Industrial Engineering and Innovation Management (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/ieim.2026.090112
ISSN 2522-6924 Vol. 9 Num. 1

A Bayesian Deep Neural Network with MC Dropout for Multi-stage Quality Prediction

Manli Liu[1], Wenhan Fu[1,2,a,*], Junxiang Li[1,2], Qiuyang Wang[1]

[1]Business School, University of Shanghai for Science and Technology, Shanghai, 200093, China
[2]School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai, 200093, China
[a]whfu@usst.edu.cn
**Corresponding author*

Keywords: Multi-stage quality prediction; Deep neural network; MC Dropout; Bayesian inference; Uncertainty modeling

Abstract: To address the challenges of complex quality evolution mechanisms and the difficulty in characterizing prediction uncertainty in multi-stage manufacturing processes, this study proposes a deep neural network-based multi-stage quality prediction method using MC Dropout for approximate Bayesian inference. First, systematic data preprocessing is performed, with SMOTE used for data augmentation and stage-wise PCA for dimensionality reduction. Subsequently, the Dropout mechanism is incorporated into the deep neural network, and MC Dropout is employed to jointly model the predictive mean and uncertainty. A five-fold cross-validation strategy is adopted for model training and evaluation, and the effectiveness of the proposed method is validated using real-world TFT-LCD manufacturing data. Experimental results demonstrate that the proposed method achieves superior prediction accuracy while providing uncertainty estimates for multi-stage quality prediction.

1. Introduction

With the continuous growth of manufacturing process data, quality prediction in multi-stage manufacturing systems has attracted increasing attention. However, the nonlinearity, strong coupling, and uncertainty of multi-stage processes make accurate and reliable prediction challenging for traditional empirical and statistical methods.

Existing studies have applied machine learning and deep learning to single-stage quality prediction. Zhao et al. [1] explored deep temporal feature modeling, while Ma et al.[2]employed optimized ensemble learning for industrial quality prediction. However, these approaches mainly consider individual stages and neglect inter-stage interactions and error propagation.

Product quality is jointly affected by multiple processing stages, motivating research on multi-stage quality prediction. Wanda et al.[3]developed a multi-stage quality prediction framework to capture spatial and temporal correlations among manufacturing stages. Other studies investigated machine learning-based quality modeling and error propagation, including additive manufacturing quality prediction [4] and small-sample multi-stage prediction using error transfer networks [5]. Deep

learning approaches have also improved feature representation and interpretability through attention and explainable learning strategies [6-7].

Optimization-based methods have also been applied to multi-stage quality prediction to improve modeling performance [8]. Probabilistic machine learning has further extended manufacturing quality modeling toward uncertainty-aware prediction [9]. However, most existing methods focus on deterministic accuracy and provide limited quantification of uncertainty arising from process fluctuations, measurement noise, and model limitations, which may compromise prediction reliability in complex industrial environments.

Bayesian methods provide a probabilistic framework for uncertainty modeling in engineering applications [10], while Bayesian deep learning combines nonlinear feature learning with uncertainty quantification [11-12].However, the complexity of such models can hinder practical deployment.

To address these limitations, this study develops an MC Dropout-based Bayesian deep neural network for multi-stage quality prediction. Stage-wise feature construction preserves process information, while MC Dropout provides uncertainty estimates through approximate Bayesian inference. The method is validated on real-world manufacturing data.

2. Problem Formulation of Multi-stage Quality Prediction

2.1 Data Structure and Notation Definition

Assume that a manufacturing process contains M stages. For the i-th stage, the process feature vector is defined as:

$$\mathbf{X}_{\mathrm{i}} = [\mathrm{x}_{\mathrm{i1}}, \mathrm{x}_{\mathrm{i2}}, \dots, \mathrm{x}_{\mathrm{id_i}}]$$

where d_i denotes the number of process variables in stage i. By integrating the feature vectors from all stages, the overall input representation is obtained as:

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{\mathrm{M}}]$$

For each sample n, the multi-stage quality prediction problem can be formulated as:

$$(\mathbf{X}^{(\mathrm{n})}, \mathrm{y}^{(\mathrm{n})})$$

where $\mathbf{X}^{(\mathrm{n})}$ denotes the integrated process features and $\mathrm{y}^{(\mathrm{n})}$ denotes the corresponding quality indicator. Stage-wise feature processing and dimensionality reduction are subsequently applied to preserve stage-specific information.

2.2 Prediction Objective and Mathematical Formulation

The prediction model aims to learn the nonlinear mapping from multi-stage process features to the quality indicator. For the n-th sample, the prediction objective is expressed as:

$$\left\{\left(\mathbf{X}^{(\mathrm{n})}, \mathrm{y}^{(\mathrm{n})}\right)\right\}_{\mathrm{n=1}}^{\mathrm{N}}$$

where $\mathbf{X}^{(\mathrm{n})}$ denotes the multi-stage process features and $\mathrm{y}^{(\mathrm{n})}$ represents the corresponding quality value. The prediction model aims to minimize the deviation between predicted and actual quality values, which can be formulated as:

$$\min \frac{1}{\mathrm{N}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathscr{L}\left(\mathrm{y}^{(\mathrm{n})}, \hat{\mathrm{y}}^{(\mathrm{n})}\right)$$

Where $\hat{y}^{(n)}$ denotes the predicted quality value. Considering process variations and

measurement noise, uncertainty estimation is incorporated into the prediction framework.

3. Bayesian Deep Neural Network Quality Prediction Model Based on MC Dropout

Figure 1 illustrates the proposed MC Dropout-based Bayesian deep neural network framework, which consists of four modules: data preprocessing, feature construction, model training, and uncertainty inference.

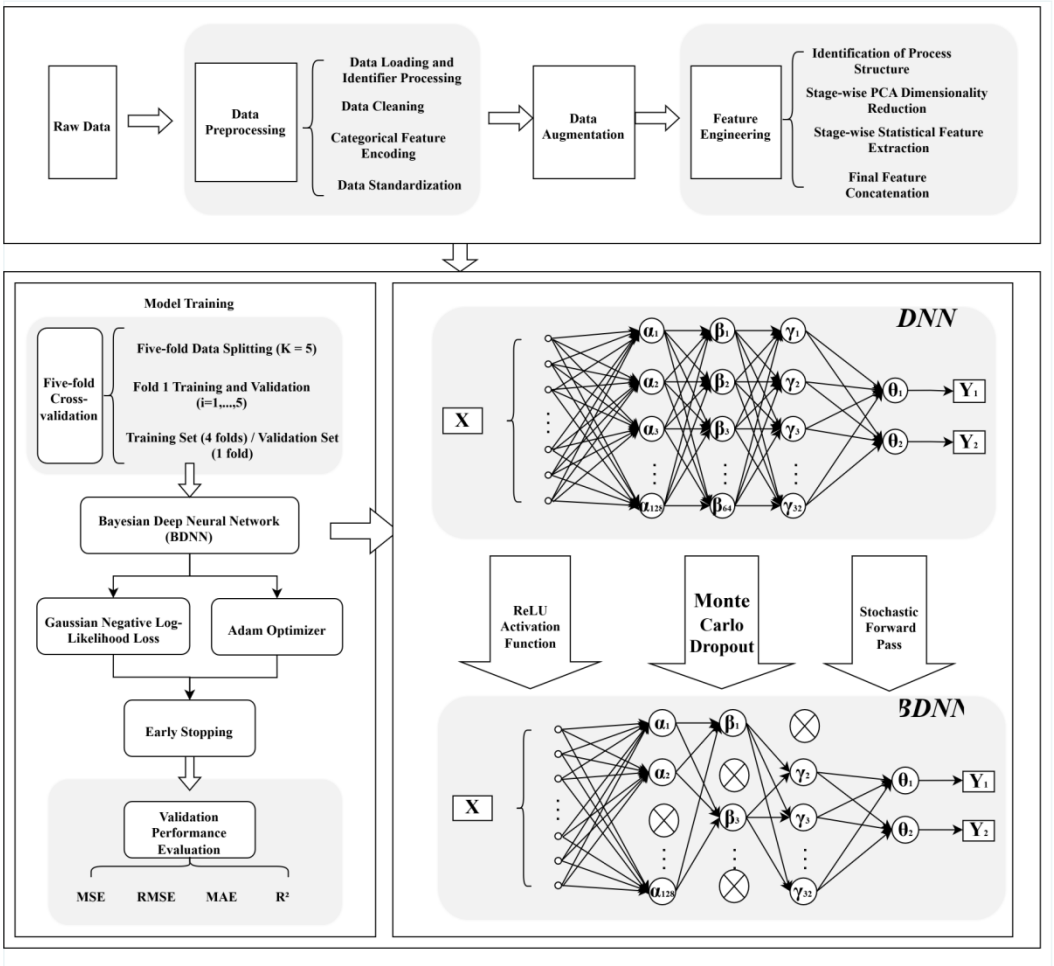

Figure 1: Framework of the Deep Bayesian Model

3.1 Data Preprocessing Procedure

3.2 Data Augmentation and Stage-wise Feature Engineering

Due to the limited number of manufacturing samples, SMOTE is employed to augment the training data. Given an original sample $\mathbf{x}_i$ and one of its nearest neighbors $\mathbf{x}_j$, a synthetic sample is generated as:

$$\mathbf{x}_{new} = \mathbf{x}_i + \delta(\mathbf{x}_j - \mathbf{x}_i) \tag{1}$$

where $\delta \in (0,1)$ is a random interpolation coefficient.

To reduce the high dimensionality and redundancy of multi-stage process features,Principal Component Analysis (PCA) is independently applied to each manufacturing stage.

For the feature subset $X_{p_j} \in R^{n \times d_j}$of each manufacturing stage, PCA is performed to obtain the

reduced feature representation $X_{p_j}^{PCA} \in R^{n \times k_j}$:

$$X_{p_j}^{PCA} = X_{p_j} W_{p_j}, \quad k_j \leq d_j \tag{2}$$

where $W_{p_j} \in R^{d_j \times k_j}$ is the PCA projection matrix and k_j is determined by the cumulative variance retention ratio:

$$\text{textVarianceRetained} = \text{frac} \sum_{i=1}^{k_j} \lambda_i \sum_{i=1}^{d_j} \lambda_i \geq \rho \tag{3}$$

where λ_i denotes the i-th eigenvalue, and ρ represents the predefined variance retention threshold. Stage-wise PCA reduces feature redundancy while preserving stage-specific information, providing a compact representation for modeling inter-stage feature coupling and error propagation.

3.3 Deep Neural Network Architecture Design

A Deep Neural Network (DNN) is employed to learn the nonlinear mapping between multi-stage process features and quality indicators. The network can be formulated as:

$$\mathbf{h}^{(l)} = f\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right) \tag{4}$$

where $\mathbf{h}^{(l)}$ denotes the output of the l-th layer,and $\mathbf{W}^{(l)}$,$\mathbf{b}^{(l)}$, and $f(\cdot)$ represent the weight matrix, bias vector, and activation function, respectively. The network consists of three hidden layers with 128, 64, and 32 neurons, respectively, using ReLU activation. The output layer jointly predicts the quality mean and predictive variance. The architecture is shown in Figure 2.

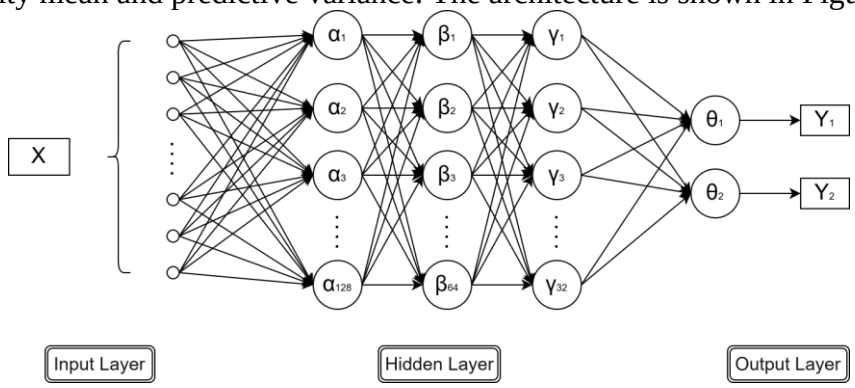

Figure 2: Architecture of the Deep Neural Network

3.4 MC Dropout-Based Approximate Bayesian Inference Method

Since point predictions cannot quantify prediction reliability, MC Dropout is incorporated into the DNN to achieve approximate Bayesian inference. During inference, the Dropout mechanism remains active, and multiple stochastic forward passes are performed to obtain predictive distributions and uncertainty estimates.

For the l-th layer, the Dropout operation is defined as:

$$\tilde{\mathbf{h}}^{(l)} = \mathbf{h}^{(l)} \odot \mathbf{z}^{(l)}, \quad z_i^{(l)} \sim \text{Bernoulli}(p) \tag{5}$$

where p denotes the probability of retaining a neuron.

Unlike conventional Dropout, MC Dropout retains the Dropout mechanism during inference and performs multiple stochastic forward passes. As demonstrated by Gal[13], this process can be

interpreted as a variational Bayesian approximation, enabling uncertainty estimation without modifying the network architecture.

The predictive distribution is approximated through repeated stochastic forward passes:

$$p(y|\mathbf{x}) \approx \frac{1}{T}\sum_{t=1}^{T} p\,(y|\mathbf{x}, \mathbf{W}_t) \tag{6}$$

where $\mathbf{W}_t$ denotes the network parameters obtained from the t-th stochastic Dropout sampling.

Unlike conventional regression networks, the proposed DNN adopts a dual-branch output layer to jointly predict the quality mean and the conditional observation variance. Let the final hidden representation be h, and the output layer is formulated as:

$$\mu = W_{\mu} h^{(L)} + b_{\mu} \tag{7}$$

$$\log \sigma^2 = W_{\sigma} h^{(L)} + b_{\sigma} \tag{8}$$

The dual-output layer jointly estimates the predictive mean μ and conditional variance σ^2. The variance is parameterized in logarithmic form during training to ensure positivity. MC Dropout sampling further captures model uncertainty through stochastic forward passes.

This formulation provides uncertainty-aware prediction through approximate Bayesian inference without explicitly placing probability distributions over all network weights.

3.5 Model Training Strategy

A five-fold cross-validation strategy is adopted for robust model evaluation. The Gaussian negative log-likelihood (NLL) is used to jointly optimize the predictive mean and variance:

$$\mathscr{L} = \frac{1}{2}\left(\log \sigma^2 + \frac{(y-\mu)^2}{\sigma^2}\right) \tag{9}$$

where μ and σ^2 enote the predicted mean and variance. The network is optimized using Adam with L2 regularization, while Dropout, early stopping, and adaptive learning-rate decay are employed to improve generalization and training stability.

4. Case Study

4.1 Dataset and Parameter Settings

The proposed method is validated using a real-world TFT-LCD manufacturing dataset containing multi-stage process parameters and quality indicators. As shown in Figure 3, process variations propagate from upstream to downstream stages, resulting in strong inter-stage coupling.

The dataset contains 5,954 fields comprising sample identifiers, the target quality variable, multi-stage process features, and equipment-related variables.

After preprocessing, 1,100 valid samples with 5,052 input features were retained, including 4,996 numerical and 54 equipment-related One-Hot features. SMOTE increased the sample size to 3,000 by generating 1,900 synthetic samples. The target mean and standard deviation changed from 2.8138 and 0.2192 to 2.8197 and 0.2047, respectively, indicating no substantial distribution shift.

The 4,996 numerical features were grouped into 15 manufacturing stages and reduced to 345 dimensions using stage-wise PCA, with approximately 94% – 96% variance retained. Six statistical features were then extracted from each stage, producing 90 features. Combined with 54

equipment-related One-Hot features, the final representation contained 489 features.

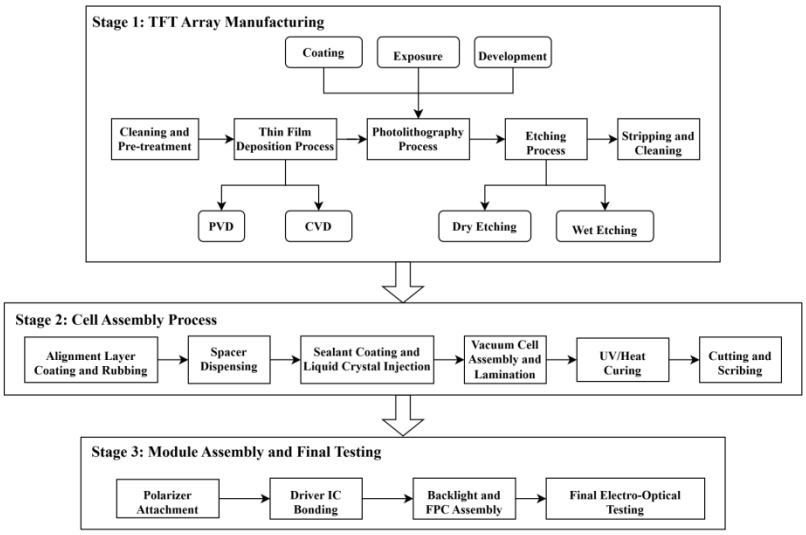

Figure 3: Partial Process Flowchart of TFT-LCD Manufacturing

The BDNN used the architecture described in Section 3, with a Dropout rate of 0.3 retained during inference. The remaining hyperparameters are listed in Table 1.

Table 1: Parameter Settings

Parameter	Value
Number of SMOTE nearest neighbors (K)	5
Number of cross-validation folds	5
Dropout rate	0.3
Hidden layer architecture	128–64–32
Learning rate	5×10^{-4}
Batch size	32
Number of MC Dropout sampling iterations	50

4.2 Model Validation Results

The BDNN was evaluated using five-fold cross-validation, achieving average MSE, RMSE, MAE, and R^2 values of 0.004928, 0.069896, 0.045492, and 0.882135, respectively. The small variation across folds indicates stable performance.

MC Dropout inference with 50 stochastic samples was used to estimate predictive uncertainty.

Figure 4 compares the predicted and actual values. Most samples are distributed close to the diagonal line, indicating high prediction accuracy. The uncertainty values increase with prediction deviations, suggesting that the estimated uncertainty is consistent with prediction reliability. A slight underestimation is observed in a small number of samples.

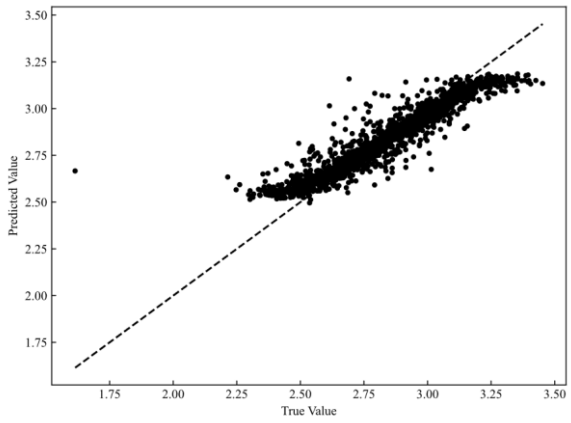

Figure 4: Scatter Plot of Predicted and Actual Values

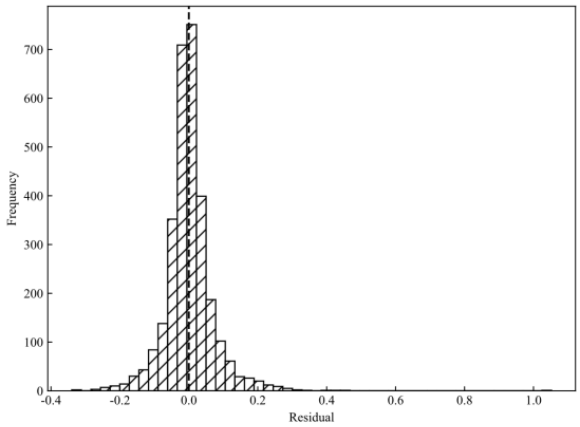

Figure 5: Residual Distribution of the Bayesian Deep Neural Network

Figure 5 presents the residual distribution of the proposed BDNN. The residuals are approximately symmetric, with a mean of 0.0013 and a standard deviation of 0.0702. Most residuals are concentrated around zero without evident long-tail characteristics, indicating no significant systematic bias.

4.3 Comparative Analysis with Different Models

Ridge Regression and Random Forest are selected as baseline models. All models are evaluated under the same preprocessing, feature engineering, and cross-validation settings.

Table 2: Performance Comparison of Different Models

Model	MSE	RMSE	MAE	R^2
BDNN	0.004928	0.069896	0.045492	0.882135
Ridge	0.012603	0.105769	0.066972	0.698573
Random Forest	0.008660	0.092747	0.066399	0.793055

As shown in Table 2, the proposed BDNN achieves the best performance across all four prediction metrics while additionally providing uncertainty estimates.

5. Conclusions

This study proposed an MC Dropout-based Bayesian deep neural network for multi-stage quality prediction, enabling joint quality prediction and uncertainty estimation. Experiments on real-world manufacturing data demonstrated superior prediction performance compared with the baseline models.

Experimental results demonstrate lower prediction errors and a higher R^2 than the baseline models, while MC Dropout provides complementary uncertainty information.In addition, the proposed method can provide uncertainty estimation for prediction results, thereby enhancing its practical value in industrial quality prediction and risk assessment applications.

Future research will focus on improving computational efficiency, further integrating uncertainty information into decision-making processes, and developing quality control optimization strategies. Moreover, the proposed framework will be extended to more complex multi-stage manufacturing environments to enhance its applicability and scalability in practical industrial applications.

References

[1] Zhao Xiaoqiang, Liu Yongyong, Hui Yongyong, et al. Quality prediction model for batch processes based on improved temporal convolutional network and multi-head self-attention mechanism [J]. Computer Applications, 2025,45(07):2245-2252.

[2] Ma Chao, Weng Zhiyi, He Fei. Quality prediction modeling of piezoelectric ceramic sintering process based on ensemble learning [J]. Computer Integrated Manufacturing Systems, 2025,31(01):147-157.

[3] Wanda Z, Yanchao Y, Jun T, et al. A method for the spatiotemporal correlation prediction of the quality of multiple operational processes based on S-GGRU [J]. Advanced Engineering Informatics, 2023,58.

[4] Xinyi X, Clarke W, Carter H, et al. Quality prediction and control in wire arc additive manufacturing via novel machine learning framework [J]. Micromachines, 2022,13(1):137.

[5] Qu D, Liang W, Zhang Y, et al. Research on machining quality prediction method based on machining error transfer network and grey neural network [J]. Journal of Manufacturing and Materials Processing, 2024,8(5):203.

[6] Guo Ziyu, Wang Ning, Wang Jianjun, et al. Research on interpretable product quality prediction based on probabilistic distribution attention mechanism [J/OL]. Chinese Journal of Management Science, 2026.

[7] Ding Long, Fu Wenhan, He Yuchun, et al. A wafer manufacturing cycle time prediction method based on interpretable deep learning [J/OL]. Industrial Engineering Journal, 2026.

[8] Yang Jianxin, Lan Xiaoping, Yao Zhiqiang, et al. Multi-stage quality prediction method based on LSSVM optimized by coyote optimization algorithm [J]. Manufacturing Automation, 2021,43(12):105-109.

[9] Mowbray M, Kay H, Kay S, et al. Probabilistic machine learning based soft-sensors for product quality prediction in batch processes [J]. Chemometrics and Intelligent Laboratory Systems, 2022,228:162-168.

[10] Wang Chengtang, Wang Hao, Qin Weimin. Collapse probability evaluation of deep foundation pits in subway stations based on polymorphic fuzzy Bayesian networks [J]. Rock and Soil Mechanics, 2020,41(05):1670-1679+1689.

[11] Li Baoluo, Xu Shiyun, Li Zonghan, et al. Interval prediction method of critical short-circuit ratio for renewable energy multi-station systems based on Bayesian deep learning [J]. Proceedings of the CSEE, 2024,44(14):5451-5463.

[12] Zhu Saihua, He Zitong, Feng Ye, et al. Prediction method of cigarette sensory evaluation indicators based on Bayesian regularized neural network with feature fusion [J]. Computer Integrated Manufacturing Systems, 2026,32(02):628-640.

[13] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning [C]//International Conference on Machine Learning. PMLR, 2016:1050-1059.