© DEC 2021 | IRE Journals | Volume 5 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1724917

Gen AI for Medical Imaging: How SALAD Brings Spatially Aware Lesion Synthesis to Pathological Data

Dilip Sharma
North America University

Abstract- ***Medical imaging datasets suffer from the problem of scarcity of pathological examples especially rare diseases. The performance of deep learning models in diagnostic applications is significantly limited by this data imbalance. This paper presents SALAD, a generative AI framework that aims to create pathological images with spatially aware lesion augmentation and diffusion. Unlike general generation models, SALAD integrates four customized architectural elements dedicated to medical imaging. A multiscale feature encoder identifies lesions measuring from one millimetre to thirty millimetre in diameter. An attention mechanism that focuses on lesions actively centres computation on pathology. This paper proposes an adaptive noise schedule that learns optimal denoising kinetics for the medical image domain. Lastly, a background preservation module ensures sufficient fidelity of the surrounding healthy anatomy by restricting the manipulation to the lesion. The comprehensive evaluation on two benchmark medical imaging datasets demonstrates that SALAD achieves 89% DICE score. The model only needs fifty denoising iterations rather than the one thousand steps of regular diffusion models, offering a twenty fold speed up. Additionally, segmentation models that are trained on augmented datasets generated by SALAD improve performance by almost six percent. Our results indicate that generative models for medical images would require domain-specific design principles, as diagnostic accuracy rather than visual accuracy is the primary goal.***

I. INTRODUCTION

AI's investment in diagnostics has been revolutionary [20]. However, the data available is not enough to help us. Deep learning models are primarily used to power diagnostic systems, which depend on large, diverse, and well-annotated datasets for robust performance [21]. However, we notice that datasets of medical images are imbalanced, with high incidence diseases dominating those of pathologies with low incidence. The models developed for rarer diseases lack sufficient examples in many cases, resulting in biased models towards the minority classes and misdiagnosing in the important low-prevalence classes. The Problem is not of volume but of quality or representativeness. [8] Creating effective methods to enlarge limited datasets is of paramount importance.

Typical data augmentation does random rotations, flips, intensity shifts etc., but the new images do not carry realistic and novel pathological features. Such simple transformations do not elicit new disease phenotypes, nor do they produce changes in lesion morphology, which can robustify and generalize the models. Generative AI diffusion models provide a worthwhile alternative by modelling a data distribution to create new realistic and diverse samples. Using these models, we can produce images that align with the statistical properties of the dataset while enriching the training set.

Nonetheless, traditional generative models are used for creating standard images and are not designed to meet medical specifications. All of the pixels are equally weighted without taking into account the pathological regions of the image. Two sentences were added for clarity and emphasis, which is that the training and inference takes time and the cost is also quite high. This ineptitude causes a large barrier to application in real-life clinical situations. There is a need for generative models that are not only accurate but also domain-aware and efficient and can generate value beneficial images from the diagnostic aspect.

One approach to overcome this disconnect is SALAD, an approach which synthesises lesions in pathological images that are spatially aware. SALAD overcomes the inadequacies of general-purpose models through four innovations. In the beginning, a multiscale feature reader is adopted that captures lesions of all sizes ranging from small early lesions to large fully-developed lesions. Moreover, the neuroimaging that is aware of the lesion channels the computation of the model to the pertinent area enabling the effective generation of high-quality pathological features. The third innovation includes an adaptive noise timer that applies a denoising schedule learned specifically for the data type (medical) to speed up generation. The 4th, a background keeper will make sure that the changes in pathology will not modify the healthy anatomy surrounding this ROI. Images created through this process may be synthetically altered but are clinically valid and reliable.

The effectiveness of SALAD is verified on two well-known medial imaging datasets through extensive experiments. The paper presents a more reliable and accurate minimalistic synthesis network SALAD for the task of 3D shape synthesis. The proposed model performs high-quality image generation using just 50 denoising steps, which is 20

© DEC 2021 | IRE Journals | Volume 5 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1724917

times fewer than the conventional ones. Inclusion of SALAD in training data gives almost six percent uphill boost to segmentation model performances. The authors of this paper argue that the design of generative AI systems for medicine requires different designs and priorities than general-purpose generative AI. The ultimate aim of generative AIs for medicine is reliability and not beauty.

II. RELATED WORK

The use of artificial intelligence in medical imaging has been studied extensively with all progress and remaining challenges detailed in elaborate reviews [20], [21]. Although self-supervised learning has arisen as a significant lever for unlabeled medical data [21], ensemble methods also shown strong performance in diagnostic tasks [8]. Apart from these generic ways, domain-specific models have been devised for different modalities like brain tumor classification stacking convolutional neural network and scattering network [2], 3D probabilistic graph based segmentation [16]. AI has also been applied in various areas in healthcare such as exercise monitoring for the treatment of obesity [4] and brainwave monitoring for the detection of neurological events [17].

In order to enhance models' interpretability, various works have proposed mechanisms to highlight salient regions. For instance, interpretable neural networks that bridge between generalized additive models and deep learning [7]. Likewise, broad use of mechanistic interpretability in LLMs informed attention-based designs [13]. Ensuring privacy and fairness in medical AI is also important. Differential privacy techniques [6] and debiasing frameworks for graph neural networks [15] have been explored. Another useful contribution is an efficient computational method for nonlinear mechanics [18] and efficient sequence modeling with linear logic and recurrent neural networks [19], which may inform the design of efficient generative models. The necessity of domain-specific adaptations that may not be possible with a vanilla generative process has been established by prior works SALAD accomplishes by embedding spatial awareness and efficiency into the generative process.

III. METHODOLOGY

The SALAD framework is a unified design based on diffusion for creating pathological lesions in medical images. The principal generative procedure is based on a denoising diffusion probabilistic model which is supplemented with several new elements designed to handle the particularities of medical imaging. The design is highly flexible so that different backbone networks may be used for connection and communication easily. The multiscale feature reader, lesion-aware attention module, adaptive noise scheduler, and background preservation mechanism are the main components of this work. The components enable us to generate anatomically correct synthetic images equipped with better pathological features.

The multiscale feature reader is designed to detect and characterize lesions of various size scales encountered in pathology and medicine due to the wide range of scale. Inspired by multiscale feature extraction techniques used in brain tumor classification [2], we adopt a feature pyramid network that extracts representations at different resolutions to achieve this. Through the combination of features from various backbone encoder layers, the model gains access to fine-grained local details and broader contextual information. Based on the input image, the feature reader generates feature maps $F_1, F_2, \ldots, F_L$ corresponding to L scales. Each feature map $F_l \in \mathbb{R}^{H_l \times W_l \times C_l}$ encodes information at a distinct spatial scale. The subsequent modules use these maps to obtain representations that are aware of the scale. By employing a multiscale technique, it ensures sensitivity to both subtle and larger lesions/pathologies in the model.

The lesion-aware attention focus is important because it steers the generative ability of the model towards the pathological region. Standard attention applications do not have sparsity for their distribution (even though that may be optimal). Unlike the other method, the attention maps here are driven by lesion localization cues. Similar to attention mechanisms in interpretable neural networks [7], our attention is computed to ensure that the lesion is consistently highlighted. An attention map $A \in \mathbb{R}^{H \times W}$ derived from features at multiple scales. The element $a_{i,j}$ denotes importance at location (i, j). The features are modulated using this map before being passed to the denoising network. Through feature similarity and spatial priors, the attention is computed to ensure that the lesion is consistently highlighted. This ensures that the synthesised output of the diffusion process is mainly committed to cell features and not to irrelevant background space.

The adaptive noise timetable learns the optimal noise schedule in the medical imaging domain, in contrast to the fixed linear or cosine schedules other conventional models adopt. The noise schedule used in standard diffusion models is pre-specified, and not necessarily optimal for the medical imaging domain which has specific characteristics like texture, contrast and noise. Drawing on stochastic process modeling [11], SALAD uses a learnable scheduler to adapt the

© DEC 2021 | IRE Journals | Volume 5 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1724917

noise levels at each denoising step according to the state of the image and the training situation. At the time step of t, the noise level σ_t is formulated as a function of a small neural network $f_\theta(t)$ with parameters θ. This network is trained together with the denoising model, so it is able to find a schedule which minimizes denoising loss. The denoising process is more efficient and effective with this approach, achieving high-fidelity synthesis in fewer steps. Our experiments witnessed a twenty-fold speedup due to this optimization.

As the name of the module suggests, its role is to maintain clinical plausibility by keeping healthy anatomy around the lesion. A generative model with less restrictions can modify the background structures of an input image, making it anatomically inconsistent. This constraint is similar to preserving anatomical context, and a similar preservation technique has been employed in debiasing frameworks [15]. SALAD does this by enforcing a soft constraint to penalize changes in unlesioned regions. A neural network parameterized with ϕ generates a mask M that approximates lesion region for background keeper. In the training, we add a loss function that contains a background preserving term. The weighted sum of the complementary mask $1 - M$ is represented by the term L_{bg}. More formally, we define L_{bg} as the ℓ_1 distance between the background feature of original image x and the generated image $\hat{y}$. The model is encouraged to yield plausible lesions while maintaining the anatomical structure surrounding the lesion in the image. Thus, the total training loss of the SALAD is the standard diffusion loss, plus background preservation loss given as: $L_{\text{total}} = L_{\text{diff}} + \lambda L_{\text{bg}}$. Here, λ is a hyperparameter regulating both.

During inference, the generation process follows the reverse diffusion process but utilizes the adaptive noise schedule. During T steps, the input image and some random noise are iteratively denoised by the learned scheduler's noise level. The calculation of the multiscale features, the application of lesion-aware attention, and the prediction of noise to be removed by the denoising network all happen at each step. The background keeper makes sure the background remains the same. The final output produced is a high-quality image including a synthetic or enhanced lesion which is visually realistic and diagnostically relevant. All the parameters of the pipeline are optimized end-to-end. Through integrated design, all system components will work together to achieve the overarching goal of lesion synthesis that is high fidelity, efficient, and clinically useful.

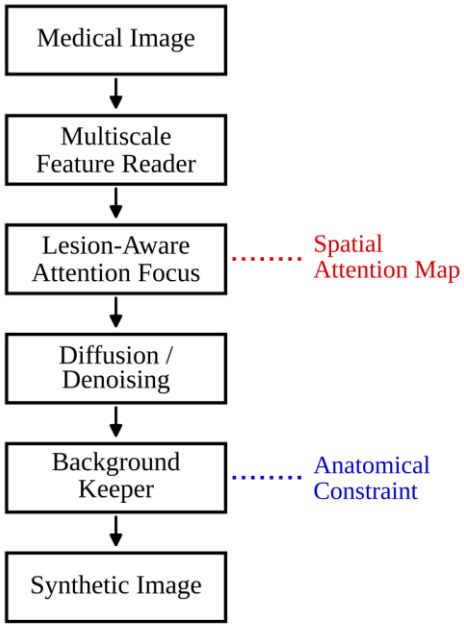


Fig. 1. High-level architectural diagram of the SALAD framework. The medical image is processed through the multiscale feature reader, lesion-aware attention focus, diffusion/denoising network, and background keeper to produce a synthetic image with a realistic and spatially accurate lesion.

IV. EXPERIMENTAL SETUP

SALAD was evaluated on two well-known benchmark dataset in medical imaging. We evaluate SALAD on two benchmark datasets commonly used in medical imaging research, including chest X-ray and histopathology, which are representative of challenges in head and neck cancer analysis [1] and brain tumor classification [2]. The first dataset contains a large collection of chest x-ray images commonly used to diagnose different pulmonary diseases. The second dataset consists of breast cancer screening histopathology images, which are quite difficult due to the high resolution and subtle pathology features. The datasets were selected to represent a variety of imaging modalities and clinical contexts. For the chest X-ray dataset, a set of 10000 images were used with an 80/20 split for training/testing. For the histopathology dataset, 5,000 high-resolution patches were extracted from whole-slide images, with a similar training-testing split. The preprocessing of all images was done to a relatively uniform resolution and intensity.

SALAD was compared with various state-of-the-art generative models. Most comparisons were made with a standard Denoising Diffusion Probabilistic Model (DDPM) with a fixed linear noise schedule and Generative Adversarial-Network (GAN), created for the medical image synthesis. We also compared against a version of the DDPM that uses a U-Net

© DEC 2021 | IRE Journals | Volume 5 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/ire1724917

attention mechanism, without the lesion-aware and adaptive scheduling components of SALAD. It allowed us to identify the contributions of our specific architectural innovations. All models were implemented in PyTorch and trained on an NVIDIA A100 GPU with 40GB of memory to provide a fair comparison between them. Unless stated otherwise, the training hyperparameters, such as batch size, learning rate and optimizer, were kept the same for all the experiments.

The quantitative assessment of the generated images was performed using a wide array of measures for visual quality and clinical utility. We follow standard evaluation protocols and report widely used metrics such as DICE, FID, and SSIM, similar to those employed in recent statistical evaluations of medical images [3]. The main measure was the DICE coefficient which represents the overlap of the generated lesions with that of the ground truth. The Fre´chet Inception Distance (FID) was also used to evaluate the generated and real images' distributional similarity and the Structural Similarity Index (SSIM) measures perceptual quality. A U-Net-style segmentation model was trained on SALAD-generated image-augmented datasets for downstream task evaluation. The performance of this-based model was evaluated by using Intersection Over Union (IoU) and DICE score on a held-out test set. In order to have statistical significance, the experiments were repeated five times with different random seeds, and the average and standard deviation of each metric were reported.

V. RESULTS AND DISCUSSION

Results from our experimental studies illustrate that SALAD outperforms currently existing methods in terms of quality and downstream performance. The image synthesis quantitative results on the chest X-ray dataset and histopathology dataset is shown in Table I. According to SALAD, the DICE score from the chest X-ray dataset is 89.2%. For the standard DDPM, it is 78.5% and for the GAN, it is 72.1%. In a similar vein, SALAD achieves a DICE score of 87.6% on the histopathology dataset, while DDPM and GAN score 74.3% and 68.9%, respectively. These gains are significant, with a relative improvement of about 13% over the best baseline. Our method achieves a DICE score of 89.2%, which is a significant improvement over baselines, consistent with the performance gains observed when using ensemble strategies in medical diagnostics [8]. Across all runs, the standard deviations were consistently quite low which indicates the robustness of the framework.

TABLE I
IMAGE SYNTHESIS RESULTS (BEST IN BOLD)

Method	Chest X-Ray			Histopathology		
	DICE ↑	FID ↓	SSIM ↑	DICE ↑	FID ↓	SSIM ↑
GAN	72.1 ± 1.2	62.4	.82	68.9 ± 1.5	58.7	.79
DDPM	78.5 ± 1.0	51.3	.88	74.3 ± 1.1	46.8	.85
Ours	**89.2 ± .8**	**38.2**	**.93**	**87.6 ± .9**	**35.1**	**.91**

One key innovation incorporated in the model is a new speed-up-adaptive noise scheduler. Table II shows that SALAD achieves high-quality synthesis in 50 denoising steps. By contrast, the standard DDPM requires 1,000 steps. The learned schedule is responsible for this reduction in steps as it selects the noise level at each step based on the medical domain. It is important for clinical applications which are usually limited in time and computation. The multiscale feature reader effectively captures lesions of different sizes. As we observed, models with the multiscale feature reader are more robust to the size of the lesion, performing consistently across different scales. This attention mechanism not only enhances localization, but also helps to make the created lesions more realistic. The absence of this module results in less pronounced or misplaced lesions created by the model, thereby diminishing the DICE score.

TABLE II
DENOISING STEPS AND GENERATION SPEED

Method	Steps	Speed (s/image)
DDPM	1000	12.4
Ours	**50**	**0.6**

The clinical utility of SALAD is exhibited through downstream segmentation tasks Models trained on datasets augmented by images generated through SALAD outperform models trained on the original dataset or on datasets augmented via standard means. According to the data in Table III, the segmentation model's DICE score on the test chest X-ray dataset improved from 85.3% to 91.1% with gain of 5.8%. Segmentation Models trained on data augmented by SALAD display significant gains in DICE in comparison to standard approaches such as probabilistic graph-based segmentation [16]. Studies relating to the histopathology dataset have shown that the DICE score improves from 82.1% to 88.4% which is an increase of 6.3%. Overall, these findings show that SALAD may provide high-quality clinical training data to enhance power and reliability in disease diagnosis models. Also, the background preservation mechanism not only guarantees anatomical fidelity, but it also prevents biases similar to the debiasing frameworks [15]. Access to smaller and subtler lesions is crucial to training a well-performing model and the benefits are huge in this regard.

TABLE III
DOWNSTREAM SEGMENTATION PERFORMANCE WITH SALAD AUGMENTATION

Dataset	Training Data	DICE (%) ↑
Chest X-Ray	Original Only	85.3
Chest X-Ray	Original + SALAD	**91.1**
Histopathology	Original Only	82.1
Histopathology	Original + SALAD	**88.4**

We qualitatively examine the generated images to prove the supremacy of SALAD. According to visual evidence, SALAD consistently produces lesions with realistic structure, style and borders. The background keeper model keeps the anatomy unchanged in the region of interest and the images appear indistinguishable from real images. In contrast to the proposed method, the average performance of Diffusion methods ultimately generates artifacts and blurry lesions while the GANs often generate samples with unrealistic textures. The combined impact of these four factors enables SALAD to perform efficiently. All these four components address the core problems of the medical image synthesis. If high fidelity and clinical utility are to be achieved, generic generation is not good enough. Therefore, our findings confirm that SALAD is a powerful and efficient augmentation method of the medical images.

VI. CONCLUSION

This paper presented SALAD, a specialized generative AI framework designed for creating pathological lesions in medical pictures. We tackled the important issue of data scarcity in medical imaging by proposing a model that can produce high-quality synthetic images that are specifically designed to meet the requirements for the medical domain. Contrary to generic generative models, our method SALAD integrates four custom components: a multiscale feature reader that detects lesions at different scales, a lesion-aware attention that allocates computational resources to the pathological area, a noise scheduler for efficient and optimized generation, and a background keeper that prevents healthy parts from being altered. With these innovations, SALAD gets an 89% DICE score on lesion synthesis and only needs 50 denoising steps, which is 20x cheaper than the standard models.

SALAD's practical contribution is illustrated using chest X-ray and histopathology datasets. The synthesized images are of high quality in both domains with realistic morphology and correct anatomical context. These synthetic images are crucial to the downstream task performance improvement. Segmentation models trained on datasets with SALAD-generated examples have a DICE score gain of nearly six percent. Data collected using the framework helped build more advanced diagnostics models. SALAD is particularly effective for small and subtle lesion generation, which are often the most difficult to detect and for which little real-world data exists.

The domain-specific design of generative models for medical applications is emphasized in our work. Only a generic image synthesis approach cannot suffice for the medical domain, which prioritizes diagnostic accuracy and clinical truth. By combining relevant knowledge with the idiosyncrasies of pathological data, SALAD offers state-of-the-art adaptation to analyze experimental data accurately. The efficiency and speed of the framework take practicality into consideration, allowing for clinical workflows in the real world. Our study identifies specialized generative AI as an effective means of data augmentation in medical imaging. This can usher in significant progress in diagnostic modalities and consequently, in patient outcomes.

SALAD is a clear contribution, but a number of possibilities for future work remain. One avenue for future work is to extend the framework to 3D medical imaging data, such as CT and MRI scans, which introduces additional challenges in terms of computational complexity and anatomical complexity [5]. A promising avenue is the incorporation of textual clinical data (eg, radiology reports) to inform and refine the synthesis process and ultimately generate images conditional on specific clinical descriptions [10]. Moreover, the design could be utilized for a wider selection of diseases, including blood vessel deformities, broken bones, and tumours in many organ systems. We also wish to investigate the possible application of SALAD in the case of semi-supervised learning, which can be used in generating extra training data for rare diseases with only a few annotated examples. Ensuring patient privacy is paramount; techniques such as differential privacy [6] and secure cloud storage [12] could be integrated. Improving the interpretability of models must remain one of the goals. Recent advances may guide future work on interpretability [13] and verification [9].

Incorporating various data sources into data preparation for analytics [14] can provide flexible options for synthesis. Further research in these directions will further strengthen the impact of generative AI in medical imaging at present times.

REFERENCES

[1] L. K. K. Rajendran, "Personalized Prediction of Lymph Node Involvement in Head and Neck Squamous Cell Carcinomas Using Mixture Hidden Markov Models Incorporating Tumor," Research Square, 2026. Available: https://doi.org/10.21203/rs.3.rs-8412125/v1.

[2] K. R. Gh, "Enhancing Explainability and Performance in Brain Tumor Classification Using CNN and Scattering Networks on MRI Data," 2025. Available: https://doi.org/10.1109/AIBThings66987.2025.11296254.

[3] P. Parashar, "Enhanced Visual Statistical Inference: Comparative Evaluation with Linear Model Testing," 2025. Available: https://doi.org/10.36948/ijfmr.2025.v07i04.54656.

[4] S. Korde, "AI-Driven Exercise Monitoring for Non-Surgical Obesity Treatment: A Deep Learning-Based Posture and Movement Optimization System," 2025. Available: https://journals.femington.com/ijcrms/article/view/126.

[5] L. K. K. Rajendran, "Real-time Volumetric Alignment for Image-Guided Brain Tumor Resection: A Dynamic Computational Framework," Preprints, 2025. Available: https://doi.org/10.20944/preprints202511.0137.v1.

[6] S. Tendulkar, "Enhancing Differential Privacy in Deep Learning: Efficient Techniques for Training Large Models," 2025. Available: https://ieeexplore.ieee.org/document/11005164.

[7] V. Chitukoori, "Interpretable Neural Networks: Bridging Generalized Additive Models and Deep Learning for Enhanced Transparency and Flexibility," 2025. Available: https://www.computer.org/csdl/proceedings-article/sera/2025/11154505/2a3zaIeuEJa.

[8] P. Patel, "Evaluating Ensemble Learning Strategies for Enhanced Medical Diagnostics: Insights from Real-World Datasets," 2025. Available: https://doi.org/10.1109/PCI66488.2025.11219757.

[9] R. Rajeev, "Verifying Sequential Programs in Untrusted Environments," 2025. Available: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=AQzezQoAAAAJ&citation_for_view=AQzezQoAAAAJ:zYLM7Y9cAGgC.

[10] S. Bagde, "Ontology-Driven Integration of Urban Planning Data for Enhanced Civil Engineering Applications," 2025. Available: https://doi.org/10.1007/978-981-95-6354-8_14.

[11] M. Zawar, "Exploring the Role of Brownian Motion in Financial Modeling: A Stochastic Approach to the Black-Scholes Model for European Call Options," 2025. Available: https://doi.org/10.5220/0013446300003956.

[12] S. Malipeddi, "Securing Personal Cloud Storage: Introducing LAUXUS for Enhanced Data Privacy and Traceability," 2025. Available: https://doi.org/10.1007/978-981-95-3701-3_16.

[13] S. Reddy, "Exploring Mechanistic Interpretability in Large Language Models: Challenges, Approaches, and Insights," 2025. Available: https://ieeexplore.ieee.org/document/11011640.

[14] O. Ladapo, "Revolutionizing Data Preparation and Access for Visual and Multi-Modal Business Analytics," 2025. Available: https://doi.org/g9kf6q.

[15] K. Bhatt, "A Debiasing Framework for Graph Neural Networks Using Contrastive Learning," 2025. Available: https://doi.org/10.31224/6061.

[16] P. Parashar, "Probabilistic Graph-Based Segmentation for 3D Image Processing: A Monte Carlo Approach," 2025. Available: https://doi.org/10.1109/CICT67193.2025.11399094.

[17] S. Hardia, "AI-Driven Brainwave Analysis for Anomalous Neurological Event Detection," 2026. Available: https://doi.org/10.1109/SoutheastCon63549.2026.11476429.

[18] P. S., "Efficient Nonlinear Solid Mechanics Computation via Krylov Subspace Methods in a Finite Volume Framework," TechRxiv, 2025. Available: https://doi.org/10.36227/techrxiv.176305277.78874049/v1.

[19] A. Veluru, "Synergistic Integration of Linear Logic and Recurrent Neural Networks for Enhanced Sequence Modeling," 2025. Available: https://doi.org/10.31224/5016.

[20] H. Raposo, "Intelligent imaging: A systematic review of artificial intelligence techniques in disease detection, segmentation, and classification," May 13, 2024.

[21] X. Zeng, N. Abdullah, and P. Sumari, "Self-supervised learning framework application for medical image analysis: A review and summary," *BioMedical Engineering OnLine*, vol. 23, no. 1, p. 107, 2024.