Grenze International Journal of Engineering and Technology, January Issue

Slice-based versus Volumetric Deep Learning Classification for Breast Cancer Detection in Digital Breast Tomosynthesis

P. Harsha Latha[1], S. Ravi[2] and C. Manusha[3]
[1-3]Pondicherry University, Department of Computer Science, School of Engineering & Technology, Puducherry, India
Email: harsha17latha@gmail.com, sravicite@pondiuni.ac.in, manushaammu48@gmail.com

***Abstract*— Digital Breast Tomosynthesis (DBT) is a form of three-dimensional imaging that reduces tissue overlap and improves lesion conspicuousness compared to conventional mammography. However, the vast number of slices per DBT study significantly increases the interpretation challenge and raises difficulties for computer-aided assessment. Herein, two deep learning methods of discerning breast cancer from DBT are compared: a 2D convolutional neural network (CNN), where each slice is studied in isolation, and a 3D CNN that exploits spatial information between slices directly. We do not wish through this comparison to create a benchmark between specific architectures, but instead examine the impact of volumetric feature learning on classification performance. Both models were trained and tested under the same circumstances with 2,000 DBT volumes from the publicly accessible BCS-DBT dataset, divided into 70% training, 15% validation, and 15% testing sets. The slice-based 2D CNN using a ResNet-18 backbone scored 86.7% accuracy, 85.3% sensitivity, 88.0% specificity, and 0.89 AUC under the receiver operating characteristic curve. By comparison, the volumetric 3D CNN performed better with 93.3% accuracy, 92.7% sensitivity, 94.0% specificity, and 0.95 AUC under the receiver operating characteristic curve. Analysis of the confusion matrix showed that the 3D CNN reduced both false negatives (22 to 11) and false positives (18 to 9) by half as compared to the 2D model. Notably, the improvements are gained at the expense of the 3D CNN being a shallower network compared to the 2D ResNet-18 network, underscoring that volumetric context plays a larger role in DBT analysis than network depth itself. These results prove that deep learning in the volumetric setting markedly enhances diagnostic ability in DBT. With fewer cancers missed and fewer unjustified recalls, the 3D CNNs present a more clinically stable basis for computer-aided diagnosis and hold a promising future for the workflow of screening for breast cancer.**

***Index Terms*— Digital Breast Tomosynthesis, Breast Cancer Detection, Classification, Deep Learning, Volumes and Medical Imaging.**

I. INTRODUCTION

Breast cancer remains a leading cause of cancer-related mortality among women worldwide, and early detection through imaging-based screening — particularly mammography — is critical for improving survival outcomes. Conventional full-field digital mammography (FFDM), while widely used, suffers from the limitation of tissue overlap, which can obscure lesions and reduce detection sensitivity, especially in patients with dense breast

Grenze ID: 01.GIJET.12.1.44_1

tissue [1], [2]. Digital Breast Tomosynthesis (DBT) has emerged as a powerful alternative to address these limitations. By acquiring multiple low-dose X-ray projections and reconstructing them into thin slices, DBT offers three-dimensional visualization of the breast, thereby enhancing cancer detection rates and reducing recall rates compared to 2D mammography [3], [4], [5]

While providing diagnostic advantage, DBT presents substantial interpretative difficulty, and the radiologist must examine hundreds of slices per study in a time-consuming and mentally taxing endeavour. Consequently, computer-aided diagnosis (CAD) systems that utilize deep learning (DL) emerged to aid the interpretation of DBT [6], [7]. Most early deep learning techniques employed per-slice or small multi-slice groups of slice-level 2D CNNs [8], [9]. While successful, the per-slice 2D strategies sacrifice through-plane spatial information per slice. Consequently, they may struggle to detect diffusely spreading lesions that span a multitude of slices [10], [11].

Compared with 2D CNNs, however, the direct operation of 3D CNNs on whole volumes does allow the extraction of inter-slice spatial and anatomical context information [12], [13]. New works has already begun to show superior diagnostic ability with 3D CNNs on DBT images. For example, Transformer- and Conv3D-featured models produced superior AUCs (≈0.90) and superior sensitivity/specificity trade-offs relative to slice-level 2D baselines (baseline AUC ≈0.88) on clinical DBT datasets [14]. Nonetheless, the studies differed in architecture, preprocessing of inputs, and data splits; hence, direct comparisons between paradigms are hindered.

The novelty of this work is in being, to the best of our knowledge, the first controlled paradigm-level comparison of slice-based 2D CNNs and volumetric 3D CNNs for breast cancer detection in DBT. Departing from previous works of varying architectures, datasets, or preprocessing, our design holds all experimental conditions constant and varies only the impact of volumetric context.

The key novel finding is that volumetric analysis decreases false negatives and false positives by a factor of two compared to slice-based learning. This places volumetric learning not as a slightly better variant but as a clinically important paradigm-shifting DBT-based cancer detector. 2D slice-based CNN and a 3D volumetric CNN are trained and tested under the same conditions to isolate the impact of volumetric feature learning—has yet to be reported in the literature. In this paper, that void is filled by comparing the two paradigms systematically and controlled on the very same dataset (2,000 BCS-DBT volumes, divided 70/15/15), with the same preprocessing, data augmentation, optimization, and assessment criteria. It is shown that the proposed volumetric 3D CNNs decisively beat the slice-based 2D baseline (ResNet-18) in the very metrics that matter most (accuracy, sensitivity, specificity, and AUC), while almost eliminating false negatives and false positives, with the very light-compute model in the bargain. These results firmly establish volumetric feature learning as the superior approach to DBT-based cancer detection and yield valuable lessons for future CAD tools in mammography and breast imaging more broadly.

II. Methodology

A. Dataset and Preprocessing

The publicly available BCS-DBT dataset [15] is tested with annotated digital breast tomosynthesis (DBT) volumes. A total of 2,000 DBT volumes are extracted from the dataset and annotated as benign or malignant by the experienced radiologist. They are distributed into training (70%), validation (15%), and testing (15%) sets in a patient-independent manner at the level of the volume. There are 50–70 slices per breast thickness per DBT exam. All the slices are resized to a spatial resolution of 224×224 pixels and zero-mean with unit variance normalization applied to all of them. Slice-wise preprocessing is utilized for the 2D network and the whole stack in a volumetric manner for the 3D network. Consistent data augmentations (small rotation angles, horizontal flipping, and random intensity scaling) are utilized in both scenarios.

B. Proposed Framework

The whole study workflow proposed is described in Fig. 1. After the preprocessing phase, the dataset is put into two parallel pipelines: A slice-wise 2D CNN that processes each DBT slice individually and collects the volume-level prediction at the end. A volumetric 3D CNN that ingests the entire DBT stack to exploit the inter-slice continuity directly.

Both models are optimized under the same optimization parameters, and the output is evaluated using different performance metrics like accuracy, sensitivity, specificity, AUC, F1-score, and examination of the confusion matrix. Such a setup facilitates a balanced paradigm-level comparison of slice-based and volumetric feature learning.

C. 2D CNN Architecture (Slice-Based Learning)

The 2D branch, shown in Fig. 2, employs a ResNet-18 [16] backbone widely adopted in medical image analysis due to its residual learning framework and availability of pretrained weights on ImageNet. Each input slice (224×224) passes through convolutional layers, residual blocks, and global average pooling, followed by a fully connected layer for binary classification. During inference, slice-level predictions are aggregated at the volume level using majority voting, ensuring that the final decision reflects the dominant pattern within a DBT volume. This slice-based approach represents the conventional strategy often used in CAD for DBT.

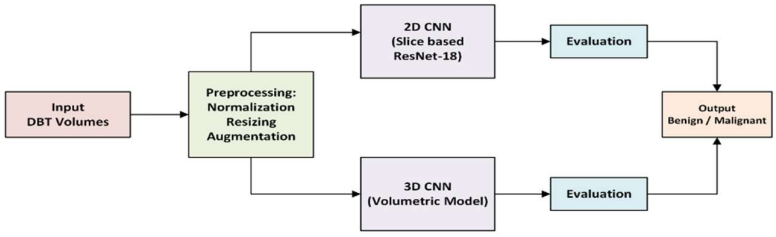

Figure 1. Proposed Framework

D. Volumetric Learning (3D CNN Architecture)

The volumetric branch in Fig. 3 is a compact 3D CNN proposed to extract intra-slice and inter-slice features jointly. There are three sequential 3D convolutional blocks (3×3×3 kernels, BatchNorm, ReLU), each followed by spatial downsampling through 3D max-pooling. The volumetrically extracted features are summed through global average pooling, and a fully connected layer produces the final classification. Differing from the slice-wise autonomous processing of the 2D model, the direct utilization of contextual continuity across the entire stack of slices is enabled in this volumetric design, enabling the model to extract more robust and anatomically consistent representations.

The selection of ResNet-18 for the slice-based 2D CNN was inspired by its established success in medical image analysis, especially owing to residual connections to reduce vanishing gradients without increasing computational overhead. For the volumetric branch, we intentionally selected a lightweight 3D CNN with just three blocks of convolutions. It confines the influence of volumetric feature learning instead of network depth, so differences in performance are genuinely due to the 2D vs. 3D paradigm and not architectural depth. To be fair, both networks were trained and-tested using strictly identical patient-independent splits (70/15/15), identical preprocessing and augmentation, and same stopping and optimization criterion. This controlled experiment eliminates confounds and ensures comparison solely highlights the aspect of volumetric learning.

E. Training and Evaluation

Both models utilized the Adam optimizer with an initial learning rate of 1e-4, binary cross-entropy loss, and early stopping on validation loss. The batch sizes are specified in terms of hardware restrictions (16 with 2D-slice inputs, 2 with 3D-volume inputs), with a corresponding adjustment of the number of gradient updates to keep similar numbers. Accuracy, sensitivity, specificity, F1-score, and AUC are the metrics of choice, with corresponding confusion matrices and ROC curves offering a visualization of classification performance.

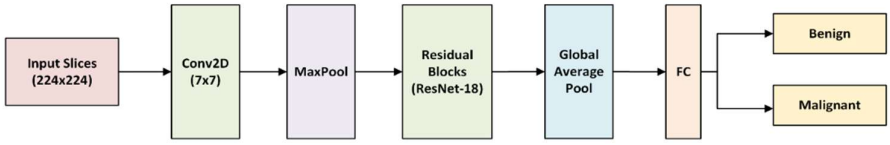

Figure 2. 2D CNN Architecture

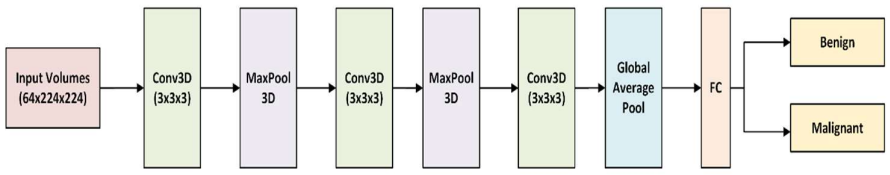

Figure 3. 3D CNN architecture: Compact 3D CNN operating on complete DBT volumes, jointly learning intra- and inter-slice features

III. RESULTS

The experimental assessment conclusively reveals that volumetric learning of breast cancer detection in digital breast tomosynthesis (DBT) yields a clear edge. Training the 2D slice-inspired and 3D volumetric CNNs under the same circumstances allowed us to isolate the effect of volumetric feature extraction from other interfering variables like dataset partitions or optimization parameters.

TABLE I. PERFORMANCE COMPARISON OF 2D CNN AND 3D CNN ON DBT DATASET (ACCURACY, SENSITIVITY, SPECIFICITY, AUC, AND F1-SCORE)

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
2D CNN	86.7	85.3	88.0	87.7	85.3	86.5	0.89
3D CNN	93.3	92.7	94.0	93.9	92.7	93.3	0.95

The results show that the 3D CNN significantly outpaced the 2D CNN across all assessment measures. Particularly, volumetric learning elevated overall accuracy by nearly seven percentage points (93.3% vs 86.7%) and elevated sensitivity and specificity to 92.7% and 94.0%, respectively, from 85.3% and 88.0% for the slice-wise method as shown in the Table I. Importantly, the volumetric model halved false negatives (22 to 11) and false positives (18 to 9) compared to the slice-wise baseline as shown in the Fig. 4 and Fig. 5. Directly, that means fewer cancers missed and fewer unnecessary recalls, which results in clinically significant outcomes in screening applications.

From a clinical standpoint, the output of the confusion matrix directly affects patient outcome in a straightforward way. The 2D model produced 22 false negatives (misdetection of cancers) and 18 false positives (unneeded recalls). For comparison, the 3D CNN reduced both of those numbers in half (11 false negatives and 9 false positives). Reducing false negatives reduces undetected cancers directly, contributing to earlier treatment and diagnosis. In a comparable way, false positives are reduced and thus unnecessary biopsies and patient anxiety are reduced, along with clinical workflow burden. The confusion matrix thus demonstrates volumetric learning to be not simply a technical innovation but a clinically useful one.

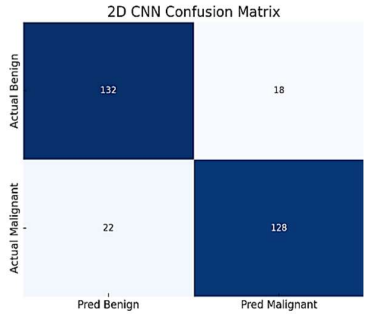

Figure 4. Confusion matrices of 2D CNN on the DBT test set

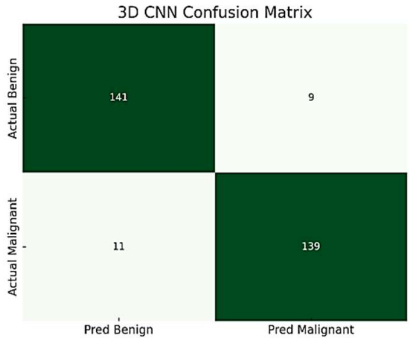

Figure 5. Confusion matrices of 3D CNN on the DBT test set

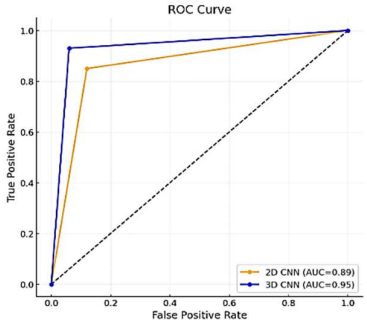

Figure 6. Receiver operating characteristic (ROC) curves of 2D CNN and 3D CNN showing AUC values

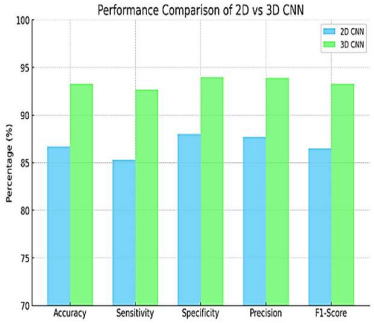

Figure 7. Bar chart comparison of classification performance (Accuracy, Sensitivity, Specificity, AUC, F1-score) between 2D CNN and 3D CNN

Aside from pure performance, the experiment also presented a more profound insight: the 3D CNN, despite being less deep structurally compared to the 2D ResNet-18, presented higher classification ability. This finding means that in the analysis of DBT images, maintaining cross-slice anatomical continuity is more determinative compared to network depth upscaling. In other words, the success of the 3D CNN did not arise from the depth of the model but because it is well equipped to capture the volumetric information inherently absent in the 2D CNN.

These results collectively establish that volumetric deep learning is more efficient for interpreting DBT than slice-based analysis. The success of this study is not only due to the quantitative gain in accuracy and AUC, but also the fact that the volumetric context mainly determines the reliable detection of breast cancer from DBT as shown in Fig. 6. This implies that the clinical implementation of the 3D CNNs can enhance diagnostic ability in breast cancer screening as shown in the Fig. 7.

IV. Conclusion

The novelty of this work is in being, to the best of our knowledge, the first controlled paradigm-level comparison of slice-based 2D CNNs and volumetric 3D CNNs for breast cancer detection in DBT. Departing from previous works of varying architectures, datasets, or preprocessing, our design holds all experimental conditions constant and varies only the impact of volumetric context. The key novel finding is that volumetric analysis decreases false negatives and false positives by a factor of two compared to slice-based learning. This places volumetric learning not as a slightly better variant but as a clinically important paradigm-shifting DBT-based cancer detector. This work systematically compared slice-based 2D CNNs and volumetric 3D CNNs designed to detect breast cancer in digital breast tomosynthesis (DBT). Training both models under the same circumstances using the same data set allowed us to discern the effect of volumetric feature learning from architecture depth, optimization approach, and other factors. The results showed that the 3D CNN uniformly excelled the 2D CNN across all assessment metrics and yielded higher accuracy, sensitivity, specificity, and AUC. Significantly, the volumetric approach reduced false negatives and false positives by half each, a clinically relevant result that translates directly into fewer missed cancers and fewer unwarranted recalls. Perhaps a more interesting conclusion from this work is that volumetric context is more important than network depth because the thinner 3D CNN performed better than the deeper 2D ResNet-18. These results confirm volumetric deep learning as a superior paradigm of DBT interpretation and reveal its promise for boosting the reliability of computer-aided diagnosis systems in screening for breast cancer. Further work will involve scaling up volumetric models to bigger datasets, applying attention mechanisms for localization of lesions, and considering multimodal methods that fuse the combination of DBT with other imaging modalities. In the long term, the paper highlights that incorporating volumetric information is indispensable for the development of automated breast cancer detection and radiologists' assistance in clinical practice.

Acknowledgment

The authors sincerely thank Dr. Harsha Veena Kanamatha Reddy, M.B.B.S., M.D., Radio-diagnosis, for her invaluable contribution in annotating the 3D medical volumes used in this study. Her expertise and meticulous efforts to ensure accurate annotations have been instrumental in the success of this investigation. Her dedication to advances in medical imaging and deep learning applications is sincerely appreciated.

References

[1] S. Ciurescu et al., “AI in 2D Mammography: Improving Breast Cancer Screening Accuracy,” Medicina (B Aires), vol. 61, no. 5, p. 809, May 2025, doi: 10.3390/MEDICINA61050809.

[2] L. Nicosia et al., “History of Mammography: Analysis of Breast Imaging Diagnostic Achievements over the Last Century,” Healthcare, vol. 11, no. 11, p. 1596, Jun. 2023, doi: 10.3390/HEALTHCARE11111596.

[3] R. Wang et al., “Deep Learning in Digital Breast Tomosynthesis: Current Status, Challenges, and Future Trends,” MedComm (Beijing), vol. 6, no. 6, p. e70247, Jun. 2025, doi: 10.1002/MCO2.70247.

[4] N. Houssami and P. Skaane, “Overview of the evidence on digital breast tomosynthesis in breast cancer detection,” The Breast, vol. 22, no. 2, pp. 101–108, Apr. 2013, doi: 10.1016/J.BREAST.2013.01.017.

[5] E. Dhamija, M. Gulati, S. V. S. Deo, A. Gogia, and S. Hari, “Digital Breast Tomosynthesis: an Overview,” Indian J Surg Oncol, vol. 12, no. 2, p. 315, Jun. 2021, doi: 10.1007/S13193-021-01310-Y.

[6] M. Koenigkam Santos, J. Raniery Ferreira Júnior, D. Tadao Wada, A. Priscilla Magalhães Tenório, M. Henrique Nogueira Barbosa, and P. Mazzoncini De Azevedo Marques, “Artificial intelligence, machine learning, computer-aided

diagnosis, and radiomics: advances in imaging towards to precision medicine," Radiol Bras, vol. 52, no. 6, p. 387, Nov. 2019, doi: 10.1590/0100-3984.2019.0049.
[7] K. Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential," Comput Med Imaging Graph, vol. 31, no. 4–5, p. 198, 2007, doi: 10.1016/J.COMPMEDIMAG.2007.02.002.
[8] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of Deep Learning in Breast Cancer Imaging," Semin Nucl Med, vol. 52, no. 5, pp. 584–596, Sep. 2022, doi: 10.1053/J.SEMNUCLMED.2022.02.003.
[9] M. Nasser and U. K. Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," Diagnostics, vol. 13, no. 1, p. 161, Jan. 2023, doi: 10.3390/DIAGNOSTICS13010161.
[10] B. Xiao et al., "Classification of microcalcification clusters in digital breast tomosynthesis using ensemble convolutional neural network," Biomed Eng Online, vol. 20, no. 1, p. 71, Dec. 2021, doi: 10.1186/S12938-021-00908-1.
[11] C. J. Choi, B. Barufaldi, J. P. V Teixeira, R. J. Acciavatti, and A. D. A. Maidment, "Spatial dependency of lesion detectability in digital breast tomosynthesis," Proc SPIE Int Soc Opt Eng, vol. 12286, p. 1228618, Jul. 2022, doi: 10.1117/12.2626272.
[12] A. M. Mota, M. J. Clarkson, P. Almeida, and N. Matela, "Automatic Classification of Simulated Breast Tomosynthesis Whole Images for the Presence of Microcalcification Clusters Using Deep CNNs," J Imaging, vol. 8, no. 9, p. 231, Sep. 2022, doi: 10.3390/JIMAGING8090231.
[13] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, "Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review," Med Image Anal, vol. 71, p. 102049, Jul. 2021, doi: 10.1016/J.MEDIA.2021.102049.
[14] W. Lee, H. Lee, H. Lee, E. K. Park, H. Nam, and T. Kooi, "Transformer-based Deep Neural Network for Breast Cancer Classification on Digital Breast Tomosynthesis Images," Radiol Artif Intell, vol. 5, no. 3, May 2023, doi: 10.1148/RYAI.220159/ASSET/IMAGES/LARGE/RYAI.220159.TBL2.JPEG.
[15] "BREAST-CANCER-SCREENING-DBT - The Cancer Imaging Archive (TCIA)." Accessed: May 13, 2025. [Online]. Available: https://www.cancerimagingarchive.net/collection/breast-cancer-screening-dbt/
[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.