IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1613~1622
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1613-1622

Enhancing imbalanced dataset diagnosis using class-based input image composition

Azzeddine Hlali, Majid Ben Yakhlef, Soulaiman El Hazzat
Engineering Sciences Polydisciplinary Laboratory, Faculty of Taza, Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco

Article Info

Article history:

Received Dec 8, 2025
Revised Jan 27, 2026
Accepted Feb 6, 2026

Keywords:

Class imbalance
False diagnostic metrics
Low-quality features
Medical image analysis
Retinal disease classification

ABSTRACT

Deep learning models often falter when faced with small, imbalanced datasets or degraded image quality, leading to unacceptably high false prediction rates. To bridge this gap, we introduce class-based image composition. This technique reformulates training inputs by fusing multiple intra-class images into unique composite input images (CoImg). By concentrating information density and amplifying intra-class variance, CoImg forces the model to discern subtle, nuanced disease patterns that might otherwise be lost. We validated this approach using the optical coherence tomography dataset for image-based deep learning methods (OCTDL), a collection of seven imbalanced retinal disease scan categories. From this, we engineered Co-OCTDL: a perfectly balanced variant where each training sample exists as a 3×1 layout composite. To measure the impact of this new representation, we benchmarked the original dataset against its composite counterpart using a VGG16 architecture. Precision was paramount. We maintained identical hyperparameters and model structures across all experiments to ensure a rigorous, fair comparison. The results were transformative. While baseline datasets struggled, the enhanced Co-OCTDL achieved a near-perfect F1-score of 0.995 and an AUC of 0.9996. The method effectively neutralized the risks of class imbalance. It didn't just improve the numbers; it refined the diagnostic reliability of the model.

Corresponding Author:

Azzeddine Hlali
Engineering Sciences Polydisciplinary Laboratory, Faculty of Taza
Sidi Mohamed Ben Abdellah University (USMBA)
Fes, Morocco
Email: azzeddine.hlali@usmba.ac.ma

1. INTRODUCTION

Deep learning models have undoubtedly been a breakthrough in medical image analysis, but their diagnostic precision largely depends on the quantity and quality of training data [1]. The domain is dealing with two challenging issues, which are the unbalanced number of classes, where the number of pathological cases is considerably less than that of healthy controls [2] and the fact that different disorders visually resemble each other very closely [3]. Such problems often result in errors in diagnosis [4]. The retinal imaging branch is a typical example where these difficulties are manifested, and for which age-related macular degeneration (AMD) and diabetic retinopathy (DR) share a common histogram of the morphological features [5], [6].

To mitigate these challenges, the research community has explored strategies across three primary dimensions: loss modulation, data augmentation, and architectural innovation. To address the imbalance, loss re-weighting techniques have been widely adopted. Approaches such as unified focal loss [7], class-balanced loss [8], and asymmetric loss [9] attempt to modulate the influence of easy versus hard examples. More

recently, batch-balanced focal loss mitigates the inherent class imbalances of medical datasets by prioritizing rare, clinically significant pathological features to sharpen diagnostic precision [10].

In addition, access to datasets has been enhanced not only through data augmentation but also through data synthesizing. It has been proven that generative adversarial networks (GANs) can be trained very effectively to generate medical lesions [11], [12]. At the same time, traditional augmentations inspired by the domain, like MixUp and CutMix did bring improvement to the model calibration [13]. Lately, the field of medical image generation has seen the introduction of diffusion models capable of generating realistic-looking scans [14].

On the architectural front, one of the methods is the integration of convolutional neural networks (CNNs) with transformers, as seen from the examples of TransUNet [15] and Swin-UNet [16]; or training a model on one domain and then fine-tuning it on another domain by using pre-trained backbones such as ResNet-50 [17] and EfficientNet-B3 [18]. However, some authors argue that transfer learning from the typical image datasets (e.g., ImageNet) may not be very beneficial in the case of the medical domain [19]. However, these methods still have large limits. Re-weighting the loss may cause the features' distribution to be inconsistent, and the synthesis of the images might introduce non-realistic artifacts. Besides, techniques for dealing with inter-class similarity, like the use of segmentation to guide classification [20] or attention-based models [21] require a large amount of computation and are extremely dependent on the availability and quality of pixel-level annotations [22]. More importantly, the methods at hand struggle to tell one type of disease apart from the other when the raw images exhibited do not contain any substantial or clear features.

We, therefore, propose a simple yet effective method to both enrich and improve the solutions. Class-based input image composition (CB-ImgComp) is a technique that essentially alters the nature of the inputs during training. CB-ImgComp departs from the traditional augmentations or oversampling methods. It goes about generating structured layouts (e.g., 3×1 grids) called composite input images (Co-Img) by combining several images from the same class. This approach has two main functions. On the one hand, it works as a targeted augmentation method that is capable of re-balancing the number of samples in the under-represented classes. On the other hand, it is a very effective way of making discriminative signals more prominent by maximizing both concentration and phenotypic variability within individual training samples. Features from various images within a composite are mixed, leading to a high degree of feature diversity, which is instrumental in preventing the over-fitting on minority classes without any loss of semantic integrity. Optical coherence tomography (OCT) images will be used to substantiate the experiments [23]. As a good proof of concept, it is shown that using CB-ImgComp, classification errors can be seriously lowered, thus providing a strong, yet low-complexity checkboard for the performance growth on small and imbalanced datasets.

2. METHOD

2.1. Research design and data acquisition

The research was mainly focused on improving deep learning models' diagnostic performance by addressing the drawbacks of small, imbalanced datasets and low-quality inputs through a novel preprocessing technique. We came up with CB-ImgComp which is an input-level method that changes the training data by combining multiple images from the same class into unified visual structures—Co-Img. Deliberate visual compositions designed to preserve class semantics while amplifying structural features. We sourced all raw samples directly from the public optical coherence tomography dataset for image-based deep learning methods (OCTDL) [23]. This dataset includes 2,064 high-resolution OCT scans of the human retina, categorized into seven different disease classes. Due to the significant class imbalance of the dataset, it was used as the baseline dataset for our experiment.

2.2. Research procedures: class-based input image composition

The CB-ImgComp procedure was applied to increase the diversity of data and the information density per training sample, as illustrated in Figure 1. The process involves generating Co-Imgs. This method aims to obtain the most relevant information content from each training sample [24]. As a result, the original dataset is transformed into a new one named Co-OCTDL, which is made up of Co-Img. The size of these Co-Img can be changed, so that different matrix configurations (e.g., 3×1, 2×5, 4×2) can be used depending on the model and the initial dataset. Here, we have chosen a 3×1 layout pattern for Co-Img creation.

2.2.1. CB-ImgComp function definition

The CB-ImgComp function operates on three core variables: the dimensional layout of the Co-Img, the source dataset, and a boolean toggle for intra-composite image repetition. In (1) formalizes this relationship. By explicitly defining these parameters, we regulate the architectural composition of each training sample to ensure maximum data diversity without redundant overlap.

$$F(L,D,r) = \{CoImg_{ci} | CoImg_{ci} = Compose(x_{c1}, \dots, x_{ck}; L), \\ i = 1, \dots, N_c, x_{ck} \in D_c, k = size(L), c \in \mathcal{C}, \\ x_{cj} \neq x_{cp} \; only \; if \; r = 0, j = 1, \dots, k, and \; p = 1, \dots, k\} \quad (1)$$

Where N_c is the size of the Co-Img of class c in the new dataset, L is the layout (e.g., 3×1), D is the original dataset, $r \in \{0,1\}$ indicates the repetition flag, and $\mathcal{D}_c \subseteq \mathcal{D}$ is the set of samples from class c.

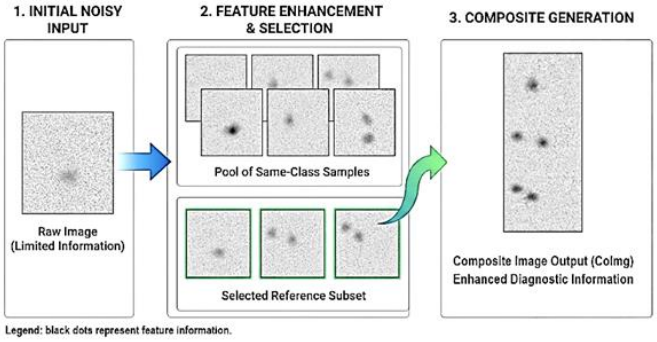


Figure 1. Synthesizing intra-class samples into high-density composite training inputs

2.2.2. CB-ImgComp function algorithm

We design a precise algorithmic pipeline to assemble the composite image dataset—clean, repeatable, and tightly controlled from start to finish. Instead of random aggregation, the method builds each composite image from class-specific samples using fixed spatial dimensions, while explicitly managing whether images may reappear across combinations or remain unique. Algorithm 1 encodes this logic in the CB-ImgComp function, where class-wise image selection and structured composition converge into a single generation flow that produces the enriched dataset D_{final}.

Algorithm 1. CB-ImgComp function

```
Input: Dataset D, dimensions m × n, repetition flag r ∈ {0,1}
Output: Enriched dataset Dfinal
Initialize Dfinal ← ∅ {Empty list for CoImg}
k ← m × n {Number of images per CoImg}
for each class c ∈ D do
    Nc ← |Dc| {Number of images in class c}
    if r = 1 then
        Combinationsc ← generate_multicombinations(Dc, k) {with repetition}
    else
        Combinationsc ← generate_combinations(Dc, k) {without repetition}
    end if
    for each combination W ∈ Combinationsc do
         CoImg ← create_CoImg(W, m, n)
         Dfinal ← Dfinal ∪ {CoImg}
    end for
end for
return Dfinal
```

Helper functions:

i) Generate combinations (without repetition)

```
function generate_combinations(S, k)
return all size-k subsets of S (no repetition)
```

ii) Generate multi-combinations (with repetition)

```
function generate_multicombinations(S, k)
return all size-k multisets from S (with repetition)
```

iii) Create composite image CoImg

```
function create_CoImg(W, m, n)
CoImg ← initialize_grid(m, n)
for i = 1 to m do
  for j = 1 to n do
    CoImg[i, j] ← W [(i - 1) · n + j]  end forend for
return CoImg
```

The schema as shown in Figure 2, outlines the process of creating a new dataset representation through the CB-ImgComp function. It has led to the generation of a fresh data collection where every picture is a blend of three pictures from the raw dataset. Each composite image is therefore a fusion of multiple elements and transfers more source signals at a time, resulting in a much stronger feature signal than any of the original single images. Figure 3 presents a few instances of images from this new data collection.

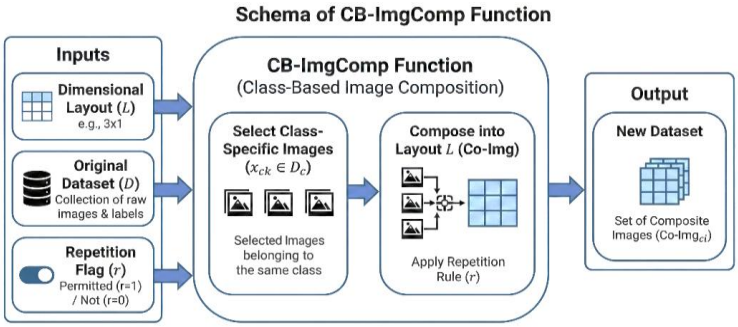


Figure 2. Redefining training inputs through class-based image synthesis

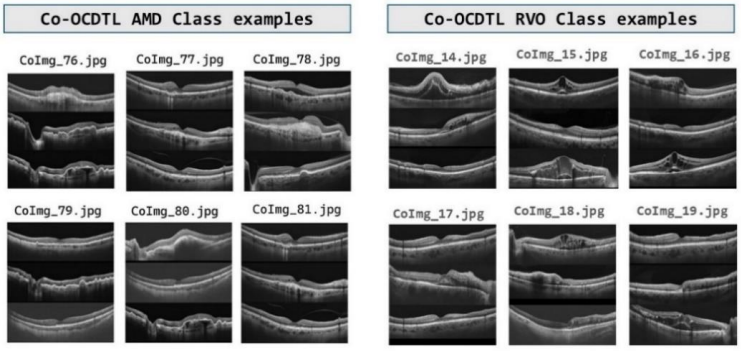


Figure 3. Co-OCDTL AMD and RVO classes examples

2.3. Perfect balancing class representation

The dataset undergoes combinatorial expansion after composite-image generation is applied. Most importantly, the number of informative samples in the minority class increases significantly, thus it gets better represented. The reason behind this is that we consider it as the baseline and all other classes must be aligned with it.

2.3.1. Balancing based on the minority class

For the majority of classes, we enforce strict non-reuse during composition. Each image appears once. No repetition; this way, the number of composites of these classes is naturally capped, while the most important information in each composite is preserved.

To balance classes, we set a target number of composites (T), which corresponds to the amount of expansion achieved for the minority class. Composites for the rest of the classes are produced under the following principle:

i) We enforce a strict "no-repetition" policy within each CoImg composite to ensure every training sample presents a unique visual context. However, data scarcity occasionally disrupts this.

ii) Should a majority class fail to reach the target T under these rigid constraints, we bridge the gap by integrating the remaining available scans, subjected to light geometric and photometric augmentations like subtle rotations, translations, or contrast shifts, to complete the layout. This balance between raw data integrity and synthetic variety.

This process guarantees that all classes eventually hit the same goal, hence a perfectly balanced data set is obtained while the content remains informative and diverse, as shown in Table 1, where each disease class

(e.g., AMD and diabetic macular edema (DME)) is either expanded or reduced up to non-repetitive composites or a slight increase.

Table 1. Distribution of generated samples across retinal disease classes

Disease class	Raw scans	CoImg	New representation of dataset
Age-related macular degeneration	1,231	310,144,295	Majority classes.
Diabetic macular edema	147	518,665	Expand/reduce up to T(1,540)
Epiretinal membrane (ERM)	155	608,685	
Normal	332	6,044,060	
Retinal vein occlusion (RVO)	101	166,650	
Vitreomacular interface disease (VID)	76	70,300	
Retinal artery occlusion (RAO)	22	1,540	Minority class. Minimal number of samples (T =1,540).
Total	2,064	317,554,195	10,780 (T×7)

2.3.2. Minority-class-driven balancing algorithm

Class imbalance distorts learning, especially when composite data amplifies the dominance of already overrepresented classes. Naive composite generation does exactly that: it multiplies what is abundant, starves what is rare, and trains the model on a skewed reality that looks statistically rich but semantically poor. Algorithm 2 breaks this pattern by enforcing a strict equilibrium: it fixes the number of composites per class using feasible k-combinations, not convenience or availability, forcing the dataset to grow symmetrically rather than organically. This strategy preserves class diversity by retaining every scan, avoiding the data loss typically required for parity. Figure 4 illustrates the resulting Co-OCTDL alignment, a perfectly balanced dataset.

Algorithm 2. Balanced co-dataset generation

```
Input: A dataset D with N classes, where class i has Nᵢ images; CoImg size k
Output: A balanced dataset where each class has T images (CoImg)
1: Use the procedure defined in Algorithm 1 as a CBCFunction
2: T ← min_{i=1,…,N} (Ni k) {Baseline: minimum possible k-combinations across classes}
3: for each class C with N_C ← |C| do
4:     M ← (Nc k)  {Total unique k-combinations in class C}
5:     if M > T then
6:         CBCFunction(k, D, 0) {Generate without intra-composite repetition}
7:     else
8:          CBCFunction(k, D, 1) {Generate all unique composites; if T is not reached,
complete with slight augmentation}
9:     end if
10:    Store the resulting T composite candidates (Co-Img) for class C
11: end for
12: Merge all class-specific T composite sets into the final balanced dataset
13: return Final balanced dataset
```

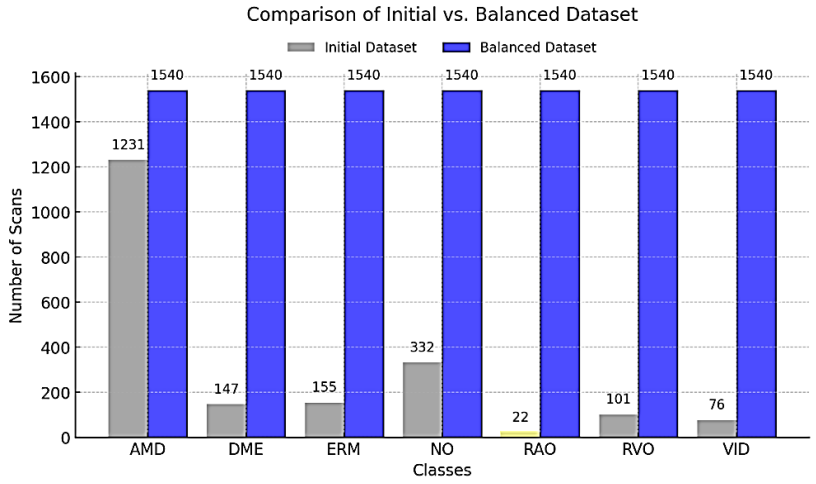


Figure 4. Benchmarking baseline performance against the balanced Co-OCTDL variant

2.4. Experimental setup and testing

To validate CB-ImgComp, we constructed a rigorous comparative framework. We stripped away confounding variables. This control guarantees that observed performance gains stem solely from superior data representation, not from architectural quirks or training anomalies.

2.4.1. Implementation framework

To guarantee computational reproducibility, we implemented the pipeline using PyTorch within the Kaggle environment. We optimized the configuration to process high-resolution composite images efficiently. Crucially, we maintained strict algorithmic consistency throughout the validation process.

2.4.2. Dataset configuration

To measure the specific impact of our approach, we compared two distinct dataset configurations:

i) Baseline configuration: we utilized the raw OCTDL dataset [23]. Comprising 2,064 OCT scans across seven imbalanced classes, this set is unmodified. We applied no balancing techniques. It functions strictly as a control benchmark.

ii) Experimental configuration (Co-OCTDL): the proposed dataset generated via the CB-ImgComp algorithm. This configuration comprises 10,780 composite images, structured in a 3×1 layout. As detailed in sub-section 2.3, this dataset is perfectly balanced, with each of the seven classes normalized to a target count of T =1,540 samples.

2.4.3. Model architecture and training

To ensure a valid comparison, the same VGG16 network was used for both dataset configurations. This architecture was chosen to be in line with the method used in the original OCTDL dataset benchmark [23], thus making it a consistent baseline for evaluation. Each of the models was trained with the same set of hyperparameters to remain the only variable for the changing input data configuration (244 as input size, 100 epochs, 64 batch size, Adam as optimization algorithm, 0.0003 as learning rate, and cross entropy as loss function).

2.4.4. Evaluation metrics

We quantified diagnostic reliability using accuracy, precision, recall, F1-score, and AUC. On top of that, for both runs, confusion matrices were created with the intention of visualizing class-specific misclassifications and evaluating the degree to which inter-class confusion was reduced. Figure 5 shows the complete experimental procedure, from initial data input and Co-OCTDL creation, through to VGG-16 training and comparison evaluation.

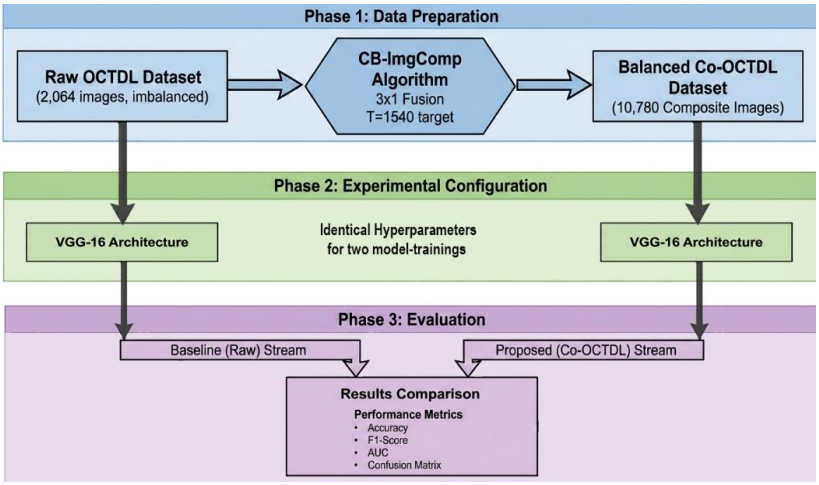


Figure 5. Experiment workflow schema for CB-ImgComp validation

3. RESULTS AND DISCUSSION

3.1. Results

As indicated in Table 2, the model trained on Co-OCTDL data reaches near-perfect performance, with accuracy climbing to 99.70% and the F1-score stabilizing at 0.997, a margin that decisively separates it from the baseline model, which remains constrained at 85.9% accuracy and 0.869 F1-score. This gap is not

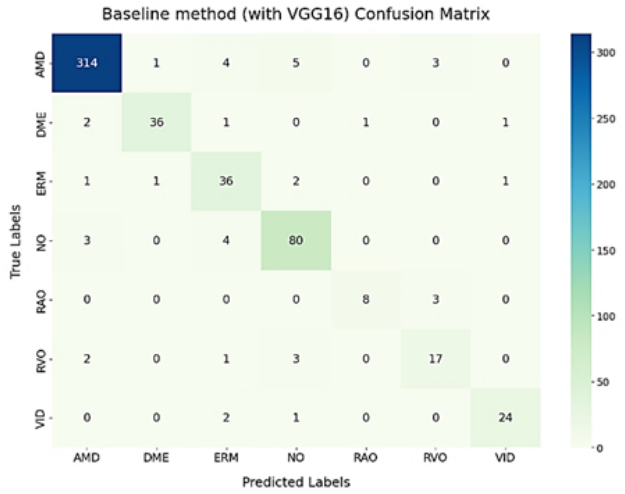
cosmetic, it reflects a structural improvement in representation learning, where Co-OCTDL reshapes the feature space rather than merely increasing sample volume. The most telling shift appears in the AUC, which rises from 0.977 to 0.9996, signaling an almost ideal separation boundary between retinal disease categories. In practical terms, the model no longer approximates class distinctions; it resolves them with near-deterministic precision.

Table 2. Comparative analysis of the proposed framework against the baseline

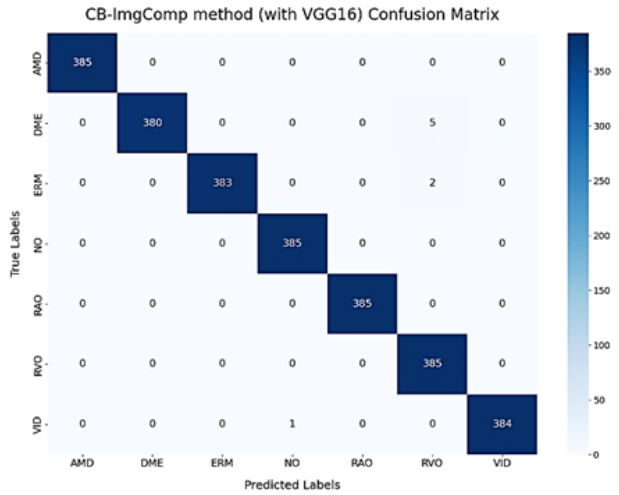
Method	Accuracy	F1-score	AUC	Precision	Recall
Baseline (VGG16-OCTDL)	0.859	0.869	0.977	0.888	0.859
CB-ImgComp (VGG16-Co-OCTDL)	0.997032	0.997036	0.999643	0.997079	0.997032

3.1.1. Analysis of misclassification errors

Figure 6 shows the comparative analysis of confusion matrices. The baseline model, as shown in Figure 6(a) [23] shows the spread of misclassifications, especially among classes that are supposedly most similar; for example, the confusion between RAO and RVO is shown, as well as confusion between AMD and normal (NO) classes. By comparison, the model with our CB-ImgComp method, as shown in Figure 6(b) has a nearly perfect diagonal pattern representing a tremendous decrease in mistakes. Practically, the false predictions were kept to negligible outliers, only a few cases like DME wrongly identified as RAO (5 samples) and ERM as RVO (2 samples) hence, the method's success in solving the baseline's limitation of feature ambiguity is thereby certified.



(a)



(b)

Figure 6. Comparative analysis of confusion matrices for (a) the baseline model and (b) the proposed CB-ImgComp approach

3.1.2. Advantages of the new dataset

We combine several images of the same class that belong to different samples through merging and thus in this way, the samples that were initially not as strong become strengthened by the strong ones and the new resulting composites that have a lot of features drastically improve the training process. This method essentially overcomes the problems of having a small number of data, or low-quality images, and at the same time, it gives the model a powerful input that results in better generalization.

i) Balanced classes of dataset: by setting every class to a fixed target count (T), the dataset has been made perfectly balanced, this way majority-class bias is completely removed, and diverse content is still preserved as can be seen clearly in (Table 1).
ii) Enhancing variability: each Co-Img fuses multiple images to consolidate relevant visual cues and enrich informative content, enabling the model to extract stronger, more generalizable features for improved learning. Figure 3 displays example composite images for the AMD and RVO classes within the new dataset representation.

3.2. Discussion

The VGG16 performance on Co-OCTDL upends the current fixation on complex architectural modifications. Input-level optimization proves itself to be the superior strategy. Standard augmentation often fails to bridge the semantic gap between minority and majority classes. CB-ImgComp succeeds. It does not simply replicate data; it constructs semantically dense representations. These force the network to learn robust class markers rather than background noise. This precision eradicates misclassification between clinically mirroring pathologies, specifically RAO and RVO.

Our composite approach preserves fine-grained vascular details often obliterated by standard down-sampling. By stabilizing the training distribution, we provide a concrete blueprint for data-scarce environments. The leap from 0.859 to 0.997 accuracy is not just a statistical gain. It marks a definitive transition from experimental uncertainty to clinical utility.

3.2.1. Clinical relevance and impact

One of the ways CB-ImgComp increases the trustworthiness of diagnostics (which medical AI trust depends on) is by greatly cutting down on the number of false positive and false negative results. The enhanced precision and recall imply that the model is quite sturdy. Hence, it is capable of safely minimizing error rates, which is of utmost importance for patients' safety [25], [26].

3.2.2. Implications of findings

Addressing small and imbalanced datasets: crucially, the technique thrives on small, unbalanced datasets. Functioning as a semantic augmentation [27] method, CB-ImgComp balances classes directly at the input level. This strategy outperforms standard algorithmic fixes like weighted loss functions. By supplying high-fidelity, feature-rich examples, we achieve stable training and superior generalization for the minority class.

Minimizing false diagnostic results: clinical utility requires minimizing diagnostic errors [26]. By sharpening pathological representations, we enabled the model to achieve superior discriminative power. This reduction in uncertainty is pivotal. It substantiates real-world deployment, strictly limiting the risks of missed diagnoses or unnecessary treatments.

4. CONCLUSION

We bypassed the constraints of small, imbalanced medical datasets by introducing CB-ImgComp. This method synthesizes information-dense, balanced training samples. The impact was distinct. Using standard VGG16 architecture on the OCTDL dataset, we achieved 99.7% accuracy and 0.9996 AUC. We deliberately prioritized input optimization over architectural complexity. These results confirm that refining data representation is a potent, model-agnostic alternative to engineering heavy networks. Our approach establishes a viable path for deploying robust AI in data-scarce environments. Future investigations will extend beyond OCT imagery. We intend to validate generalizability on magnetic resonance imaging (MRI) and histopathology and assess integration with emerging Transformer-based architectures.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Azzeddine Hlali	✓	✓	✓		✓		✓	✓	✓	✓	✓			
Majid Ben Yakhlef		✓		✓		✓		✓		✓		✓	✓	
Soulaiman El Hazzat		✓		✓		✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that supports the findings of this study are openly available in Kaggle.

– The experimental dataset is accessible at https://www.kaggle.com/datasets/azdineh/c-dataset-2025
– The baseline dataset (OCTDL) is available at https://www.kaggle.com/datasets/orvile/octdl-optical-coherence-tomography-dataset

REFERENCES

[1] M. J. Willemink *et al.*, "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, pp. 4–15, Apr. 2020, doi: 10.1148/radiol.2020192224.
[2] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0192-5.
[3] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," *Medical Image Analysis*, vol. 54, pp. 10–19, May 2019, doi: 10.1016/j.media.2019.02.010.
[4] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *Journal of Medical Systems*, vol. 42, no. 11, Nov. 2018, doi: 10.1007/s10916-018-1088-1.
[5] J. Morano, Á. S. Hervella, J. Rouco, J. Novo, J. I. F.-Vigo, and M. Ortega, "Improving AMD diagnosis by the simultaneous identification of associated retinal lesions," in *Image Analysis and Processing – ICIAP 2022*, Springer, Cham, 2022, pp. 148–159, doi: 10.1007/978-3-031-06427-2_13.
[6] H. Ghebrechristos, G. Alaghband, and R. Y. Hwang, "RetiNet — feature extractor for learning patterns of diabetic retinopathy and age-related macular degeneration from publicly available datasets," in *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2017, pp. 1643–1648, doi: 10.1109/CSCI.2017.286.
[7] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, Jan. 2022, doi: 10.1016/j.compmedimag.2021.102026.
[8] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2019, pp. 9260–9269, doi: 10.1109/CVPR.2019.00949.
[9] T. Ridnik *et al.*, "Asymmetric loss for multi-label classification," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 82–91, doi: 10.1109/ICCV48922.2021.00015.
[10] T. Huang *et al.*, "Masked distillation with receptive tokens," in *International Conference on Learning Representations (ICLR 2023)*, 2023.
[11] M. F.-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.
[12] M. N. Fekri, A. M. Ghosh, and K. Grolinger, "Generating energy data for machine learning with recurrent generative adversarial networks," *Energies*, vol. 13, no. 1, Dec. 2019, doi: 10.3390/en13010130.
[13] A. Rao, J.-Y. Lee, and O. Aalami, "Studying the impact of augmentations on medical confidence calibration," in *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2023, pp. 2454–2464, doi: 10.1109/ICCVW60793.2023.00260.
[14] S. Pan *et al.*, "Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model," *Medical Physics*, vol. 51, no. 4, pp. 2538–2548, Apr. 2024, doi: 10.1002/mp.16847.
[15] J. Chen *et al.*, "TransUNet: transformers make strong encoders for medical image segmentation," 2021, arXiv: 2102.04306.
[16] H. Cao *et al.*, "Swin-Unet: Unet-like pure transformer for medical image segmentation," in *Computer Vision – ECCV 2022 Workshops*, 2023, pp. 205–218. doi: 10.1007/978-3-031-25066-8_9.
[17] E. Hassan *et al.*, "Enhanced deep learning model for classification of retinal optical coherence tomography images," *Sensors*, vol. 23, no. 12, Jun. 2023, doi: 10.3390/s23125393.
[18] N. J. Bilal and R. Z. Yousif, "Neutrosophic-inspired data augmentation and multi-architecture ensemble model for robust retinal disease classification," *Engineering Research Express*, vol. 7, no. 3, Sep. 2025, doi: 10.1088/2631-8695/adfc29.

[19] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: understanding transfer learning for medical imaging," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019, pp. 3347–3357.
[20] W. Li *et al.*, "Interpretable detection of diabetic retinopathy, retinal vein occlusion, age-related macular degeneration, and other fundus conditions," *Diagnostics*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/diagnostics14020121.
[21] Z. Li *et al.*, "A deep-learning pipeline for the diagnosis and grading of common blinding ophthalmic diseases based on lesion-focused classification model," *Frontiers in Artificial Intelligence*, vol. 7, Sep. 2024, doi: 10.3389/frai.2024.1444136.
[22] M. Driban, A. Yan, A. Selvam, J. Ong, K. K. Vupparaboina, and J. Chhablani, "Artificial intelligence in chorioretinal pathology through fundoscopy: a comprehensive review," *International Journal of Retina and Vitreous*, vol. 10, no. 1, Apr. 2024, doi: 10.1186/s40942-024-00554-4.
[23] M. Kulyabin *et al.*, "OCTDL: optical coherence tomography dataset for image-based deep learning methods," *Scientific Data*, vol. 11, no. 1, Apr. 2024, doi: 10.1038/s41597-024-03182-7.
[24] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016, doi: 10.1109/TMI.2015.2508280.
[25] S. Chatio, N. A. Ansah, D. A. Awuni, A. Oduro, and P. O. Ansah, "Community acceptability of seasonal malaria chemoprevention of morbidity and mortality in young children: a qualitative study in the upper West Region of Ghana," *PLOS ONE*, vol. 14, no. 5, May 2019, doi: 10.1371/journal.pone.0216486.
[26] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.
[27] Y. Zhu, X. Cai, X. Wang, X. Chen, Z. Fu, and Y. Yao, "BSDA: Bayesian random semantic data augmentation for medical image classification," *Sensors*, vol. 24, no. 23, Nov. 2024, doi: 10.3390/s24237511.

BIOGRAPHIES OF AUTHORS

Azzeddine Hlali a Ph.D. student in Intelligent Systems and Biomedical Informatics, he is a Master's in Intelligent and Mobile Systems from Sidi Mohamed Ben Abdellah University (2021). His research targets the intersection of AI and healthcare, specifically examining biomedical imaging and electronic health records. Previously, he deployed deep learning models for COVID-19 pneumonia screening using radiological data. His doctoral work now aims to engineer intelligent solutions for massive healthcare datasets to sharpen diagnostic accuracy and automate clinical decisions. For this study, he led conceptualization and methodology. He developed the software framework, validation, and investigation, managed data curation, and authored the original draft. He can be contacted at: azzeddine.hlali@usmba.ac.ma.

Majid Ben Yakhlef holding a Doctorate from the Faculty of Sciences at Sidi Mohamed Ben Abdellah University (USMBA), currently, he serves as a professor of Computer Science at the Faculty of Multidisciplinary Studies in Taza. He is a permanent member of the Engineering Sciences Polydisciplinary Laboratory. His research creates a nexus between computer engineering, automation systems, and artificial intelligence, specifically investigating intelligent decision-making technologies. He engineers AI-based solutions for complex, real-world applications. Beyond theoretical inquiry, his focuses on advanced computational methods and intelligent control systems. He balances these scholarly pursuits with the rigorous mentorship of students, actively driving innovation within engineering sciences while integrating emerging technologies into modern computational frameworks. He can be contacted at email: majid.benyakhlef@usmba.ac.ma.

Soulaiman El Hazzat earning his Ph.D. in Computer Science from Sidi Mohamed Ben Abdellah University in 2018, currently, he serves as a professor at the Polydisciplinary Faculty of Taza. A core member of the Engineering Sciences Polydisciplinary Laboratory, he drives research in artificial intelligence and advanced interactive technologies. His focus is sharp. It spans computer vision, deep learning, and immersive environments like VR and AR. Crucially, he investigates deepfake detection, engineering innovative methods to identify manipulated media and bolster digital security. For this specific study, he validated and interpreted the experimental results. He ensured scientific rigor by reviewing and refining the final manuscript. Beyond technical analysis, his role extended to high-level supervision and project administration, including the acquisition of necessary resources. He can be contacted at email: soulaiman.elhazzat@usmba.ac.ma.