applied sciences

MDPI

Article

Three-Dimensional Automated Breast Ultrasound (ABUS) Tumor Classification Using a 2D-Input Network: Soft Voting or Hard Voting?

Shaode Yu [1], Xiaoyu Liang [1], Songnan Zhao [1], Yaoqin Xie [2] and Qiurui Sun [3,*]

[1] School of Information and Communication Engineering, Communication University of China, Beijing 100024, China; yushaodecuc@cuc.edu.cn (S.Y.); 202211133019@cuc.edu.cn (X.L.); 2021211123002@cuc.edu.cn (S.Z.)
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; yq.xie@siat.ac.cn
[3] Center of Information & Network Technology, Beijing Normal University, Beijing 100875, China
* Correspondence: qiuruisun@bnu.edu.cn

Abstract: Breast cancer is a global threat to women's health. Three-dimensional (3D) automated breast ultrasound (ABUS) offers reproducible high-resolution imaging for breast cancer diagnosis. However, 3D-input deep networks are challenged by high time costs, a lack of sufficient training samples, and the complexity of hyper-parameter optimization. For efficient ABUS tumor classification, this study explores 2D-input networks, and soft voting (SV) is proposed as a post-processing step to enhance diagnosis effectiveness. Specifically, based on the preliminary predictions made by a 2D-input network, SV employs voxel-based weighting, and hard voting (HV) utilizes slice-based weighting. Experimental results on 100 ABUS cases show a substantial improvement in classification performance. The diagnosis metric values are increased from ResNet34 (accuracy, 0.865; sensitivity, 0.942; specificity, 0.757; area under the curve (AUC), 0.936) to ResNet34 + HV (accuracy, 0.907; sensitivity, 0.990; specificity, 0.864; AUC, 0.907) and to ResNet34 + SV (accuracy, 0.986; sensitivity, 0.990; specificity, 0.963; AUC, 0.986). Notably, ResNet34 + SV achieves the state-of-the-art result on the database. The proposed SV strategy enhances ABUS tumor classification with minimal computational overhead, while its integration with 2D-input networks to improve prediction performance of other 3D object recognition tasks requires further investigation.

Keywords: breast cancer diagnosis; automated breast ultrasound; soft voting; hard voting; convolutional neural network; deep learning

check for updates

Citation: Yu, S.; Liang, X.; Zhao, S.; Xie, Y.; Sun, Q. Three-Dimensional Automated Breast Ultrasound (ABUS) Tumor Classification Using a 2D-Input Network: Soft Voting or Hard Voting? *Appl. Sci.* **2024**, *14*, 11611. https://doi.org/10.3390/app142411611

Academic Editor: Thomas Lindner

Received: 30 October 2024
Revised: 30 November 2024
Accepted: 6 December 2024
Published: 12 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Breast cancer (BC) is the most diagnosed cancer among women. It caused about 2.26 million new BC cases globally in 2020 as the most prevalent cancer [1]. The incidence of female BC continued a slow increase during 2014 through 2018 by 0.5% annually [2]. In developed countries, such as in the United States, BC incidence rates have risen in most of the past four decades, largely driven by localized-stage and hormone receptor-positive disease [3]. Compared to developed countries, the rates are relatively lower in China, while the actual number of new cases and deaths has been considerable [4]. BC poses a serious threat to women's health and places a heavy burden on the finance and healthcare systems. Improvements in medical imaging, cancer screening and diagnosis, therapeutic planning and delivery, and follow-up monitoring contribute to the wide coverage of the health system, insurance, and advances in treatment and management [1–4].

Medical imaging devices are routinely used for BC screening and diagnosis, including mammography (MAM) and ultrasound (US). MAM remains the gold standard due to its high-resolution imaging of internal anatomy and its sensitivity for early-stage BC

detection [5]. Its effectiveness has been demonstrated in large-scale clinical trials, leading to improved treatment outcomes, higher survival rates, and reduced mortality through early intervention [2,3]. However, it is associated with issues such as over-diagnosis, radiation exposure, and decreased sensitivity in patients with dense breast tissue. MAM has been found to be less suitable for young women and Asian women, who often have denser breast tissue [6]. Hand-held US (HHUS) eliminates the risk of radiation exposure and provides more detailed imaging for women with dense breast tissue. It effectively differentiates between solid tumors and fluid-filled cysts, thereby reducing unnecessary biopsies [7]. To address operator dependence, limited field of view, and to enhance imaging quality, three-dimensional (3D) ultrasound imaging has been developed [8–11].

As an emerging technique, 3D automated breast ultrasound (ABUS) serves as a supplementary method for evaluating women with heterogeneously and extremely dense breasts [11]. It offers several advantages in screening and diagnostic settings, including an increased BC detection rate, improved workflow efficiency, and reduced examination time. Notably, ABUS separates image acquisition from image interpretation, thereby decreasing operator dependence and time cost. Vourtsis and Kachulis investigated the performance of ABUS and HHUS in a large cohort of 1886 women and found that ABUS enhances the sensitivity of cancer detection [12]. Additionally, Klein et al. conducted a retrospective clinical study comparing the performance of ABUS and HHUS in cancer diagnosis, identifying that ABUS results in lower recall and biopsy rates, as it provides multiple perspectives of suspicious regions for examination [13]. Therefore, ABUS has significant potential to be routinely used as a standardized, reproducible, and reliable tool for whole-breast visualization, screening, and diagnosis [12–14], offering added value for patients with dense breasts [10,11].

Accurate diagnosis of breast lesions observed in ABUS enables the determination of tumor malignancy and the formulation of treatment plans. However, few deep learning models have been developed for this purpose [15,16]. Tan et al. extracted spiculation patterns in coronal planes and designed spiculation and other characteristic features for classifying lesions as malignant or benign using support vector machine [17]. Wang et al. modified the Inception-v3 architecture for efficient feature extraction, integrating features from both transverse and coronal views for cancer diagnosis [18]. Xiang et al. combined residual blocks, capsule neural structures, and group normalization for ABUS tumor classification [19]. Zhou et al. designed a multi-task learning framework for joint ABUS tumor segmentation and classification, incorporating multi-scale feature extraction and iterative feature refinement [20]. Wang et al. added an automatic segmentation network for morphological analysis along with ResNet-based tumor diagnosis [21]. Ding et al. proposed a multi-view attention network that utilizes a localization unit for lesion region cropping and a classification unit for malignancy prediction based on the Transformer architecture [22]. Yang et al. developed a 2.5D deep model that fine-tunes a pre-trained network for tumor classification, using the ten slices with the largest lesion regions along with adjacent slices as input [23]. Therefore, there is an urgent need to explore deep learning networks for ABUS tumor classification.

In 3D high-resolution medical image analysis using deep learning, a key challenge in ABUS tumor classification is achieving a balance between time efficiency and classification accuracy. However, 3D-input deep networks face several significant problems [20]. At first, high dimensionality of 3D images increases data processing complexity, computational demands, and resource intensity compared to 2D images. Secondly, ABUS tumors exhibit a wide range of shapes, sizes, and characteristics, and therefore, it is difficult to develop a model that can accurately classify all types of lesions. Thirdly, the quality of ABUS images is influenced by the imaging devices used, and these images often suffer from noise and various artifacts [15]. Fourthly, many clinical settings require rapid image processing and analysis for timely decision-making, which adds pressure to achieve high classification accuracy while maintaining fast processing speeds. One of the most significant challenges is the limited availability of annotated datasets; very few ABUS cases are publicly accessible

for training [16]. Specifically, an overview [15] of recent advancements in BC image analysis indicates that one 3D ABUS database with 100 volumetric cases is accessible for algorithm development and fair comparison.

In practice, using volumetric images as input for deep networks necessitates iterative optimization of hyper-parameters, which requires a large-scale, high-quality database and leads to significant time costs [19,20]. While noisy data can be utilized for model training, the robustness of these models must be thoroughly examined in the context of medical image analysis [24–26]. On the other hand, while using slice images as input allows for faster slice-wise lesion predictions, effectively combining the slice-wise probabilities for benign and malignant predictions remains an open question. In summary, deep learning-based 3D-input ABUS tumor classification faces challenges related to high time costs, a lack of sufficient training samples, and the complexity of hyper-parameter optimization.

To the best of our knowledge, few studies have specifically addressed the aforementioned issues in ABUS tumor classification [16]. In this study, a soft voting (SV) strategy is proposed using voxel-level weighting of slice prediction. It performs as a post-processing step after image slice prediction via a 2D-input deep learning model. It should be noted that, in pediatric brain tumor classification, Bianchessi et al. also propose soft voting for per-slice class prediction [27]. However, there are key differences between the SV strategies in our work and that in Ref. [27]. Firstly, our SV is directly related to the prediction for per-volume classes, while in Ref. [27], it pertains to slice-level class prediction. Secondly, our SV operates on the predicted probabilities of the benign and malignant classes for each voxel, whereas that in [27] aims to predict each slice as belonging to one of three classes. Thirdly, our SV uses only a single classification model, while the SV in [27] involves multiple classification models. To verify the effectiveness, the proposed SV strategy is compared to the hard voting (HV) strategy, and the latter uses slice-based weighting. Furthermore, the baseline deep networks are compared to those with proposed strategies. Experimental results suggest that the proposed SV strategy improves tumor classification performance due to the utilization of tumor sizes and slice-level predicted probabilities.

2. Materials and Methods

This section introduces the ABUS database, the proposed post-processing strategy, the deep learning models for evaluation, the experiment design, the used performance metrics, and the implementation details.

2.1. The Database

The database is the training set of the Tumor Detection, Segmentation and Classification Challenge on ABUS 2023 (TDSC-ABUS2023). It contains 100 ABUS volume images. Currently, the database is the only ABUS database available online to the community, paving the way for improved data availability and 3D ultrasound image analysis [15].

In the database, the matrix size of volumetric images ranges between [843, 546, 270] and [865, 682, 354], the physical in-plane spacing is [0.200 mm, 0.073 mm], and the between-slice spacing is around 0.476 mm. The volumes are acquired by using an Invenia ABUS system (GE Healthcare) at Harbin Medical University Cancer Hospital, Harbin, China. The data are stored in nrrd format, and the pixel intensity ranges from 0 to 255. An experienced radiologist checked the data cases and annotated the tumor regions.

Figure 1 shows the distribution of case numbers in terms of the voxel numbers of tumor regions. The horizontal axis presents the values of the base-10 logarithm of the voxel numbers v ($\log_{10} v$). The values are equally divided into seven bins. The vertical axis displays the number of ABUS cases. It shows that most cases (90 out of 100 cases) contain around 10^4 to 10^6 voxels in the annotated tumor regions.

Moreover, the database provides the volume images and corresponding biopsy labels (42 benign and 58 malignant). The voxel number of tumor regions is $(1.06 \pm 1.36) \times 10^5$ (benign) and $(3.91 \pm 9.38) \times 10^5$ (malignant), and the voxel intensity is 67.10 ± 11.96 (benign) and 70.19 ± 10.58 (malignant). In addition, the smallest tumor contains 3539 voxels, and

the largest tumor contains 6,863,915 voxels. The uniform voxel intensity and varying tumor shapes and sizes present significant challenges for accurate lesion classification.

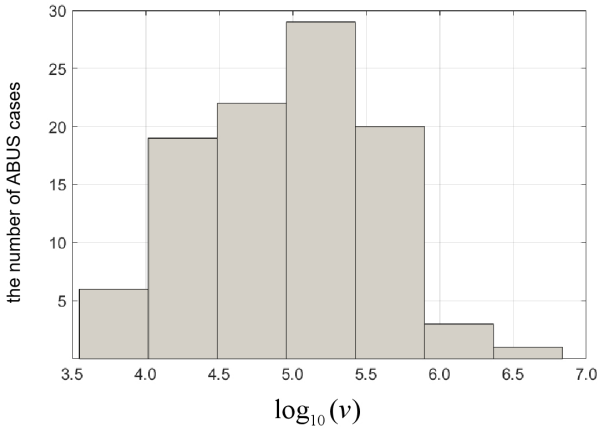


Figure 1. The distribution of voxel numbers of tumor regions.

2.2. The Proposed Soft-Voting Strategy

The proposed SV strategy is a post-processing step following a 2D-input image classification model. As shown in Figure 2, after slice-wise malignancy prediction, the proposed voxel weighting strategy takes both tumor sizes and predicted probabilities (benign and malignant) into consideration for tumor classification.

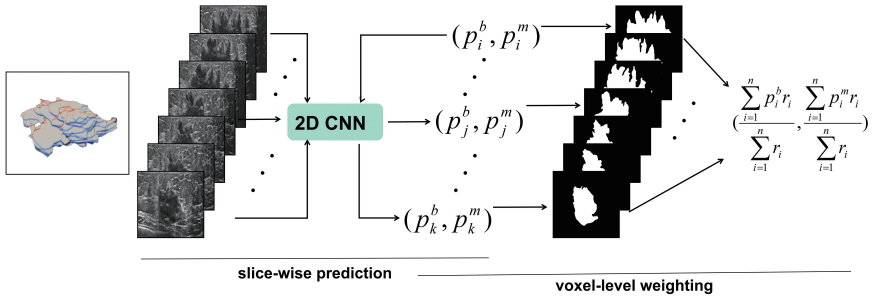


Figure 2. The proposed soft-voting strategy. After slice-wise prediction, the strategy accounts both tumor sizes and voxel probabilities for tumor classification.

At first, a convolutional neural network (CNN) is employed for slice-wise prediction. After the CNN model is trained, to an unseen input image slice (s_i), its output is two prediction probabilities (benign, p_i^b; malignant, p_i^m; and $p_i^b + p_i^m = 1$).

Then, the proposed SV strategy is implemented with voxel-level weighting. To a voxel in slice s_i of tumor region, its predicted probability is p_i^b and p_i^m. Since a tumor is a volume with many slices and voxels, we assume that each voxel contributes to the prediction of tumor malignancy, and thus, voxel weighting becomes reasonable.

Specifically, assume that a volumetric tumor v contains n slices $\{s_i\}_{i=1}^n$, and the voxel numbers in slices is $\{r_i\}_{i=1}^n$ correspondingly. And thereby, the benign probability p_v^b and the malignant probability p_v^m of the tumor can be defined in Equation (1). The numerators stand for the contribution of voxel-weighted benign or malignant probabilities, and the denominators denote the number of voxels or the tumor size. Since the denominators are the same $\sum_{i=1}^n r_i$, the values of the numerators are directly related to the final tumor classification.

$$\begin{cases} p_v^b = \dfrac{\sum_{i=1}^n p_i^b r_i}{\sum_{i=1}^n r_i} \propto \displaystyle\sum_{i=1}^n p_i^b r_i \\ p_v^m = \dfrac{\sum_{i=1}^n p_i^m r_i}{\sum_{i=1}^n r_i} \propto \displaystyle\sum_{i=1}^n p_i^m r_i \end{cases} \tag{1}$$

In the end, the classification p_v^{sv} of the volumetric tumor is determined by the larger value of p_v^b and p_v^m as shown in Equation (2). The core idea of voxel weighting or the SV strategy is that the voxel numbers of a tumor and the voxel probabilities of malignancy derived from slice-wise prediction are utilized.

$$p_v^{sv} = \max\{p_v^b, p_v^m\} \tag{2}$$

Compared to the baseline CNN model used in the workflow, as a post-processing strategy, voxel weighting slightly increases the computing time in tumor classification, while it enhances the prediction robustness and decision-making confidence.

In contrast to voxel weighting, a more straightforward strategy is HV, or slice-level weighting. It mainly depends on the number of slices predicted as benign or as malignant. In other words, among the tumor with n slices of lesion regions, if k slices are predicted as malignant ($p^m > p^b$) and $k > (n-k)$, the volumetric tumor is voted as malignant, and vice versa. Equation (3) shows the comparison of the number of benign (k) and malignant ($n-k$) slices in the volumetric tumor.

$$p_v^{hv} = \max\{n-k, k\} \tag{3}$$

2.3. Involved CNN Models

To verify the effectiveness and efficiency of the proposed strategy, several CNN models are explored. This part briefly describes the four 2D models (ResNet34 [28], MedViT [29], HiFuse [30], and MedMamba [31]) and three 3D models (M3T3D [32], ResNeXt3D [33], DenseNet3D [34], and ResNet3D [35]). The proposed post-processing strategy is added to 2D CNN models for evaluating its effectiveness in ABUS lesion classification.

2.3.1. 2D CNN Models

The first 2D CNN model is ResNet34 [28], which is widely used as the backbone of many advanced networks in image classification and medical diagnosis [5,36]. Figure 3 shows the repeated residual blocks that use convolution layers for hierarchical representation and skip connections to mitigate the vanishing gradient problem. And consequently, deep neural networks can be straightforward, computationally efficient, and trained with fast convergence.

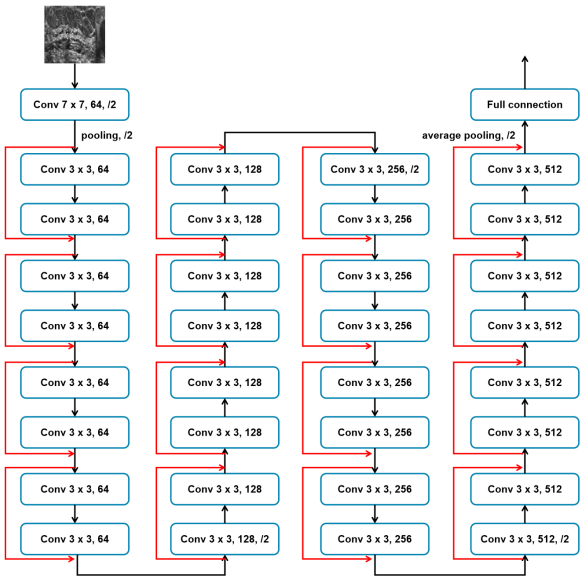


Figure 3. ResNet34. It uses skip connections of convolutional layers to avoid vanishing gradient. The structure of skip connections (red arrows) enables the building and training of very deep networks.

The second 2D CNN model is MedMamba [31] that combines both CNNs and Transformers. It is made up of patch embedding layers, stacked SS-Conv-SSM blocks, patch merging layers, and a feature classifier, as shown in Figure 4. It is similar to Vision Transformers [37] in that MedMamba splits the input image into non-overlapping patches. It builds hierarchical representations using four SS-Conv-SSM blocks for image downsampling. Specifically, the basic block SS-Conv-SSM includes channel-split, convolutional layers, structured state-space model (SSM) layers, and channel-shuffle in two branches of Conv-Branch and SSM-Branch for local and global information processing.

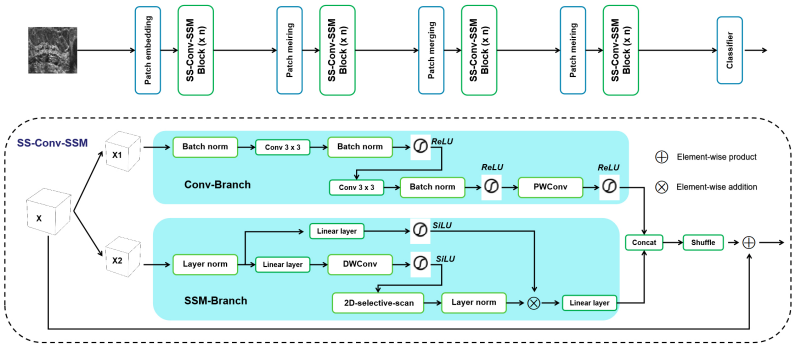


Figure 4. MedMamba. It designs two branches to form the SS-Conv-SSM block, and tumor lesions are progressively represented by the blocks for malignancy prediction.

The third 2D CNN model is MedViT [29]. It introduces the multi-head convolutional attention (MHCA), local feed-forward network (LFNN), and efficient self-attention (ESA), and the model is built based on the efficient convolutional blocks (ECBs) and local-token block (LTB), as shown in Figure 5. Specifically, MHCA decomposes an image into multiple regions or tokens and captures long-range dependencies, LFNN token-wisely rearranges the feature maps and token sequences converted by Seq2Img and Img2Seq, and ECBd benefit residual blocks of detail preservation while integrating the Transformer for deep feature representation. At the same time, LTB combines both local features from ECBs and global features from MHCA and ESA, and patch momentum changer is used for data diversity augmentation and improved model robustness. After progressive feature extraction and fusion, the malignancy is predicted by using simple batch normalization, global average pooling, and a full connection layer.

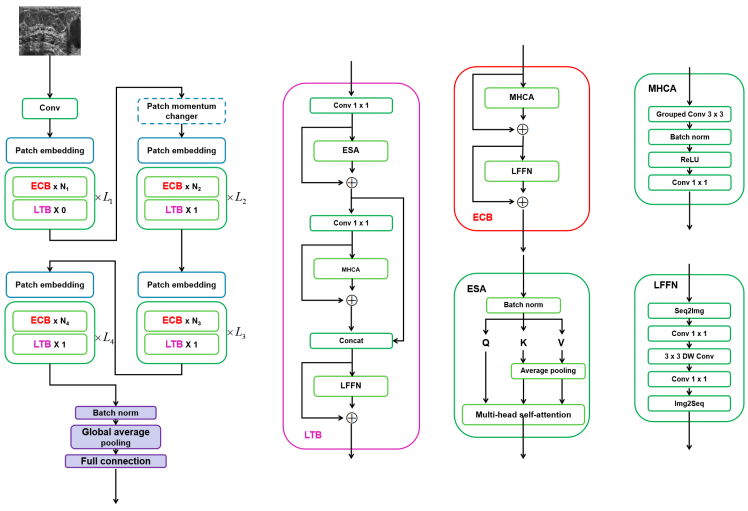


Figure 5. MedViT. It forms ECBs and LTB by using LFNN, MHCA, and ESA for progressively local and global feature extraction, long-range dependency capturing, and information fusion.

The fourth 2D CNN model is HiFuse [30]. It develops three-branch hierarchical integration of multi-scale features. Self-attention-based Transformer and CNN are combined without destroying the respective modeling. As shown in Figure 6, a parallel hierarchy structure with local and global feature blocks is designed for efficient representations of local and global semantic cues, and an adaptive hierarchical feature fusion (HFF) block is proposed to integrate the multi-scale features comprehensively. Specifically, the HFF block uses small modules, including spatial attention, channel attention, residual inverted multi-layer perceptron (MLP), and shortcut, to integrate the semantic features of each branch. At last, global average pooling and a layer-normalized linear classifier are used for lesion malignancy prediction.

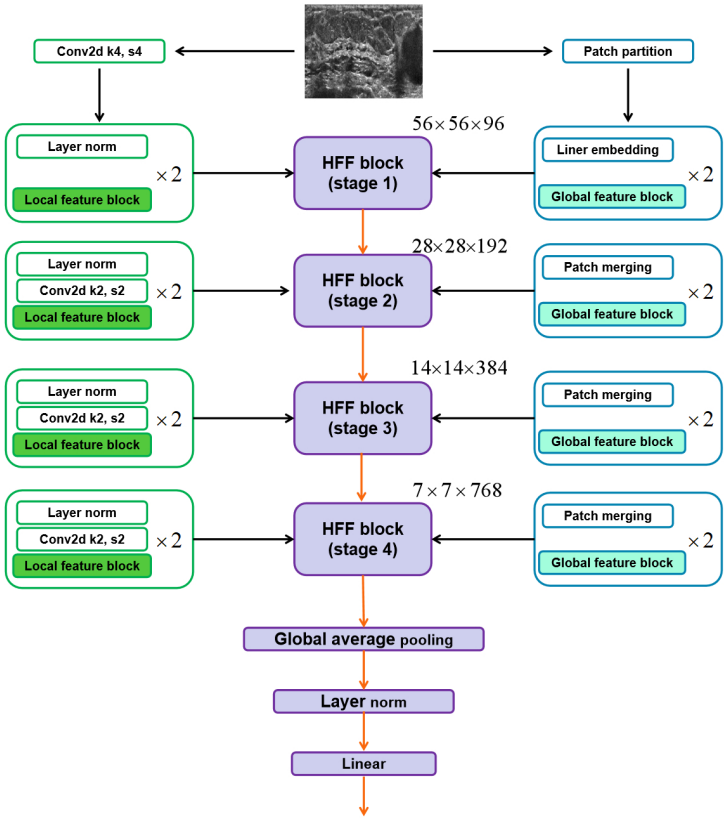


Figure 6. HiFuse. It designs two branches to extract local features and global features, and the HFF blocks are used to fuse dual-stream features progressively for tumor malignancy classification.

It is observed that ResNet34 uses four different sizes of residual blocks (Figure 3), and MedViT, HiFuse, and MedMamba repeat four specific designs of blocks (Figures 4–6) for hierarchical data representation, progressive feature fusion, and object classification. Full technical details of these 2D-input networks can be found in the relevant publications and code implementations.

2.3.2. 3D CNN Models

The first 3D CNN model is M3T3D [32] that extracts feature representation of the input 3D data samples by using two convolution layers, batch normalization, and ReLU activation functions. Meanwhile, it extracts 2D features from each slice of coronal, sagital, and axial planes. After that, 2D features are concatenated, projected, and embedded as tokens passed to the Transformer encoder [38] for global information integration and long-range dependency capture. In the end, all the features are aggregated into global feature representation for malignancy prediction.

The second 3D CNN model is ResNeXt3D [33], which is designed for Alzheimer's disease classification. It combines ResNeXt [39] and Bi-LSTM [40] for 3D magnetic reso-

nance brain image analysis by replacing 2D convolution kernels with 3D kernels. The 3D representation features are flattened as 1D signal input of Bi-LSTM, and thereby, the spatial information of the 3D medical images are thoroughly learned for disease classification.

The third 3D CNN model is DenseNet3D [34] that combines both global and local features by using 3D densely connected CNN and prior shape information. Specifically, it upgrades the representation capacity by connecting each layer with the other convolution layers, and consequently, low-level features and high-level shape features are connected. Meanwhile, vanishing gradients are relieved by feature fusion.

The fourth 3D CNN model is ResNet3D [35] that designs spatio-temporal convolutions for action recognition. It proposes the mixed convolution to model object motion with a low- and mid-level operation in the early layers. Most importantly, the spatio-temporal variant divides the 3D convolution block explicitly into a 2D spatial image convolution and a 1D temporal or time-scale convolution, and an additional nonlinear rectification is embedded between these 2D and 1D convolution operations. Therefore, this bi-connection enables ResNet3D representation learning of complex functions.

2.4. Experiment Design

The database contains 100 volumes and 2028 slices with tumor regions. To assess the generalization performance, 5-fold cross-validation is employed. The database is randomly partitioned into five mutually exclusive folds with equal size. Each model is then trained and evaluated five times. In other words, four folders of 80 ABUS volumes (and the slices) are selected for 3D (and 2D) model training, and the remaining folder of volumes (and slices) is used for performance evaluation. The kind of splitting avoids data leakage and potential over-fitting, providing a more reliable estimation of classification performance.

After random data splitting, to 2D CNN models, the proposed SV strategy as well as the HV approach are verified by re-computing the prediction results for volumetric tumor classification. Moreover, for fair comparison, a total of 100 epochs are conducted for training these 2D and 3D deep networks.

2.5. Performance Metrics

Assume that true positive (TP) is the number of correctly predicted positive samples, true negative (TN) is the number of correctly predicted negative samples, false positive (FP) is the number of incorrectly predicted positive samples, and false negative (FN) is the number of incorrectly predicted negative samples.

In the current study, six metrics, namely accuracy (ACC), sensitivity (SEN), specificity (SPE), AUC (area under the curve), the training time in minute (time), and the score, are used to measure the classification performance. Equation (4) shows how to compute these metrics. The metric *score* is the official metric for the challenge.

$$\begin{cases} ACC = \dfrac{TP + TN}{TP + TN + FP + FN} \\ SEN = \dfrac{TP}{TP + FN} \\ SPE = \dfrac{TN}{TN + FP} \\ score = \dfrac{ACC + AUC}{2} \end{cases} \tag{4}$$

2.6. Implementation Details

The codes of the CNN models are available online. The deep learning models are evaluated without code modification. The networks run on the platform with GPU (NVIDIA, Santa Clara, California, USA) RTX 4090 (24 GB), CPU (15 vCPU Intel(R) Xeon(R) Platinum 8474C) and RAM 80 GB. For iterative hyper-parameter optimization, 100 epochs are conducted to 2D-input CNN models, and due to the much greater number of more parameters, 300 epochs are used for 3D-input CNN models.

3. Results

This section presents and compares the performance of the 2D-input models, 2D-input models with different post-processing strategies, and 3D-input models on tumor classification by using different metrics. The receiver operating characteristic (ROC) curves and *t*-distributed stochastic neighbor embedding (*t*-SNE) visualization [41] are used for visual comparison of feature learning. In the end, the state-of-the-art results (the top-15 scores) on the database are presented.

3.1. Performance of 2D-Input CNN Models

Table 1 shows the metric values of the 2D-input CNN models with and without HV and SV strategies on average after five-fold cross validation. It indicates that both the voting strategies improve the prediction results, and the SV strategy achieves better performance than the HV approach. Specifically, the SV strategy increases the SPE of ResNet34 from 0.757 to 0.963 (0.206 ↑), the SEN of MedViT from 0.808 to 0.987 (0.179 ↑), the SEN of HiFuse from 0.685 to 0.864 (0.179 ↑), and the SEN of MedMamba from 0.726 to 0.975 (0.249 ↑). It is also found that the HV approach causes slightly inferior AUC values when ResNet34 or HiFuse perform as the baseline for lesion classification.

Table 1. Tumor classification using 2D-input models with and without the strategies.

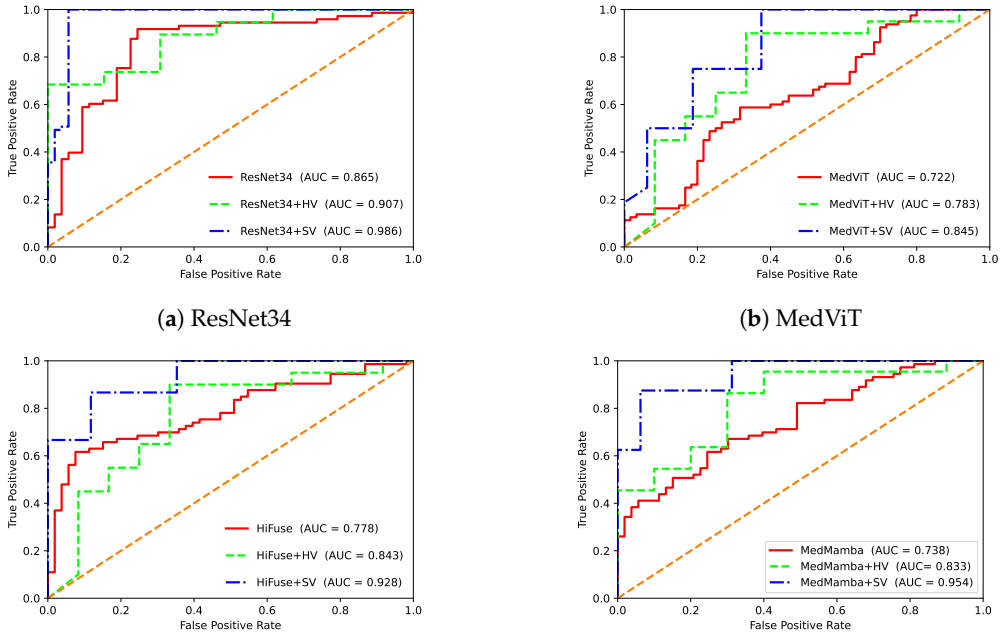
	Epoch	ACC	SEN	SPE	AUC	Score	Time (min)
ResNet34 [28]	100	0.865	0.942	0.757	0.936	0.901	63.2
ResNet34 + HV	100	0.907	0.990	0.864	0.907	0.907	
ResNet34 + SV	100	0.986	0.990	0.963	0.986	0.986	
MedViT [29]	100	0.699	0.808	0.547	0.772	0.736	88.3
MedViT + HV	100	0.783	0.955	0.611	0.783	0.783	
MedViT + SV	100	0.845	0.987	0.683	0.845	0.845	
HiFuse [30]	100	0.778	0.685	0.906	0.883	0.831	71.7
HiFuse + HV	100	0.843	0.735	0.951	0.843	0.843	
HiFuse + SV	100	0.928	0.864	0.980	0.928	0.928	
MedMamba [31]	100	0.738	0.726	0.755	0.799	0.769	76.7
MedMamba + HV	100	0.833	0.824	0.842	0.833	0.833	
MedMamba + SV	100	0.954	0.975	0.943	0.954	0.954	

Among the 2D-input CNN models, ResNet34 achieves the highest AUC value (AUC, 0.936), followed by HiFuse (AUC, 0.883) and MedMamba (AUC, 0.799), while MedViT causes the lowest AUC value (AUC, 0.772) and SPE value (SPE, 0.547) on the classification task. In addition, ResNet34 requires the least time consumption on model training (63.2 min per 100 epochs). The training stage of all the 2D-input CNN models takes more than 1 h, and MedViT lasts for about 88.3 min.

3.2. ROC Curves of 2D-Input CNN Models

Figure 7 presents a one-time experimental ROC curve for the four 2D-input CNN models by using the baseline (red, solid line), the baseline with HV (green, dashed line), and the baseline with SV (blue, dot-dashed line) for identification.

According to the ROC curves, the baseline ResNet34 achieves the best performance, while both the voting strategies, HV and SV, enable further improvement in the classification results. In addition, the SV strategy can enhance the prediction performance of HiFuse and MedMamba from AUC < 0.80 to AUC > 0.92, and the values are dramatically increased.



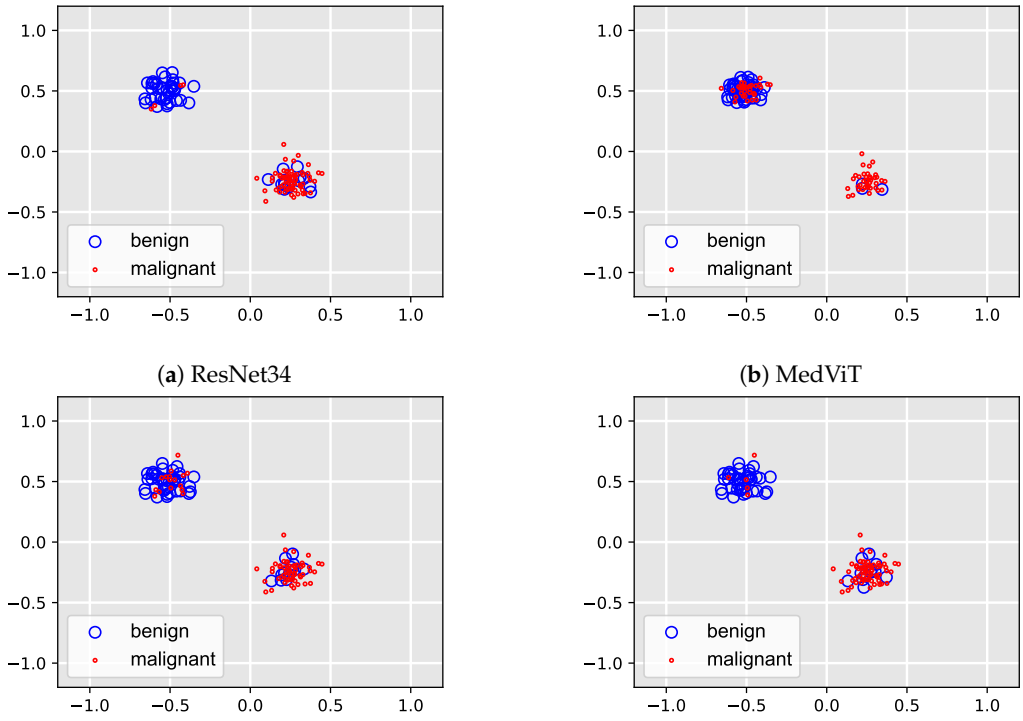
(**a**) ResNet34 (**b**) MedViT

(**c**) HiFuse (**d**) MedMamba

Figure 7. ROC curves of the four 2D-input CNN models from the baseline (red, solid line), to the baseline with HV (green, dashed line) and that with SV (blue, dot-dashed line).

3.3. Visualization with t-SNE of 2D-Input CNN Models

Figure 8 presents the *t*-SNE visualization of learned features of 2D-input CNN models. In each plot, large blue circles represent benign slices, and small red circles indicate malignant slices.



(**a**) ResNet34 (**b**) MedViT

(**c**) HiFuse (**d**) MedMamba

Figure 8. Visualization with *t*-SNE of learned features of the 2D-input CNN models. In each plot, large blue circles stand for benign slices, and small red circles denote malignant slices.

In the projection space, the learned features from the 2D-input models can effectively separate benign and malignant ones, since the blue and red circles are visually separable. Figure 8a and Figure 8d, respectively, indicate that 4 and 5 malignant slices are misclassified as benign by ResNet34 and MedMamba, while Figure 8b shows that three benign slices are wrongly predicted as malignant by MedViT.

3.4. Performance Comparison to 3D-Input CNN Models

The performance of tumor classification using the 2D-input CNN models with SV strategy and the 3D-input CNN models is summarized in Table 2. The best values and the worst values of the metrics are in boldface and underlined, respectively.

Table 2. Tumor classification using voxel-weighted 2D-input models and 3D-input models.

	Epoch	ACC	SEN	SPE	AUC	Score	Time (min)
ResNet34 [28] + SV	100	**0.986**	**0.990**	**0.963**	**0.986**	**0.986**	**63.2**
MedViT [29] + SV	100	0.845	0.987	0.683	0.845	0.845	88.3
HiFuse [30] + SV	100	0.928	0.864	0.980	0.928	0.928	71.7
MedMamba [31] + SV	100	0.954	0.975	0.943	0.954	0.954	76.7
M3T3D [32]	100	0.598	<u>0.078</u>	0.962	<u>0.525</u>	0.562	<u>511.4</u>
ResNeXt3D [33]	100	0.548	0.667	0.411	0.596	0.572	246.8
DenseNet3D [34]	100	<u>0.468</u>	0.642	<u>0.294</u>	0.622	0.545	197.5
ResNet3D [35]	100	0.512	0.496	0.528	0.554	<u>0.533</u>	181.1
M3T3D [32]	300	0.633	0.294	0.905	0.605	0.619	1482.5
ResNeXt3D [33]	300	0.581	0.727	0.553	0.634	0.608	656.4
DenseNet3D [34]	300	0.522	0.708	0.457	0.653	0.588	583.3
ResNet3D [35]	300	0.564	0.642	0.638	0.639	0.602	553.7

When the models are trained for 100 epochs, the metric values of 2D-input models with the SV strategy are much better than those using 3D-input models. In general, all the metric scores of 2D-input CNN models are larger than 0.84, and the training times are less than 90 min. Among the 3D-input CNN models, DenseNet3D achieves the highest AUC 0.622, M3T3D leads to the worst SEN 0.078, and ResNeXt3D causes relatively good balance of SEN and SPE values. Comparatively, the metric values of 3D-input CNN models are much worse than those of the ResNet34 + SV framework.

For further understanding the effect of training epoch numbers on the classification performance, the results of 300 epochs of 3D-input networks are shown. In comparison to the 3D-input models trained with 100 epochs, the metric values are correspondingly increased but slightly. For instance, ResNeXt3D achieves 0.033 ↑ in ACC, 0.060 ↑ in SEN, 0.142 ↑ in SPE, and 0.038 ↑ in AUC. However, its minor performance improvement is at the cost of 200 additional epochs of iteration and 409.6 min in model training.

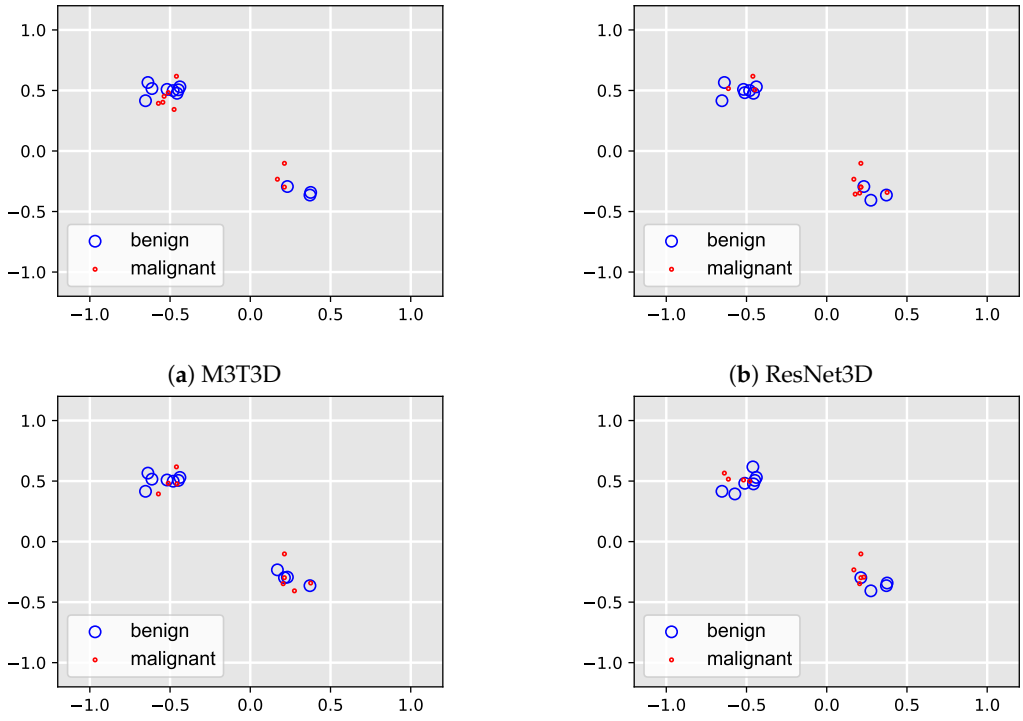
In addition, the 3D-input CNN models take over 3 h (>180 min) to complete the 100 epochs of model training, which are significantly longer than those 2D-input CNN models (≈60 min). ResNet3D, an 3D extension of ResNet18 model, is the fastest among the 3D-input models, requiring 181.1 min, which is three times longer than the time taken by ResNet34 for model training.

3.5. Visualization with t-SNE of 3D-Input Networks

The *t*-SNE of learned features of 3D-input CNN models is shown in Figure 9. In each plot, a large blue circle stands for a benign volume case, and a small red circle denotes a malignant volume case. Because of the limited number of volumetric samples, five-fold cross validation leads to around 20 cases in the plots.

The projection space of the one-time experiment shows that both blue circles and red circles are mixed with each other. This indicates that the benign and malignant volumetric

cases are difficult to be separated, and the intrinsic features of benign and malignant lesions are not well learned.



(**a**) M3T3D (**b**) ResNet3D

(**c**) DeseNet3D (**d**) ResNet3D

Figure 9. Visualization with *t*-SNE of learned features of the 3D-input models. In each plot, large blue circles stand for benign volumetric cases, and small red circles denote malignant volumetric cases.

3.6. The State-of-the-Art Achievement on the Database

Figure 10 illustrates the state-of-the-art performance on the database, displaying the results of top 15 teams in the ABUS challenge and our ResNet + SV model. It should be noted that these teams use the 100 samples for model training and the trained models are evaluated on the unreleased testing set. The horizontal axis represents the rankings of the top teams, with the 16th position designated for the ResNet + SV model, while the vertical axis indicates the score values.

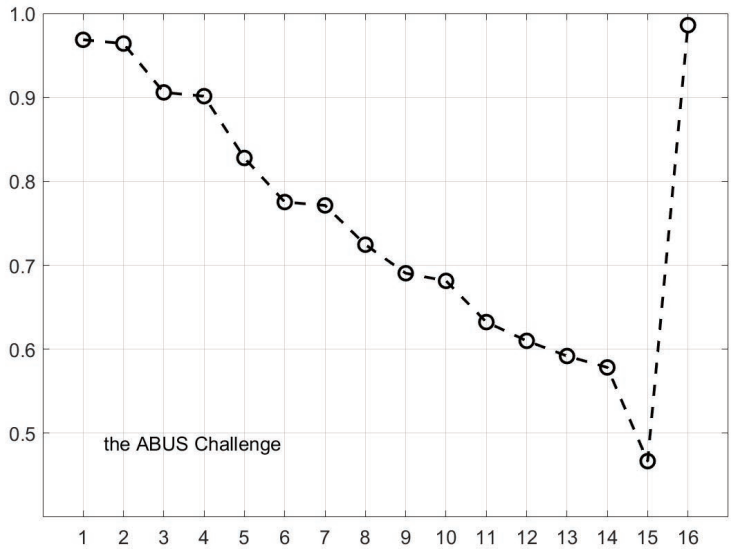


Figure 10. The achievement of tumor classification scores on the ABUS challenge.

Four teams achieved scores greater than 0.90, with the top team reaching 0.9686, which is slightly lower than the ResNet + SV model (score = 0.986). The fifth-ranked team obtained a score of 0.8278, while the remaining teams scored below 0.80. This finding indicates that the majority of the prediction models (11 out of 15) are struggling to effectively classify the ABUS volumetric images into benign and malignant groups.

4. Discussion

BC remains the leading cause of cancer-related deaths among women worldwide, underscoring the urgent need for effective screening and diagnostic tools. Three-dimensional ABUS has emerged as a promising method for improving the screening and diagnosis of women with dense breast tissue, offering numerous advantages over traditional imaging techniques. While a limited number of studies have begun to explore deep learning-based approaches for tumor classification in ABUS, the challenge of balancing time efficiency with predictive accuracy has yet to be adequately addressed.

This study introduces a novel strategy, termed the SV strategy, which employs voxel weighting to enhance classification performance. This method can be seamlessly integrated into any 2D-input CNN models. To the best of our knowledge, this is the first application of voxel-level probabilities for distinguishing between benign and malignant tumors in the context of volumetric tumor classification, marking a significant advancement in the use of deep learning for BC diagnosis.

The SV strategy significantly enhances tumor classification performance when applied to 2D-input CNN models. As shown in Table 1, this strategy leads to notable improvements in various metrics compared to the baseline models. For instance, MedMamba achieves reasonable results (ACC, 0.738; SEN, 0.726; SPE, 0.755; AUC, 0.799), while the SV-powered MedMamba demonstrates excellent performance (ACC, 0.954; SEN, 0.975; SPE, 0.943; AUC, 0.954). Several factors may contribute to this improvement. First and foremost, the SV strategy incorporates voxel probabilities into the computation of volumetric tumor classification (see Figure 2). This approach effectively balances both the number of voxels and slice-wise probabilities in the final prediction. Secondly, the baseline 2D-input CNN models are proficient at slice-wise tumor classification. When a slice contains sufficient voxels for deep learning-based hierarchical feature representation, these models excel in slice-level classification. This is supported by the *t*-SNE visualization (Figure 8), which illustrates a clear separation between benign and malignant image slices. However, tumors exhibit diverse shapes, sizes, and textures, and the baseline models may struggle with smaller slices, leading to inaccurate predictions. Furthermore, the SV strategy outperforms the HV strategy, indicating that voxel-level weighting is more effective than slice-level weighting. Both the number of voxels and predicted probabilities play crucial roles in the accurate classification of tumor volumes (Figure 7). It is important to note that this concept has parallels in the artificial intelligence community [42,43]. Unlike approaches that weight multiple classifiers or networks [27], the proposed strategy re-weights voxel importance to enhance prediction performance.

The 3D-input CNN models demonstrate poor performance in ABUS tumor classification, as indicated by their metrics (AUC < 0.65) and as shown in Table 2. Although these models have witnessed success in disease classification using multi-plane multi-slice Transformer (M3T3D [32]), fusing ResNeXt and Bi-LSTM (ResNeXt3D [33]), and densely connecting global and local features (DenseNet3D [34]), several factors contribute to their shortcomings in this context. Firstly, a limited number of 300 training epochs hinders the iterative optimization of representation learning. The training parameters, such as learning rate, patch size, and loss function, are critical for prediction performance [44]. However, grid searching of optimal parameters is bound to dramatically increase the time cost in model training. As shown in Table 2, when the number of iteration epochs increases, the time cost is correspondingly and linearly increased. Compared to 2D-input CNN models, 3D data input increases computational complexity, requiring high memory and longer training time. Secondly, the insufficient number of ABUS tumor cases restricts the effective

training of 3D-input CNN models. Deep 3D-input networks involve numerous hidden parameters that necessitate a large-scale database for optimal hyper-parameter tuning. Whether data augmentation, transfer learning, and domain adaptation can address this issue [45–47] requires further investigation. It is important to note that the ABUS database, containing 100 volumetric cases, remains the only publicly available dataset in the field of 3D US imaging. The release of significantly larger databases in the future could greatly advance research in this domain, facilitating everything from algorithm development to fair performance comparisons, and ultimately paving the way for more accurate and robust 3D US image analysis. Lastly, due to the imaging principles involved, the quality of ABUS images is generally inferior to that of magnetic resonance images and computerized tomographic images. This poor image quality is also the reason of poor prediction performance. The quality of ABUS images heavily relies on the device and acquisition procedure and is further affected by artifacts, noise, and nipple shadow [16]. Enhancing ABUS imaging quality remains a significant challenge that warrants further attention.

Several limitations remain in the current study. Firstly, the proposed SV strategy can only be integrated into 2D-input networks for 3D object classification. It benefits the ABUS lesion prediction, while its potential for improving performance of other 3D object classification tasks is scheduled in our future studies. Secondly, a limited number of epochs were used for hyper-parameter optimization, which may hinder the deep representation learning process [48]. And consequently, the performance of the 3D-input CNNs involved may be underestimated, despite the fact that their predictive performance is not related to our finding that the proposed SV strategy improves ABUS lesion classification. However, increasing the number of training epochs, along with hyper-parameter optimization (e.g., learning rate and batch size), requires substantial additional time, effort, and funding. Thirdly, the number of training cases is insufficient due to the challenge in ABUS data collection. To relieve this issue, five-fold cross validation is conducted for robust evaluation and selection of better 2D-input CNN models. Most importantly, the release of larger databases is highly desirable, since diverse and large-scale databases can significantly enhance research progress, improve model generalization, and facilitate more comprehensive evaluations. To reduce labor costs, incorporating noisy and augmented data could be considered, although their effects on model training are still under investigation [24]. Fourthly, on the TDSC-ABUS2023 challenge, the prediction score of the first-ranked team is slightly inferior to our ResNet34 + SV model. However, due to limited open resource of the challenge results, the details of these models are not yet available. Finally, data samples could be stratified into different groups based on factors such as tumor size and image quality, which would help elucidate the elements contributing to tumor classification.

5. Conclusions

ABUS provides high-resolution imaging of internal anatomy in a standardized, reproducible, and reliable manner, significantly enhancing BC screening and diagnosis. However, when applying deep learning-based volumetric image analysis to ABUS, a critical challenge remains on how to effectively balance computational effectiveness and efficiency. This study introduces a novel soft voting strategy that integrates both tumor sizes and voxel-level malignancy probabilities, offering a more nuanced approach to classification. This strategy can serve as a useful post-processing step for 2D-input deep networks, with the potential to improve the accuracy and reliability of tumor assessments, thereby supporting diagnostic performance.

Author Contributions: Conceptualization, S.Y., X.L. and Q.S.; Data curation, S.Y., X.L. and S.Z.; Formal analysis, S.Y., Y.X. and Q.S.; Funding acquisition, Y.X. and Q.S.; Investigation, S.Y., Y.X. and Q.S.; Methodology, S.Y., X.L. and S.Z.; Project administration, Y.X. and Q.S.; Software, S.Y., X.L. and S.Z.; Supervision, Q.S.; Validation, S.Z. and Y.X.; Visualization, X.L and S.Z.; Writing—original draft, S.Y. and X.L.; Writing—review and editing, Y.X. and Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work was in part supported by National Key Research and Develop Program of China (Grant No. 2022ZD0115901, and 2022YFC2409000), National Natural Science Foundation of China (Grant No. 62177007, U20A20373, and 82202954), China-Central Eastern European Countries High Education Joint Education Project (Grant No. 202012), Medium- and Long-term Technology Plan for Radio, Television and Online Audiovisual (Grant No. ZG23011), and Architecture Design and Implementation of Building Common Business Services based on Rapid Development Platform (Grant No. 2222000247). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset supporting the current study is available online (the Tumor Detection, Segmentation and Classification Challenge on Automated 3D Breast Ultrasound (ABUS) 2023, namely TDSC-ABUS 2023, https://tdsc-abus2023.grand-challenge.org, accessed on 5 December 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BC	breast cancer
MAM	mammography
US	ultrasound
HHUS	hand-held ultrasound
3D	three-dimensional
ABUS	automated breast ultrasound
TDSC-ABUS2023	Tumor Detection, Segmentation and Classification Challenge on ABUS 2023
SV	soft voting
CNN	convolutional neural network
MHCA	multi-head convolutional attention
LFFN	local feed-forward network
ESA	efficient self-attention
ECB	efficient convolutional block
LTB	local-token block
HFF	hierarchical feature fusion
MLP	multi-layer perceptron
SSM	structured state-space model
TP	true positive
TN	true negative
FP	false positive
FN	false negative
ACC	accuracy
SEN	sensitivity
SPE	specificity
AUC	area under the curve
ROC	receiver operating characteristic
t-SNE	*t*-distributed stochastic neighbor embedding

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [CrossRef] [PubMed]
3. Giaquinto, A.N.; Sung, H.; Miller, K.D.; Kramer, J.L.; Newman, L.A.; Minihan, A.; Jemal, A.; Siegel, R.L. Breast cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 524–541. [CrossRef] [PubMed]

4. Tao, X.; Li, T.; Gandomkar, Z.; Brennan, P.C.; Reed, W.M. Incidence, mortality, survival, and disease burden of breast cancer in China compared to other developed countries. *Asia-Pac. J. Clin. Oncol.* **2023**, *19*, 645–654. [CrossRef]
5. Zou, L.; Yu, S.; Meng, T.; Zhang, Z.; Liang, X.; Xie, Y. A technical review of convolutional neural network-based mammographic breast cancer diagnosis. *Comput. Math. Methods Med.* **2019**, *2019*, 6509357. [CrossRef]
6. Richman, I.B.; Long, J.B.; Soulos, P.R.; Wang, S.-Y.; Gross, C.P. Estimating breast cancer overdiagnosis after screening mammography among older women in the United States. *Ann. Intern. Med.* **2023**, *176*, 1172–1180. [CrossRef]
7. Yang, S.; Gao, X.; Liu, L.; Shu, R.; Yan, J.; Zhang, G.; Xiao, Y.; Ju, Y.; Zhao, N.; Song, H. Performance and reading time of automated breast US with or without computer-aided detection. *Radiology* **2019**, *292*, 540–549. [CrossRef]
8. Littrup, P.J.; Mehrmohammadi, M.; Duric, N. Breast tomographic ultrasound: The spectrum from current dense breast cancer screenings to future theranostic treatments. *Tomography* **2024**, *10*, 554–573. [CrossRef]
9. Yu, S.; Wu, S.; Zhuang, L.; Wei, X.; Sak, M.; Neb, D.; Hu, J.; Xie, Y. Efficient segmentation of a breast in B-mode ultrasound tomography using three-dimensional GrabCut (GC3D). *Sensors* **2017**, *17*, 1827. [CrossRef]
10. Allajbeu, I.; Hickman, S.E.; Payne, N.; Moyle, P.; Taylor, K.; Sharma, N.; Gilbert, F.J. Automated breast ultrasound: Technical aspects, impact on breast screening, and future perspectives. *Curr. Breast Cancer Rep.* **2021**, *13*, 141–150. [CrossRef]
11. Boca, I.; Ciurea, A.I.; Ciortea, C.A.; Dudea, S.M. Pros and cons for automated breast ultrasound (ABUS): A narrative review. *J. Pers. Med.* **2021**, *11*, 703. [CrossRef] [PubMed]
12. Vourtsis, A.; Kachulis, A. The performance of 3D ABUS versus HHUS in the visualisation and BI-RADS characterisation of breast lesions in a large cohort of 1,886 women. *Eur. Radiol.* **2018**, *28*, 592–601. [CrossRef] [PubMed]
13. Wolterink, K.F.; Mumin, N.A.; Appelman, L.; Derks-Rekers, M.; Imhof-Tas, M.; Lardenoije, S.; van der Leest, M.; Mann, R.M. Diagnostic performance of 3D automated breast ultrasound (3D-ABUS) in a clinical screening setting—A retrospective study. *Eur. Radiol.* **2024**, *34*, 5451–5460. [CrossRef] [PubMed]
14. Rahmat, K.; Mumin, N.A.; Ng, W.L.; Taib, N.A.M.; Chan, W.Y.; Hamid, M.T.R. Automated breast ultrasound provides comparable diagnostic performance in opportunistic screening and diagnostic assessment. *Ultrasound Med. Biol.* **2024**, *50*, 112–118. [CrossRef]
15. Zhang, J.; Wu, J.; Zhou, X.S.; Shi, F.; Shen, D. Recent advancements in artificial intelligence for breast cancer: Image augmentation, segmentation, diagnosis, and prognosis approaches. *Semin. Cancer Biol.* **2023**, *96*, 11–25. [CrossRef]
16. Pengiran Mohamad, D.N.F.; Mashohor, S.; Mahmud, R.; Hanafi, M.; Bahari, N. Transition of traditional method to deep learning based computer-aided system for breast cancer using automated breast ultrasound system (ABUS) images: A review. *Artif. Intell. Rev.* **2023**, *56*, 15271–15300. [CrossRef]
17. Tan, T.; Platel, B.; Huisman, H.; Sánchez, C.I.; Mus, R.; Karssemeijer, N. Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation. *IEEE Trans. Med. Imaging* **2012**, *31*, 1034–1042. [CrossRef]
18. Wang, Y.; Choi, E.J.; Choi, Y.; Zhang, H.; Jin, G.Y.; Ko, S.-B. Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound Med. Biol.* **2020**, *46*, 1119–1132. [CrossRef]
19. Xiang, H.; Huang, Y.-S.; Lee, C.-H.; Chien, T.-Y.C.; Lee, C.-K.; Liu, L.; Li, A.; Lin, X.; Chang, R.-F. 3-D Res-CapsNet convolutional neural network on automated breast ultrasound tumor diagnosis. *Eur. J. Radiol.* **2021**, *138*, 109608. [CrossRef]
20. Zhou, Y.; Chen, H.; Li, Y.; Liu, Q.; Xu, X.; Wang, S.; Yap, P.-T.; Shen, D. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Med. Image Anal.* **2021**, *70*, 101918. [CrossRef]
21. Wang, Q.; Chen, H.; Luo, G.; Li, B.; Shang, H.; Shao, H.; Sun, S.; Wang, Z.; Wang, K.; Cheng, W. Performance of novel deep learning network with the incorporation of the automatic segmentation network for diagnosis of breast cancer in automated breast ultrasound. *Eur. Radiol.* **2022**, *32*, 7163–7172. [CrossRef] [PubMed]
22. Ding, W.; Zhang, H.; Zhuang, S.; Zhuang, Z.; Gao, Z. Multi-view stereoscopic attention network for 3D tumor classification in automated breast ultrasound. *Expert Syst. Appl.* **2023**, *234*, 120969. [CrossRef]
23. Yang, Z.; Fan, T.; Smedby, Ö.; Moreno, R. 3D breast ultrasound image classification using 2.5 D deep learning. *Int. Workshop Breast Imaging* **2024**, *13174*, 443–449.
24. Karimi, D.; Dou, H.; Warfield, S.K.; Gholipour, A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Med. Image Anal.* **2020**, *65*, 101759. [CrossRef] [PubMed]
25. Algan, G.; Ulusoy, I. Image classification with deep learning in the presence of noisy labels: A survey. *Knowl.-Based Syst.* **2021**, *215*, 106771. [CrossRef]
26. Yu, S.; Chen, M.; Zhang, E.; Wu, J.; Yu, H.; Yang, Z.; Ma, L.; Gu, X.; Lu, W. Robustness study of noisy annotation in deep learning based medical image segmentation. *Phys. Med. Biol.* **2020**, *65*, 175007. [CrossRef]
27. Bianchessi, T.; Tampu, I.E.; Blystad, I.; Lundberg, P.; Nyman, P.; Eklund, A.; Haj-Hosseini, N. Pediatric brain tumor type classification using deep learning on MR images from the children's brain tumor network. *medRxiv* **2023**. [CrossRef]
28. Yue, Y.; Li, Z. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; Volume 1, pp. 770–778.
29. Manzari, O.N.; Ahmadabadi, H.; Kashiani, H.; Shokouhi, S.B.; Ayatollahi, A. MedViT: A robust vision transformer for generalized medical image classification. *Comput. Biol. Med.* **2023**, *157*, 106791. [CrossRef]
30. Huo, X.; Sun, G.; Tian, S.; Wang, Y.; Yu, L.; Long, J.; Zhang, W.; Li, A. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification. *Biomed. Signal Process. Control* **2024**, *87*, 105534. [CrossRef]
31. Yue, Y.; Li, Z. Medmamba: Vision mamba for medical image classification. *arXiv* **2023**, arXiv:2403.03849.

32. Jang, J.; Hwang, D. M3T: Three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; Volume 1, pp. 20718–20729.
33. Wang, X.; Yi, J.; Li, Y. Application of fusion model of 3D-ResNeXt and Bi-LSTM network in Alzheimer's disease classification. In Proceedings of the 2022 6th International Conference on Communication and Information Systems, Chongqing, China, 14–16 October 2022; pp. 136–140.
34. Cui, R.; Liu, M. Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 2099–2107. [CrossRef]
35. Tran, D.; Wang, H.; Torresani, L.; Ray, J.; LeCun, Y.; Paluri, M. A closer look at spatiotemporal convolutions for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6450–6459.
36. Shafiq, M.; Gu, Z. Deep residual learning for image recognition: A survey. *Appl. Sci.* **2022**, *12*, 8972. [CrossRef]
37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations, Virtual, 26 April–1 May 2020.
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 1–15.
39. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
40. Zhang, S.; Zheng, D.; Hu, X.; Yang, M. Bidirectional long short-term memory networks for relation classification. In Proceedings of the Pacific Asia Conference on Language, Information and Computation, Shanghai, China, 30 October–1 November 2015; pp. 73–78.
41. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
42. Yang, Y.; Lv, H.; Chen, N. A survey on ensemble learning under the era of deep learning. *Artif. Intell. Rev.* **2023**, *56*, 5545–5589. [CrossRef]
43. Mahajan, P.; Uddin, S.; Hajati, F.; Moni, M.A. Ensemble learning for disease prediction: A review. *Healthcare* **2023**, *11*, 1808. [CrossRef] [PubMed]
44. Xiao, T.; Liu, L.; Li, K.; Qin, W.; Yu, S.; Li, Z. Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *BioMed Res. Int.* **2018**, *1*, 4605191. [CrossRef]
45. He, W.; Zhang, C.; Dai, J.; Liu, L.; Wang, T.; Liu, X.; Jiang, Y.; Li, N.; Xiong, J.; Wang, L.; et al. A statistical deformation model-based data augmentation method for volumetric medical image segmentation. *Med. Image Anal.* **2024**, *91*, 102984. [CrossRef]
46. Yu, S.; Liu, L.; Wang, Z.; Dai, G.; Xie, Y. Transferring deep neural networks for the differentiation of mammographic breast lesions. *Sci. China Technol. Sci.* **2019**, *62*, 441–447. [CrossRef]
47. Zheng, B.; Zhang, R.; Diao, S.; Zhu, J.; Yuan, Y.; Cai, J.; Shao, L.; Li, S.; Qin, W. Dual domain distribution disruption with semantics preservation: Unsupervised domain adaptation for medical image segmentation. *Med. Image Anal.* **2024**, *97*, 103275. [CrossRef]
48. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.