ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Probabilistic Bayesian Deep Learning Approach for Online Forecasting of Fed-Batch Fermentation

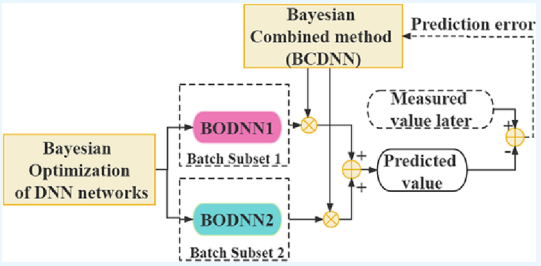
Tao Wang,* Jiebing You, Xiugang Gong, Shanliang Yang, Lei Wang, and Zheng Chang

Cite This: *ACS Omega* 2023, 8, 25272−25278

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The microbial fermentation process often involves various biological metabolic reactions and chemical processes. The mixed bacterial culture process of 2-keto-L-gulonic acid has strong nonlinear and time-varying characteristics. In this study, a probabilistic Bayesian deep learning approach is proposed to obtain a highly accurate and robust prediction of product formation. The Bayesian optimized deep neural network (BODNN) is utilized as basic model for prediction, the structural parameters of which are optimized. Then, the training datasets are classified into different categories according to the prior evaluation of prediction error. The final forecasting is a weighted combination of BODNN models based on the Bayesian hybrid method. The weights can be interpreted as Bayesian posterior probabilities and are computed recursively. The validation of 95 industrial batches is carried out, and the average root mean square errors are 1.51 and 2.01% for 4 and 8 h ahead prediction, respectively. The results illustrate that the proposed approach can capture the dynamics of fermentation batches and is suitable for online process monitoring.



INTRODUCTION

Fed-batch fermentation is a typical production mode in chemical process industries such as pharmaceutical and food processing, with flexible production operations that can adapt to changing market conditions. For instance, L-ascorbic acid, also known as vitamin C, is mainly produced through a two-step fermentation process.[1] In the first fermentation stage, L-sorbose is produced from D-sorbitol by batch culture of *Acetobacter melanogenum* with a high molar yield. In the second stage, the mixed culture of *Ketogulonicigenium vulgare* (*K. vulgare*) and *Bacillus megaterium* (*B. megaterium*) is practiced to produce 2-keto-L-gulonic acid (2-KGA) with L-sorbose as substrate. 2-KGA is then converted to L-ascorbic acid by catalytic reactions.

Great attention has been paid to batch process optimization,[2,3] which is motivated by the increasing pressure to improve process efficiency with reduced operating costs. The optimization objective may take many forms, such as cutting down the time to market, maximizing profit, or minimizing production costs.[4] An online monitoring and optimization approach for a 2-KGA fermentation workshop is essential to increase the economic efficiency of the process. The accurate prediction of the product formation is the critical factor in process monitoring and optimization, which influences the scheduling strategy for improving the allocation of L-sorbose resources with the aim of profit maximization in a multi-bioreactor workshop.[5,6]

With the widespread use of computer control systems in the fermentation industry, stored online measurement and offline assay data are growing rapidly. The process data implies information on the operation status of the fermentation process. Therefore, using data-driven modeling techniques to predict and control key variables in the fermentation process can effectively improve plant productivity.[7−9] Deep learning has been an important research area in machine learning and artificial intelligence research in recent years. Deep learning methods are composed of multiple layers to learn nonlinear features of data with multiple levels of abstraction,[10] which have been widely used for supervised or unsupervised feature learning and representation, classification, and pattern recognition.

As the performance of deep learning models depends critically on a good set of hyperparameters, several approaches for hyperparameter optimization have been proposed.[11,12] A deep neural network-based hybrid model is developed by integrating the kinetic model with a DNN trained with time-series process data to predict sensitive and uncertain model parameters.[13] Bose et al. proposed a trustworthy hybrid model by cascading multivariate adaptive regression splines and deep neural network to predict closing prices of stock.[14] Among them, the most commonly used approach is Bayesian optimization.[15] To further enhance the performance of a single deep model, a series of hybrid models has been proposed.[16−20] Grover et al. studied

Received: April 9, 2023
Accepted: June 22, 2023
Published: July 4, 2023

specifically the power of making predictions via a hybrid approach that combines discriminatively trained predictive models with a deep neural network that models the joint statistics of a set of weather-related variables.[21] Guo et al. propose a novel locally supervised deep hybrid model (LS-DHM) that effectively enhances and explores the convolutional features for scene recognition.[22] A hybrid model is proposed based on RNNs and CNNs to extract biomedical relation.[23] A hybrid model-based emotion contextual recognition approach for cognitive assistance services in ubiquitous environments is proposed.[24] Likewise, Bayesian approaches play an important role in the hybrid models. Hao et al. present a hybrid optimization methodology with Bayesian inference for the probabilistic finite element model updating of structural systems.[25] Zhang et al. proposed a Bayesian hybrid collaborative filtering-based electricity plan recommender system (BHCF-EPRS), which is constructed in a two-stage model integrated with model-based and memory-based collaborative filtering methods.[26] Lessan et al. developed a hybrid Bayesian network model to tackle delays in train operations.[27]

In biological fermentation, the production process is affected by the uncertainty of the fermentation cycle, contaminated bacteria, and other factors. Moreover, process variables in the fermentation process (e.g., the concentrations of the product, substrate, and biomass) are challenging to measure directly. During the 2-KGA fermentation process, only *K. vulgare* converts L-sorbose to 2-KGA, while the concomitant bacteria *B. megaterium* stimulates the growth of *K. vulgare*.[28] The co-cultured relationships between cells in synthetic microbial consortia are dynamically balanced, leading to greater adaptability and stability to variable environments.[29] Xu et al. proposed an integrated approach for VFA production prediction in an anaerobic fermentation process,[30] which learns nonlinear process information and consequently enlarges the application range of DNN models. However, for the 2-KGA industrial fermentation, the data labels are noisy since some key process variables are manually assayed offline. Furthermore, the mixed culture process increases the time-variability and uncertainty of industrial production.[31] Therefore, it is necessary to construct a data-driven model of the 2-KGA fermentation process using both Bayesian-optimized deep learning models and hybrid modeling.

This study proposes an approach for obtaining highly accurate and robust prediction of 2-KGA product formation. The fermentation batch data are processed with a moving window technique and put into a Bayesian optimized deep neural network (BODNN). The optimal parameters of the BODNN model are searched for during the model's training process. The fermentation batches are then classified based on the model's prediction results. Batches with high prediction accuracy are classified into one category. Other batches with a high degree of nonlinearity that a single model could not accurately predict are classified into another category.

Based on a large amount of process data collected over a long period, different categories of fermentation batch data are used to train separate BODNN models. The output of BODNN models represents the characteristics of the corresponding batch category. Due to the highly time-varying and nonlinear nature of the mixed culture process, the status of the current batch under online monitoring needs to be re-evaluated at each sampling time point. Finally, a hybrid model is constructed by dynamically adjusting the combination weights of each BODNN model. This approach has effectively improved the results of 4 and 8 h ahead forecasting of product formation in the 2-KGA mixed culture process.

MATERIALS AND METHODS

Process Description. The industrial fermentation data in this study come from a Chinese pharmaceutical factory. The critical precursor of Vitamin C, 2-KGA, is produced by a mixed culture of *K. vulgare* and *B. megaterium* in air-lift bioreactors. The maintenance of a stable pH value of approximately 7.0 during 2-KGA cultivation is achieved through the automatic addition of alkali solution. The distributed control system (DCS) online continuously monitors and regulates process variables, including temperature, pH, aeration rate, and liquid level. Nevertheless, crucial parameters such as the concentration of product 2-KGA and substrate L-sorbose require manual, offline assaying. The fermentation process ceases when the residual L-sorbose concentration falls below 1 kg/m^3.

Preprocessing of Training Data Pair. To forecast the 2-KGA product formation P in advance, relevant process variables were screened. A correlation between the consumption of alkali solution (Alk) and P was identified through statistical analysis of historical data. This relationship is potentially attributed to the DCS system's ongoing addition of alkali during fermentation to regulate medium pH at approximately 7.0.

A potential correlation between substrate consumption and 2-KGA production has also been noted. However, in current industrial production, the concentration of critical substrate L-sorbose, is solely assayed manually during the latter stages of cultivation and the data obtained is used to determine the termination time of the current batch.

As a result, solely the time series of P and AlK serve as input variables for the deep learning model. To produce the input−output data pairs required for model training, the moving data window method is employed,[32] as depicted in Figure 1.

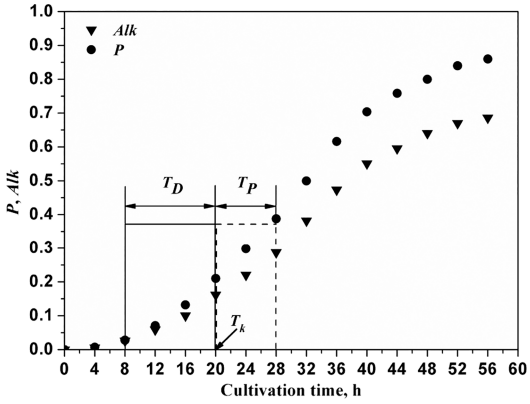


Figure 1. Moving data window method (the data have been normalized).

The initiation of product formation prediction commences at T_1 when the minimal required data quantity is obtained. The input window's width, T_D, and the practical prediction window width, T_P, determine the time frame for the forecasts. A T_P of 4 or 8 h signifies a prediction for a 4 or 8 h lead-time, respectively.

The data pair of the kth data window $\{X(T_k), Y(T_k)\}$ is presented with eqs 1−3. The input vector $X(T_k)$ consists of the normalized sampling time T_k and all the sampling process data located within the corresponding input window T_D. The

https://doi.org/10.1021/acsomega.3c02387
ACS Omega 2023, 8, 25272−25278

temporal resolution of the discrete-time system is represented by τ.

$$X(T_k) = [T_k\ x_{T_k}\ x_{T_{k-\tau}}\ x_{T_{k-2\tau}\cdots\cdots X_{T_{k-T_D}}}]^T \quad (1)$$

$$x_{T_k} = [P(T_k)\ Alk(T_k)]^T \quad (2)$$

$$Y(T_k) = P(T_k + T_p) \quad (3)$$

BODNN: Bayesian Optimized Deep Neural Network. Deep neural network (DNN), as one of the deep learning approaches, is mainly composed of neurons and connections. The architectures of DNN usually have many hidden layers, which can be used to extract features from the inputs and compute complex functions. With the emergence of large-scale datasets, DNN has exhibited impressive performance in numerous machine learning tasks. To obtain better model performance, the hyperparameters of the DNN model, such as the number of layers, neurons in each layer, and the learning rate, are considered for exploration and optimization. Therefore, Bayesian optimized deep neural network (BODNN) is proposed to predict the 2-KGA product formation.

In the typical DNN model, the weighted combination of neurons in one layer is calculated and then used as input to another neuron in a subsequent layer. Assuming that there are m neurons in layers l-1, the output $O_j^l$for the jth neuron in layer l is defined in eq 4, where w and b are the corresponding weight and bias.

$$O_j^l = \sigma\left(\sum_{k=1}^{m} w_{jk}^l O_k^{l-1} + b_j^l\right) \quad (4)$$

The nonlinear activation function $\boldsymbol{\sigma}$ is applied to the weighted sum of neurons to catch the nonlinearity of the data. The exponential linear unit (ELU) function is utilized here since it alleviates the vanishing gradient problem via the identity for positive values and improves learning characteristics compared to other activation functions of the deep neural network.[33] The ELU function is shown in eq 5, where parameter $\boldsymbol{\beta}$ is a constant that is generally set to 1.0.

$$\sigma(x) = \begin{cases} x & \text{if } x > 0 \\ \beta(\exp(x) - 1) & \text{if } x \le 0 \end{cases} \quad (5)$$

The gradient descent learning process is expressed in eq 6, where $\boldsymbol{\eta}$, $\boldsymbol{t}$, and $\boldsymbol{H}$ are the learning rate, number of epochs, and loss function, respectively.

$$W(t+1) = W(t) - \eta\frac{\delta H(t)}{\delta W(t)} \quad (6)$$

Considering the continuous optimization of the 2-KGA fermentation techniques and the dynamic changes in the production environment, the BODNN model needs to be retrained periodically. Therefore, the maximum number of layers is limited to 6, and neurons for each layer are 50 during the model structural optimization. Table 1 illustrates the hyperparameters with their corresponding range values under concern.

Bayesian optimization (BO) is one of the gradient-free optimization algorithms, which wants to obtain the optimal global solution of the optimization proposition with as few trials as possible.

Table 1. Hyperparameters Ranges for Exploration and Optimization

hyperparameter	number of layers	number of neurons	learning rate
range	$\mathbb{Z}[2, 6]$	$\mathbb{Z}[5, 50]$	$\mathbb{R}[10^{-1}, 10^{-6}]$

The input−output data pairs of all batches are split into training set and validation set. Each DNN model constructed from hyperparameters $\boldsymbol{a}$ is trained with the training set. The structural optimization problem is defined in eq 7, where A is the search space, and $f(\boldsymbol{a})$ denotes the objective metric aimed at minimizing the error rate as assessed on the validation set.

$$a^* = \arg\min_{a \in A} f(a) \quad (7)$$

The BO process has two key ingredients: surrogate model and acquisition function. The surrogate model is used to approximate f. Gaussian processes are chosen to be the surrogate model here,[34] for their expressiveness and well-calibrated. A Gaussian process $\mathcal{G}(m(\lambda), k(\lambda, \lambda'))$ is fully specified by a mean $m(\lambda)$ and a covariance function $k(\lambda, \lambda')$.

The mean function is assumed to be constant, and then the quality of the Gaussian process depends solely on the covariance function. The Matérn 5/2 function is used as covariance function here, as shown in eq 8.

$$k(\alpha_i, \alpha_j) = \sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)\exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right) \quad (8)$$

where σ_l is the characteristic length scale, σ_f is the signal standard deviation, and parameter r is the Euclidean distance between α_i and α_j.

In each iteration of the BO process, the surrogate model is fitted to all observations of $f(a)$ made so far. Then, an acquisition function determines the sequence of possible points and performs the exploration and exploitation. The expected improvement (EI) is used as the acquisition function, which is a common choice for its lower computational complexity and better performance.[35] After running an n number of trials, based on the values of the sampled points, the probability distribution of $f(a)$ is calculated, and the mathematical expectation of the improved values of next step can be obtained. The expected improvement function is defined in eq 9.

$$\mathrm{EI}_n(\alpha) = \mathrm{E}_n[[f(\alpha) - f_n^*]^+] \quad (9)$$

where f_n^* is the minimal value of the function in all trials, and the best α is chosen for the next step. It is usually assumed that $f(a)$ follows a normal distribution.

Combining eqs 4−9, the Bayesian optimization method is established, with which the hyperparameters are optimized for better performance. With an AMD Ryzen 3700X 3.6 GHz computer, 172 min is needed for this optimization process.

Batch Classification. After the optimized hyperparameters are found, both the batches in training set and validation set together constitute a batch dataset. This dataset is utilized to retrain a single BODNN model for the batch classification, which acts as a batch classifier. The prediction error of each batch is evaluated with this model, and the root mean square error (RMSE) is utilized as the index of prediction error. All batches in the dataset are sorted in ascending order by RMSE metrics, and the data of top 70% batches are used to construct subset θ_1. The data of the remaining batches are used to construct subset θ_2, which holds the nonlinear characteristics

https://doi.org/10.1021/acsomega.3c02387
ACS Omega 2023, 8, 25272−25278

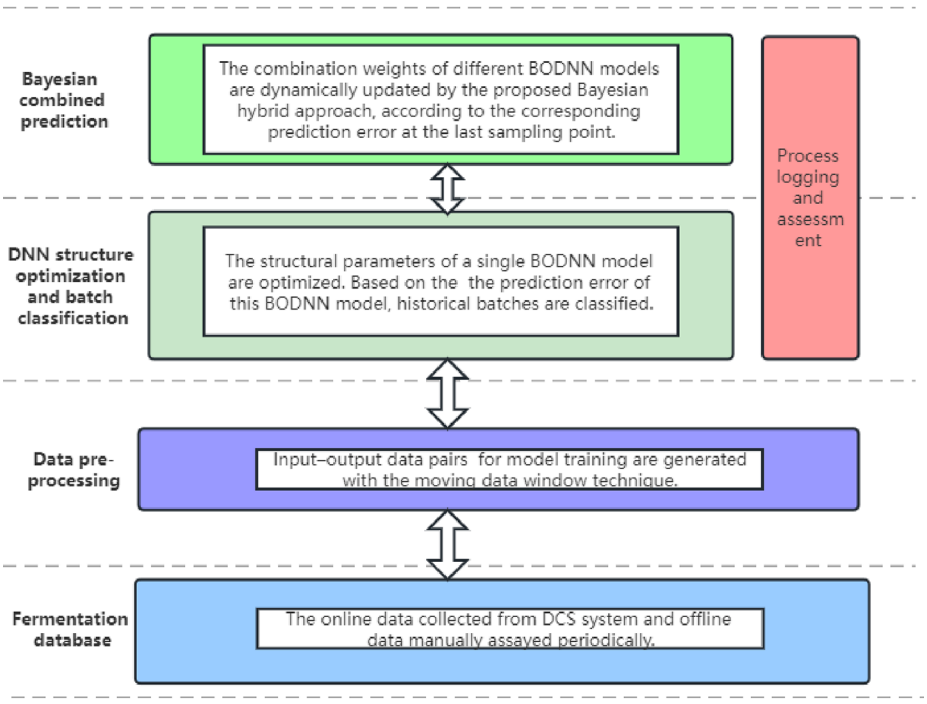


Figure 2. Block diagram of the Bayesian hybrid approach.

that cannot be modeled very well with the single BODNN model.

BCDNN: Bayesian Combination of BODNN Models.

With the divided subsets $\theta_1$and θ_2, two new BODNN models are trained separately as the basic predictors. Both models make use of the hyperparameters obtained by Bayesian optimization in BODNN: Bayesian Optimized Deep Neural Network, which means that their network structures are identical and the only difference is the training dataset.

When the data of a new online batch is collected at cultivation time T_k, the Bayesian combined prediction of the two models at time $T_k + T_P$ can be obtained, as shown in eq 10.

$$\hat{P}(T_k + T_P) = \sum_{r=1}^{2} p_{T_k}^r \hat{P}_r(T_k + T_P)\ r = 1, 2 \tag{10}$$

$\hat{P}_r(T_k + T_P)$ is the output of rth BODNN predictor at T_k. The dynamic weight $p_{T_k}^r$denotes the posterior probability of selecting the rth output as the hybrid output. The calculation of $p_{T_k}^r$can be determined utilizing eq 11, with the stochastic process Z's definition as follows: Z is equivalent to r at T_k when the rth BODNN model is selected to execute the prediction.

$$p_{T_k}^r = p(Z = r|P(T_k), P(T_{k-1}), \cdots, P(T_1)) \tag{11}$$

The following equations can be obtained based on the Bayes rule.

$$p_{T_k}^r = \frac{p(P(T_k), Z = r|P(T_{k-1}), P(T_{k-2}), ..., P(T_1))}{\sum_{j=1}^{2} p(P(T_k), Z = j|P(T_{k-1}), P(T_{k-2}), ..., P(T_1))} \tag{12}$$

$$\begin{aligned} &p(P(T_k), Z = r|P(T_{k-1}), P(T_{k-2}), ..., P(T_1)) \\ &= p(P(T_k)|Z = r, P(T_{k-1}), P(T_{k-2}), ..., P(T_1)) \cdot p_{T_{k-1}}^r \end{aligned} \tag{13}$$

The prediction error between the measured value and the output of rth BODNN model is defined as $e_{T_k}^r$, which obeys a Gaussian distribution $(e_{T_k}^r \sim N(\mu_r, \sigma_r^2))$, and then eq 14 is obtained.

$$\begin{aligned} &p(P(T_k)|Z = r, P(T_{k-1}), P(T_{k-2}), ..., P(T_1)) \\ &= p(e_{T_k}^r = P(T_k) - \hat{P}_r(T_k)|Z = r, P(T_{k-1}), P(T_{k-2}), ... \\ &\quad , P(T_1)) \\ &= \frac{1}{\sqrt{2\pi}\sigma_r} \exp(-[(e_{T_k}^r - \mu_r)/\sqrt{2}\sigma_r]^2) \end{aligned} \tag{14}$$

The distribution parameters μ_r and σ_r can be estimated from the prediction errors of the rth predictor, using the corresponding batch subset. Combining eqs 10−14, the iterative process of posterior probability $p_{T_k}^r$is shown in eq 15.

$$p_{T_k}^r = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_r}\right)^{k-1} \exp\left(\sum_{i=0}^{k-1} - [(e_{T_{k-1}}^r - \mu_r)/\sqrt{2}\sigma_r]^2\right)}{\sum_{j=1}^{2} \left(\frac{1}{\sqrt{2\pi}\sigma_j}\right)^{k-1} \exp\left(\sum_{i=0}^{k-1} - [(e_{T_{k-1}}^j - \mu_j)/\sqrt{2}\sigma_j]^2\right)} \tag{15}$$

The recursive calculation of hybrid forecasting can be performed at each sampled time. As the industrial data is constantly being collected, the fermentation database will be periodically updated with data from recent batches, requiring the entire Bayesian optimization process to be conducted anew. Figure 2 displays the block diagram of the proposed methodology.

https://doi.org/10.1021/acsomega.3c02387
ACS Omega 2023, 8, 25272−25278

RESULTS

To validate the performance of the proposed approach, the industrial data of 95 online batches are used to make comparison tests. During the data preprocessing, the input window T_D is set to 12 h. Prediction window T_P values are set to 4 and 8 h, which means 4 and 8 h ahead forecasting, respectively. The performance of the single BODNN model, Bayesian combined model (BCDNN), and other soft sensor modeling methods are compared.

Support vector machine (SVM) is a widely used method for the fermentation process, which can approximate the nonlinear process.[36] The polynomial $K(X_i, X) = (X_i^T X + 1)^3$ is used as the kernel function, while C and ε are set to 100 and 0.001, respectively.

The long short-term memory (LSTM),[37] gated recurrent unit (GRU),[38] and bidirectional LSTM(Bi-LSTM)[39] have been shown to achieve state-of-the-art results in many applications with time series or sequential data.[40,41] These algorithms can capture long-term temporal dependencies and variable-length observations. For training of these models, the ADAM algorithm is used as the gradient-based optimizer,[42] the number of neurons is 50 for the hidden layer, and the batch size is 32.

As the index of the prediction error, the average RMSE of 95 online batches using the methods mentioned above are compared in Table 2. Both 4 and 8 h forecasting errors are evaluated.

Table 2. Average RMSE of 95 Batches with Different Methods

model	SVM	LSTM	GRU	Bi-LSTM	BODNN	BCDNN
+4 h (%)	2.71	2.27	2.38	2.15	2.17	1.51
+8 h (%)	3.26	2.98	3.11	2.83	2.59	2.01

Table 2 demonstrates that the proposed BCDNN method gets better results for these fermentation batches compared to other mentioned models. To further compare the performance between the BODNN and BCDNN models, forecasting errors for the 95 batches are demonstrated separately.

Figures 3 and 4 show the RMSE of 4 and 8 h ahead prediction for each batch, respectively. The BCDNN method not only outperforms BODNN on average but also outperforms it on individual batches (with a few exceptions). The variance of 8 h ahead prediction error is more significant than 4 h due to the larger prediction window and uncertainty of the fermentation process.

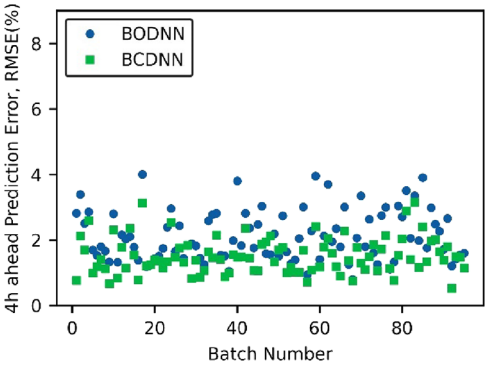


Figure 3. Prediction error distribution of industrial batches, 4 h ahead results.

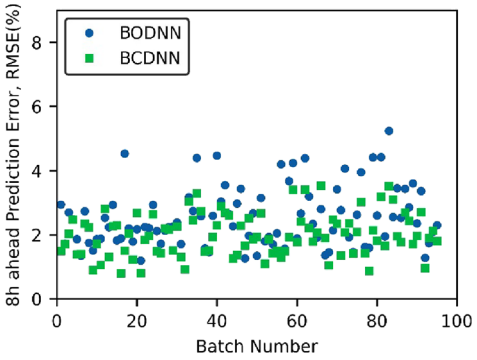


Figure 4. Prediction error distribution of industrial batches, 8 h ahead results.

DISCUSSION

The dynamics of different fermentation batches during cultivation are illustrated here. Figures 5 and 6 present the 4 h ahead predicted and measured product formation for two batches. The measurements sampled at 4 h intervals are represented by dotted lines. Batch no. 68 terminated in 52 h when residual L-sorbose was less than 1 kg/m^3, whereas batch no. 63 is an abnormal batch that terminated early. The BCDNN method can dynamically adjust the weights of the combined models according to the prediction error of the previous sampling point, thus achieving better results than a single BODNN predictor in the following time.

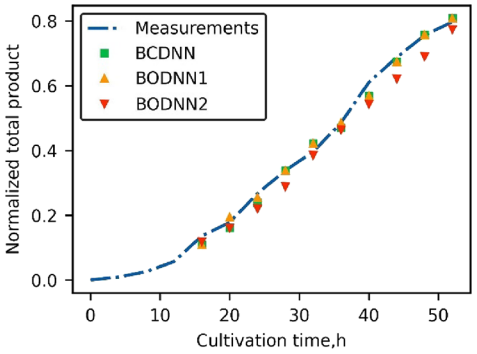


Figure 5. Actual and 4 h ahead predicted of batch no. 68.

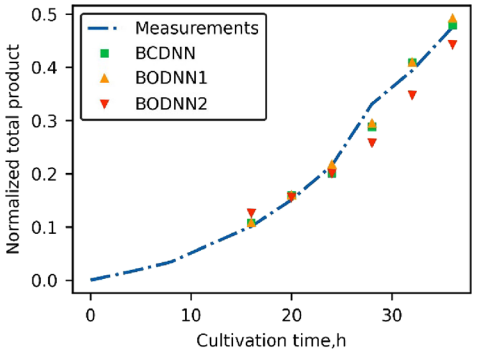


Figure 6. Actual and 4 h ahead prediction of batch no. 63.

Figures 7 and 8 present the 8 h ahead measurements and prediction results of representative batches. Because the prediction window is larger than that of the 4 h task, there is an increased difficulty for a single BODNN model to obtain accurate 8 h ahead perdition results. In the early stages of batch

https://doi.org/10.1021/acsomega.3c02387
ACS Omega 2023, 8, 25272−25278

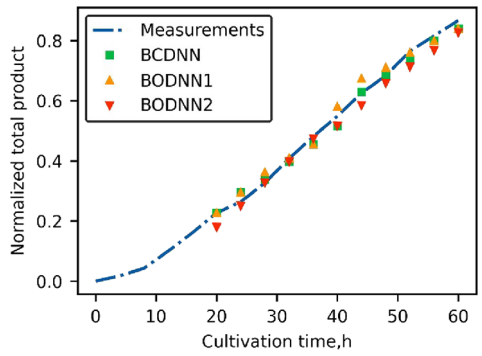


Figure 7. Actual and 8 h ahead prediction of batch no. 38.

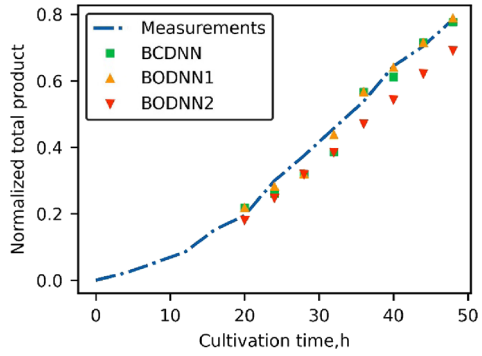


Figure 8. Actual and 8 h ahead prediction of batch no.79.

nos. 38 and 79, BODNN2 had a better performance. However, during the middle and late stages of the fermentation process, the prediction results of BODNN1 were getting closer to the measurements, and its weights in the BCDNN predictor became larger.

Specifically, in batch no. 79, the prediction accuracy of the BODNN1 at 32 h became significantly higher than that of the BODNN2. The larger weights of BODNN2 in the BCDNN predictor led to a large prediction error at this sampling time. The source of this sudden variation was mainly the disturbance of the dynamic process in the mixed culture of *B. megaterium* and *K. vulgare*. However, at the next sampling time 36 h, the weights of BODNN1 immediately became larger after the Bayesian iterative calculation, which led to a rapid reduction in the prediction error of the BCDNN predictor and maintained it until the end of the batch. Therefore, the proposed BCDNN method can produce more accurate predictions than a single BODNN model through adaptively changing posterior probabilities of different models.

CONCLUSIONS

This study proposes a probabilistic Bayesian deep learning approach for obtaining highly accurate and robust prediction of 2-KGA product formation. First, the input−output data pairs for predictor training are generated with the moving data window technique. The structural parameters such as neurons and connections of BODNN model are optimized by the Bayesian approach. Then, the process data of historical batches are classified into different categories, according to the prediction error. Each predictor is trained with the data of corresponding category. The final forecasting is a weighted combination of predictors based on the Bayesian hybrid approach. The weights can be interpreted as Bayesian posterior probabilities and are computed recursively. The validation results illustrate that the dynamic and iterative calculation makes this approach more suitable for online process monitoring.

AUTHOR INFORMATION

Corresponding Author

Tao Wang − *School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China;* orcid.org/0000-0001-5956-8440; Email: wangtaovc@sdut.edu.cn

Authors

Jiebing You − *Department of Neurology, Zibo Central Hospital, Zibo, Shandong 255036, China*

Xiugang Gong − *School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China*

Shanliang Yang − *School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China*

Lei Wang − *School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China*

Zheng Chang − *School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c02387

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research is supported by Shandong Provincial Natural Science Foundation (grant numbers: ZR2020QF069 and ZR2021MF017).

REFERENCES

(1) Hancock, R. D.; Viola, R. Biotechnological approaches for L-ascorbic acid production. *Trends Biotechnol.* **2002**, *20*, 299−305.

(2) Fonseca, J. D.; Latifi, A. M.; Orjuela, A.; Rodríguez, G.; Gil, I. D. Modeling, analysis and multi-objective optimization of an industrial batch process for the production of tributyl citrate. *Comput. Chem. Eng.* **2020**, *132*, No. 106603.

(3) Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M.; Aspuru-Guzik, A.; Hein, J. E. Data-science driven autonomous process optimization, Communications. *Chemistry* **2021**, *4*, 1−12.

(4) Zeng, Z.; Hong, M.; Man, Y.; Li, J.; Zhang, Y.; Liu, H. Multi-object optimization of flexible flow shop scheduling with batch process-Consideration total electricity consumption and material wastage. *J. Cleaner Prod.* **2018**, *183*, 925−939.

(5) Grisales Díaz, V. H.; von Stosch, M.; Willis, M. J. Butanol production via vacuum fermentation: An economic evaluation of operating strategies. *Chem. Eng. Sci.* **2019**, *195*, 707−719.

(6) Mears, L.; Stocks, S. M.; Sin, G.; Gernaey, K. V. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *J. Biotechnol.* **2017**, *245*, 34−46.

(7) Gan, J.; Parulekar, S. J. Multi-rate data-driven models for lactic acid fermentation-Parameter identification and prediction. *Comput. Chem. Eng.* **2019**, *128*, 405−416.

(8) Yuan, X.; Li, L.; Wang, Y. Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network. *IEEE Trans. Industr. Inform.* **2020**, *16*, 3168−3176.

(9) Wang, B.; Shahzad, M.; Zhu, X.; Rehman, K. U.; Ashfaq, M.; Abubakar, M. Soft-Sensor Modeling For L-Lysine Fermentation Process Based on Hybrid Ics-Mlssvm. *Sci. Rep.* **2020**, *10*, 11630.

(10) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436−444.
(11) Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J.; Blum, M.; Hutter, F. Auto-Sklearn: Efficient and Robust Automated Machine Learning. *Automated Machine Learning* **2019**, 113−134.
(12) Sagingalieva, A.; Kurkin, A.; Melnikov, A.; Kuhmistrov, D.; Perelshtein, M.; Melnikov, A. A.; Skolik, A., Dollen, D. V. Hyperparameter optimization of hybrid quantum neural networks for car classification, *ArXiv*, **2022**, 2205.04878.
(13) Shah, P.; Sheriff, M. Z.; Bangi, M. S. F.; Kravaris, C.; Kwon, J. S. I.; Botre, C.; Hirota, J. Deep neural network-based hybrid modeling and experimental validation for an industry-scale fermentation process: Identification of time-varying dependencies among parameters. *Chem. Eng. J.* **2022**, *441*, No. 135643.
(14) Bose, A.; Hsu, C. H.; Roy, S. S.; Lee, K. C.; Mohammadi-Ivatloo, B.; Abimannan, S. Forecasting stock price by hybrid model of cascading Multivariate Adaptive Regression Splines and Deep Neural Network. *Computers and Electrical Engineering* **2021**, *95*, No. 107405.
(15) Imani, M.; Ghoreishi, S. F. Scalable Inverse Reinforcement Learning Through Multifidelity Bayesian Optimization. *IEEE Transactions on Neural Networks and Learning Systems* **2022**, *33*, 4125−4132.
(16) Petridis, V.; Kehagias, A.; Petrou, L.; Bakirtzis, A.; Kiartzis, S.; Panagiotou, H.; Maslaris, N. A Bayesian Multiple Models Combination Method for Time Series Prediction. *Journal of Intelligent and Robotic Systems* **2001**, *31*, 69−89.
(17) Yang, Y.; Li, W.; Gulliver, T. A.; Li, S. Bayesian Deep Learning Based Probabilistic Load Forecasting in Smart Grids. *IEEE Transactions on Industrial Informatics* **2020**, *16*, 4703−4713.
(18) Abedinia, O.; Bagheri, M. Execution of synthetic Bayesian model average for solar energy forecasting. *Iet Renewable Power Generation* **2022**, *16*, 1134−1147.
(19) Abedinia, O.; Bagheri, M.; Agelidis, V. G. Application Of An Adaptive Bayesian-Based Model For Probabilistic And Deterministic Pv Forecasting. *Iet Renewable Power Generation* **2021**, *15*, 2699−2714.
(20) Noruzi, A.; Banki, T.; Abedinia, O.; Ghadimi, N. A new method for probabilistic assessments in power systems, combining monte carlo and stochastic-algebraic methods. *Complexity* **2015**, *21*, 100−110.
(21) Grover, A.; Kapoor, A.; Horvitz, E. J. A Deep Hybrid Model for Weather Forecasting. *Knowledge Discovery and Data Mining* **2015**, 379−386.
(22) Guo, S.; Huang, W.; Wang, L.; Qiao, Y. Locally Supervised Deep Hybrid Model for Scene Recognition. *IEEE Trans. Image Process.* **2017**, *26*, 808−820.
(23) Zhang, Y.; Lin, H.; Yang, Z.; Wang, J.; Zhang, S.; Sun, Y.; Yang, L. A hybrid model based on neural networks for biomedical relation extraction. *J. Biomed. Inf.* **2018**, *81*, 83−92.
(24) Ayari, N.; Abdelkawy, H.; Chibani, A.; Amirat, Y. Hybrid Model-Based Emotion Contextual Recognition for Cognitive Assistance Services. *IEEE Transactions on Cybernetics* **2022**, *52*, 3567−3576.
(25) Sun, H.; Betti, R. A Hybrid Optimization Algorithm with Bayesian Inference for Probabilistic Model Updating. *Computer-Aided Civil and Infrastructure Engineering* **2015**, *30*, 602−619.
(26) Zhang, Y.; Meng, K.; Kong, W.; Dong, Z. Y.; Qian, F. Bayesian Hybrid Collaborative Filtering-based Residential Electricity Plan Recommender System. *IEEE Transactions on Industrial Informatics* **2019**, *15*, 4731.
(27) Lessan, J.; Fu, L.; Wen, C. A hybrid Bayesian network model for predicting delays in train operations. *Computers & Industrial Engineering* **2019**, *127*, 1214−1222.
(28) Zhou, J.; Ma, Q.; Yi, H.; Wang, L.; Song, H.; Yuan, Y.-J. Metabolome Profiling Reveals Metabolic Cooperation between Bacillus megaterium and Ketogulonicigenium vulgare during Induced Swarm Motility. *Appl. Environ. Microbiol.* **2011**, *77*, 7023−7030.
(29) Wang, Y.; Li, H.; Liu, Y.; Zhou, M.; Ding, M.; Yuan, Y. Construction of synthetic microbial consortia for 2-keto-L-gulonic acid biosynthesis. *Synthetic and Systems Biotechnology* **2022**, *7*, 481−489.
(30) Xu, R.-Z.; Cao, J.-S.; Wu, Y.; Wang, S.-N.; Luo, J.-Y.; Chen, X.; Fang, F. An integrated approach based on virtual data augmentation and deep neural networks modeling for VFA production prediction in anaerobic fermentation process. *Water Res.* **2020**, *184*, No. 116103.
(31) Song, H.; Kim, M.; Park, D.; Lee, J.-G. Learning from Noisy Labels with Deep Neural Networks: A Survey. *IEEE transactions on neural networks and learning systems* 2022, DOI: 10.1109/TNNLS.2022.3152527.
(32) Wang, T.; Sun, J.; Zhang, W.; Yuan, J. Prediction of product formation in 2-keto-L-gulonic acid fermentation through Bayesian combination of multiple neural networks. *Process Biochem.* **2014**, *49*, 188−194.
(33) Clevert, D. A.; Unterthiner, T., Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), arXiv preprint arXiv:1511.07289 2015.
(34) Lee, J.; Bahri, Y.; Novak, R.; Schoenholz, S. S.; Pennington, J., Sohl-Dickstein, J. Deep Neural Networks as Gaussian Processes, 2018, ArXiv, abs/1711.00165.
(35) Daulton, S.; Balandat, M.; Bakshy, E. Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement. *Conference on Neural Information Processing Systems* **2021**, 2187−2200.
(36) Liu, G.; Zhou, D.; Xu, H.; Mei, C. Model optimization of SVM for a fermentation soft sensor. *Expert Systems with Applications* **2010**, *37*, 2708−2713.
(37) Ma, X. L.; Tao, Z. M.; Wang, Y. H.; Yu, H. Y.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C-emerging Technologies* **2015**, *54*, 187−197.
(38) Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks For Multivariate Time Series With Missing Values. *Computing Research Repository* **2018**, *8*, 6085.
(39) Ma, X., Hovy, E. End-To-End Sequence Labeling Via Bi-Directional Lstm-Cnns-Crf, 2016, arXiv preprint arXiv: abs/1603.01354.
(40) Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation* **2019**, *31*, 1235−1270.
(41) Shewalkar, A.; Nyavanandi, D.; Ludwig, S. A. Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN LSTM and GRU. *Journal of Artificial Intelligence and Soft Computing Research* **2019**, *9*, 235−245.
(42) Reddi, S. J.; Kale, S., Kumar, S. On the Convergence of Adam and Beyond, arXiv preprint arXiv:1904.09237, 2019.

https://doi.org/10.1021/acsomega.3c02387
ACS Omega 2023, 8, 25272−25278