Deep Learning-Enhanced Prediction of Small Intestinal Bleeding Points Using Long Short-Term Memory Networks: Supplementary Material

S1. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a powerful deep learning model capable of processing millions of parameters. By taking 2D images as input and applying convolution operations with filters/kernels, CNNs significantly reduce computational costs. CNNs are widely applied in various domains, including facial recognition, biometric systems, autonomous driving, emotion detection, image restoration, and robotics [1]. The structure of a CNN is illustrated in Figure S1, with each layer explained below:

Convolutional Layer

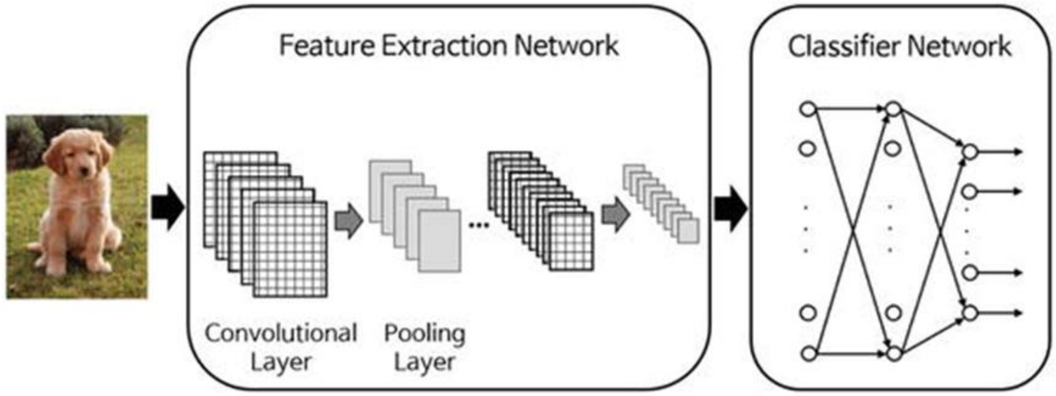
As illustrated in Figure S2 [2], the convolutional operation starts at the top-left corner and progresses sequentially from left to right using specific feature detectors (filters). The obtained feature maps capture distinctive patterns, as shown in Figure S2. In most deep learning models, the feature extraction stage utilizes this principle to extract and train various feature maps from images.

Pooling Layer

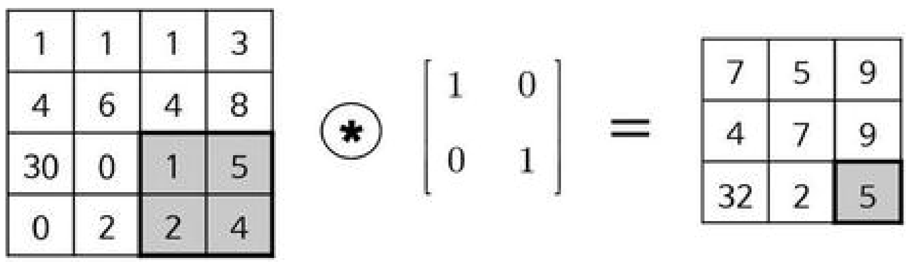
The pooling layer in a CNN serves to reduce the dimensionality of feature maps, thereby alleviating computational costs. The most commonly used pooling method is max pooling, which selects the maximum value from each region as its representative. Other pooling methods, such as average pooling, are also utilized. These operations help in downsampling feature maps while preserving crucial features, enhancing the efficiency of subsequent processing steps.

Fully Connected Layer

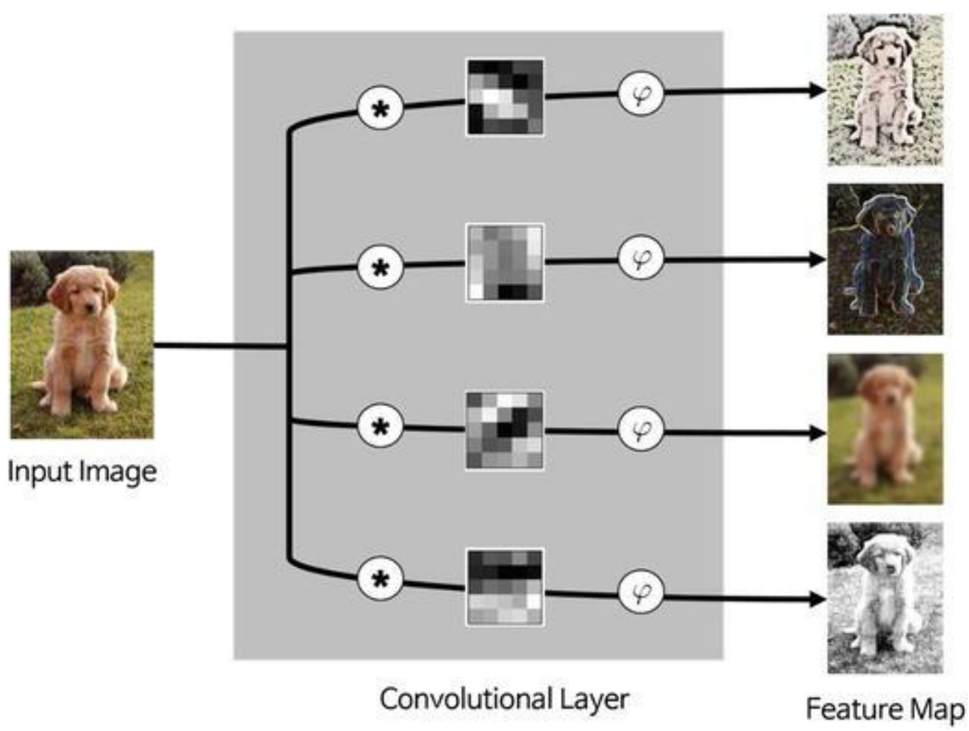
In the fully connected layer of a CNN, the output of the preceding layers is flattened into a one-dimensional vector. This vector is then fed into a neural network for training. During training, the network adjusts its weights based on the input image data to optimize classification performance. By fine-tuning the parameters of this layer, the predictive capability of the model can be effectively enhanced. The fully connected layer integrates and transforms the extracted features from previous layers to capture higher-level patterns, ultimately improving classification accuracy.



Supplementary Figure 1 Schematic diagram of CNN structure [2]



Supplementary Figure 2 Schematic diagram of CNN convolution operation [2]

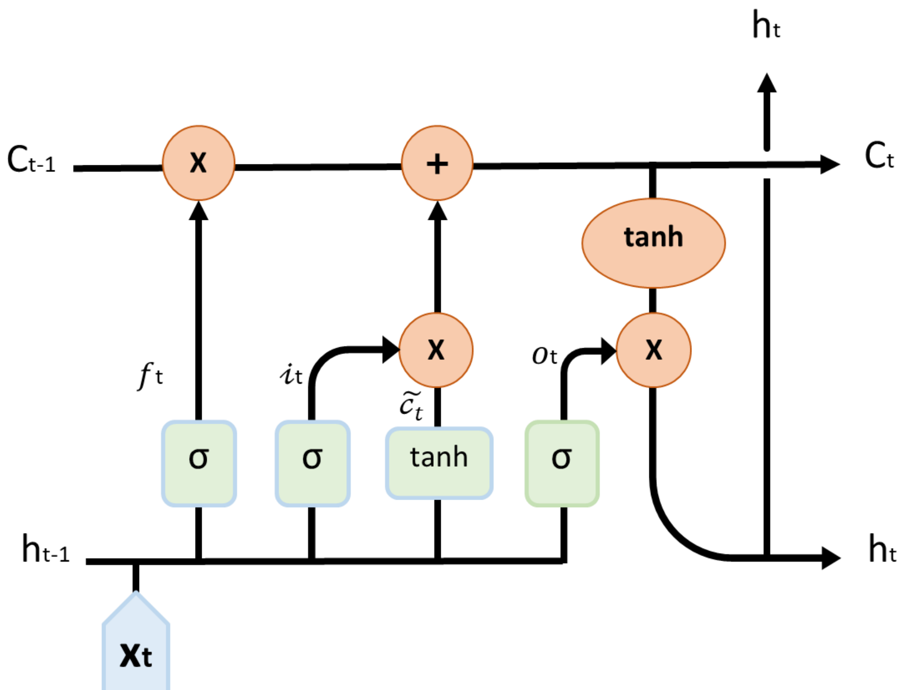


Supplementary Figure 3 CNN convolutional layer feature extraction flow chart [2]

S2. Long Short-Term Memory Network (LSTM)

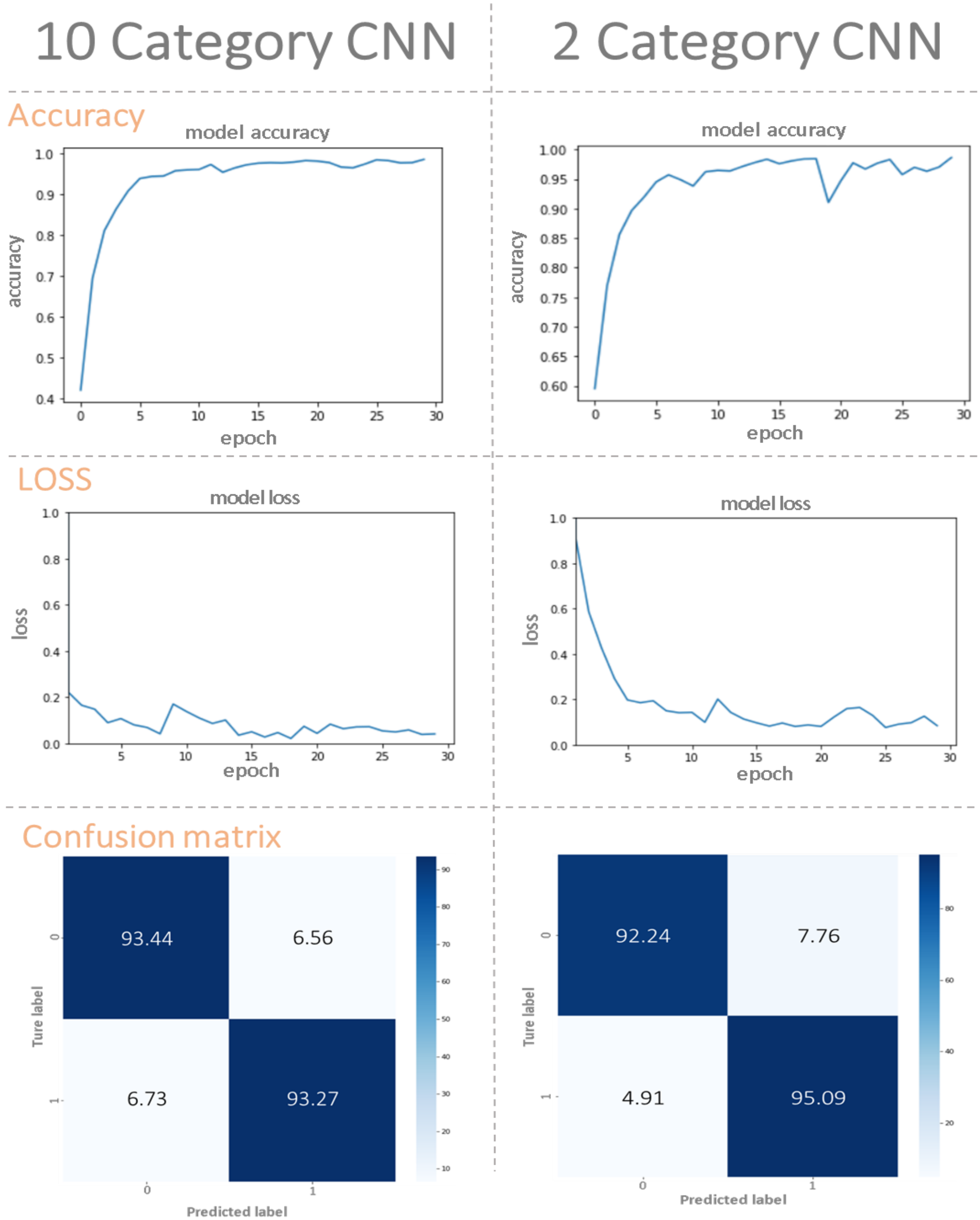
Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural networks (RNNs) designed to learn long-term dependencies. These networks effectively address issues such as vanishing and exploding gradients in traditional RNNs by utilizing forget gates (sigmoid functions) to retain only relevant information.

LSTM networks function similarly to information conveyors. As shown in Figure S4, their structure facilitates minor linear interactions, allowing information to propagate across states. Various arrows in the structure represent gates that interact with cell states, filtering relevant information for transmission.

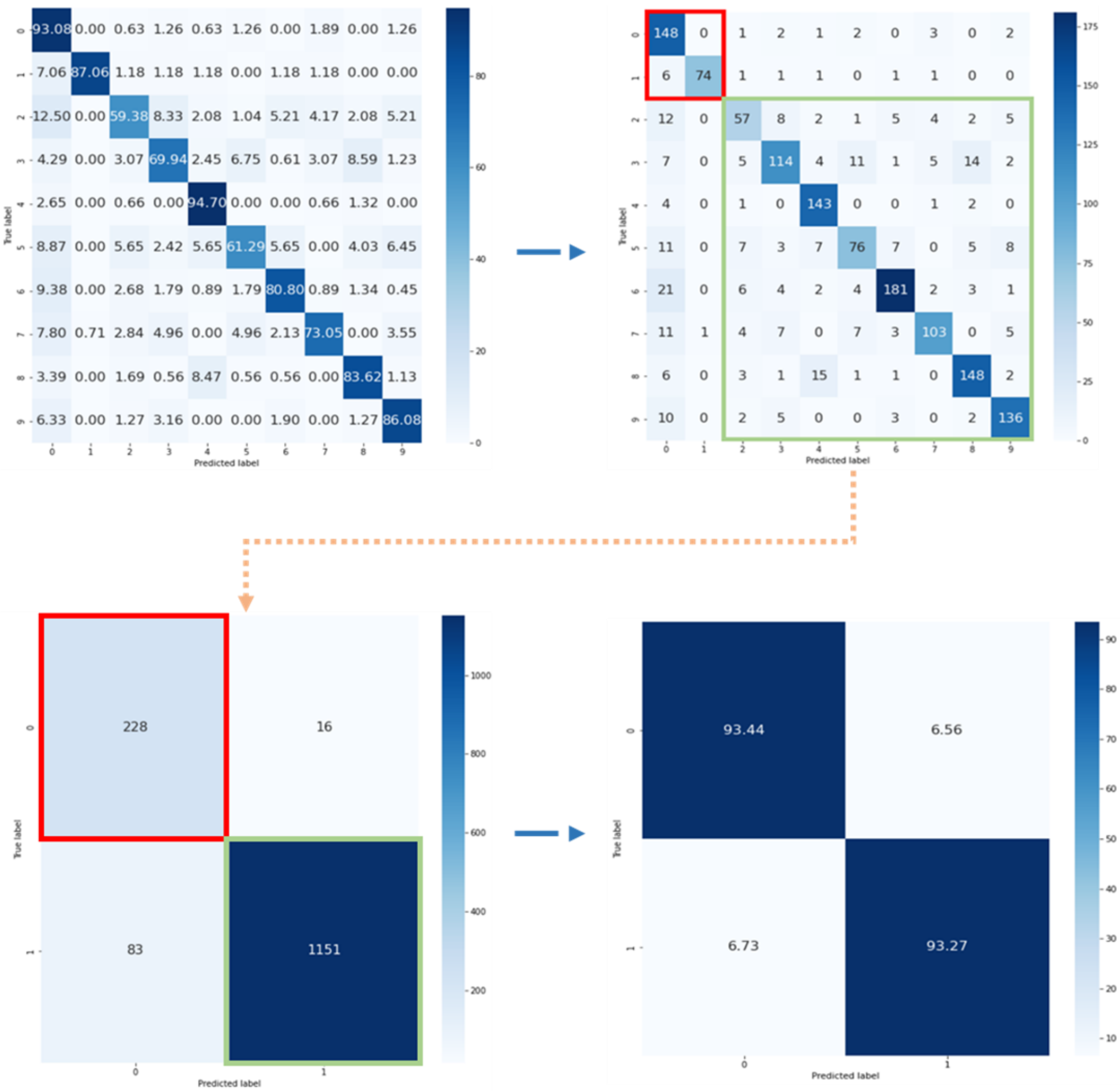


Supplementary Figure 4 LSTM interaction layer architecture [3,4].

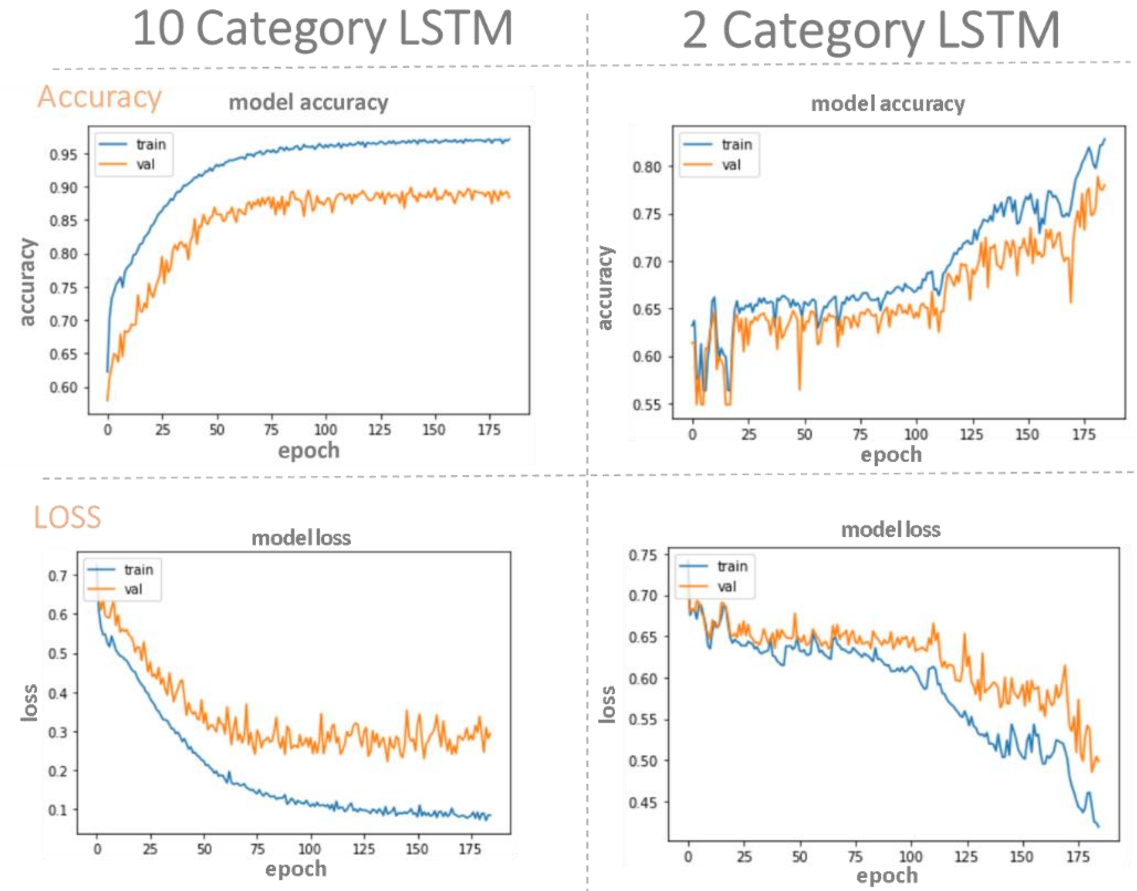
S3. Training Data for CNN and LSTM



Supplementary Figure 5 CNN training data.



Supplementary Figure 6 Confusion matrix 10 categories to 2 categories.



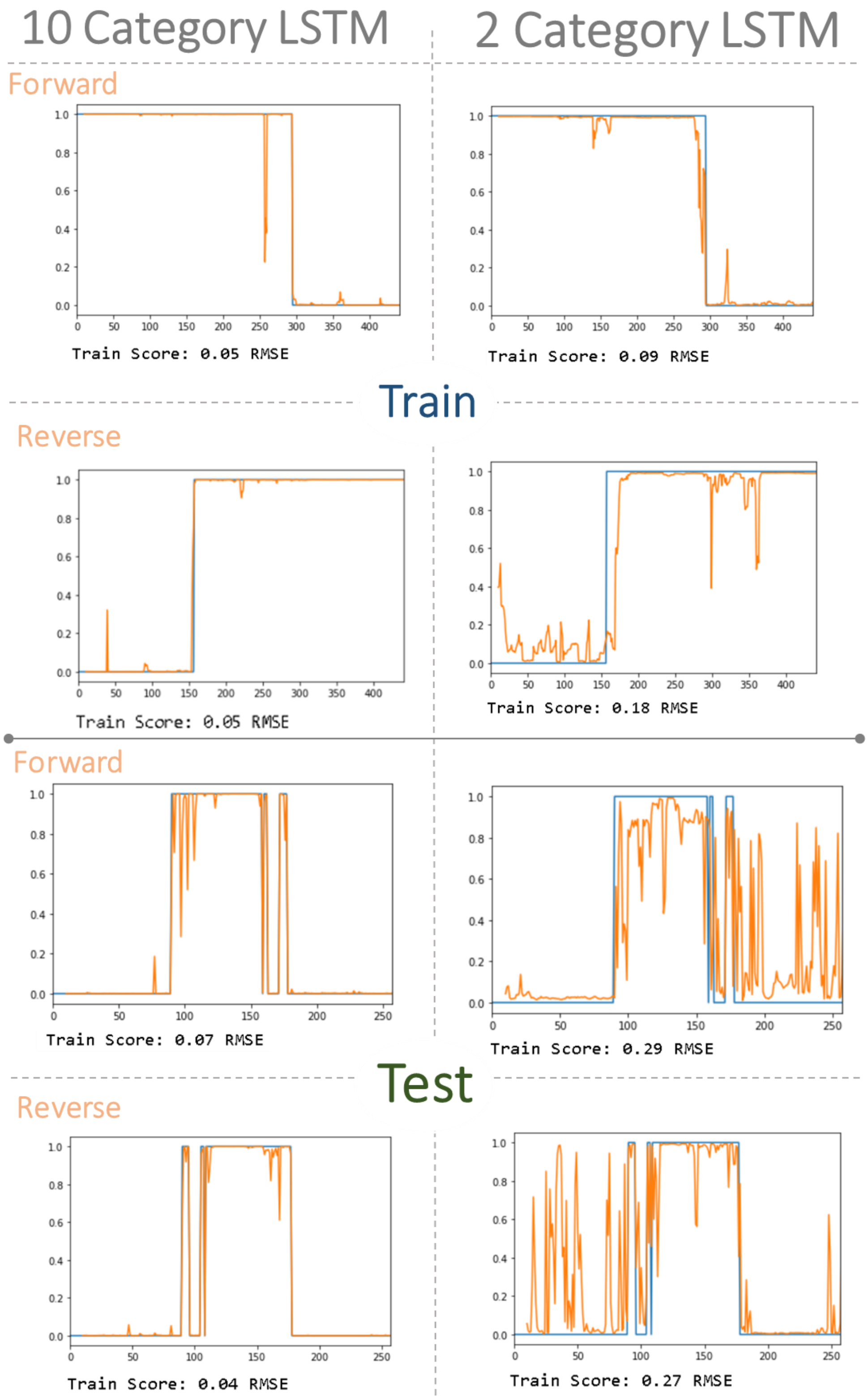
Supplementary Figure 7 LSTM training data.

Supplementary Table 1 CNN Performance Metrics for 10-class and 2-class Configurations

Model	Configuration	Accuracy (%)	Loss (%)
CNN	10-class	98.6	5.45
CNN	2-class	96.7	8.32

Supplementary Table 2 LSTM Performance Metrics for 10-class and 2-class Configurations

Model	Configuration	Training Accuracy (%)	Testing Accuracy (%)	Training Loss (%)	Testing Loss (%)	RMSE (Train)	RMSE (Test)
LSTM	10-class	96.2	87.1	9.8	31.2	0.05 (F), 0.05 (B)	0.07 (F), 0.04 (B)
LSTM	2-class	82.6	76.7	41.3	52.6	0.09 (F), 0.18 (B)	0.29 (F), 0.27 (B)



Supplementary Figure 8 Comparison of LSTM prediction performance for 10-category and 2-category configurations in both training and testing phases. The plots illustrate forward and reverse sequence predictions for capsule endoscopy frames. The blue line represents the ground-truth labels, while the orange line shows the LSTM model's predicted probability sequence. For the 10-category LSTM, forward training and testing achieved low errors with RMSE values of 0.05 and 0.07, respectively, while reverse predictions yielded RMSE values of 0.05 (train) and 0.04 (test). In contrast, the 2-category LSTM demonstrated higher variability, with forward RMSE values of 0.09 (train) and 0.29 (test), and reverse RMSE values of 0.18 (train) and 0.27 (test). These results highlight the superior stability and predictive accuracy of the 10-category model compared to the binary configuration, especially in test sequences.

References

1. R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," Proceedings of the First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282. DOI: 10.1109/ICSCCC.2018.8703316.

2. Kim, P., Convolutional Neural Network, in MATLAB Deep Learning. Springer, 2017, pp. 121-147.

3. B. Nath Saha and A. Senapati, "Long Short-Term Memory (LSTM) Based Deep Learning for Sentiment Analysis of English and Spanish Data," 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2020, pp. 442-446. DOI: 10.1109/ComPE49325.2020.9200054.

4. M. Kaur and A. Mohta, "A Review of Deep Learning with Recurrent Neural Network," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 460-465. DOI: 10.1109/ICSSIT46314.2019.8987837.