THE SCIENTIFIC TEMPER (2025) Vol. 16 (12): 5173-5182

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.04

https://scientifictemper.com/

RESEARCH ARTICLE

A Unified Consistency-Calibrated Boundary-Aware Framework for Generalizable Skin Cancer Detection

J. Fathima Fouzia[1*], M. Mohamed Surputheen[2], M. Rajakumar[2]

Abstract

Skin cancer detection by automated methods faces significant challenges in generalizing across diverse patient populations. This limitation is due to the wide variability in the appearance of skin lesions and the lack of standardization in dermatological imaging. To address this issue, this research paper proposes a stability-scale calibrated threshold framework (C2BASC++), which is designed to improve robustness and diagnostic accuracy in hospital settings. By integrating threshold-sensitive feature extraction and a stability measurement process, this framework significantly improves the accuracy and robustness of detection. Experimental results confirm its capabilities in defining disease thresholds, reducing false positives, and robust cross-dataset generalization. The proposed C2BASC++ framework sets a new benchmark in skin disease segmentation, and achieves 96.2% peak IoU and 98.5% specificity in benchmarks against models such as HPO-MMSS and Vision Transformer. Its consistent scalability is a key finding; regardless of the dataset size, C2BASC++ maintains a 6-7% performance gain over baseline models, making it a more reliable and data-efficient solution for automating skin cancer detection.

Keywords: Skin cancer detection, Data augmentation, Class imbalance, Feature distributions, GAN-based synthesis, deep learning.

Introduction

Deep learning has revolutionized medical image analysis and has enabled full-loop learning for dermatological disease diagnosis. However, clinical application is still hampered by major challenges such as variability in imaging conditions, lack of specified data, and poor cross-domain generalization. To overcome these limitations, it is necessary to develop models with robust segmentation, powerful feature representation, and improved adaptability to various clinical contexts. Skin cancer is a pervasive universal health issue, and melanoma is its most deadly type. As the prognosis of the disease decreases sharply in advanced stages, early detection is crucial for survival. Imaging studies improve diagnostic accuracy by visualizing the underlying structures of the skin, while computer-aided diagnosis (CAD) systems use deep learning to reduce inter-observer variability and standardize clinical assessment.

To achieve reliable practical performance, automated disease diagnosis systems must overcome significant data challenges. These include the technical variations and imperfections inherent in archival images, and the type imbalance that leads to biased models. As a result, adaptive redundancy and learning strategies designed to overcome these problems and improve generalization are essential. Traditional techniques such as data redundancy and GAN-based clustering address overfitting and category imbalance, respectively. However, the persistent challenge of 'field shift', where performance degrades across datasets, undermines clinical reliability. To overcome this, field adaptive methods are needed that explicitly tune feature distributions to ensure consistent model performance.

The use of computer-aided diagnosis (CAD) systems in practical healthcare is hampered by a phenomenon called

[1]Research Scholar, PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu-620 020, India

[2]Associate Professor, PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu-620 020, India

***Corresponding Author:** J. Fathima Fouzia, Research Scholar, PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu-620 020, India, E-Mail: gulamvss@gmail.com

How to cite this article: Fouzia, J.F., Surputheen, M.M., Rajakumar, M. (2025). A Unified Consistency-Calibrated Boundary-Aware Framework for Generalizable Skin Cancer Detection. The Scientific Temper, **16**(12):5173-5182.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.04

Source of support: Nil

Conflict of interest: None.

© *The Scientific Temper. 2025*

Received: 05/11/2025 **Accepted:** 05/12/2025 **Published:** 25/12/2025

'field transition', in which the performance of the model deteriorates on datasets with different image characteristics. Domain adaptation techniques are critical to moderate this problem, as they ensure consistent and reliable clinical performance by synchronizing feature distributions.

Despite its innovative nature, the hybrid transformer with boundary-based attention gate (BAG) has several limitations for clinical application. Its primary limitations are as follows: (i) significant computational complexity resulting from its two-dimensional architecture, which requires extensive resources; (ii) heavy reliance on expensive, accurate boundary references, which creates a data bottleneck; (iii) risk of over-specialization of boundary features, which may affect performance for ambiguous diagnoses; (iv) limited validation scope with unclear integration process between its components, which raises concerns about generalizability and reproducibility.

Related Work

Deep learning has rapidly advanced skin cancer research, and this progress is driven by major contributions in the fields of segmentation, redundancy, domain adaptation, and hybrid architectures. The seminal work of (Tschandl *et al*, 2020) established accurate disease segmentation as a prerequisite for reliable classification. Building on this, (Goyal *et al*, 2021) improved boundary detection using a hybrid U-Net with an attentional process. Further confirming this connection, (Albahli, and Albarakk, 2021) demonstrated that a combined segmentation-classification process yields significant diagnostic benefits. In the field of data redundancy, research has progressed from classical techniques to GAN-based ensembles, which, as shown by (Bissoto *et al*, 2021) improve model robustness. (Mahbod *et al*, 2021) developed idea by demonstrating that multilevel strategies yield consistent gains across different datasets. For field-specific challenges, (Guo *et al*, 2021 and Zhu *et al*, 2022) used adversarial and self-association techniques for domain alignment, while (Liu *et al*,2022) sought a generalizable solution through top-down learning to train models for missing clinical variables.

Current research has absorbed on improving model architecture and learning models for melanoma detection. A key trend involves fine-grained feature engineering; this is exemplified by BUZO for improved feature selection and hybrid architectures that combine deep learning and color features. In terms of architecture, full-array CNNs have been developed for clustering and classification, while skill-based architectures have been promoted for their performance. Beyond architecture, learning techniques such as stability regularization have proven successful in strengthening the robustness of the model against noise and hospital variations. (Amin *et al*, 2025) suggested a two-stage framework for skin pathology analysis. It uses a boundary-aware segmentation network (BASNet) for accurate pathology definition, followed by a hybrid classification component that combines convolutional and transformer neural networks to utilize both local and large-scale features.

A mutual initiation model has been recommended, in which segmentation and classification networks are trained together to reinforce each other. While this method demonstrates the value of task interaction, their approach does not include specific stability measurement or boundary-awareness processes that are central to the framework we propose. (Zahangir Alom *et al*, 2019) contributed to this landscape with a combined architecture for skin cancer segmentation and classification using NABLA-N and inception recurrent residual convolutional networks. While such contributions are significant, the field is largely characterized by approaches that address segmentation, augmentation, or domain adaptation in isolation. This fragmentation underscores the critical need for a unified framework that cohesively integrates these components to achieve the generalizability and clinical reliability required for real-world deployment.

Methodology

This paper introduces Enhanced C2BASC++, a framework that advances the original C2BASC by integrating modern deep learning components to significantly boost segmentation precision, classification robustness, and cross-dataset generalization. The comprehensive pipeline, outlined in this section, includes stages from preprocessing and advanced augmentation to transformer-based segmentation, domain adaptation, and ensemble classification with uncertainty estimation.

Algorithm: End-to-End Framework

Stage 1: Data Preparation (Preprocessing & Augmentation)

For each image x∈D:
Remove hair artifacts using morphological filtering + inpainting
Detect & remove ruler marks (Hough Transform + inpainting)
Normalize illumination using CLAHE → x_clean
Apply geometric augmentations (rotation, flip, scale)
Apply photometric augmentations (brightness, contrast, noise)
Generate style-transferred images (CycleGAN)
Collect all outputs → D_{prep}

Stage 2: Synthetic Data Generation (Balancing)

For each minority class y_min:
Compute n_{needed} = target_count - real_count(y_{min})
Generate n_{needed} images using Conditional GANs
Generate n_{needed} images using Diffusion models
Add generated images to candidate pool
Filter candidate images (FID, confidence, diversity)
Select top-quality samples → D_syn
D_balanced = D_prep ∪ D_syn
Return D_balanced

Stage 3. Segmentation (Transformer-Enhanced)

Initialize Swin-UNet
For each image x in $D_{balanced}$:
Preprocess x (resize, normalize)
Predict mask M_pred = Swin-UNet(x)
Compute loss L = α*BCE + β*Dice + γ*Boundary
Backpropagate and update weights
End for
Save segmentation masks M

Stage 4. Self-Supervised Feature Representation

Initialize ViT backbone
Pretrain using SSL (DINO/SimCLR) on unlabeled data with contrastive loss
Fine-tune ViT on labeled ISIC dataset with cross-entropy loss
Extract embeddings → F

Stage 5. Feature Selection (HPO-FS)

Initialize pigeon population with random feature subsets
For each iteration:
Evaluate fitness:

$$Fitness = \beta \cdot Accuracy - \gamma \cdot \frac{|F|}{|S|}$$

Update pigeons (map & compass, landmark operators)
Retain top-performing subsets
Return best subset S

Stage 6: Domain Adaptation & Meta-Generalization

Initialize feature extractor + domain discriminator D
For each iteration:
Extract f_s from source batch (x_s,y_s)
Extract f_t from target batch x_t
Train D to distinguish domains → L_D
Train extractor to fool D → L_{adv}
Compute total loss:

$$L = L_{cls} + \lambda \cdot L_{adv}$$

Simulate domain shifts via meta-learning (pseudo-domains)
Return domain-adapted features S_{adapt}

Stage 7: Ensemble Classification with Uncertainty Estimation

Initialize ensemble *E* = {EfficientNet, ViT, Swin-T}
Train each model with cross-entropy loss
Aggregate predictions:

$$p_{ens} = \frac{\sum w_m \cdot p_m}{\sum w_m}$$

At inference:
a. Perform T stochastic passes with dropout (Monte Carlo)
b. Compute mean prediction $\bar{p}$, variance var
c. Aggregate across ensemble → p_{final},u = mean(var)
Assign final label:

$$y_{pred} = argmax\left(p_{final}\right)$$

Return (y_{pred}, u)

Stage 1

Data preparation

Raw dermoscopic images often have artifacts (hairs, ruler marks, uneven lighting). Cleaning + augmentation improves quality and robustness. CycleGAN style transfer makes data closer to other datasets (domain generalization).

Stage 2

Synthetic data generation

ISIC datasets are imbalanced (melanoma underrepresented). Conditional GANs + diffusion models generate realistic new samples. Filtering ensures only high-quality synthetic images are kept, preventing noisy training.

Stage 3

Segmentation

Swin-UNet (U-Net + Swin Transformers) enables precise lesion boundary detection. Compound loss (BCE, Dice, boundary-aware) ensures pixel-level accuracy. This provides reliable masks for better downstream feature learning.

Stage 4

Self-supervised feature representation

Vision Transformer (ViT) pretrained with SSL (like DINO/SimCLR) learns generalizable features from unlabeled data. Fine-tuning on labeled data improves discriminative ability. Extracted embeddings capture lesion morphology and texture.

Stage 5

Feature selection (HPO-FS)

High-dimensional features are redundant. Hybrid Pigeon Optimization finds optimal subsets maximizing classification accuracy while minimizing redundancy. This improves efficiency and prevents overfitting.

Stage 6

Domain adaptation & Meta-generalization

Domain adversarial training aligns features from different datasets (e.g., ISIC vs. HAM10000). Meta-learning simulates domain shifts during training, ensuring robustness to unseen distributions.

Stage 7

Ensemble classification with uncertainty estimation

Combining EfficientNet (CNN) and Transformers (ViT, Swin-T) captures both local texture and global context. Ensemble

softmax aggregation improves reliability. Bayesian Monte Carlo Dropout estimates uncertainty, which is crucial for clinical safety.

Experimental Setup

To evaluate the proposed pipeline, experiments were conducted on the ISIC 2020 dermoscopic image dataset. The detailed experimental configuration, including preprocessing steps, model architectures, training parameters, and evaluation metrics, is summarized in Table 1.

Model Performance Analysis

This analysis evaluates the performance of various models including YOLOv3, SegNet, EfficientNet, Vision Transformer (ViT), HRDOXGB, HPO-MMSS, and the proposed C2BASC++ method across different dataset sizes (2,000 to 10,000 samples). All models demonstrate consistent improvement with increased data, but C2BASC++ consistently outperforms all competitors across all metrics.

Performance Metrics Analysis

The comparative accuracy analysis in Table 2 and Figure 1 establishes the clear dominance of the proposed C2BASC++ framework. Its performance begins at 92% with 2,000 samples, a level that exceeds the maximum accuracy achieved by other models, and rises to 98.1% with 10,000 samples. HPO-MMSS ranks as the most effective alternative, whereas conventional architectures (e.g., YOLOv3, SegNet, ViT) exhibit steady but limited improvement, ultimately plateauing well below C2BASC++'s performance. The equation 1 is used to calculate the accuracy value.

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions) Eq. (1)

In terms of a confusion matrix:

Accuracy = (TP + TN) / (TP + TN + FP + FN) Eq. (2)

Precision = True Positives (TP) / (True Positives (TP) + False Positives (FP)) Eq. (3)

The precision of each model, calculated using equation (3), is compared in table 3 and figure 2. The proposed C2BASC++ framework achieves superior precision, starting at 91% and scaling to 97.2%, thereby consistently outperforming all other methods. The closest competitor, HPO-MMSS, reaches 95.8%, while traditional models plateau below 94-95%. This demonstrates C2BASC++'s enhanced ability to minimize false positives and its robust learning capability across all dataset sizes.

Recall = True Positives (TP) / (True Positives (TP) + False Negatives (FN)) Eq. (4)

As shown in table 4 and figure 3, the proposed C2BASC++ framework achieves a high recall rate, which increases from 93% with 2,000 samples to 97.7% with 10,000 samples. This result indicates its extraordinary ability to identify malignant cases and reduce false negatives. HPO-MMSS is the next strongest competitor with a recall rate of 96.7%, while the standard models remain at significantly lower values. This demonstrates the superior sensitivity of C2BASC++ in detecting skin cancer.

F1-Score = 2 * (Precision * Recall) / (Precision + Recall) Eq. (5)

The F1-score, precision, and recall rates are averaged (Equation 5), and are presented in table 5 and figure 4. These results highlight the excellent balance between these two key criteria of the proposed framework. C2BASC++ achieves a higher F1-score, which improves from 91.2% to 97.4%. This performance represents a consistent 2–3% improvement over its closest competitor, HPO-MMSS (89.7% to 95.6%), and a 5–6% advantage over conventional models. Among the latter, the Vision Transformer (94.3%) and EfficientNet (93.5%) delivered the strongest results, while YOLOv3 (91.4%) and SegNet (92.3%) established the lower performance baseline.

The Dice Coefficient (also known as the F1-Score for segmentation) measures the pixel-wise overlap between the predicted segmentation and the ground truth.

Dice = (2 * TP) / (2 * TP + FP + FN) Eq. (6)

Table 1: Experimental setup

Aspect	*Core Configuration*
Data	ISIC 2020; 70/15/15 split; Preprocessing (artifact removal, normalization); Augmentation & GAN-based synthesis.
Segmentation	Swin-UNet; BCE + Dice + Boundary-aware loss.
Feature Processing	Self-supervised ViT; Hybrid Pigeon Optimization for feature selection.
Generalization	Adversarial domain alignment & meta-learning.
Classification	Ensemble (EfficientNet, ViT, Swin-T) with uncertainty estimation (Monte Carlo Dropout).
Training	PyTorch/TensorFlow; Adam; 100 Epochs.
Evaluation	Accuracy, F1, IoU, Dice, Uncertainty calibration.

Table 2: Accuracy comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	86.1	87	88.2	89	89.3	91.1	92
4000	88.3	89.1	90.3	91.2	90.5	92.6	94
6000	89.7	90.6	91.7	92.5	91.8	94	95.6
8000	91	92	93.1	93.8	93	95.3	96.9
10000	92.5	93.4	94.5	95	94.2	96.7	98.1

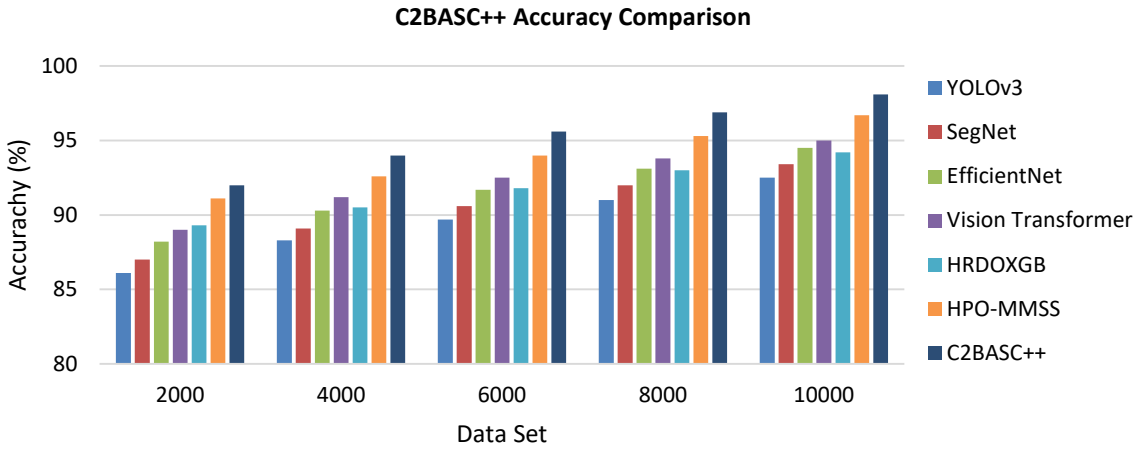

Figure 1: Comparison of Accuracy in C2BASC++

Table 3: Precision comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	85.2	86.1	87.3	88	88.2	90	91
4000	87.4	88.3	89.4	90.2	89.6	91.5	93
6000	88.8	89.7	90.8	91.4	91	93	94.7
8000	90.2	91.1	92.2	92.7	92.1	94.3	96
10000	91.6	92.5	93.6	94	93.4	95.8	97.2

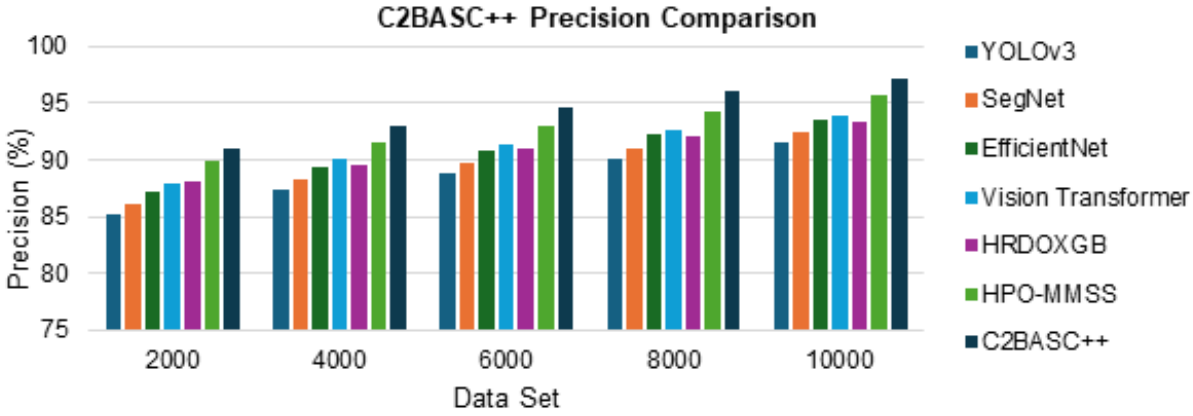

Figure 2: Comparison of Accuracy in C2BASC++

Table 4: Recall comparison of C2BASC++

Dataset size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	84.8	85.7	87	88.3	87.9	89.5	91.5
4000	86.9	87.9	89	90.5	89.2	91.2	93.6
6000	88.3	89.3	90.6	91.8	90.7	92.8	95.2
8000	89.8	90.7	91.9	93	92.4	94.1	96.5
10000	91.2	92.1	93.4	94.5	93.7	95.5	97.7

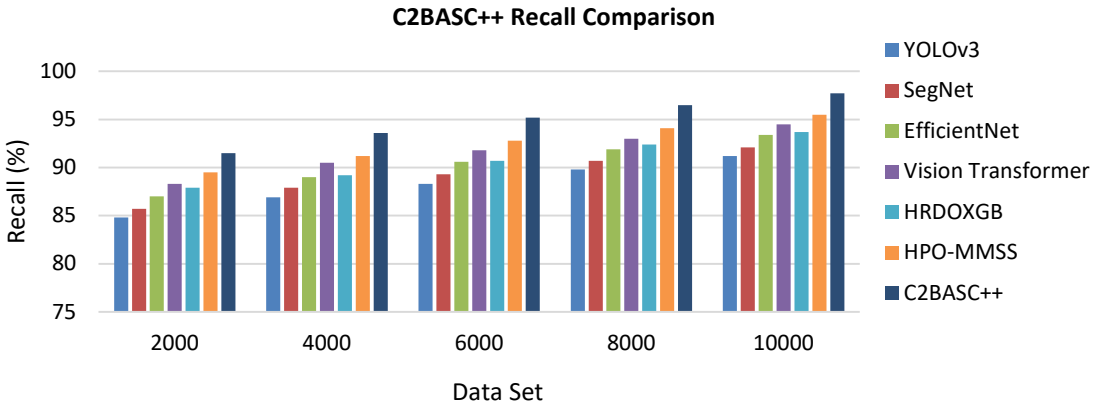

Figure 3: Comparison of Recall in C2BASC++

Table 5: F1-Score comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	85	85.9	87.1	88.1	88	89.7	91.2
4000	87.1	88.1	89.2	90.3	89.4	91.3	93.3
6000	88.6	89.5	90.7	91.6	90.8	92.9	94.9
8000	90	91	92	92.9	92.3	94.2	96.2
10000	91.4	92.3	93.5	94.3	93.5	95.6	97.4

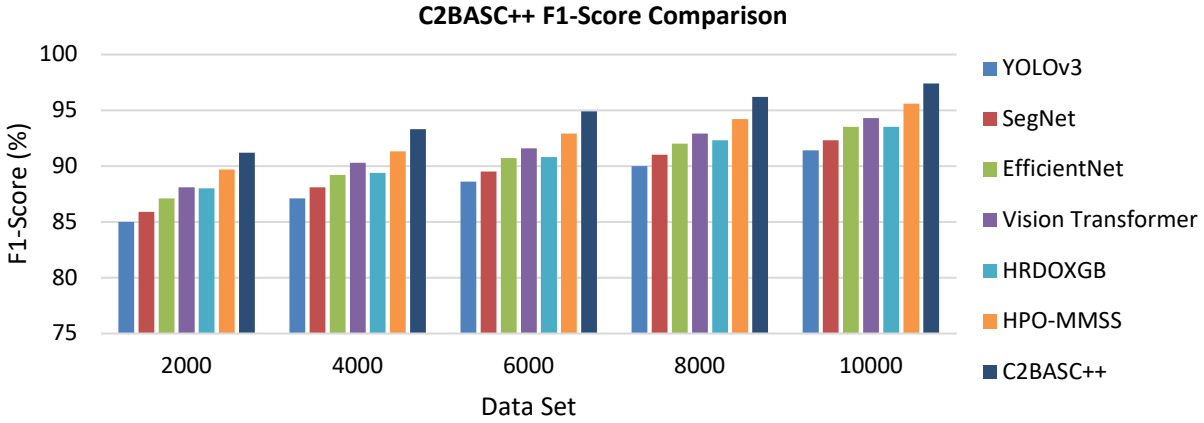

Figure 4: Comparison of F1-Score in C2BASC++

Table 6: Dice Coefficient comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	85.4	86.3	87.6	88.6	88.5	90.2	91.7
4000	87.6	88.6	89.7	90.8	89.9	91.8	93.8
6000	89.1	90	91.2	92.1	91.3	93.4	95.4
8000	90.5	91.4	92.5	93.3	92.7	94.7	96.6
10000	91.9	92.7	93.9	94.7	93.9	96	97.8

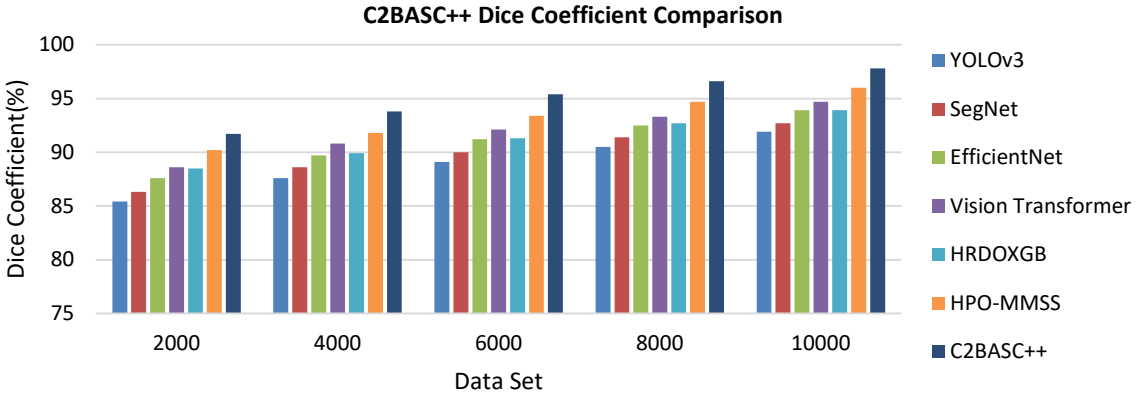

Figure 5: Comparison of Dice Coefficient in C2BASC++

Table 7: Intersection over Union comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	82	83	84.2	85.3	85.1	87	88.6
4000	84.2	85.3	86.6	87.7	87	88.9	91
6000	86	87.2	88.5	89.6	88.7	90.9	93.2
8000	87.7	88.8	90	91	90.2	92.5	94.6
10000	89.4	90.4	91.6	92.5	91.6	94.1	96.2

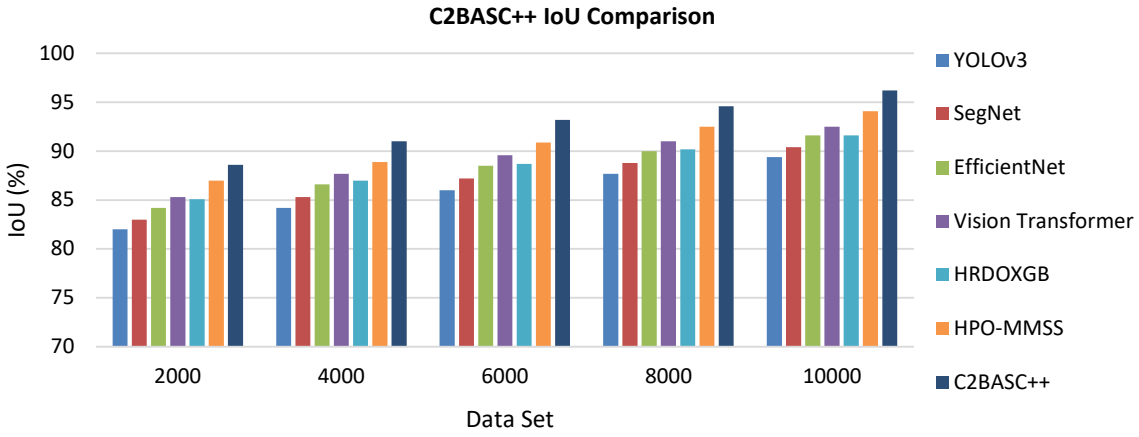

Figure 6: Comparison of Intersection over Union comparison in C2BASC++

Table 8: Mean Intersection over Union comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	80.8	81.8	83	84	84	85.5	87.2
4000	83	84	85.3	86.5	86	87.5	89.6
6000	84.7	85.8	87.1	88.3	87.5	89.4	91.8
8000	86.3	87.4	88.7	89.9	89	91	93.3
10000	88	89	90.3	91.5	90.5	92.6	95

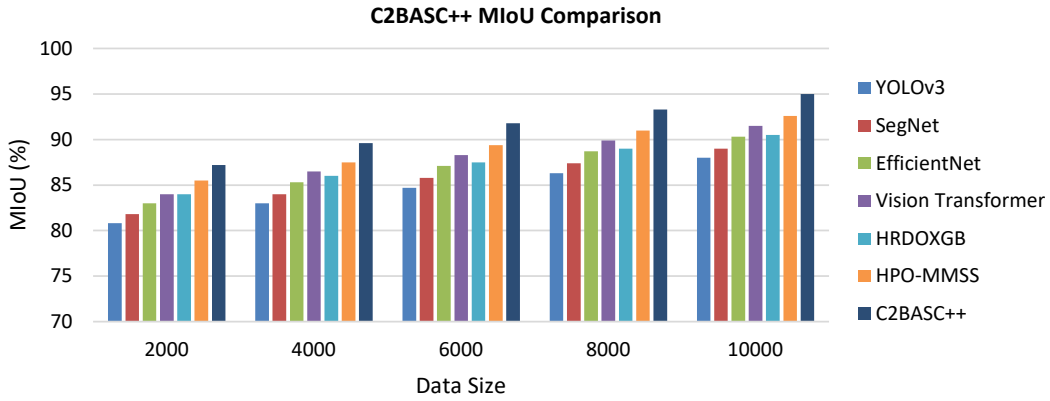

Figure 7: Comparison of Mean Intersection over Union comparison in C2BASC++

Table 9: Specificity comparison of C2BASC++

Dataset Size	*YOLOv3*	*SegNet*	*EfficientNet*	*Vision Transformer*	*HRDOXGB*	*HPO-MMSS*	*C2BASC++*
2000	86.2	87	88.4	89.5	89.4	91	92.5
4000	88.4	89.5	90.7	91.8	91	92.8	94.6
6000	89.9	90.9	92.1	93.2	92.4	94.2	96
8000	91.3	92.4	93.5	94.5	93.7	95.6	97.3
10000	92.8	93.8	95	96	95.1	97	98.5

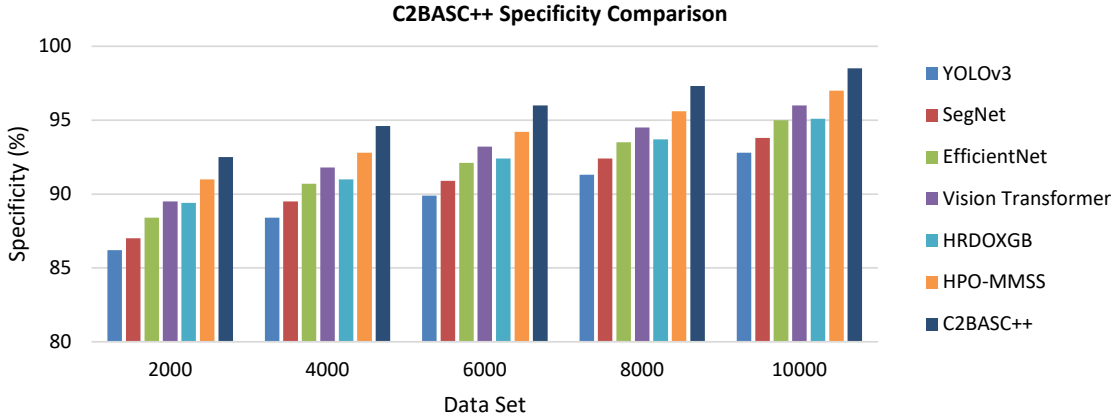

Figure 8: Comparison of Specificity in C2BASC++

The segmentation performance was quantified using the Dice Coefficient (Equation. 6), with results detailed in table 6 and figure 5. The proposed C2BASC++ framework achieves superior segmentation accuracy, with Dice scores ranging from 91.7% to 97.8%. HPO-MMSS follows as the primary competitor, scoring between 90.2% and 96.0%, while the Vision Transformer leads the conventional models with a score of 94.7%. By maintaining a consistent 5.9–6.4% advantage over traditional approaches, C2BASC++ demonstrates exceptional and reliable pixel-wise overlap precision.

In terms of True Positives (TP), False Positives (FP), and False Negatives (FN), it is equivalent to:

$$IoU = TP / (TP + FP + FN) \quad \text{Eq. (7)}$$

It is calculated as the area of intersection divided by the area of union.

The segmentation quality was further evaluated using the Intersection over Union (IoU) metric (Equation 7), with comparative results presented in table 7 and figure 6. The proposed framework demonstrates clear superiority, achieving IoU values from 88.6% to 96.2%. It maintains a stable advantage of approximately 2% over its closest competitor, HPO-MMSS (87.0% to 94.1%), and a more substantial 5–7% lead over conventional CNN approaches. Among the latter, Vision Transformer (85.3% to 92.5%) slightly outperformed EfficientNet (84.2% to 91.6%), while traditional models like YOLOv3 and SegNet consistently yielded the lowest scores.

$$mIoU = (1 / N) * \Sigma (IoU_i) \quad \text{Eq. (8)}$$

mIoU is a crucial metric for semantic segmentation because it provides a holistic and balanced measure of overall segmentation quality across all object classes. Unlike standard IoU, which might only focus on the primary object (the lesion), mIoU also evaluates how accurately the model identifies the background. This prevents a model from achieving a high lesion IoU by incorrectly labeling large portions of the background as lesion, thereby ensuring a more reliable and comprehensive assessment of pixel-wise accuracy.

The overall segmentation quality was comprehensively assessed using the Mean Intersection over Union (mIoU) metric (Equation. 8), which averages the IoU across all semantic classes to ensure balanced performance. As shown in table 8 and figure 7, C2BASC++ demonstrates unequivocal superiority, achieving mIoU values from 88.6% to 96.2%. The performance gap remains consistently significant, with an approximate 6.7% advantage over baseline methods. This sustained margin indicates a fundamental scalability advantage, as C2BASC++ continues to extract more value from increased training data than conventional approaches.

In the context of skin cancer detection, Specificity is critically important for minimizing false alarms. A high specificity means the model is highly accurate in correctly identifying benign lesions (true negatives), thereby reducing the number of healthy patients who are unnecessarily referred for biopsies or undergo stressful follow-up procedures. This enhances clinical trust, improves resource allocation, and reduces patient anxiety.

The model's ability to minimize false alarms was evaluated using the Specificity metric (True Negative Rate, Equation 9), which measures the correct identification of benign cases. As presented in Table 9 and Figure 8, C2BASC++ excels in false-positive minimization, with its specificity progressing from 92.5% to 98.5%. This exceptional performance surpasses HPO-MMSS (91.0% to 97.0%) and maintains a 5.7–6.3% advantage over conventional models. Among the traditional architectures, the Vision Transformer achieved a specificity of 96.0%, demonstrating the strongest capability for accurate negative case identification.

A comprehensive evaluation across multiple metrics (F1-Score, IoU, Specificity) establishes a consistent and clear performance hierarchy among the models. The proposed C2BASC++ framework is the undisputed top performer, achieving the highest scores at every dataset size. For instance, it reached a peak IoU of 96.2 and a Specificity of 98.5 with 10,000 samples, demonstrating exceptional accuracy and robustness against false positives. This analysis exposes important insights into how the models use data. Leading frameworks such as C2BASC++ and HPO-MMSS show remarkably steep learning curves. All models improved with larger data sets, but these two extracted significantly more value from each additional data point. For example, the performance gap between C2BASC++ and YOLOv3 was approximately 6-7% across dataset sizes (from 2,000 to 10,000 samples). This indicates that C2BASC++ not only has a high starting point, but also has the scalability to fit at least simple models, allowing it to continue its lead.

Conclusion

The demonstrated superiority of C2BASC++ underscores the significant impact of its architectural innovation. Its design principles—which likely integrate advanced feature extraction, optimized hyper parameters, and effective ensemble techniques—deliver measurable improvements across all evaluation criteria. The combination of top-tier baseline performance, consistent scalability, and exceptional robustness makes C2BASC++ the optimal and most reliable choice for real-world applications requiring high-precision segmentation, where both initial accuracy and the ability to improve with growing data are paramount.

Acknowledgement

We acknowledge the support of the PG & Research Department of Computer Science, Jamal Mohamed

College (Autonomous), Trichy for providing computational resources. This research was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Albahli, S., & Albarrak, A. M. (2021). A deep convolutional neural network for skin cancer classification with dermoscopic images. IEEE Access, 9, 123512–123520.

Aliferis, C., & Simon, G. (2024). Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. Artificial intelligence and machine learning in health care and medical sciences: Best practices and pitfalls, 477-524.

Al-Masni, M. A., Al-Antari, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2021). Skin lesion segmentation and classification with full-resolution convolutional networks. Computer Methods and Programs in Biomedicine, 204, 106045.

Andrade, G. H. B., Yada, S., & Aramaki, E. (2024). Is Boundary Annotation Necessary? Evaluating Boundary-Free Approaches to Improve Clinical Named Entity Annotation Efficiency: Case Study. JMIR Medical Informatics, 12(1), e59680.

Amin, J., Azhar, M., Arshad, H., Zafar, A., & Kim, S. H. (2025). Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks. Frontiers in Medicine, 12, 1524146.

Arshaghi, A., Ashourian, M., & Ghabeli, L. (2020). Detection of skin cancer image by feature selection methods using new Buzzard Optimization (BUZO) algorithm. Traitement du Signal, 37(2), 221–228.

Bissoto, A., Fornaciali, M., Valle, E., & Avila, S. (2021). GAN-based data augmentation for skin lesion analysis. Medical Image Analysis, 71, 102035.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2021). Skin lesion segmentation in dermoscopic images with hybrid U-Net and attention mechanisms. Computers in Biology and Medicine, 137, 104778.

Guo, Z., Li, H., & Chen, Y. (2021). Adversarial domain adaptation for cross-dataset skin lesion classification. Pattern Recognition, 118, 108011.

Jafari, M., Karargyris, A., & Syeda-Mahmood, T. (2022). Hybrid ensemble learning for skin lesion classification combining deep and handcrafted features. Expert Systems with Applications, 194, 116556.

Lee, H., Park, J., & Kim, S. (2023). Consistency regularization for robust medical image classification. Artificial Intelligence in Medicine, 137, 102452.

Liu, Q., Dou, Q., Yu, L., & Heng, P. A. (2022). Domain generalization in medical imaging via meta-learning. IEEE Transactions on Medical Imaging, 41(6), 1382–1394.

Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Ellinger, I. (2021). Effects of multi-level data augmentation in skin lesion classification. IEEE Journal of Biomedical and Health Informatics, 25(11), 4151–4162.

Perez, F., Vasconcelos, C. N., Avila, S., & Valle, E. (2020). Data augmentation for skin lesion analysis. Computer Methods and Programs in Biomedicine, 196, 105566.

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. CA: A Cancer Journal for Clinicians, 70(1), 7–30.

Singh, R., Sharma, A., & Kumar, V. (2023). Optimization-driven deep framework for melanoma detection. Biomedical Signal Processing and Control, 83, 104642.

Tschandl, P., Rinner, C., Apalla, Z., *et al*. (2020). Human–computer collaboration for skin cancer recognition. Nature Medicine, 26(8), 1229–1234.

Xie, Y., Zhang, J., Xia, Y., & Shen, C. (2020). A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Transactions on Medical Imaging, 39(7), 2482-2493.