



OPEN ACCESS

EDITED BY

Rajesh Kumar Tripathy,
Birla Institute of Technology and
Science, India

REVIEWED BY

Kyungsang Kim,
Harvard Medical School, United States
Guang Han,
Tianjin Polytechnic University, China

*CORRESPONDENCE

Ming Yu,
✉ yuming_1990@outlook.com
Hang Wu,
✉ wuhang1991@nankai.edu.cn

RECEIVED 21 November 2024

ACCEPTED 14 May 2025

PUBLISHED 27 May 2025

CITATION

Qiang X, Wang Q, Liu G, Song L, Zhou W, Yu M
and Wu H (2025) Use video comprehension
technology to diagnose ultrasound
pneumothorax like a doctor would.
Front. Physiol. 16:1530808.
doi: 10.3389/fphys.2025.1530808

COPYRIGHT

© 2025 Qiang, Wang, Liu, Song, Zhou, Yu and
Wu. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Use video comprehension technology to diagnose ultrasound pneumothorax like a doctor would

Xiaoyong Qiang¹, Qiang Wang¹, Guanjun Liu², Limei Song²,
Weibin Zhou¹, Ming Yu^{2*} and Hang Wu^{2,3*}

¹College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China, ²Systems Engineering Institute, Academy of Military Sciences, People's Liberation Army, Tianjin, China, ³School of Artificial Intelligence, Nankai University, Tianjin, China

Introduction: Emergency rescue scenes and pre-hospital emergency stages commonly encounter trauma victims. Life-saving measures must be taken at the scene if a trauma patient has pneumothorax; if the patient is not evaluated and diagnosed right away, their life may be in jeopardy. Ultrasound, which has the advantages of being non-invasive, non-radioactive, portable, rapid, and repeatable, can be used to diagnose pneumothorax. However, those who interpret ultrasound images must undergo extensive, specialized, and rigorous training. Deep learning technology allows for the intelligent diagnosis of ultrasound images, allowing general healthcare professionals to quickly and with minimal training diagnose pneumothorax in lung ultrasound patients.

Methods: Previous studies focused primarily on the lung-sliding characteristics of M-mode images, neglecting other key features in lung ultrasonography pneumothorax, and used similar technological techniques. Our study team used video understanding technology for medical ultrasound imaging diagnostics, training the TSM video understanding model on the ResNet50 network with 657 clips and testing the model with untrained 164 lung ultrasound clips.

Results: The model's sensitivity was 99.21%, specificity was 89.19%, and average accuracy was 96.95%. The F1 score was 0.929, and the AUC was 0.97.

Discussion: This study is the first to apply video understanding models to the multi-feature fusion diagnosis of pneumothorax, demonstrating the feasibility of using video understanding technology in medical image diagnosis.

KEYWORDS

pneumothorax, lung ultrasound, medical imaging, intelligent diagnostics, deep learning, video understanding

1 Introduction

Trauma is one of the most prevalent reasons for emergency room visits, as well as a leading cause of death (Qian and Zhang, 2019). Pneumothorax (PTX), or lung collapse, is a life-threatening respiratory emergency that can occur in trauma patients as well as individuals with acute and chronic diseases

(Duclos et al., 2019; Weissman and Agrawal, 2021) and must be treated immediately (Lichtenstein, 2015). Ultrasonography has the virtue of being quick, repeatable, and noninvasive, and it has become an essential injury detection tool in the emergency room (Valenzuela et al., 2020). PTX can be recognized early and promptly using the E-FAST (Extended Focused Assessment with Sonography in Trauma) technique and treated symptomatically (Hefny et al., 2019). Despite the fact that lung ultrasound (LUS) has higher diagnostic accuracy than chest X-ray for PTX (Nagarsheth and Kurek, 2011; Alrajhi et al., 2012) and offers additional benefits, a lack of training opportunities and clinician-centered workflows continue to impede widespread use (Brady et al., 2021).

AI-powered ultrasound interpretation eliminates training and workflow bottlenecks, allowing non-specialists to use the technology to provide fast, portable imaging on a large scale (VanBerlo et al., 2022). This study is built on this objective, and it uses deep learning networks to learn and recognize pneumothorax features in ultrasound images, as well as to achieve automatic recognition and diagnosis of pneumothorax ultrasound images. An auxiliary tool is supplied to healthcare workers to ease the process of interpreting ultrasound images, shorten the learning cycle, and lower the barrier to using ultrasound equipment. Our goal is for even general healthcare workers with basic training to be able to operate ultrasound equipment proficiently for pneumothorax diagnosis and obtain diagnostic accuracy comparable to that of specialized imaging physicians. This will not only help improve accessibility and efficiency of healthcare but will also improve patient outcomes by providing rapid and reliable diagnostic information in emergencies.

Because PTX assessment is a critical component in the identification of complex life-threatening patient situations such as trauma (Blaivas et al., 2005), cardiac arrest (Hernandez et al., 2008), and respiratory distress (Wallbridge et al., 2018), detecting the existence of PTX is a crucial component of lung ultrasonography. At the time of writing, there is a limited amount of relevant literature on the diagnosis of pneumothorax by AI ultrasound, and no publicly available datasets on ultrasound pneumothorax have been collected. The majority of studies on PTX are based on publicly available CT chest radiograph datasets or X-ray chest radiograph datasets, with a few based on private ultrasound datasets (animal experimental data, simulated material data, real-life ultrasound data).

Some recognized research claims to have produced relatively good results for automated PTX detection in animal ultrasonography investigations (Mehanian et al., 2019; Kulhare et al., 2018; Summers et al., 2017; Lindsey et al., 2019), but human accuracy is unknown. Boice et al. (2022) used synthetic gelatin models cast in 3D-printed rib molds and a simulated lung. M-mode ultrasound pneumothorax simulation images were obtained and used to train the pneumothorax detection model, which was tested using animal ultrasound data and ended up with an overall accuracy of 93.6%. Jaščur et al. (2021) employed a convolutional neural network (CNN) model on a restricted dataset with 82% sensitivity to missing lung sliding. VanBerlo et al. (2022) used a large labeled LUS dataset from two academic institutions to classify images after converting B-mode movies to M-mode images, and the model had a sensitivity of 93.5% to lung sliding, specificity of 87.3%, and AUC of 0.973. Kim et al. (2023) proposed an AI-assisted pneumothorax diagnosis framework that simulates

clinical workflows through sequential steps, including pleural line localization under B-mode ultrasound, B-to-M-mode image reconstruction, and lung sliding detection. Utilizing lightweight models (<3 million parameters), they achieved pleural line quality assessment (Dice coefficient: 89%) and sliding classification individually, with single-model AUCs exceeding 95% and an overall workflow AUC of 89%.

However, Lichtenstein and Menu (Lichtenstein and Menu, 1995) clearly stated that although the disappearance of lung sliding is observed in 100% of PTX cases, the disappearance of lung sliding may have other causes. Additionally, the BLUE procedure (Lichtenstein, 2015), which is commonly used by doctors, does not validate the diagnosis of PTX based just on the absence of lung sliding; other diagnostic techniques, such as lung point detection, are required to validate the diagnosis. Thus, the binary classification of seashore/barcode signs in M-mode images, as well as the classification of visceral and parietal pleural motion in B-mode videos, is not pneumothorax detection, but rather a test for the presence or absence of lung-sliding motion (Jaščur et al., 2021).

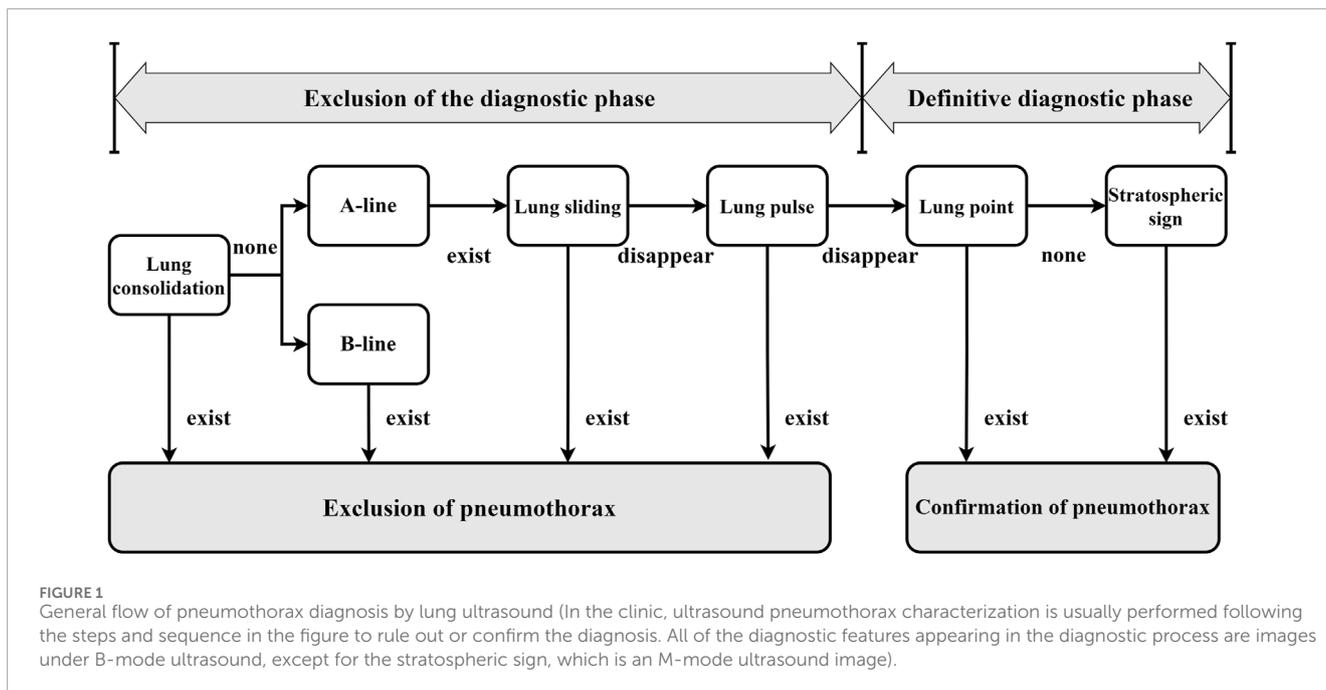
Currently, the authors' known ultrasound pneumothorax studies primarily determine the presence of PTX by detecting the presence or absence of lung sliding, whereas our team, after discussion, concluded that using the absence of lung sliding as the sole basis for determining the positivity of PTX was insufficient. Therefore, after conducting the research and discussion, we attempted to introduce video understanding technology and applied it for the first time in the detection of pneumothorax through ultrasound. By learning the sliding characteristics of the lungs in pneumothorax ultrasound and other related features, we were able to classify the ultrasound video clips of pneumothorax and non-pneumothorax, thereby achieving the goal of pneumothorax detection. Subsequent experiments have proved that this cross-domain application has certain research value in achieving multi-feature detection of medical images.

2 Methods and materials

2.1 Selection of PTX diagnostic features

To accurately diagnose lung ultrasound pneumothorax, our team initially investigated the ultrasound pneumothorax diagnostic procedure. The process of diagnosing lung ultrasound pneumothorax (Zhao et al., 2023a) is divided into two stages. ① Diagnostic exclusion stage: Sweep both lungs one by one at each intercostal gap, check ultrasonography signs in the order of "solid lung lesion → B-line → pleural sliding sign → pleural pulsation sign," and exclude pneumothorax if one of the signs is present. ② Definitive diagnosis stage: If none of the above four indicators exist, use the "lung point sign → stratospheric sign" in order to check for the presence of one or both signals, which can indicate pneumothorax. Refer to Figure 1 for the specific approach.

The diagnostic accuracy of the lung ultrasound pneumothorax diagnostic procedure for PTX was 99.1%, sensitivity was 97.8%, and specificity was 100.0% (Zhao et al., 2023b). There is no pleural slide or pulsation symptom because PTX occurs in only one layer of wall pleura. As a result, the pleural pulsation sign is an essential sign in ruling out PTX and is classified as grade A evidence in the lung ultrasonography evidence-based guidelines (Zhao et al.,



2023b). In this study, we feel that when a pleural sliding sign or a pleural pulsation sign is present, we can rule out the potential of a positive PTX and identify it as a negative PTX. As a result, we chose pleural sliding and pleural pulsation signs as crucial criteria for diagnosing negative pneumothorax. When pleural sliding and pulsation disappeared, as well as the existence of lung points or other signs (in clinically verified pneumothorax-positive patients), we used the aforementioned characteristics to diagnose a positive pneumothorax.

2.2 Database establishment

This study's database was created using real clinical data. We formed a collaboration with two hospitals to collect video data from lung ultrasound cases (ultrasound video images from convex and line array probes) accumulated over the previous few years, and all positive and negative instances have been clinically identified and confirmed. We rigorously filtered these clinical data using pre-selected diagnostic features of pneumothorax-negative and pneumothorax-positive to retrieve genuine data that satisfied the study requirements. Given the limited number of clinical cases, we used a data segmentation approach to increase the size of the dataset.

A healthy adult breathes approximately 12–20 times per minute (Wan and Chen, 2015). As a result, we can estimate that each breath lasts three to 5 seconds. Based on this concept, when processing lung ultrasound video clips, we used a 5-second sliding window for segmentation with a 3-second time interval to trim numerous 5-second video clips. As a result, each video segment used for deep learning network training is 5 s long and can ensure that at least one breathing cycle is included in each data segment, allowing for pneumothorax diagnosis. Because the time interval is set to 3 s, there is a 2-second overlap between the two adjacent video clips before and after, but their contents are not exactly the same. This

strategy decreases the risk of overfitting while maximizing training data, effectively increasing the amount of data available to us.

By segmenting the raw video data, we were able to collect additional training samples, which improved the model's training and diagnosis accuracy. After the data segmentation was done, we cleansed the segmented data again to verify its usability and dependability in the research results. We invited two specialized sonographers with more than 8 years of experience in the field to independently screen and clean the data (one with 12 years of experience). They carefully analyzed each segmented video clip to confirm that both positive and negative PTX data utilized for training had acceptable diagnostic features, and they eliminated 17 positive invalid clips and 16 negative invalid clips. Following these data processing processes, we were able to create a dataset with 188 positive clips and 633 negative clips.

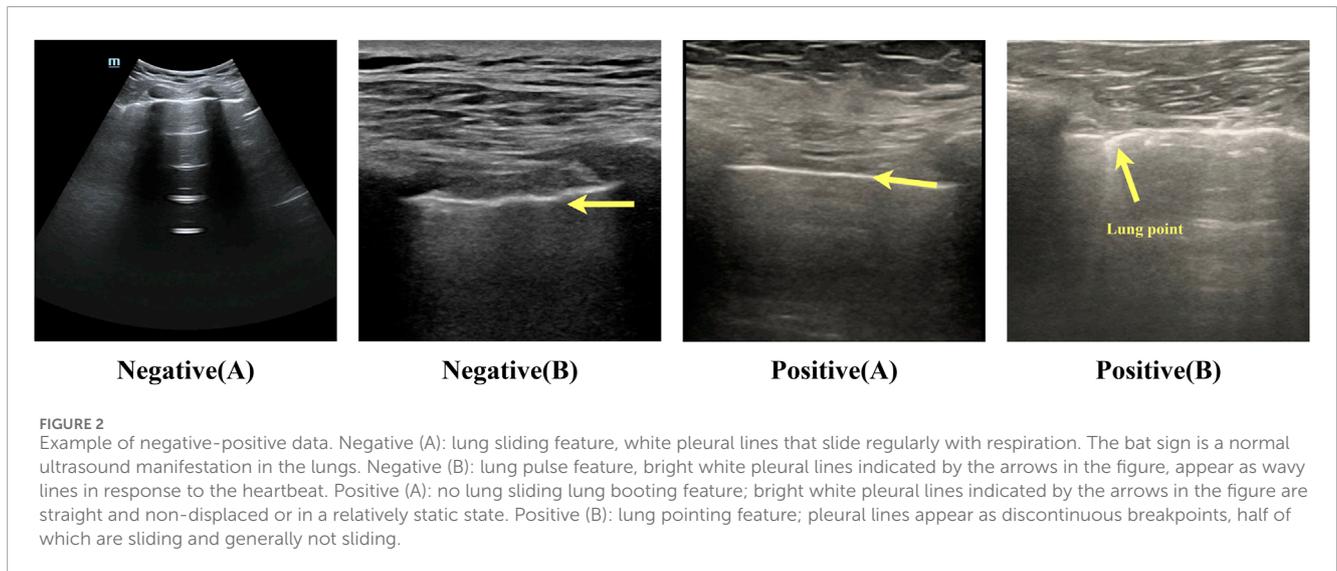
To ensure the model's training efficacy and generalization capacity, we partition the dataset into training, validation, and test sets in a 6:2:2 ratio for further model training and evaluation. This type of data split method allows the model to learn a wider range of characteristics during the training process while also testing its performance on the validation and test sets. Table 1 lists information about the data set. Figure 2 shows many examples of data. Figure 3 details the unique data processing flow, clearly showing each step from data collection to data cleaning to data segmentation.

2.3 Network structure

Over the years, deep learning has become the norm for video comprehension (Tran et al., 2015; Wang et al., 2016; Carreira and Zisserman, 2017; Wang et al., 2018; Zolfaghari et al., 2018; Xie et al., 2018; Zhou et al., 2018). One important difference between image and video recognition is the need for temporal modeling. For example, altering the order of opening and shutting a box yields the

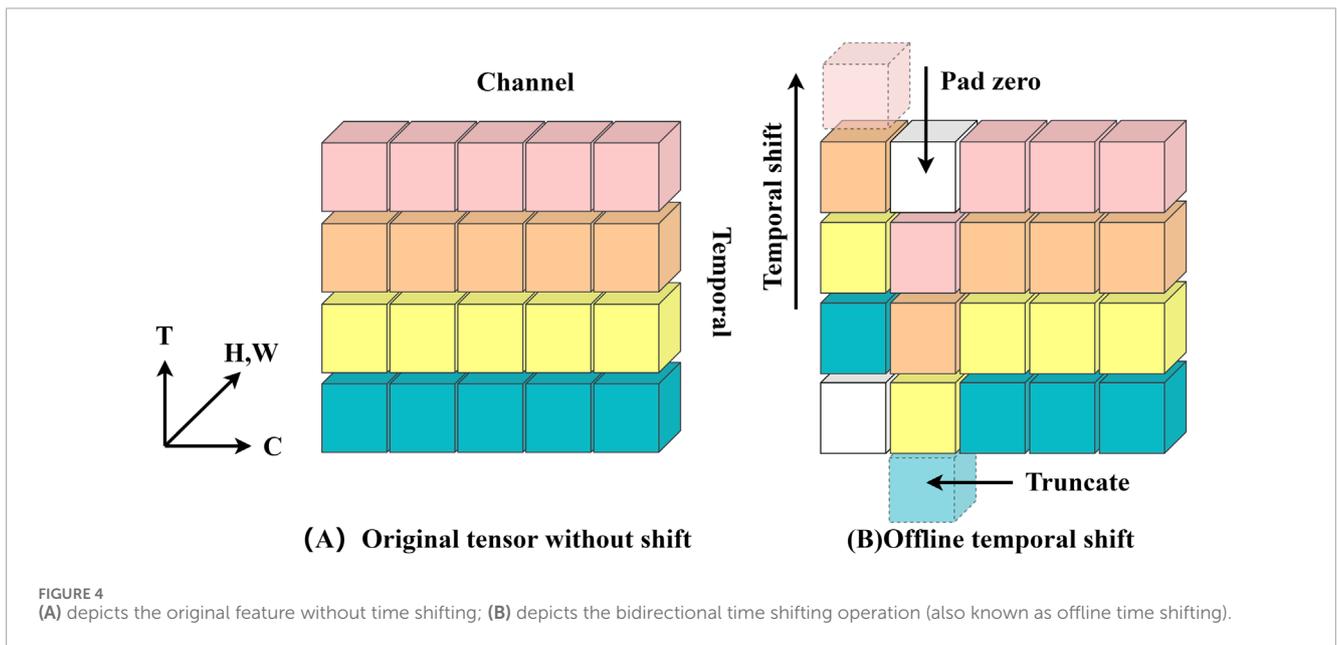
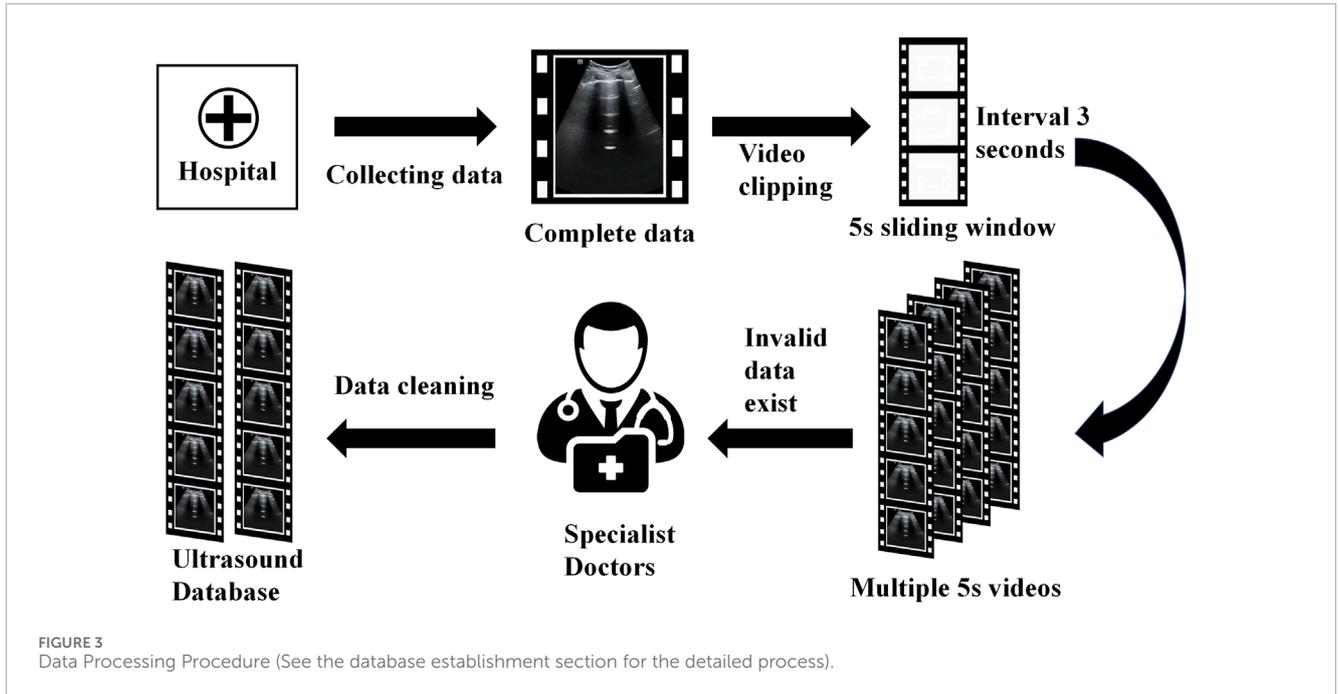
TABLE 1 Data Information. The table contains information about the dataset such as the number of datasets, the division ratio of the training validation test data, the features corresponding to the positive data and their number (Multiple features may exist for the same instance of data), the features corresponding to the negative data and their number (Multiple features may exist for the same instance of data) as well as the image resolution, the type of ultrasound probe used, the frame rate of the video, and the device information.

Clips label	Class	Training data (Train and Val)		Holdout data (Test)	
Number of clips	Negative	379 (train)	127 (val)	127(Test)	
	Positive	113 (Train)	38(Val)	37(Test)	
Number of features (There is a crossover of features)	Negative	B-line	Lung sliding	Lung pulse	Lung consolidation
		22	573	70	41
	Positive	A-line	Lung point	No Lung sliding or pulse	No lung consolidation
		136	53	161	1
Duration of data clips	All data segments have a duration of 5 s				
Total frames per clip	MAX: 62 frames*5 s; MIN: 16 frames*5 s				
Resolution (of a frame)	800*600; 1,200*900; 1960*910				
Transducers	Curved and Linear				
Machine Vendors	Mindray				



opposite effect, highlighting the importance of temporal modeling (Lin et al., 2019). TSM (Temporal Shift Module) is a generalized and effective time-shift module that may be based on an existing network model to incorporate the temporal shift module technique (the base model used in this paper is ResNet-50, and ResNet-50 is utilized by default in all subsequent sections). It has the performance of a 3D CNN while keeping the intricacy of a 2D CNN. Information can move across adjacent frames thanks to TSM's transmission of a piece of the channel along the time dimension. For temporal modeling, it may be added to a 2D CNN with no parameters and no processing (Lin et al., 2019).

The video model's activation is written as $A \in R^{N \times C \times T \times H \times W}$, where N is the batch size, C is the number of channels, T is the temporal dimension, and H and W are the spatial resolutions (Lin et al., 2019). Conventional 2D CNNs function independently in the T dimension and so do not exhibit temporal modeling effects, as shown in Figure 4A. The Time Shift Module (TSM), on the other hand, alternately transfers the channels along the time dimension. Following time-shifting, the information from nearby frames is combined with the information from the present frame, as seen in Figure 4B (Lin et al., 2019). To represent offline video recognition, a bidirectional TSM is used. Given a video V, the first T



frames F_i, F_1, \dots, F_T are sampled. After sampling the frames, the 2D CNN baseline processes each frame separately before averaging the logarithm of the outputs to generate the final prediction.

The model's data input consists of a batch of ten RGB images with a tensor form of $(10, 3, 224, 224)$, indicating that each image is 224×224 pixels with three color channels. Following a series of convolution, batch normalization, ReLU activation, and maximum pooling, the data successfully passes through the initial feature extraction stage and enters the first residual block, where it is reshaped to $(10, 64, 56, 56)$. The information fusion between frames is then realized by performing "right shift" and "left shift" operations

on this tensor, which replace the corresponding part of the first nine frames with one-eighth of the data chosen from each channel of the last nine frames, and then replacing the corresponding part of the last nine frames with one-eighth of the data of the first nine frames. This tactic creates the impression that "you are in me, I am in you" and improves the information exchange across frames. The residual block is then used to learn from the tensor following the shift. Notably, the model's sensory field doubles with each embedding of the time-shift module in the temporal dimension, much like when a convolution with a kernel size of three is applied to the time series. **Figure 5** illustrates the model's structure. Therefore, the

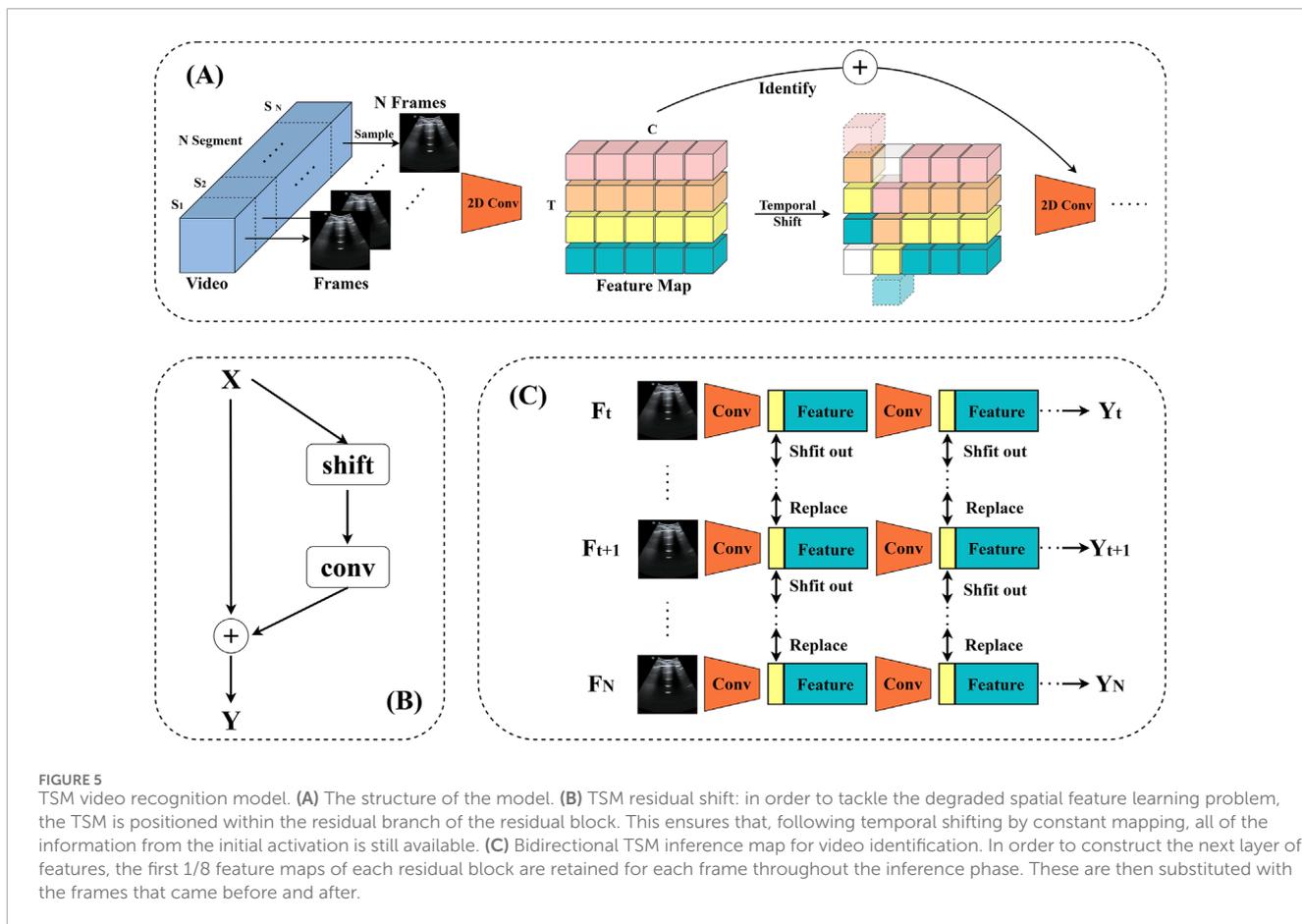


FIGURE 5 TSM video recognition model. **(A)** The structure of the model. **(B)** TSM residual shift: in order to tackle the degraded spatial feature learning problem, the TSM is positioned within the residual branch of the residual block. This ensures that, following temporal shifting by constant mapping, all of the information from the initial activation is still available. **(C)** Bidirectional TSM inference map for video identification. In order to construct the next layer of features, the first 1/8 feature maps of each residual block are retained for each frame throughout the inference phase. These are then substituted with the frames that came before and after.

model using TSM has a broad spatio-temporal sensory field and is capable of modeling fine spatio-temporal relationships, which is an extremely effective technological tool for video analytics and other tasks that require simultaneous consideration of spatial and temporal dimensions.

2.4 Model training framework

The training structure is depicted in Figure 6. The model used in the code is a TSM network structure based on a single RGB picture, and the backbone network is the traditional ResNet-50 design. Each training video is divided into 10 segments (each lasting 0.5 s) during the data preparation step. One frame is randomly selected from each segment (10 frames total), and the data is then fed into the training model following uniform processing (random scaling, cropping, flipping). The model initially extracts features from the input ultrasound image data using a convolutional neural network, which yields a rich feature representation. These features are then fed into a classifier, which calculates the probability distribution of each video feature class. A loss function is built during training based on the discrepancy between the model's output probability values and the real sample labels, and the model is then optimized. In order to reduce the discrepancy between the expected and actual values, this stage is essential to model learning. In the inference

stage, the model will produce the final prediction for the category with the highest probability, ensuring that ultrasound image data is accurately classified.

2.5 Training configuration and parameters

Training was performed on a server configured with Ubuntu 16.04 operating system, using a single GPU, model TITAN RTX. Model construction was based on the PaddlePaddle 2.6 framework. In order to improve the starting performance of the model, we used the ResNet-50 model weights pre-trained on the ImageNet dataset as initialization parameters for the backbone network.

We established the following parameters for the dataset's training configuration: A total of 100 training cycles (epochs) were carried out, with eight samples in each batch (batch size). To avoid overfitting, the dropout ratio was set to 0.8, and the optimizer's momentum parameter was set to 0.9. With an initial learning rate of 0.001, the decay strategy for the learning rate (LR) was set up as follows. It was carried out on a boundary of 10–20 epochs, and anytime the decay boundary was reached, the learning rate was decreased to 0.1 times its initial value. Additionally, we froze the Batch Normalization (BN) layer (Zolfaghari et al., 2018) and adjusted the pre-trained weights from ImageNet throughout model training to ensure stability.

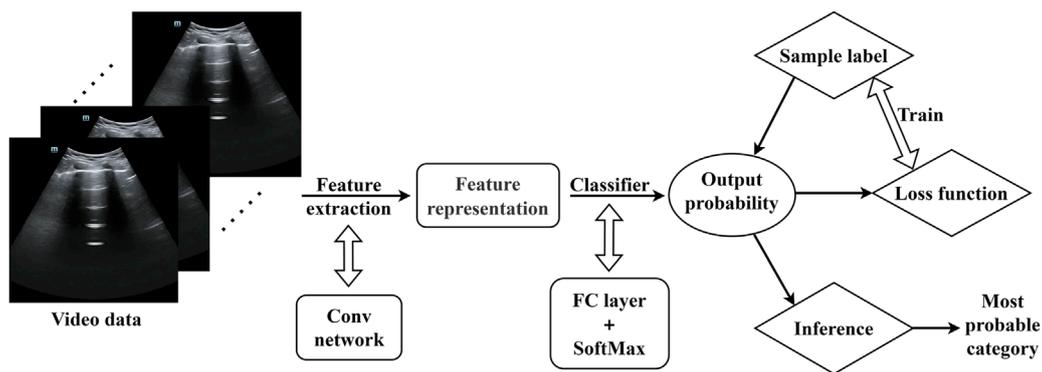


FIGURE 6 Model training framework (The video data were randomly scaled, cropped, and flipped before being entered into the model).

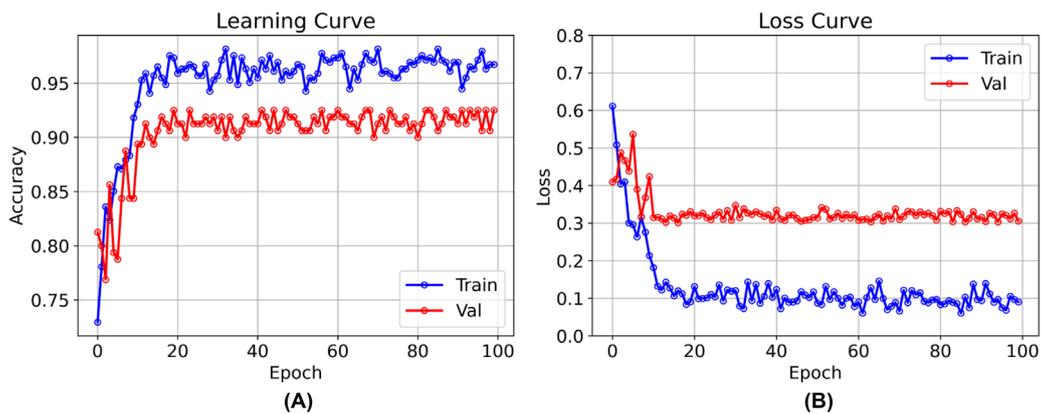


FIGURE 7 The training and validation curves after 100 epochs of training using the TSM-ResNet50 model [(A) shows the process of accuracy change, and (B) shows the process of loss value change].

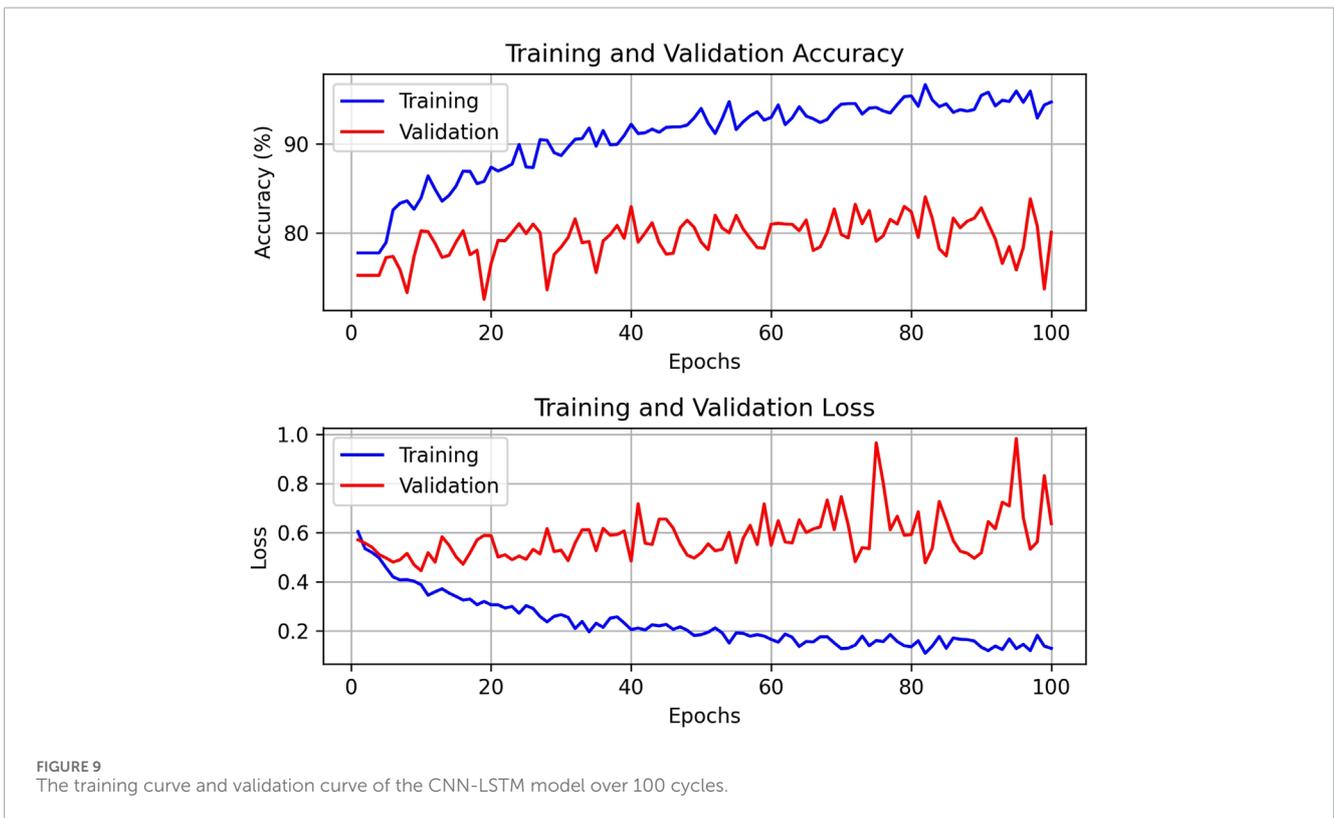
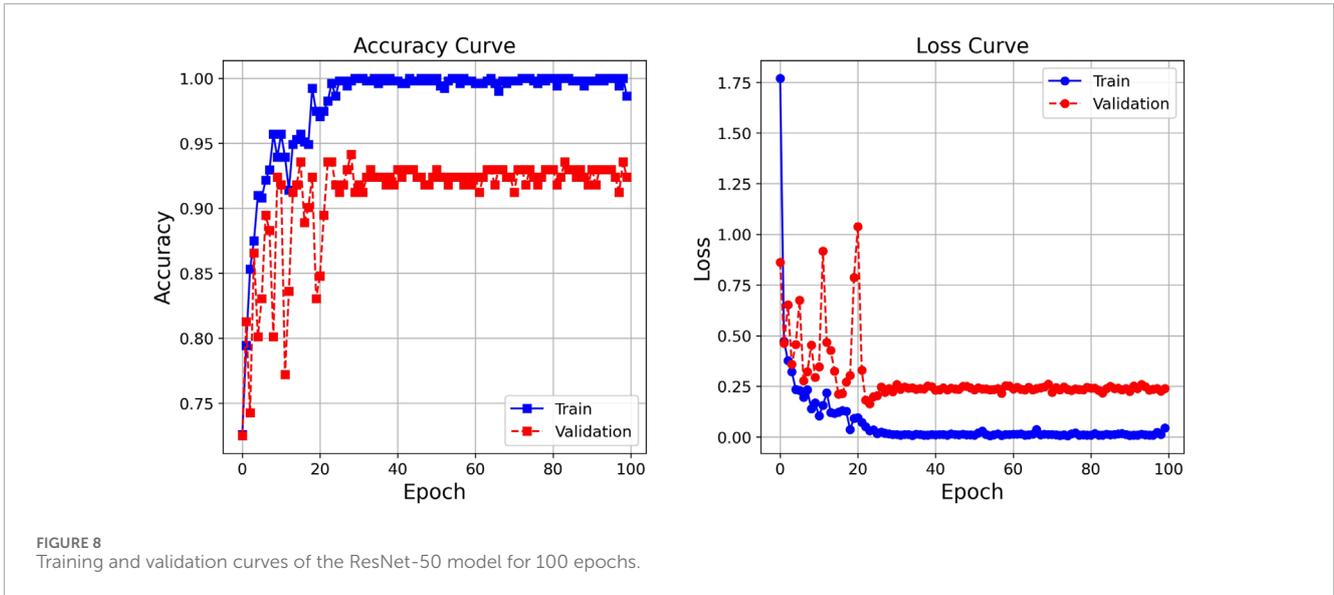
3 Result

3.1 Result of training

After 100 rounds of training using the above training parameter settings, the learning curve of the TSM-ResNet50 model is shown in Figure 7A and the loss curve is shown in Figure 7B. The training accuracy of the model is up to 98.15% and the model also achieves 92.5% accuracy on the validation set. The value of the loss value for the training of the model stabilized around 0.1, and the value of the loss for the validation set stabilized around 0.3, and from the training metrics, the model learned the features we picked as we expected it to. For comparison, the same dataset was used to train the ResNet-50 model without adding the TSM module and the CNN-LSTM model, and the tests were conducted on the same test set. The training curve of the ResNet-50 model is shown in Figure 8, and its highest accuracy on the validation set was 94.15%. The training curve of the CNN-LSTM model is shown in Figure 9, and its highest accuracy on the validation set was 86.03%.

3.2 Result of test

In order to the actual performance of the TSM-ResNet50 model and to evaluate it, we tested the trained model using 164 cases of data (37 positive and 127 negative) from the test set (data not involved in the training) that was kept before the training. Of the 164 cases of test data, 159 cases of data were correctly predicted, 33 cases were positive and 126 cases were negative. There were 5 cases of incorrectly predicted data, one negative and four positive cases, and the model achieved an overall recognition accuracy of 96.95%, a precision (check rate) of 97.06%, a recall (sensitivity) of 89.18%, a specificity of 99.21%, and an F1_score of 0.9692, the confusion matrix of the TSM-ResNet50 model is shown in Figure 10, and the ROC curve is shown in Figure 11. The confusion matrix of the ResNet-50 model without using the TSM module on this test set is shown in Figure 12, while the confusion matrix of the CNN-LSTM model on this test set is shown in Figure 13. The test performance of all the trained models on the independent test set is listed in Table 2. Taking all these evaluation indicators into consideration, the test

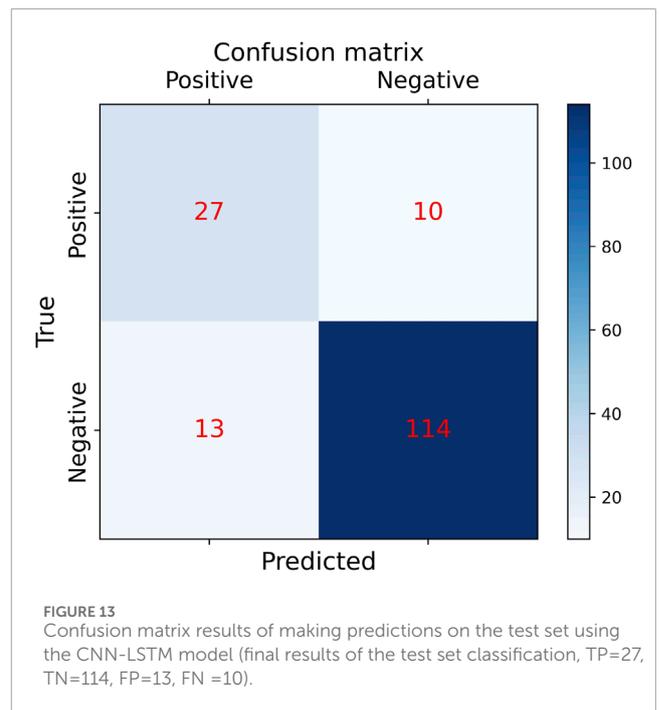
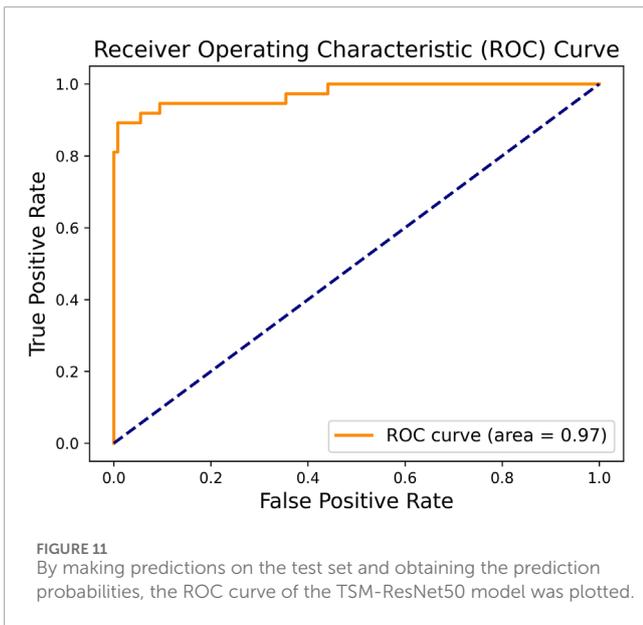
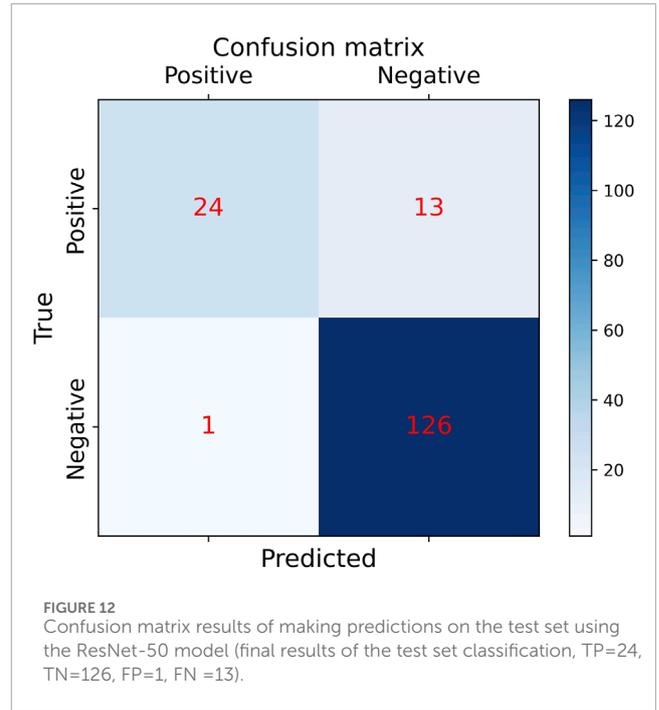
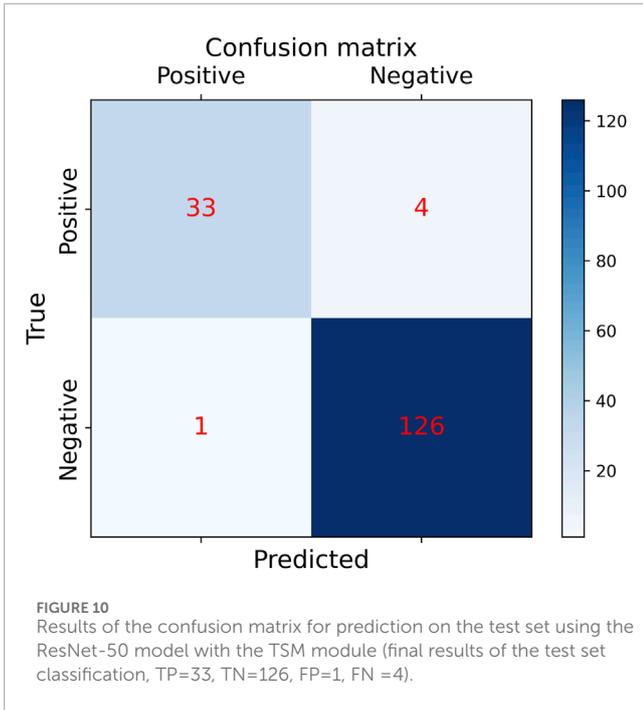


results of the TSM-ResNet50 model are all superior to those of the ResNet-50 model and the CNN-LSTM model.

3.3 Misanalysis

In the test, there was one case where negative data was predicted to be positive. Our model considered the likelihood of the data being negative to be 29.63% and the likelihood of the data being positive

to be 70.37%, so the data was labeled as positive, and we extracted and analyzed the data from that case. The image of the example data is shown in [Figure 14](#). The feature of this data that was labeled as negative by the expert was the presence of lung sliding, and we found that there was indeed lung sliding in this data after reviewing the raw data, but the lung sliding appeared between 3.5 and 5 s, and the pleural line was almost at a standstill before 3.5 s, which was a large percentage of the positive features and an inconspicuous percentage of the negative features, which led to the prediction error.



There were four cases of positive data in the test that were predicted to be negative, and we also analyzed them separately. **Figure 15A**, positive data labeled by the expert as no lung sliding and no pleural pulsation but predicted by our model to be negative with a 94.47% probability, was analyzed to show that the pleural line in the picture did not undergo lung sliding but produced a large movement (a movement similar to lung sliding), which resulted in the model incorrectly judging the movement of the pleural line to be lung sliding. **Figure 15B**, again positive data without lung sliding and lung pulsation, but the pleural line produced a wavy line of pulsation (very similar to the pleural pulsation sign), and the presence of pleural pulsation was negative, so the model incorrectly

judged it as negative. **Figure 15C**, the expert gave this data the classification feature of A line, but the pleural line in the picture is too blurred, the contrast with other lung tissues is low, there is a large movement in the picture, and the model did not accurately recognize the diagnostic feature and misjudged it as negative. **Figure 15D**, the data situation of this case is similar to that in **Figure 15A**, the pleural line in the picture showed a large degree of lateral movement, which was recognized as negative by the model. By analyzing the data of the above prediction errors, the following conclusion can be deduced: when the position of the pleural line in the positive

TABLE 2 The test performance of all models on the same test set.

Model Indicator	TSM-ResNet50	ResNet-50	CNN-LSTM
Accuracy	0.9695	0.9146	0.8597
Precision	0.9706	0.9600	0.6750
Recall	0.8919	0.6486	0.7297
Specificity	0.9921	0.9921	0.8976
F1_score	0.9296	0.7742	0.7013
Model parameter	94,289,074	94,367,681	192,999,414

data frame is shifted to a large extent, it is easy for the model to misdiagnose. The reason for this may be that the weight of positive data is small compared to negative data, the weight of negative features is too high, and the model is not fine enough to recognize the positive features.

4 Discussion

Interpretation of medical images is often performed by professionals, and compared to other medical images, ultrasound images are more difficult to interpret. Ultrasound plays an important role in the initial screening of medical conditions. Therefore, if the diagnostic threshold of ultrasound can be lowered, ultrasound technology can play a greater role in more fields, and in the future, ultrasound can be operated remotely in real-time or unmanned automated diagnosis. Our work is based on the above objectives and focuses on the intelligent diagnosis of ultrasound pneumothorax.

In this report, we propose a technical solution for accurately recognizing ultrasound pneumothorax, a first attempt to directly use video comprehension techniques for ultrasound pneumothorax diagnosis, and the results of the study show that AI models also have an extremely great potential for the interpretation of ultrasound images. Through deep learning algorithms, the model is able to learn the abstract features of PTX in ultrasound images, thus realizing a more accurate diagnosis. This finding is of great significance in lowering the threshold of ultrasound image diagnosis, especially in scenarios where medical resources are limited. The AI-assisted diagnosis system is expected to improve the diagnostic capability of primary care organizations.

From the current point of view, deep learning research in ultrasound diagnostic PTX still lags behind more traditional chest imaging techniques such as CT or X-ray, in which organized and labeled data are easier to acquire (Thian et al., 2021; Taylor et al., 2018; Röhrich et al., 2020) while lung ultrasound data acquisition is still difficult. Although some scholars have begun to engage in research related to lung ultrasound pneumothorax, they have adopted much the same approach, starting from B-mode ultrasound and focusing on the interpretation of lung sliding of M-mode images after reconstruction of B-mode ultrasound images into pseudo-M-mode images (VanBerlo et al., 2022; Boice et al., 2022;

Jaščur et al., 2021). The difference is what kind of model or strategy is used to reconstruct and categorize M-mode images from B-mode ultrasound. Our work starts from another angle, applying the video understanding technology in the field of ultrasound diagnosis, directly classifying and diagnosing the B-mode ultrasound images, just like a doctor who has been practicing medicine for many years, directly reading the ultrasound images without the need of reconstructing M-mode ultrasound images, and our experimental results show that this technical route is feasible. With the advantage of this technique, we can save a lot of time and labor costs in data annotation, and this advantage will be more obvious in large databases. VanBerlo et al. believe that it is better to include pleural pulsation sign in the detection of PTX (VanBerlo et al., 2022), but their study focuses on a single judgment of pleural sliding like previous studies. Our work not only focuses on pleural sliding but also adds some other key features used to diagnose pneumothorax, such as pleural pulsation sign, lung points (Zhao et al., 2023a), and so on. The diagnosis of ultrasound pneumothorax by fusion of multiple features has been realized using video comprehension techniques, and good results have been achieved. By learning the fusion of multiple key features in the diagnostic process of ultrasound pneumothorax, the positive and negative abstract features in the ultrasound pneumothorax video are directly categorized, which further improves the accuracy of pneumothorax diagnosis.

Extrapolating from the existing findings of our current work, as well as the research results related to video understanding in the field of natural imagery, theoretically, if we have a large enough variety of cases in our database and a large enough amount of image data, we can perform the diagnostic classification work of multiple diseases at the same time, and the diagnostic accuracy will be further improved, and this research will become even more meaningful. However, as with other medical models, the lack of interpretability of the models is an issue that needs to be addressed to ensure that clinicians can understand and trust the diagnostic results of the models. Therefore, currently AI models are only used as auxiliary tools to help doctors do their jobs more efficiently. How to combine AI models with doctors' expertise to achieve human-machine collaboration is also another important direction for future research. However, the amount of data available for training is still limited by the size of the dataset. As a result, we think that future studies will concentrate on growing the database, including additional disease categories, enhancing model performance, and resolving the model's interpretability. Despite the fact that the diagnosis of ultrasonic pneumothorax was the main focus of our work, the model and technique that were developed may have wider uses. In the future, we can use this method to diagnose additional ultrasound images, like liver and heart diseases, expanding the applications of ultrasound technology and giving patients faster, more precise medical care.

5 Conclusion

In conclusion, the development of a straightforward and user-friendly ultrasound image acquisition and diagnosis system will tremendously aid general medical staff in the pre-hospital emergency stage for prompt diagnosis, hence increasing patient survival rates. Ultrasound technology's ease of use and effectiveness in emergency medical settings will guarantee that patients can

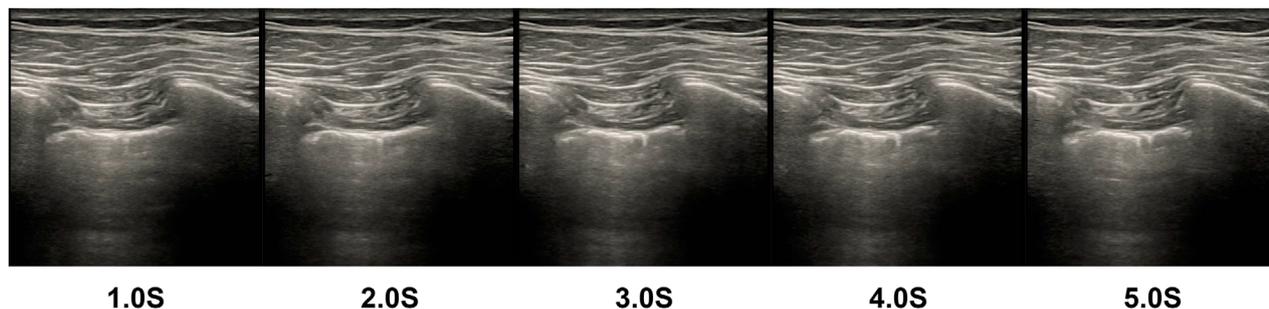


FIGURE 14

Cases predicted as positive by the model in negative data (Picture frames from the ultrasound video were intercepted at 1-second intervals and stitched together into 1- to 5-second screenshots, with the bright white pleural line in the center of each image almost at rest).

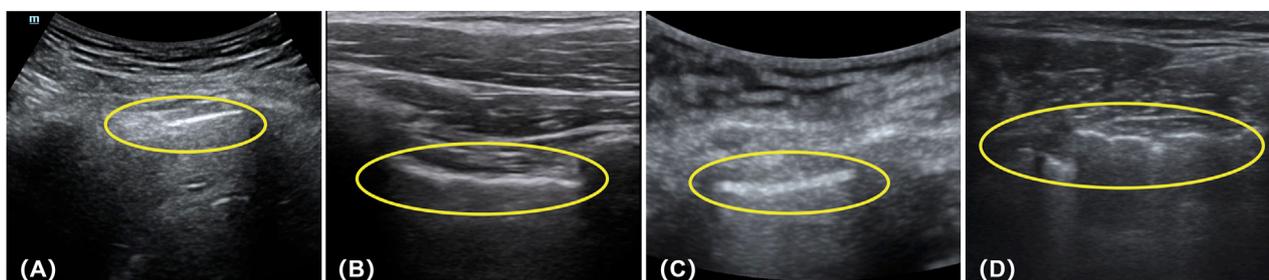


FIGURE 15

The data of (A–D) in the figure are all positive, but the model previously predicted their results as negative (the circled parts in the figure represent the pleural lines present in each image).

be quickly evaluated, screened, and triaged, improving medical institutions' emergency response capabilities. The current study is a component of the intelligent diagnostic system called E-FAST (Extended Focused Assessment with Sonography for Trauma), which is designed to determine whether a patient has a PTX. Looking ahead, our next plan is to optimize the training strategy and expand the database in order to build diagnostic models with more generalization ability for diagnosis of a wide range of diseases. To give users a more practical and effective diagnostic experience, we will also investigate integration with portable handheld ultrasound instruments. We anticipate that these initiatives will transform the healthcare industry and advance the adoption of diagnostic ultrasonography technology.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author/s.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation

and institutional requirements. Written informed consent from the [patients/ participants OR patients/participants legal guardian/next of kin] was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

XQ: Conceptualization, Formal Analysis, Investigation, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review and editing. QW: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. GL: Conceptualization, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review and editing. LS: Resources, Software, Validation, Writing – original draft, Writing – review and editing. WZ: Writing – original draft, Writing – review and editing. MY: Conceptualization, Formal Analysis, Investigation, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing. HW: Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

References

- Alrajhi, K., Woo, M. Y., and Vaillancourt, C. (2012). Test characteristics of ultrasonography for the detection of pneumothorax: a systematic review and meta-analysis. *Chest* 141 (3), 703–708. doi:10.1378/chest.11-0131
- Blaivas, M., Lyon, M., and Duggal, S. (2005). A prospective comparison of supine chest radiography and bedside ultrasound for the diagnosis of traumatic pneumothorax. *Acad. Emerg. Med.* 12 (9), 844–849. doi:10.1197/j.aem.2005.05.005
- Boice, E. N., Hernandez Torres, S. I., Knowlton, Z. J., Berard, D., Gonzalez, J. M., Avital, G., et al. (2022). Training ultrasound image classification deep-learning algorithms for pneumothorax detection using a synthetic tissue phantom apparatus. *J. Imaging* 8 (9), 249. doi:10.3390/jimaging8090249
- Brady, A. K., Spitzer, C. R., Kelm, D., Brosnahan, S. B., Latifi, M., and Burkart, K. M. (2021). Pulmonary critical care fellows' use of and self-reported barriers to learning bedside ultrasound during training: results of a national survey. *Chest* 160 (1), 231–237. doi:10.1016/j.chest.2021.01.068
- Carreira, J., and Zisserman, A. (2017). "Quo vadis, action recognition? A new model and the kinetics dataset," in 2017 IEEE conference on computer vision and pattern recognition (CVPR). 21–26 July 2017, Honolulu, HI, USA. 4724–4733.
- Duclos, G., Bobbia, X., Markarian, T., Muller, L., Cheyssac, C., Castillon, S., et al. (2019). Speckle tracking quantification of lung sliding for the diagnosis of pneumothorax: a multicentric observational study. *Intensive Care Med.* 45 (9), 1212–1218. doi:10.1007/s00134-019-05710-1
- Hefny, A. F., Kunhivalappil, F. T., Paul, M., Almansoori, T. M., Zoubeidi, T., and Abu-Zidan, F. M. (2019). Anatomical locations of air for rapid diagnosis of pneumothorax in blunt trauma patients. *World J. Emerg. Surg.* 14 (1), 44. doi:10.1186/s13017-019-0263-0
- Hernandez, C., Shuler, K., Hannan, H., Sonyika, C., Likourezos, A., and Marshall, J. C. A. U. S. E. (2008). C.A.U.S.E.: cardiac arrest ultra-sound exam—a better approach to managing patients in primary non-arrhythmic cardiac arrest. *Resuscitation* 76 (2), 198–206. doi:10.1016/j.resuscitation.2007.06.033
- Jašćur, M., Bundzel, M., Malík, M., Dzian, A., Ferenčík, N., and Babič, F. (2021). Detecting the absence of lung sliding in lung ultrasounds using deep learning. *Appl. Sci.* 11 (15), 6976. doi:10.3390/app11156976
- Kim, K., Macruz, F., Wu, D., Bridge, C., McKinney, S., Saud, A. A. A., et al. (2023). Point-of-care AI-assisted stepwise ultrasound pneumothorax diagnosis. *Phys. Med. and Biol.* 68 (20), 205013. doi:10.1088/1361-6560/acfb70
- Kulhare, S., Zheng, X., Mehanian, C., Gregory, C., Zhu, M., Gregory, K., et al. (2018). "Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks," in *Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation*. Editors D. Stoyanov, Z. Taylor, S. Aylward, J. M. R. S. Tavares, Y. Xiao, A. Simpson, et al. Cham, Springer International Publishing, 65–73.
- Lichtenstein, D. A. (2015). BLUE-protocol and FALLS-protocol: two applications of lung ultrasound in the critically ill. *Chest* 147 (6), 1659–1670. doi:10.1378/chest.14-1313
- Lichtenstein, D. A., and Menu, Y. (1995). A bedside ultrasound sign ruling out pneumothorax in the critically ill: lung sliding. *Chest* 108 (5), 1345–1348. doi:10.1378/chest.108.5.1345
- Lin, J., Gan, C., and Han, S. (2019). "Temporal shift module for efficient video understanding," in 2019 IEEE/CVF international conference on computer vision (ICCV), 7082–7092. Available online at: <https://ieeexplore.ieee.org/document/9008827>.
- Lindsey, T., Lee, R., Grisell, R., Vega, S., and Veazey, S. (2019). "Automated pneumothorax diagnosis using deep neural networks," in *Progress in pattern recognition, image analysis, computer vision, and applications*. Editors R. Vera-Rodriguez, J. Fierrez, and A. Morales Cham: Springer International Publishing, 723–731.
- Mehanian, C., Kulhare, S., Millin, R., Zheng, X., Gregory, C., Zhu, M., et al. (2019). "Deep learning-based pneumothorax detection in ultrasound videos," in *Smart ultrasound imaging and perinatal, preterm and paediatric image analysis*. Editors Q. Wang, A. Gomez, J. Hutter, K. McLeod, V. Zimmer, O. Zettinig, et al. Cham, Springer International Publishing, 74–82.
- Nagarsheth, K., and Kurek, S. (2011). Ultrasound detection of pneumothorax compared with chest X-ray and computed tomography scan. *Am. Surgeon* 77 (4), 480–483. doi:10.1177/000313481107700427
- Qian, A. Y., and Zhang, M. (2019). Actively participating in the construction of trauma centers to accelerate the development of emergency medicine. *Chin. J. Emerg. Med.* 28, 550–552. doi:10.3760/cma.j.issn.1671-0282.2019.05.002
- Röhrich, S., Schlegel, T., Bardach, C., Prosch, H., and Langa, G. (2020). Deep learning detection and quantification of pneumothorax in heterogeneous routine chest computed tomography. *Eur. Radiol. Exp.* 4 (1), 26. doi:10.1186/s41747-020-00152-7
- Summers, S. M., Chin, E. J., April, M. D., Grisell, R. D., Lospinoso, J. A., Kheirabadi, B. S., et al. (2017). Diagnostic accuracy of a novel software technology for detecting pneumothorax in a porcine model. *Am. J. Emerg. Med.* 35 (9), 1285–1290. doi:10.1016/j.ajem.2017.03.073
- Taylor, A. G., Mielke, C., and Mongan, J. (2018). Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study. *PLoS Med.* 15 (11), e1002697. doi:10.1371/journal.pmed.1002697
- Thian, Y. L., Ng, D., Hallinan, JTPD, Jagmohan, P., Sia, S. Y., Tan, C. H., et al. (2021). Deep learning systems for pneumothorax detection on chest radiographs: a multicenter external validation study. *Radiol. Artif. Intell.* 3 (4), e200190. doi:10.1148/ryai.2021200190
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015). Learning spatiotemporal features with 3D convolutional networks. In: 2015 IEEE international conference on computer vision (ICCV), 4489–4497. Available online at: <https://ieeexplore.ieee.org/document/7410867/keywords#keywords>.
- Valenzuela, J., Stilson, B., Patanwala, A., Amini, R., and Adhikari, S. (2020). Prevalence, documentation, and communication of incidental findings in focused assessment with sonography for trauma (FAST) examinations. *Am. J. Emerg. Med.* 38 (7), 1414–1418. doi:10.1016/j.ajem.2019.11.040
- VanBerlo, B., Wu, D., Li, B., Rahman, M. A., Hogg, G., VanBerlo, B., et al. (2022). Accurate assessment of the lung sliding artefact on lung ultrasonography using a deep learning approach. *Comput. Biol. Med.* 148, 105953. doi:10.1016/j.combiomed.2022.105953
- Wallbridge, P., Steinfurt, D., Tay, T. R., Irving, L., and Hew, M. (2018). Diagnostic chest ultrasound for acute respiratory failure. *Respir. Med.* 141, 26–36. doi:10.1016/j.rmed.2018.06.018
- Wan, X. H., and Chen, H. (2015). *Clinical diagnostics*. 3rd ed. Beijing: People's Medical Publishing House, 178–180.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2025.1530808/full#supplementary-material>

Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., et al. (2016). Temporal segment networks: towards good practices for deep action recognition. in *Computer Vision – ECCV 2016*. Editors B. Leibe, J. Matas, N. Sebe, and M. Welling Cham: Springer International Publishing, 20–36.

Wang, X., Girshick, R., Gupta, A., and He, K. (2018). Non-local neural networks. Available online at: https://openaccess.thecvf.com/content_cvpr_2018/html/Wang_Non-Local_Neural_Networks_CVPR_2018_paper.html.

Weissman, J., and Agrawal, R. (2021). Dramatic complication of pneumothorax treatment requiring lifesaving open-heart surgery. *Radiol. Case Rep.* 16 (3), 500–503. doi:10.1016/j.radcr.2020.12.034

Xie, S., Sun, C., Huang, J., Tu, Z., and Murphy, K. (2018). Rethinking spatiotemporal feature learning: speed-accuracy trade-offs in video classification. in *Computer Vision – ECCV 2018*. Editors V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss Cham, Springer International Publishing, 318–35.

Zhao, H. T., Liu, Y., Liu, Y. L., Niu, H. M., Wang, X. N., Xue, H. Y., et al. (2023b). Application value of lung ultrasound pleural effusion diagnosis process and three-point localization method guided pleural effusion puncture. *Chin. J. Ultrasound Med.* 39 (11), 1239–1242. doi:10.3969/j.issn.1002-0101.2023.11.013

Zhao, H. T., Long, L., Ren, S., and Zhao, H. L. (2023a). Research progress on the value of bedside lung ultrasound in the diagnosis of pneumothorax. *Chin. J. Emerg. Med.* 39 (9), 892–897. doi:10.3969/j.issn.1002-1949.2019.09.018

Zhou, B., Andonian, A., Oliva, A., and Torralba, A. (2018). Temporal relational reasoning in videos. in *Computer Vision – ECCV 2018*. Editors V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss Cham, Springer International Publishing, 831–46.

Zolfaghari, M.M., Singh, K.K., and Brox, T. (2018). Efficient convolutional network for online video understanding. in *Computer Vision – ECCV 2018*. Editors V.V. Ferrari, M.M. Hebert, C.C. Sminchisescu, and Y.Y. Weiss, 11206, 713–730.