdatawg.nist.gov/MGI_big_data_full_report.pdf.

[6] Przytula, K. W., and Choi, A., 2008, "An Implementation of Prognosis with Dynamic Bayesian Networks," *Proc. IEEE Aerospace Conference*, Big Sky, MT, USA, pp. 1–8.

[7] Atsalakis, G., 2014. "New Technology Product Demand Forecasting Using a Fuzzy Inference System". *Operational Research*, **14**(2), pp. 225–236.

[8] Benardos, P. G., and Vosniakos, G.-C., 2003. "Predicting Surface Roughness in Machining: a Review". *International Journal of Machine Tools and Manufacture*, **43**(8), pp. 833–844.

[9] Nannapaneni, S., and Mahadevan, S., 2014, "Uncertainty Quantification in Performance Evaluation of Manufacturing Processes," *Proc. IEEE International Conference on Big Data Conference*, Washington DC, USA, pp. 996–1005.

[10] Park, J., Law, K. H., Bhinge, R., Srinivasan, A., Dornfeld, D., Helu, M., and Rachuri, S., 2015, "A Generalized Data-Driven Energy Prediction Model with Uncertainty for a Milling Machine Tool Using Gaussian Process," *Proc. ASME 2015 International Manufacturing Science and Engineering Conference*, Charlotte, North Carolina, USA.

[11] Mukherjee, I., and Ray, P. K., 2006. "A Review of Optimization Techniques in Metal Cutting Processes". *Computers & Industrial Engineering*, **50**(1–2), pp. 15–34.

[12] Gungor, V. C., Sahin, Kocak, D. T., Ergut, S., Buccella C., Cecati, C., and Hancke, G. P., 2011. "Smart Grid Technologies: Communication Technologies and Standards". *IEEE Transactions on Industrial Informatics*, **7**(4), pp. 529–539.

[13] Zhou, Z.-H., Wu, J., and Tang, W., 2002. "Ensembling Neural Networks: Many Could Be Better Than All". *Artificial Intelligence*, **137**(1–2), pp. 239–263.

[14] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A. M. T., 2002. "A fast and elitist multiobjective genetic algorithm: NSGA-II". *IEEE Transactions on Evolutionary Computation*, **6**(2), pp. 182-197.

[15] Ak, R., Li, Y, Vitelli, V., Zio, E., Droguett E. L., and Jacinto, C. M. C., 2013, "NSGA-II-Trained Neural Network Approach to the Estimation of Prediction Intervals of Scale Deposition Rate in Oil & Gas Equipment". *Expert Systems with Applications*, **40**(4), pp. 1205–1212.

16] Ak, R., Vitelli, V., and Zio, E., 2015, "An Interval-Valued Neural Network Approach for Prediction Uncertainty Quantification in Short-term Wind Speed Prediction". *IEEE Transactions on Neural Networks and Learning Systems*, DOI. 10.1109/TNNLS.2015.2396933.

[17] Hornik, K., Stinchcombe, M., and White, H., 1989, "Multilayer Feedforward Networks are Universal Approximators". *Neural Networks*, **2**(5), pp. 359–366.

[18] Zhang, G., Patuwo, B. E., and Hu, M. Y., 1998, "Forecasting with Artificial Neural Networks: The State of the Art". *International Journal of Forecasting*, **14**(1), pp. 35–62.

[19] Rojas, R., 1996, Neural Networks: A Systematic Introduction, Springer-Verlag, Berlin.

[20] Bishop, C. M., 1996, Neural Networks for Pattern Recognition, Oxford University Press Inc., Newyork, USA.

[21] Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F., 2011. "Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances". *IEEE Transactions on Neural Networks*, **22**(9), pp. 1341–1356.

[22] Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F., 2011. "Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals". *IEEE Transactions on Neural Networks*, **22**(3), pp. 337–346.

[23] Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning*, 2nd ed., Springer New York, Chap. 2, pp. 9-39.

[24] Diaz, N., Redelsheimer, E., and Dornfeld, D., 2011, "Energy Consumption Characterization and Reduction Strategies for Milling Machine Tool Use," Glocalized Solutions for Sustainability in Manufacturing, Springer Berlin Heidelberg, pp. 263-267.

[25] Helu, M. M., 2013. "Leveraging Manufacturing Precision to Reduce Product Life Cycle Environmental Impacts," PhD Thesis, University of California Berkeley, See also URL http://gradworks.umi.com/36/19/3619691.html.