

A HYBRID TCN-LSTM MODEL FOR ACCURATE TOBACCO CURING STATE RECOGNITION

一种用于烤房烟叶状态精准识别的混合 TCN-LSTM 模型

Chengyu YIN¹⁾, Hanchao ZHU¹⁾, Lei ZHOU^{*1)}

¹⁾ School of Electronics and Information Engineering, Hangzhou Dianzi University, Zhejiang /China;

Tel: +8613868139686; E-mail: zhoulei@hdu.edu.cn

Corresponding author: Lei Zhou

DOI: <https://doi.org/10.35633/inmateh-77-34>

Keywords: Tobacco curing, state recognition, Kernel Density Estimation, Long Short-Term Memory, Temporal Convolutional Network

ABSTRACT

The curing of tobacco is a critical process that determines the quality of the final product. Accurate recognition of tobacco curing states is essential for ensuring optimal quality. Existing recognition models mostly focus on the transient states within the curing barn. In contrast, this study incorporates multiple time steps to capture dynamic feature changes in the curing barn over time, providing a more accurate state recognition. A hybrid deep learning model combining Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM) networks, and a novel Density-aware Channel Redistribution Unit (DCRU) based on Kernel Density Estimation is proposed. The model integrates the global feature extraction capability of TCN, the long-term dependency modeling strength of LSTM, and the complex channel feature extraction ability of DCRU, thereby enhancing the model's performance in recognizing the stages of tobacco leaf curing. Tests conducted on a real-world tobacco dataset demonstrate that the model achieves a prediction accuracy of 0.989 and outperforms baseline models as well as existing tobacco curing state recognition methods. These results validate the effectiveness of the hybrid TCN-LSTM model in recognizing tobacco leaf curing states, with promising applications in agricultural automation.

摘要

烟叶烘烤是决定最终产品质量的关键工序，准确识别烟叶烘烤状态对于保障其品质至关重要。现有的识别模型多基于烘烤过程中烤房内的瞬时状态，而本研究通过引入多个时间步，捕捉烤房内动态变化的特征，从而实现更为准确的状态识别。本文提出了一种融合 Temporal Convolutional Network (TCN)、Long Short-Term Memory (LSTM) 与基于核密度估计的 Density-aware Channel Redistribution Unit (DCRU) 的混合深度学习模型。该模型结合了 TCN 的全局特征提取能力、LSTM 对长期依赖的建模能力，以及 DCRU 在复杂通道特征分布提取方面的优势，从而有效提升了对烟叶烘烤阶段的识别性能。在真实烟叶数据集上的测试结果表明，该模型的预测准确率达到 0.989，优于基线模型及现有的烟叶烘烤状态识别方法。研究结果验证了该混合 TCN-LSTM 模型在烟叶烘烤状态识别中的有效性，为农业自动化应用提供了有前景的解决方案。

INTRODUCTION

Tobacco has long been an economically significant crop, contributing substantially to the economies of many countries. Globally, tobacco production supports millions of livelihoods and contributes significantly to tax revenues, especially in developing countries. Besides contributing to tax revenues, tobacco cultivation also plays a vital role in poverty alleviation in economically disadvantaged regions (Ahsan et al., 2022). The production of tobacco involves several stages, each contributing to the final product's quality. Among the various stages, the curing process stands out as a particularly critical determinant of the final product's quality. Improper curing techniques or incorrect temperature settings in bulk curing barns can degrade leaf quality, ultimately influencing the market value of tobacco products.

The tobacco curing process is complex, typically taking about five to seven days (Abubakar et al., 2000). During curing, mature tobacco leaves undergo gradual dehydration, along with a series of physiological and biochemical changes (Pang et al., 2024). As shown in Fig. 1, the curing process in bulk curing barns is typically segmented into six stages based on changes in the color and texture of the leaves (Wang et al., 2016).

Farmers constantly monitor the state of tobacco leaves, carefully adjusting the temperature, humidity, and airflow based on the recognized tobacco curing state within the bulk curing barn to ensure optimal curing conditions. However, this manual adjustment can lead to inconsistent tobacco quality, which highlights the need for automated recognition methods to achieve precise and consistent control during the curing process.

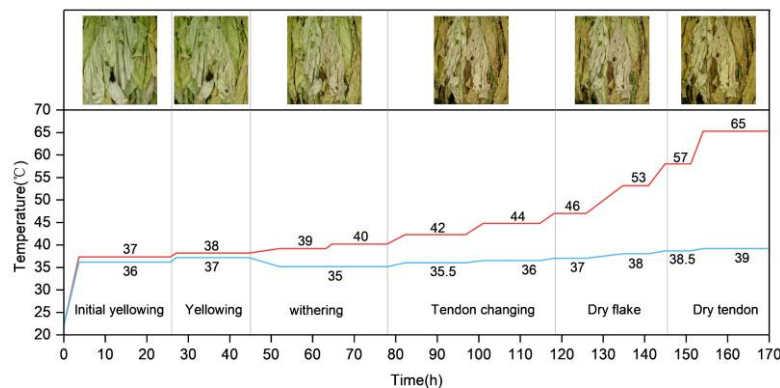


Fig. 1 - Methodology flowchart of tobacco curing state recognition.

With the advancement of machine learning and deep learning technologies, numerous scholars have explored their applications in the tobacco processing industry. These approaches have been employed to optimize various aspects of the production process, including maturity identification, leaf position recognition, tobacco grading, and automatic tobacco curing. In the tobacco curing process, the dry-bulb temperature (DT), wet-bulb temperature (WT), and temperature-raising time are key parameters that require precise control at each stage (Wu *et al.*, 2017). Consequently, many studies have focused on predicting these key parameters to ensure optimal process control (Wang *et al.*, 2017; Wu & Yang, 2019; Wu *et al.*, 2014).

Compared to work on predicting those process parameters, there has been relatively little research on recognizing the curing state of tobacco leaves. Color is a key indicator of curing extent, since moisture loss and color change progress in coordination during an ideal curing process (Meng *et al.*, 2024). Consequently, numerous studies investigated the dynamics of color change in tobacco during the curing process. Wu (2016) extracted color features from multiple color spaces (RGB, HSV, L*a*b*) and then applied a genetic algorithm-support vector machine (GA-SVM) framework for feature selection. This approach removed redundant features and reduced computational complexity. Li *et al.* (2022) extended this method by incorporating texture features alongside color features, thereby improving recognition accuracy. However, this iterative process results in longer inference times, and the genetic algorithm is prone to getting stuck in local optima. Wang and Qin (2022) considered weight variation of leaves during curing and proposed a fusion model combining Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) to exploit both curing-room environmental data and the leaves' multi-modal features. Pei *et al.* (2024) observed varying degrees of feature change across curing stages and introduced an XGBoost model incorporating feature-weighting preprocessing, which assigns weights based on each feature's change intensity. Given the ability of Convolutional Neural Networks (CNNs) to efficiently extract complex features, some scholars have used CNNs to directly extract features from images (Song *et al.*, 2024). Xiong *et al.* (2024) designed a state-recognition network combining standard convolutional layers with short-term dense connections to achieve high-precision prediction. However, it neglects curing-barn environmental factors, which significantly affect recognition performance.

Unlike the aforementioned static state recognition methods, tobacco curing is considered a dynamic process due to its time-evolving nature. Consequently, modeling this process requires techniques capable of capturing temporal dependencies in sequential data. In recent years, Long Short-Term Memory (LSTM) networks have emerged as a widely adopted deep learning model for sequential data, including applications in agriculture (Attri *et al.*, 2023). LSTM is a variant of recurrent neural networks (RNNs), designed to address long-term dependency and vanishing gradient problems via its gating mechanisms (Graves & Schmidhuber, 2005). To leverage the feature extraction capabilities of CNNs while effectively handling sequential data, Bai *et al.* (2018) proposed the temporal convolutional network (TCN). TCN uses dilated convolutions to expand the receptive field and causal convolutions to prevent information leakage from future time steps, enabling effective extraction of temporal features from time series data (Liu *et al.*, 2024).

Inspired by the STar Aggregate-Redistribute (STAR) module (Han *et al.*, 2024), a Density-aware Channel Redistribution Unit (DCRU) is proposed in this study. Unlike traditional attention mechanisms that often introduce considerable computational complexity, STAR utilizes a centralized structure built from

lightweight multilayer perceptron (MLP). However, STAR suffers from two major limitations: (1) it samples only the most prominent channels, potentially overlooking complex feature structures with multi-modal or skewed distributions; and (2) it uses different strategies during training and inference, which may lead to performance instability. To address these limitations, the proposed DCRU integrates a channel redistribution mechanism based on Kernel Density Estimation (KDE). Unlike STAR, which samples only the most prominent channels, the KDE-based approach adaptively models the distribution of channel importance, enabling the capture of more complex structures such as multi-peaked or skewed patterns. Specifically, KDE is applied along the channel dimension to estimate the density distribution of feature importance. The resulting probability distribution is then used to generate sampling weights through sampling. This allows the model to consider the entire channel distribution, rather than focusing solely on the most salient feature, thus enhancing the expressiveness of feature representations across curing stages. Furthermore, the KDE-based strategy maintains consistency between training and inference, reducing the performance fluctuations caused by STAR's inconsistent sampling policy.

In summary, considering the complex and massive data involved in the curing process and the limitations of previous studies, a hybrid TCN-LSTM model that integrates TCN, LSTM, and DCRU is proposed. In the proposed model, the TCN is leveraged to extract global features from the raw data, while the LSTM effectively captures long-term dependencies within the data. The DCRU captures both salient features and distributional characteristics of the data. This model effectively represents features by preserving detailed local information and integrating global information. To verify the effectiveness of the proposed method in practical scenarios, its recognition performance is validated using data collected from tobacco curing barns. The final results demonstrate that the proposed model outperforms conventional baseline neural network models in tobacco curing state recognition.

MATERIALS AND METHODS

Dataset

The dataset used in this study was collected from eight tobacco curing barns located at tobacco stations in Jinjiang, Jiangxi Province, and Nanping, Fujian Province, between June and October 2022. Images of the curing barns, along with DT and WT readings, were recorded every five minutes. A total of 19,776 data records were collected across twelve curing batches. Of these, eight batches were used for model training, two batches for validation, and the remaining two batches for testing. The curing stages were classified based on expert assessment in tobacco curing. As shown in Figure 2, Panel a) illustrates the tobacco leaf image acquisition device, Panel b) presents a sample of the acquired tobacco leaf images, and Panel c) shows the overall workflow of data processing, model training, and testing.

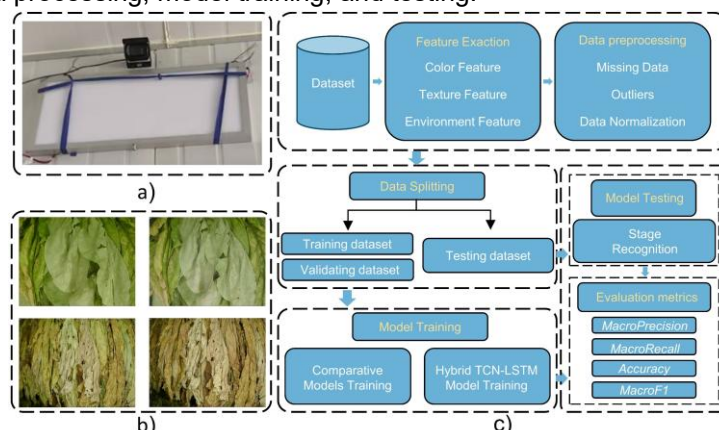


Fig. 2 - Methodology flowchart of tobacco curing state recognition.

a) Tobacco image acquisition equipment; b) Tobacco image; c) Data processing and training process;

Feature extraction of tobacco

Various color models exist, each offering distinct advantages and application scenarios. To effectively simulate human color perception, the color model was selected to correspond closely with the characteristics of the human visual system. Among these, the RGB (Red, Green, Blue) color space serves as a fundamental model in digital imaging and is commonly used as the basis for conversion to other color spaces. Because the yellow-brown coloration of tobacco leaves primarily results from a mixture of red and green components, the R and G channels of the RGB model were selected for analysis (Pei et al., 2024). In addition, the a component

of the $L^*a^*b^*$ color space (where L^* denotes lightness, a represents the red-green axis, and b the blue-yellow axis) was utilized to capture color variations along the red-green spectrum relevant to tobacco leaf color changes.

Four statistical measures derived from the Gray-Level Co-occurrence Matrix were used for texture feature extraction: Entropy (ENT), Angular Second Moment (ASM), and Inverse Difference Moment (IDM) (Haralick, 1979; Li et al., 2022). To comprehensively capture the texture information of the image, these features were calculated as the average values across four directions: 0° , 45° , 90° , and 135° .

ENT quantifies the complexity of image textures and describes the uncertainty of image information. ASM emphasizes the consistency and uniformity of image textures, while IDM reflects the local contrast and smoothness. By incorporating these features, the model gains sensitivity to the subtle visual cues that are indicative of different curing stages, enabling more accurate and robust classification. The formulas for the calculation of texture features are as follows:

$$ENT = \sum_{i,j} p(i,j) \log(p(i,j)) \quad (1)$$

$$ASM = \sum_{i,j} p(i,j)^2 \quad (2)$$

$$IDM = \sum_{i,j} \frac{p(i,j)}{1+(i-j)^2} \quad (3)$$

where $P(i,j)$ represents the probability of occurrence of the gray level (i,j) . μ is the mean of gray levels and σ is the standard deviation of the gray levels.

In total, eight features were extracted for model input, including three color features (R, G, and a from the $L^*a^*b^*$ color space), three texture features (ENT, ASM, and IDM), and two environmental features (DT and WT).

Model Architecture

As illustrated in Fig. 3, a hybrid framework integrating TCN, LSTM, and a DCRU module was constructed. This architecture is designed to effectively capture long-term dependencies and extract deeper, more complex features from the input data. The details of each component will be discussed in the following sections.

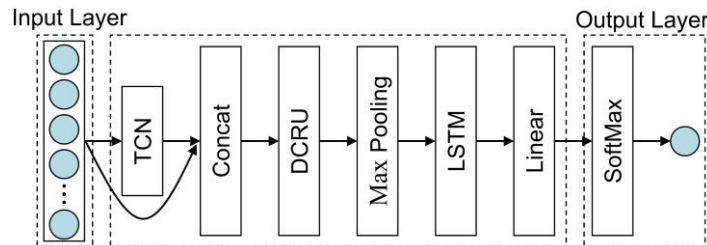


Fig. 3 - Network structure of the TCN-LSTM model

Temporal Convolutional Network (TCN)

The TCN module enables the model to capture intricate features from input data and addresses challenges such as exploding or vanishing gradients. Unlike traditional RNN, TCN is based on the architecture of CNN. Of particular importance, TCN supports parallel computing and, furthermore, a feedforward model specifically designed for sequence modelling (Fan et al., 2023). Its architecture incorporates key components such as causal convolution, dilated convolution, and residual connections, which collectively enhance its overall performance.

To process time series data, TCN adopts causal convolution, which is exclusively based on elements from time t and earlier. This strictly time-constrained scheme ensures that information from the future is not incorporated into past computations. Although causal convolutions require an increase in the number of layers to capture more information, resulting in a deeper and more complex network, the hierarchical structure of TCN captures long-range patterns in sequential data.

Dilated convolution refers to an enhanced version of traditional convolution, achieved through interval sampling, which enables the TCN model to obtain an exponentially large receptive field for aggregating diverse information, particularly addressing the challenge of making predictions with long-term memory (Cheng et al., 2021). The dilated convolution operation on elements of the 1-D sequence input $x \in \mathbb{R}^n$ is defined as shown in Equation (4) for a given filter $f: \{0, \dots, k-1\} \rightarrow \mathbb{R}$.

$$F(s) = (x *_{d} f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (4)$$

where k is the convolution kernel size, d is the dilation rate, $*$ denotes the convolution operation, and $s - d \cdot i$ indicates the past direction. The architecture of dilated causal convolution is illustrated in Fig. 4a).

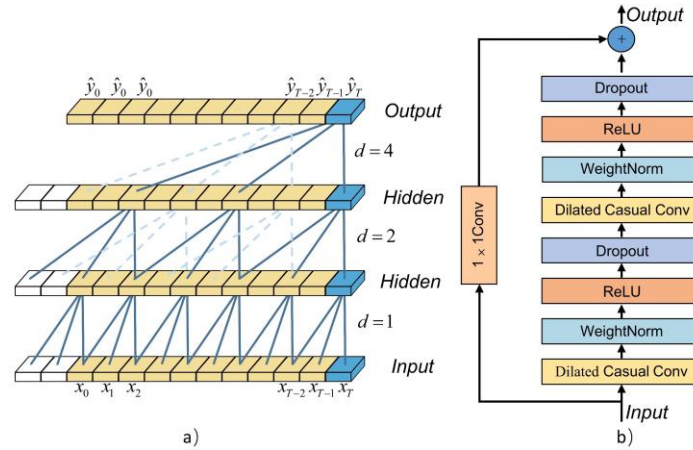


Fig. 4 - The network structure of TCN

a) Dilated causal convolution structure of TCN; b) Residual block of TCN;

As neural networks become deeper, optimization may degrade, a phenomenon known as network degradation. The residual block mitigates network degradation and the vanishing gradients problem. The residual connection operation is mathematically represented as shown in Equation (5).

$$o = x + F(x, W) \quad (5)$$

where o is the output of the residual block, and $F(x, W)$ is the residual part.

The residual block of TCN is shown in Fig. 4b). Besides the elements mentioned above, the block includes both WeightNorm and Dropout layers. The former accelerates the convergence speed, while the latter helps prevent overfitting and improves the model's generalization ability.

Long Short-Term Memory Network (LSTM)

To solve the problems of vanishing and exploding gradients, which render RNNs inadequate for capturing long-term dependencies, LSTM was developed to overcome these limitations. LSTM can be regarded as an extension of RNNs that introduces memory cells and gates. The memory cells retain the output of the current time step and manage long-term information, while the gates regulate the flow of information into and out of the memory cell.

The architecture of LSTM is illustrated in Fig.5. The forget gate f_t controls the extent to which information from the previous cell state (c_{t-1}) is retained or discarded. The input gate (i_t) determines which new information is added to the cell state (c_t). The output gate (o_t) regulates the amount of information from the updated cell state that is used to compute the hidden state (h_t), which serves as the output and is transmitted to the next LSTM unit. This gating mechanism allows selective updating or retention of information, supporting the handling of long-term dependencies.

The formulas for the LSTM cell are represented from Equation (6) to Equation (11).

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (6)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (7)$$

$$\tilde{c} = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (8)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c} \quad (9)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

where:

$\sigma(\cdot)$ is the sigmoid activation function, W is the weight matrix, b represents the bias vectors, and $\odot$ denotes the dot product.

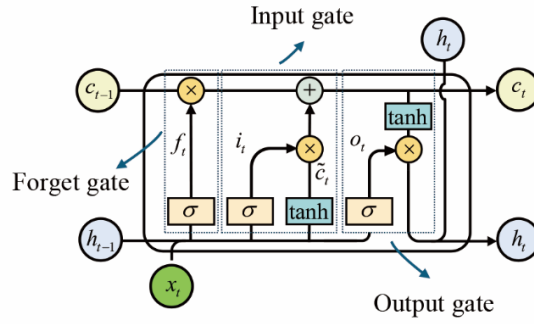


Fig. 5 - Structure of the LSTM

Density-aware Channel Redistribution Unit (DCRU)

Throughout the different stages of tobacco curing, both the curing environment and the physical attributes of tobacco leaves undergo dynamic changes. A Density-aware Channel Redistribution Unit (DCRU) is employed to extract salient features and the main feature distribution at each stage.

As shown in Fig. 6, DCRU first aggregates information from different channels to obtain the global core representation. Then the local series representation is fused with the core representation to achieve the indirect interaction between channels. This interaction method takes advantage of both channel independence and aggregated information, which enhances the representation of important features.

Given a multivariate series with C channels, DCRU first constructs a global core representation by aggregating information across all channels. Each channel is processed through a multilayer perceptron (MLP), which consists of two linear layers with GELU activation, to project the input from the original dimension d to core dimension d' . To overcome the limitations of the STAR module, particularly its tendency to sample only the most salient features while ignoring complex feature patterns, DCRU introduces a dual-path sampling strategy to generate a more expressive core representation.

KDE sampling is applied along the channel dimension to estimate the distribution of channel-wise feature importance. This density-aware approach captures complex patterns such as multi-modal and skewed distributions, resulting in a probabilistic part of core representation $Core_{kde} \in \mathbb{R}^{\tilde{d}}$, with further details provided in the following section. In parallel, stochastic sampling adopts the same strategy as STAR to capture key discriminative features from channels and produces another part of core representation $Core_{stoch} \in \mathbb{R}^{\tilde{d}}$. These two representations are concatenated to form the final core representation $Core = [Core_{kde}; Core_{stoch}] \in \mathbb{R}^{\tilde{d}+\tilde{d}}$ which effectively integrates both distributional characteristics and salient features to enhance the model's representational capacity.

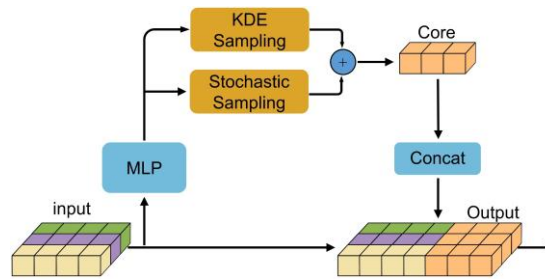


Fig. 6 - Architecture of the DCRU module

A key component of DCRU is the KDE-based sampling strategy, which estimates the distributional patterns of channel-wise feature importance. The formulation and application of this mechanism are described below. To estimate the distribution of channel-wise feature importance, KDE is applied along the channel dimension. Given a feature vector $\{x_1, x_2, \dots, x_c\} \in \mathbb{R}^C$ representing the channel-wise activations of a sample, the density at each value x is estimated by treating it as the query point. The KDE equation is presented in Equation (12).

$$\hat{f}(x) = \frac{1}{C} \sum_{i=1}^C \frac{1}{h_i} K\left(\frac{x-x_i}{h_i}\right) \quad (12)$$

The bandwidth h_i is treated as a learnable parameter in the model. This enables the KDE mechanism to adaptively adjust its sensitivity to different channels. The kernel function $K(\cdot)$ is chosen to be the Gaussian kernel.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \quad (13)$$

Substituting into the KDE formula yields the following expression for the density estimate at channel x .

$$\hat{f}(x) = \frac{1}{C} \sum_{i=1}^C \frac{1}{\sqrt{2\pi}h_i} \exp\left(-\frac{1}{2}\left(\frac{x-x_i}{h_i}\right)^2\right) \quad (14)$$

The raw density scores are converted into a valid probability distribution by applying a softmax operation over the channel density.

$$p_i = \frac{\exp(\hat{f}(x_i))}{\sum_{k=1}^C \exp(\hat{f}(x_k))}, i = 1, 2, \dots, C \quad (15)$$

This distribution P is then used to guide the sampling of multiple informative channels. Specifically, multinomial sampling is performed to select $\tilde{d}$ channels, forming a new representation vector $Core_{kde}$.

$$Core_{kde} = [x_1, x_2, \dots, x_{\tilde{d}}], \text{ where } x_i \sim P \text{ for } i = 1, \dots, \tilde{d} \quad (16)$$

The core representation is subsequently fused with the local series representation, facilitating indirect interactions across channels. This design preserves the independence of local features while enabling the model to benefit from global contextual cues. Importantly, the KDE-based sampling strategy ensures consistency between training and inference phases, thereby mitigating the performance instability observed in STAR. Furthermore, by modeling the entire distribution of channel importance rather than focusing solely on the most prominent channels, DCRU enhances the expressiveness and robustness of feature representations—particularly under dynamic curing conditions.

RESULTS

Model evaluation index

In this paper, accuracy (ACC) is used to evaluate the recognition performance of the hybrid model proposed. To account for the performance across all states equally, the macro-average of the precision (Macro-PR), recall (Macro-RE), and F1-scores (Macro-F1) is used, as described in Equation (17) to Eq.(20).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

$$MacroPrecision = \frac{\sum_{i=1}^N TP_i / (TP_i + FP_i)}{N} \quad (18)$$

$$MacroRecall = \frac{\sum_{i=1}^N TP_i / (FP_i + FN_i)}{N} \quad (19)$$

$$MacroF1 = \frac{2 * MacroPrecision * MacroRecall}{MacroPrecision + MacroRecall} \quad (20)$$

where N represents the total number of categories, i represents the index for each category, TP stands for true positive, TN stands for true negative, FP stands for false positive, and FN stands for false negative.

Experimental configuration

The experiments were conducted on a PC running Windows, equipped with an Intel Core Ultra5 125H CPU and an NVIDIA GeForce RTX 3080 GPU. Python 3.9.2 and the open-source software library PyTorch were used to conduct the experiments.

The flowchart of the hybrid TCN-LSTM model is presented in Fig. 3. Three stacked of TCN are employed to directly extract input features. Each block shares the same kernel size of 3, while the number of filters increases across blocks: 32, 64, and 128, respectively. The dilation factors for the three blocks are set to 1, 2, and 4, respectively. The core dimension of DCRU is set to 128, after which a max-pooling layer is applied to reduce the feature dimensionality. The LSTM module consists of two layers, with the first layer containing 64 units and the second layer 32 units. The model is optimized using the Adam optimizer, with an initial learning rate of 5×10^{-4} . A cosine learning rate scheduler is employed to adjust the learning rate dynamically during training.

Architectural Analysis of the Hybrid TCN-LSTM model

To evaluate the comparative effectiveness of different model architectures in recognizing tobacco curing states, a series of experiments were conducted using multiple deep learning frameworks. As shown in Table 1, the standalone TCN and LSTM models achieved comparable performance, with TCN attaining an accuracy of 0.963 and LSTM slightly lower at 0.962. When the two models were combined into a hybrid TCN-LSTM framework, performance improved notably across all metrics, achieving an accuracy of 0.975 and a

Macro-F1 score of 0.975. Further incorporating the STAR attention mechanism into the TCN-LSTM architecture (TCN-STAR-LSTM) yielded additional gains, with accuracy increasing to 0.983 and Macro-F1 reaching 0.981, indicating the effectiveness of attention-based feature refinement. Finally, the proposed TCN-DCRU-LSTM model outperformed all other models, achieving the highest accuracy of 0.989 and a Macro-F1 score of 0.988. These results demonstrate the superiority of our method in capturing complex temporal and distributional features during tobacco curing state recognition.

Table 1

Different Models architecture				
Model	ACC	Macro-PR	Macro-RE	Macro-F1
TCN	0.963	0.953	0.951	0.952
LSTM	0.962	0.961	0.958	0.959
TCN-LSTM	0.975	0.974	0.977	0.975
TCN-STAR-LSTM	0.983	0.980	0.982	0.981
TCN-DCRU-LSTM	0.989	0.989	0.987	0.988

Comparison with different model

To further evaluate the effectiveness of the proposed hybrid TCN-LSTM model for curing state recognition, experiments were conducted on the proposed and compared models in this section. To ensure a comprehensive evaluation, six baseline models were selected for comparison. The deep learning baselines include Artificial Neural Networks (ANN), RNN, CNN, and the Transformer model. Additionally, the proposed model was compared against traditional machine learning algorithms, including XGBoost and Support Vector Machines (SVM). The ANN model was implemented as a multilayer perceptron (MLP) with three hidden layers comprising 128, 64, and 32 units, respectively. The RNN model consists of three hidden layers with 64, 64, and 32 units. The CNN model includes two convolutional layers followed by a fully connected layer. The Transformer model adopts a two-layer encoder structure, with each layer containing four attention heads and an embedding dimension of 128. For XGBoost, the maximum tree depth was set to 4, and the number of estimators was 90. For SVM, the radial basis function kernel was used, with the regularization parameter C set to 1.

Table 2

Result of different models				
Model	ACC	Macro-PR	Macro-RE	Macro-F1
XVM	0.905	0.895	0.879	0.887
XGboost	0.915	0.894	0.901	0.897
ANN	0.935	0.927	0.912	0.919
CNN	0.955	0.945	0.946	0.945
RNN	0.947	0.935	0.931	0.933
Transformer	0.960	0.956	0.957	0.956
Hybrid TCN-LSTM	0.989	0.989	0.987	0.988

Table 2 presents the result of the evaluation metrics for the prediction performance of each model. In terms of prediction performance, it is evident that the proposed model outperforms all other models. The SVM achieves favorable results with 0.905 ACC, 0.895 Macro-PR, 0.879 Macro-RE, and 0.887 Macro-F1. The XGboost performs slightly better than SVM, achieving an ACC of 0.915, a Macro-PR of 0.894, a Macro-RE of 0.901, and a Macro-F1 of 0.897. The deep learning models significantly outperform the machine learning models. The ANN achieves an ACC of 0.935, a Macro-PR of 0.927, a Macro-RE of 0.912, and a Macro-F1 of 0.919. It can be observed that all deep learning models perform well in recognizing the curing state. However, a considerable number of incorrect recognitions still occur. For instance, the RNN achieves an ACC of 0.947, a Macro-PR of 0.935, a Macro-RE of 0.931, and a Macro-F1 of 0.933. Similarly, the CNN achieves an ACC of 0.955, a Macro-PR of 0.945, a Macro-RE of 0.946, and a Macro-F1 of 0.945. The Transformer performs slightly better, reaching an ACC of 0.960, a Macro-PR of 0.956, a Macro-RE of 0.957, and a Macro-F1 of 0.956.

The test results are shown in Fig. 7b), where the misclassified categories are mainly distributed among those with similar ground true classes, although there are still a considerable number of errors. The results of the Hybrid TCN-LSTM model are shown in Fig. 7a), with only 38 misclassifications across all labels.

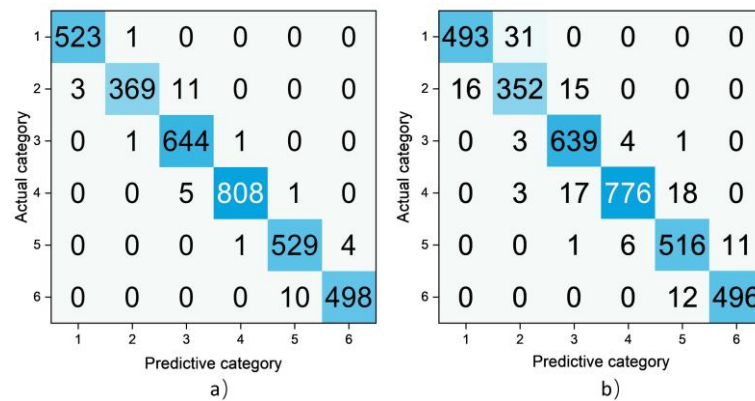


Fig. 7 - Confusion matrix of test result

a) Confusion matrix of Hybrid TCN-LSTM; b) Confusion matrix of Transformer;

In summary, the results presented in Table 2 clearly demonstrate the superior performance of the proposed hybrid TCN-LSTM model. Compared to both traditional machine learning models and advanced deep learning architectures, the hybrid TCN-LSTM consistently achieves the highest scores across all evaluation metrics. With an accuracy of 0.989 and a Macro-F1 score of 0.988, the model not only captures long-term dependencies and distributes patterns but also effectively integrates fine-grained local features. These findings confirm the robustness, expressiveness, and practical applicability of the proposed approach in the complex task of tobacco curing state recognition.

Comparison with related studies

Table 3 presents the results of the comparison conducted between the proposed model and other advanced methods, which were based on different sets of original data. Specifically, the recognition performance of hybrid TCN-LSTM model was compared with that of SRFM, GA-SVM, TFSNet, and CPBM. The results show that GA-SVM had the lowest ACC, with a value of 0.965. This was followed by the SRFM model, which attained an ACC of 0.974. Next, the CPBM model achieved better performance, with an ACC of 0.984. In comparison, the TFSNet model achieved an ACC of 0.987. The hybrid TCN-LSTM outperformed the other methods and achieved the highest ACC of 0.989. The results illustrate that the introduction of DCRU module can improve the recognition accuracy for tobacco curing state. In summary, the proposed hybrid TCN-LSTM can exhibit better recognition performance.

Table 3

Comparison with related studies				
Model	ACC	Macro-PR	Macro-RE	Macro-F1
SRFM(LSTM-XGboost)(Wang & Qin, 2022)	0.974	0.952	0.936	0.943
GA-SVM(Wu, 2016)	0.965	/	/	/
CBPM (data fusion based XGboost)(Pei et al., 2024)	0.984	0.986	0.982	0.984
TFSNet (CNN) (Xiong et al., 2024)	0.987	/	/	/
Hybrid TCN-LSTM	0.989	0.989	0.987	0.988

Practical Applicability Demonstration and Testing

To further demonstrate the practical applicability of the proposed method, the trained hybrid TCN-LSTM model was deployed on an upper-computer platform designed for tobacco curing state monitoring. The system was implemented in a Windows environment, featuring a graphical interface for real-time data visualization. Test data streams, identical to those used in the evaluation phase, were fed into the system to validate its end-to-end recognition performance.

The results showed that the model could process input sequences in real time, with an average inference latency of 146 ms per sample. The recognized curing states were automatically displayed on the interface, enabling operators to intuitively monitor stage transitions throughout the curing process.

Fig.8 illustrates an example of the upper-computer interface output, where the predicted states closely match the ground truth labels provided by experts.

This system-level validation demonstrates that the proposed model is not only theoretically effective but also practically deployable in real-world operational environments, enabling automated recognition of tobacco curing stages when deployed on cloud or edge devices.

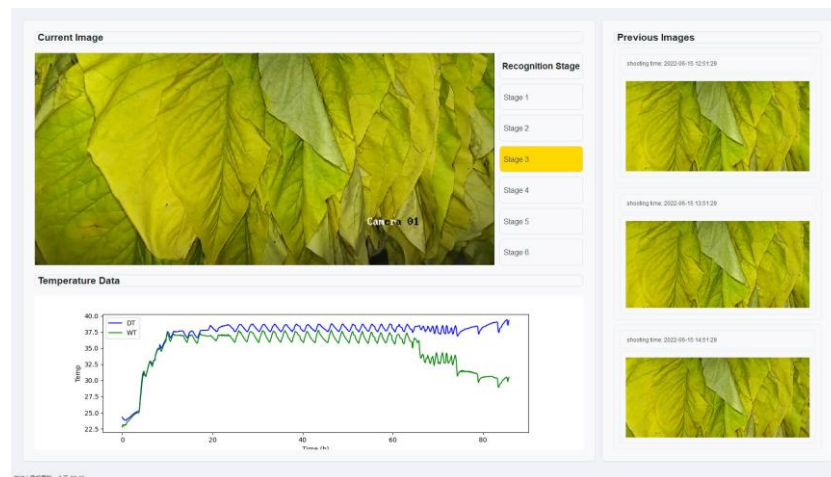


Fig. 8 - Real-time interface displaying tobacco curing state recognition result.

CONCLUSIONS

This study investigates the effectiveness of leveraging sequential data and hybrid models to recognize the states of tobacco curing. To address the challenges posed by sequential data and multiple features, a hybrid model based on TCN and LSTM was developed, demonstrating an effective ability to capture intricate features from data collected in the curing barn. Furthermore, the introduction of the DCRU module provides additional improvements in accuracy and robustness, underscoring the potential of the proposed approach in complex prediction tasks.

Experiments were conducted on real data from tobacco stations and the results were compared with other models. The experimental results demonstrate that the ACC of the hybrid TCN-LSTM model reached 0.989, demonstrating superior performance compared to traditional deep learning methods and existing tobacco leaf curing prediction models. At the same time, the effectiveness of the model was validated during practical testing. These findings indicate that the proposed tobacco curing states recognition model can effectively extract meaningful features from the complex data of the curing room and tobacco leaves. Compared to traditional methods, the proposed approach exhibits a significant advantage in recognition accuracy, offering high practical value and contributing to the automation and standardization of tobacco leaf curing.

Future research could explore the use of multi-source data, such as the concentration of chemicals inside the curing room and the humidity at the ventilation outlets of the curing room. Additionally, the potential applications of the model across different domains, such as the drying of fruits, vegetables, and herbs, could be further investigated.

ACKNOWLEDGEMENT

This work was funded by Hangzhou Dianzi University Science and Technology Project (KYH043122055).

REFERENCES

- [1] Abubakar Y., Young J.H. & Johnson, W. (2000). Changes in moisture and chemical composition of flue-cured tobacco during curing (烤烟在烘烤过程中水分和化学成分的变化). *Chinese Tobacco Science*, Vol. 122 1, pp. 51-58. <https://doi.org/10.3381/0082-4623-44.1.51>.
- [2] Ahsan, A., Handika, R., Qureshi, F., Martokoesoemo, D.R.S., Hindriyani, M., Amalia, N. & Mariz, K. (2022). Does tobacco affect economy? *Asian Pacific Journal of Cancer Prevention*, Vol. 23, pp. 1873.
- [3] Attri, I., Awasthi, L.K., Sharma, T.P. & Rathee, P. (2023). A review of deep learning techniques used in agriculture. *Ecological Informatics*, Vol. 77, pp. 102217. <https://doi.org/10.1016/j.ecoinf.2023.102217>.
- [4] Bai, S., Kolter, J.Z. & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. <https://doi.org/10.48550/arXiv.1803.01271>.
- [5] Cheng, W., Wang, Y., Peng, Z., Ren, X., Shuai, Y., Zang, S., Liu, H., Cheng, H. & Wu, J. (2021). High-efficiency chaotic time series prediction based on time convolution neural network. *Chaos, Solitons &*

- Fractals*, Vol. 152, pp. 111304. <https://doi.org/10.1016/j.chaos.2021.111304>.
- [6] Fan, J., Zhang, K., Huang, Y., Zhu, Y. & Chen, B. (2023). Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Computing and Applications*, Vol. 35, pp. 13109-13118. <https://doi.org/10.1007/s00521-021-05958-z>.
- [7] Graves, A. & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, Vol. 18, pp. 602-610. <https://doi.org/10.1016/j.neunet.2005.06.042>.
- [8] Han, L., Chen, X.Y., Ye, H.J. & Zhan, D.C. (2024). Softs: Efficient multivariate time series forecasting with series-core fusion. *Advances in Neural Information Processing Systems*. Vol. 37, pp. 64145-64175. <https://doi.org/10.48550/arXiv.2404.14197>.
- [9] Haralick, R.M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, Vol. 67, pp. 786-804. <https://doi.org/10.1109/PROC.1979.11328>.
- [10] Li, Z., MENG, L., WANG, S., GAO, J., XU, X., ZHU, X., YANG, C., WANG, B., WANG, A., MENG, L., LIU, Z., DU, H., LIU, H. & SUN, F. (2022). Selection of optimum discriminant model in tobacco curing stage based on image processing(基于图像处理的烟叶烘烤阶段判别模型优选). *Acta Tabacaria Sinica*, Vol. 28, pp. 65-76. <https://doi.org/10.16472/j.chinatobacco.2021.178>.
- [11] Liu, S., Xu, T., Du, X., Zhang, Y. & Wu, J. (2024). A hybrid deep learning model based on parallel architecture TCN-LSTM with Savitzky-Golay filter for wind power prediction. *Energy Conversion and Management*, Vol. 302, pp. 118122. <https://doi.org/10.1016/j.enconman.2024.118122>.
- [12] Meng, Y., Xu, Q., Chen, G., Liu, J., Zhou, S., Zhang, Y., Wang, A., Wang, J., Yan, D., Cai, X., Li, J., Chen, X., Li, Q., Zeng, Q., Guo, W. & Wang, Y. (2024). Regression prediction of tobacco chemical components during curing based on color quantification and machine learning. *Scientific Reports*, Vol. 14, pp. 27080. <https://doi.org/10.1038/s41598-024-78426-y>.
- [13] Pang, S., Jia, J., Xin, X., Zhang, Y., Chen, C., Wang, L. & Ding, X. (2024). Enhancing prior-knowledge-based control with deep attention neural network for outlet moisture content of cut-tobacco. *Drying Technology*, Vol. 42, pp. 1650-1663. <https://doi.org/10.1080/07373937.2024.2377725>.
- [14] Pei, W., Zhou, P., Huang, J., Sun, G. & Liu, J. (2024). State recognition and temperature rise time prediction of tobacco curing using multi-sensor data-fusion method based on feature impact factor. *Expert Systems with Applications*, Vol. 237, pp. 121591. <https://doi.org/10.1016/j.eswa.2023.121591>.
- [15] Song, X., Cheng, Z., Guo, W. & Cao, S. 2024. Recognition of tobacco yellowing degree in curing process based on deep learning. In: Proc. 2024 4th Int. Conf. Consum. Electron. Comput. Eng., Guangzhou, China. 348-352. <https://doi.org/10.1109/ICCECE61317.2024.10504148>
- [16] Wang, H., Jiang, H., Xu, C., Wang, D., Yang, C. & Wang, B. (2016). Three-Stage Six-Step flue-curing technology for Virginia tobacco leaves and Its application in China. *Journal of Agricultural Science and Technology A*, Vol. 6, pp. 232-238. <https://doi.org/10.17265/2161-6256/2016.04.003>.
- [17] Wang, L., Cheng, B., Li, Z., Liu, T. & Li, J. (2017). Intelligent tobacco flue-curing method based on leaf texture feature analysis. *Optik*, Vol. 150, pp. 117-130. <https://doi.org/10.1016/j.ijleo.2017.09.088>.
- [18] Wang, Y. & Qin, L. (2022). Research on state prediction method of tobacco curing process based on model fusion. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 13, pp. 2951-2961. <https://doi.org/10.1007/s12652-021-03129-5>.
- [19] Wu, J. (2016). Recognition of Tobacco Flue-Curing Phases Based on Image Features and GA-SVM Algorithm. *Journal of Southwest China Normal University(Natural Science Edition)*, Vol. 41, pp. 100-106. <https://doi.org/10.13718/j.cnki.xsxb.2016.09.017>.
- [20] Wu, J. & Yang, S.X. (2019). Intelligent control of bulk tobacco curing schedule using LS-SVM- and ANFIS-Based multi-sensor data fusion approaches. *Sensors*, Vol. 19, pp., <https://doi.org/10.3390/s19081778>.
- [21] Wu, J., Yang, S.X. & Tian, F. (2014). A novel intelligent control system for flue-curing barns based on real-time image features. *Biosystems Engineering*, Vol. 123, pp. 77-90. <https://doi.org/10.1016/j.biosystemseng.2014.05.008>.
- [22] Wu, J., Yang, S.X. & Tian, F. (2017). An adaptive neuro-fuzzy approach to bulk tobacco flue-curing control process. *Drying Technology*, Vol. 35, pp. 465-477. <https://doi.org/10.1080/07373937.2016.1183211>.
- Xiong, J., Hou, Y., Wang, H., Tang, K., Liao, K., Yao, Y., Liu, L. & Zhang, Y. (2024). Research on the Recognition Method of Tobacco Flue-Curing State Based on Bulk Curing Barn Environment. Vol. 14, pp. 2347. <https://doi.org/10.3390/agronomy14102347>.