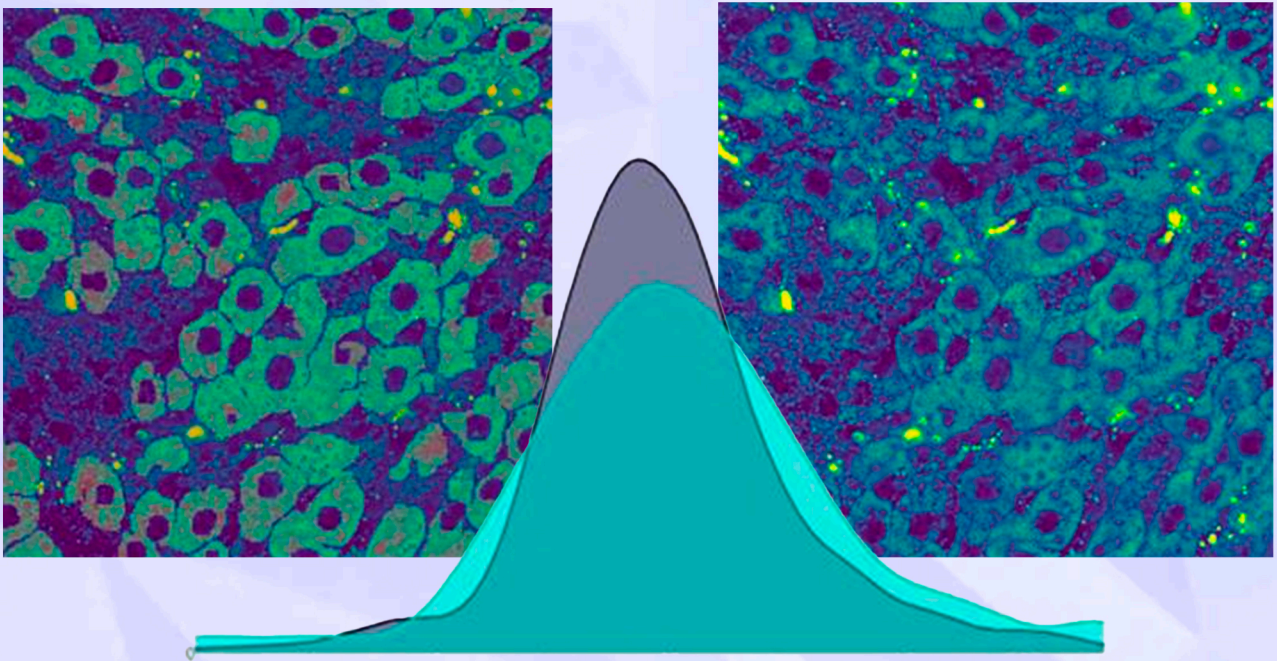


Journal

of Biomedical Photonics & Engineering



Editor-in-Chief:
V V Tuchin

© 2025 J-BPE

Journal of Biomedical Photonics & Engineering

by Samara National Research University

Volume 11, Issue 1 (March 2025)

Editor-In-Chief

Valery V. Tuchin – Saratov National Research State University, Russian Federation;
Institute of Precision Mechanics and Control RAS, Russian Federation; a corr.-member of RAS

Deputy Editor-In-Chief

Valery P. Zakharov – Samara National Research University, Russian Federation

Postal address:
39B Lukacheva st., room #301
Samara 443086, Russian Federation

Used Figures from this issue on the Cover

Ilya D. Shchekhin, Recent Trends of Fluorescence Lifetime Imaging Microscopy Analysis Using Machine Learning, 010201, Fig. 5.

Vandana Joshi and Nirali Nanavati, ADHD Subtype Diagnosis for Multiclass Imbalanced Augmented EEG Data, 010302, Fig. 6(a).

Contents

Reviews

Recent Trends of Fluorescence Lifetime Imaging Microscopy Analysis Using Machine Learning 010201
Ilya D. Shchechkin

Articles

Design of a Photonic Integrated Circuit for Swept-Source OCT for Simultaneous Operation in Two Distinct Wavelength Ranges 010301
Ivan V. Stepanov, Vladimir S. Lyubopytov, Anton A. Ivanov, and Elizaveta P. Grakhova

ADHD Subtype Diagnosis for Multiclass Imbalanced Augmented EEG Data 010302
Vandana Joshi and Nirali Nanavati

Seismocardiograph Monitoring Using SMS Fiber Structure with PDMS Enclosure 010303
Frans Rizal Agustiyanto, Agus M. Hatta, Dhany Arifianto, Mahenda Radityo, Maulana Putra Santoso, Selfi Stendafity, and Budi Susetio Pikir

Dynamic Conditional Encoding and Feature Frequency Parsing in Diffusion Probabilistic Models for Diabetic Foot Ulcer Detection Using Thermographic Imaging 010304
Kousar Nikhath A, Sireesha C, Swapna M, Vijayetha Thoday, and Esther Rani D

Non-Contact Neonatal Heart Rate Monitoring Using Blue Filter Technique with Photoplethysmography Imaging under UV Illumination 010305
Huda Ali Hashim

Optimized Classification with CBGRU-CapNet on Thermographic Analysis to Detect Diabetic Foot Ulcers with Precise Diagnosis 010306
Ashok Babu Ane, Gireesh Babu C N, Karthik Govindan Manoharan, Rafeeq Md., Bhupathi Rajarao, and Kadiyala Vijaya Kumar

Recent Trends of Fluorescence Lifetime Imaging Microscopy Analysis Using Machine Learning

Ilya D. Shchechkin^{1,2}

¹Privolzhsky Research Medical University, Institute of Experimental Oncology and Biomedical Technologies, 10/1 Minin and Pozharsky sq., Nizhny Novgorod 603950, Russian Federation

²N.I. Lobachevsky Nizhny Novgorod National Research State University, 23 Gagarina ave., Nizhny Novgorod 603022, Russian Federation

e-mail: liahasa1992@gmail.com

Abstract. Fluorescence Lifetime Imaging Microscopy (FLIM) is an advanced imaging technique that provides quantitative information about molecular interactions and changes in the microenvironment by measuring the emission decay lifetimes of molecules undergoing fluorescence. The integration of FLIM with machine learning holds great promise for advancing biomedical research and clinical diagnostics by enabling precise quantification of cellular processes and microenvironmental changes. As FLIM instrumentation becomes more accessible and analytical methods continue to improve, the technique is poised to have a significant impact across various fields of biology and medicine. This review considers a range of methods of applying Machine Learning to FLIM data processing. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: deep learning; segmentation; denoising; upscaling.

Paper #9185 received 24 Oct 2024; revised manuscript received 21 Jan 2025; accepted for publication 19 Feb 2025; published online 23 Mar 2025. doi: [10.18287/JBPE25.11.010201](https://doi.org/10.18287/JBPE25.11.010201).

1 Introduction

Fluorescence Lifetime Imaging Microscopy (FLIM) is a cutting-edge imaging technique that leverages the unique fluorescence lifetime of molecules to provide valuable insights into biological processes at the molecular level. When combined with Machine Learning (ML), FLIM imaging opens up new frontiers in biomedical research and diagnostics by enabling the precise analysis and interpretation of this type of complex fluorescence data.

FLIM is an advanced imaging technique that utilizes the fluorescence lifetime of molecules to generate images based on differences in the decay rates of their excited states. Unlike traditional imaging methods that rely on emission spectra, FLIM creates contrast based on the lifetimes of individual fluorophores. The fluorescence lifetime represents the average time a molecule remains in an excited state before returning to the ground state by emitting a photon [1].

FLIM works by capturing a 2D grid of pixels in the sample, where the fluorescence lifetime of each pixel is determined using time-correlated single photon counting (TCSPC). This involves aiming an excitation laser at each spot, collecting emission photon arrival times, and sorting them into a histogram to extract the fluorescence lifetimes mathematically. By including fluorescence

lifetime information, FLIM not only enhances image contrast but also provides additional insights into the local environment, molecular interactions, and various cellular processes that are not accessible through intensity imaging alone [1]. The principle of obtaining FLIM data is presented in Fig. 1.

The advantage of FLIM over intensity-based methods lies in its robustness against factors like concentration, sample absorption, and excitation intensity, making it a powerful tool for functional imaging across different biological contexts. By using picosecond timescales for excitation and detection, FLIM achieves high temporal resolution, enabling the study of processes that occur on the scale of a few nanoseconds. Furthermore, fluorescence lifetime is sensitive to the local environment of the fluorophore, including factors such as pH, temperature, viscosity, and the presence of quenching molecules [2].

By mapping the fluorescence lifetimes across a sample, FLIM can reveal important information about the distribution and interactions of biomolecules within cells and tissues. For example, FLIM has been used to study protein-protein interactions via Förster resonance energy transfer (FRET), to monitor changes in cellular metabolism, and to image oxygen concentrations [3].

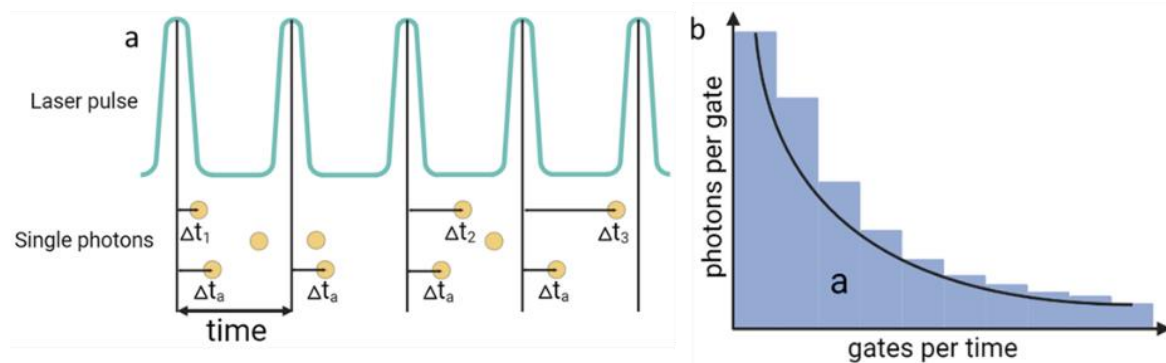


Fig. 1 The principle of obtaining FLIM microscopy data: (a) the principle of operation of TCSPC; (b) approximation of the fluorescence decay curve.

FLIM visualization has further advantages. FLIM provides quantitative data on molecular interactions and changes in the microenvironment by enabling the detection of minor changes in the lifetimes of fluorescence, and allowing simultaneous visualization of several fluorophores with different lifetimes. FLIM can even be used for label-free imaging based on endogenous fluorophores.

Multiphoton microscopy and FLIM are powerful techniques for investigating cellular metabolic dynamics in both *in vivo* and *ex vivo* settings. FLIM is particularly notable for its non-invasive nature, high sensitivity, and the ability to visualize cellular physico-chemical parameters—such as viscosity, temperature, and pH—without the need for exogenous stains [1, 4].

As an example, FLIM has been particularly useful for evaluating metabolic activity by analyzing the fluorescence lifetimes of endogenous fluorophores such as nicotinamide adenine dinucleotide (NAD(P)H). This molecule exists in two forms: NADH, which is the reduced form, and NAD⁺, its oxidized counterpart. These forms play crucial roles in metabolic pathways, including glycolysis, the tricarboxylic acid cycle, and oxidative phosphorylation. Specifically, NADH is linked to ATP production in the cytosol through glycolysis, and in mitochondria via oxidative phosphorylation. Additionally, its phosphorylated form, NADPH, is integral to biosynthetic pathways involving lipids, amino acids, and nucleotides, as well as to the glutathione cycle, which protects against reactive oxygen species [5].

Recent advancements in FLIM have enhanced the assessment of metabolic permutations, allowing for a more nuanced understanding of cellular metabolism. For instance, one study demonstrated that expanding the traditional two-lifetime model of NADH to a four-lifetime model evaluating the four components of fluorescence decay and, accordingly, a larger number of studied molecules significantly improved metabolic assessment. This approach allowed the researchers to quantitatively map changes in mitochondrial and cytosolic NADH concentrations and binding states, providing insights into the metabolic state of individual cells in response to pharmacological stimuli, such as during the browning of white adipose tissue [6].

Furthermore, FLIM can evaluate multiple metabolic parameters simultaneously, such as to provide the fluorescence-lifetime redox ratio (FLIRR), which reflects the relative concentrations of NADH and flavin adenine dinucleotide (FAD) in cells. This capability enables the detection of distinct metabolic states and variations within cellular compartments, thereby facilitating the investigation of metabolic dysfunctions in various biological contexts, including cancer and aging [7].

The integration of advanced computational tools for analyzing FLIM data also allows for the quantification of spatial heterogeneity in metabolic activity, revealing patterns that are critical for understanding complex biological processes [1, 8].

Additionally, FLIM has shown promise for clinical diagnostics, as changes in tissue fluorescence lifetimes can be indicative of pathological conditions such as cancer. A key concept that enhances the utility of FLIM in oncology is the Warburg effect. This phenomenon describes the tendency of cancer cells to preferentially rely on glycolysis for energy production, even when oxygen is available. This metabolic shift results in distinct changes in the cellular microenvironment, including variations in pH, ion concentrations, and the presence of specific metabolites. Consequently, these alterations affect the fluorescence characteristics of the tissue, including significant changes in fluorescence lifetimes [9]. FLIM has shown remarkable potential for early cancer detection, as it can differentiate between healthy and malignant tissues based on their fluorescence lifetimes. Research has also demonstrated its ability to identify cancerous lesions in various tissues, including oral and skin cancers, with high sensitivity and specificity [10]. The technique's ability to provide label-free, non-invasive imaging makes it an attractive option for *in vivo* applications [9].

However, FLIM visualization also has some limitations. The necessary data analysis requires specialized equipment and experience. FLIM may work more slowly than other image acquisition methods. Noise sensitivity can limit the accuracy of measurements over the equipment's entire service life.

While FLIM has traditionally been limited by the complexity of the instrumentation and the time-

consuming data analysis required, recent advances in detector technology and the development of deep-learning-based analysis methods have helped to address these challenges [11]. As a result, FLIM is becoming more accessible and practical for a wider range of biomedical applications.

Machine learning is a widely used approach for analyzing various types of biomedical images. When implementing the FLIM method, there are several techniques and neural network architectures available for image enhancement and analysis. This article will discuss methods of machine learning that can be used to improve and interpret the results of FLIM.

2 FLIM-Image Preprocessing

The raw data obtained from FLIM can be noisy and require significant preprocessing to enhance image quality and ensure accurate analysis. The primary objective of FLIM data preprocessing is to improve signal quality through denoising and other enhancement techniques.

2.1 Denoising FLIM Images

There are many techniques for the denoising of FLIM images using non-convolutional, convolutional filters, machine- and deep-learning methods.

Non-convolutional filters are techniques used in image processing that do not rely on the convolution operation to modify or enhance images. Instead, these filters often involve direct manipulation of pixel values based on specific algorithms or criteria. An example of non-convolutional filter is presented in Fig. 2.

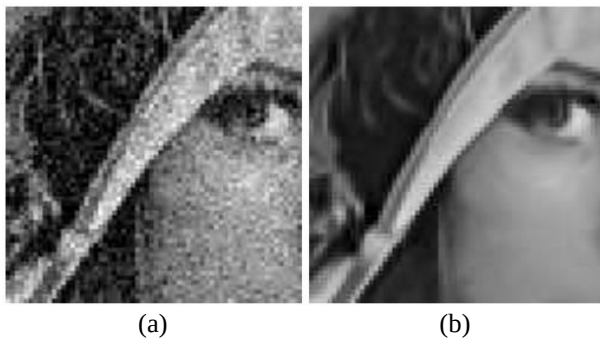


Fig. 2 NL-means denoising experiment with a natural image [12]: (a) noisy image with standard deviation 20, (b) restored image. Reprinted under the Creative Commons CC BY license from Ref. [12]).

Median filtering is a powerful technique for denoising FLIM images. The median filter operates by considering a local window around each pixel in the image, collecting the pixel values within that window, sorting them in ascending order, and replacing the center pixel value with the median value. This process is repeated for all pixels in the image. The median filter is particularly effective in removing small spatial disruptions to a signal while preserving the overall shape. It is also useful for

removing outliers and anomalous jumps in values, making it a suitable choice for denoising FLIM data [13].

Non-local Means (NL-Means) filtering is a further effective image denoising algorithm that computes the filtered value of a pixel as a weighted average of all pixels in the image, determined by the similarity of small image patch centered at pixel of interest, effectively reducing Gaussian noise while preserving edges and details by adding “method noise”. That noise looks like white noise. The use of the rotationally invariant patch distance and guided image filtering for efficiency, make the technique particularly useful for denoising scientific images such as microscopy data in various applications unlike the use of local filters that take the mean value of a group of surrounding pixels [12, 13].

Moving Average Filter is a filter that replaces each pixel's value with the average of its neighboring pixels. It is a simple way to smooth an image and reduce noise. Commonly used in time-series data smoothing and basic image noise reduction [14].

Bilateral Filter is a further filter smooths images while preserving edges by considering both the spatial distance and pixel intensity difference when averaging pixel values. It is often used in medical imaging to enhance structures without losing details, because it is effective at reducing noise while maintaining sharp edges [15].

Adaptive Filters is a specific filter that adjust their behavior based on the characteristics of the local image area. For example, adaptive median filters change their size based on the amount of noise detected. It is useful in scenarios where noise characteristics vary across an image due to its more effectiveness than static filters in non-uniform noise conditions [16, 17].

Convolutional filters, or kernels, are small matrices applied to an image to perform various operations such as blurring, sharpening, edge detection, denoising and more. The process involves sliding the filter over the image and computing a weighted sum of the pixels in the neighborhood defined by the filter. The result is a new image where each pixel value is derived from its surrounding pixels.

Convolutional filters can be categorized based on their function:

- Edge detection filters, such as Sobel and Prewitt filters, which highlight areas of rapid intensity change. The Sobel filter is one of the most widely used edge detection methods. It consists of two 3×3 kernels: one for detecting horizontal edges and another for vertical edges. The horizontal kernel emphasizes changes in intensity from left to right, while the vertical kernel focuses on changes from top to bottom. This method calculates the gradient magnitude and direction at each pixel, effectively highlighting edges. Similar to the Sobel filter, the Prewitt filter uses two 3×3 kernels for horizontal and vertical edge detection. It operates by convolving the input image with these kernels to approximate the gradient, helping to identify edges [18, 19].
- Blur filters, like Gaussian or Box filters that smooths images by averaging pixel values. The Box Blur

filter calculates the average of all pixels in a defined neighborhood (kernel) around each pixel. This filter treats all pixels equally and is simple to implement but can produce a loss of detail. Gaussian Blur uses a Gaussian function to assign weights to the pixels in the kernel, giving more importance to the central pixel and less to those further away [20].

- Sharpening filters, like the Laplacian, High Boost or Unsharp filters which enhances edges and fine details. The Laplacian filter is a second-order derivative operator that detects edges by calculating the second derivative of the image intensity. It highlights regions of rapid intensity change, making it effective for sharpening. High boost filtering enhances high-frequency components while preserving low-frequency information. Unsharp masking involves subtracting a blurred version of the image from the original image, which enhances edges. The process typically involves three steps: blurring the original image, subtracting this blurred image from the original, and then combining the results to enhance sharpness [21–23].

Convolutional filters are utilized in various applications, to improving visual quality through sharpening or blurring, to identifying edges and shapes for object detection and recognition, for separating different regions within an image for analysis.

Phasor analysis provides a graphical representation of FLIM data, where each pixel is represented as a point in the phasor plot, enabling the application of image processing techniques like median filtering to the phasor component images for effective denoising, while preserving spatial resolution. This helps in revealing hidden fluorescence components obscured by noise or background, and allowing for segmentation, based on clustering pixels with similar phasor coordinates, using unsupervised methods like *K*-means. It offers a “fit-free” approach that avoids complex curve-fitting procedures and provides quantitative information about the sample, with the linear properties of the phasor transformation enabling simple yet powerful denoising and analysis of the FLIM data [24].

Recent studies have demonstrated the application of Convolutional Neural Networks (CNNs) for denoising FLIM images. These networks are trained to enhance the signal-to-noise ratio (SNR) in FLIM data, which is crucial for accurate phasor measurements. By integrating CNNs with fast data acquisition systems, researchers have achieved significant improvements in image clarity and segmentation accuracy [25].

CNNs are utilized for denoising FLIM images after training on pairs of noisy and clean FLIM images to improve the signal-to-noise ratio and to handle the mixed Poisson-Gaussian noise common in FLIM data acquired using TCSPC. This CNN-based denoising approach, that includes transfer learning with pre-trained models like DnCNN, effectively enhances FLIM measurements, providing faster processing speeds and accurate, phasor-based segmentation for improved analysis, leveraging the benefits of transfer learning to avoid the need for

extensive FLIM-specific training datasets, while still efficiently handling noise in the FLIM data [26, 27].

A transfer learning scheme that involves training a denoiser with synthetic noise data and then transferring this knowledge to denoise real-noise images, enhances the generalization ability of the denoising method. The approach leverages both adaptive instance normalization and transfer learning to improve denoising performance, allowing a denoiser trained with synthetic noise to effectively denoise real-noise images. This transfer learning scheme demonstrates robust performance even with a small number of labeled data, showcasing the effectiveness of transfer learning in FLIM denoising tasks. Such an approach leverages the powerful feature extraction capabilities of CNNs trained on large datasets like the Fluorescence Microscopy Dataset (FMD) [28].

These denoising techniques, combined with machine learning and deep learning approaches have shown promising results in enhancing the quality of FLIM images and enabling more accurate analysis of fluorescence lifetime data, which can be performed using transfer learning from pre-trained models on large datasets like the Fluorescence Microscopy Dataset (FMD).

2.2 Spatial Resolution Improvement

The conversion of low-resolution (LR) FLIM images to high-resolution (HR) images involves various techniques, including mathematical algorithms and NN-powered methods.

One approach to such resolution improvement is “single-sample image-fusion up-sampling” (SiSIFUS) that combines data from a low-resolution time-resolved detector and a high-resolution camera to achieve super-resolution in FLIM. Key points in this respect are the ability of FLIM to provide detailed information on molecular interactions but its limited resolution at high speeds due to time-resolved imaging constraints. The method merges data from a low-resolution time-resolved detector (measuring photon arrival times) and a high-resolution camera (measuring intensity) to computationally achieve super-resolution. It incorporates statistically informed priors to address the inverse retrieval problem, avoiding the hallucination risks, incorrect or misleading results that models generate, resulting from traditional data-driven methods. The resulting enhanced images surpass those from standard bilinear interpolation, offering a versatile solution applicable to various image super-resolution challenges where there are two distinct datasets available [29]. An example of method implementation is presented in Fig. 3.

Another approach is sensor fusion implementation that combines data from a low spatial resolution but high temporal precision single-photon avalanche diode (SPAD) array with data from a CMOS camera that has high spatial resolution but no temporal resolution. This allows the reconstruction of time-resolved images at significantly higher spatial resolution than the SPAD input, upsampling by factors up to 12×12 in simulations and 4×4 experimentally. The method works by optically

blurring the image on the SPAD array to ensure temporal information from the entire field of view is captured, despite the low fill-factor of such SPAD arrays. Computational sensor fusion and optimization are then used to reconstruct a high resolution time-resolved image cube. The technique has been demonstrated for both LIDAR applications and FLIM studies of fluorescent cancer cells. It can fill in holes from dead SPAD pixels and remains robust against noise and ambient light. The main limitation is the long reconstruction time required post-measurement [30].

A further approach is a technique involving a two-channel attention network (TCAN) that leverages spatial representations and frequency contents to map low-resolution images to high-resolution ones effectively. This approach is designed to overcome limitations in current super-resolution fluorescence microscopy, such as the requirement for special fluorophores, complex optical systems, and lengthy acquisition and computational processes. The TCAN model is robust and can adapt to changes in pixel size and imaging setups, allowing it to generalize across different fluorescence microscopy modalities not included in the training data. The algorithm has been tested on various biological

structures and dual-color confocal images, successfully improving resolution from around 230 nm to approximately 110 nm. This technique enables live-cell super-resolution imaging, and has provided detailed insights into the structures and dynamic behaviors of micro-tubules [31].

A separate method has been described that involves training neural networks to generate massive semi-synthetic FLIM data sets representing various cellular morphologies, sizeable dynamic lifetime ranges, and complex decay components. A degrading model is then used to obtain LR-HR pairs, and a hybrid neural network, the improved spatial resolution FLIM net (SRI-FLIMnet), is developed to simultaneously estimate both the fluorescence lifetimes and to realize the nonlinear transformation from LR to HR images. The evaluative results demonstrate SRI-FLIMnet's superior performance in reconstructing spatial information from limited pixel resolution. This approach has been verified using experimental images of bacterial-infected raw macrophage cells from mice, showing that the proposed data generation method and SRI-FLIMnet efficiently achieve superior spatial resolution for FLIM applications [32].

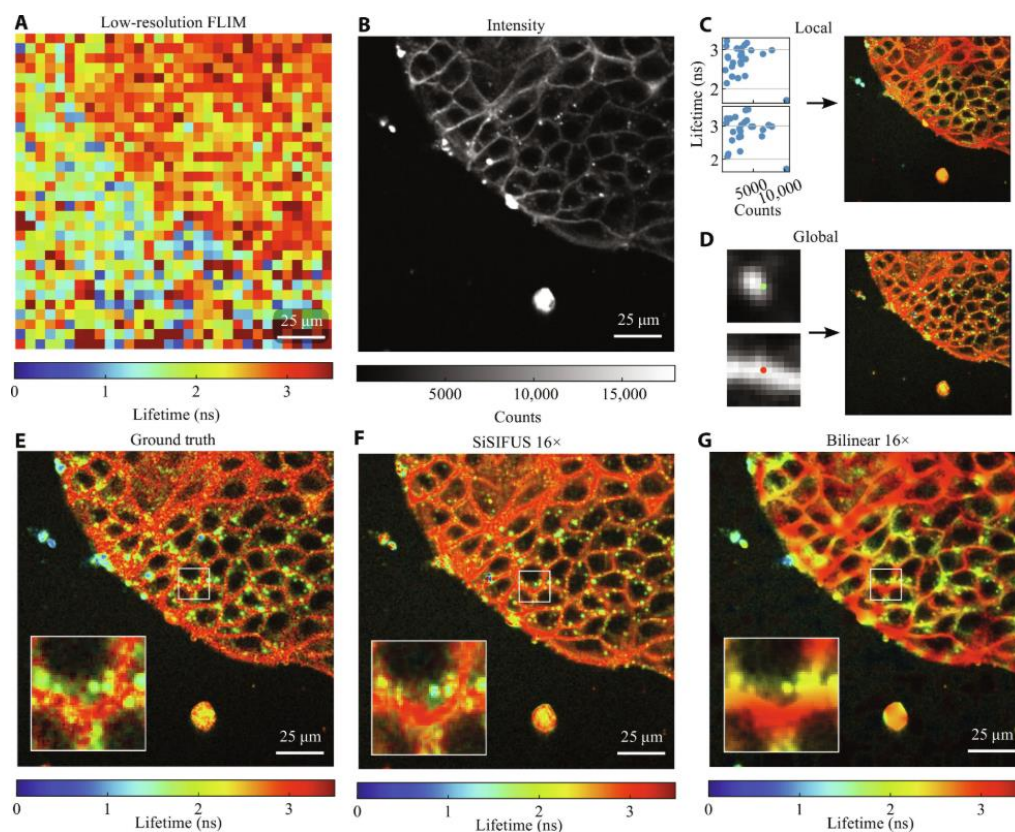


Fig. 3 Upsampling (16×16) of Madin-Darby canine kidney cells. (a) Low-resolution fluorescence lifetime image (32×32) of MDCK cells expressing Flipper-TR dye. (b) Corresponding high-resolution intensity image (512×512) of the sample. (c) Windows (5×5) of low-resolution FLIM are fitted to corresponding intensity values to generate an LP image (two example windows are shown). (d) A GP image is generated from 13×13 intensity patches with central FLIM measurements (two examples are shown). (e) The GT high-resolution FLIM target, intensity-weighted for visualization. (f) The proposed method, upsampling the low-resolution measurement by a factor of 16×16 . (g) Bilinear interpolation upsampling the FLIM measurement by 16×16 [29]. Reprinted under the Creative Commons CC BY license from Ref. [29].

3 FLIM Analysis Using Machine Learning Techniques

FLIM image analysis includes obtaining fluorescence decay parameters for selected objects in the image in order to evaluate changes in the lifetimes, and accordingly the distribution of the fluorescent molecule being studied. The decay parameters are calculated for each pixel in the image, this being followed by the selection of a region of interest (ROI) for calculation of its own decay parameters as an average of the corresponding pixels.

The extraction of true lifetime values from the raw data takes place using traditional methods such as Maximum Likelihood Estimation and Least Squares fitting of fluorescence decay curves.

Maximum Likelihood Estimation (MLE) is a statistical method that estimates the parameters of a probability distribution by maximizing the likelihood function. For FLIM data, MLE is used to estimate the parameters of the fluorescence decay curve, such as the lifetime and amplitude. MLE is particularly useful when dealing with low photon counts, as it provides unbiased estimates of the decay parameters [33].

Least Squares (LS) fitting is a method that minimizes the sum of the squared differences between the observed data and the fitted model. In FLIM data analysis, LS fitting is therefore used to estimate the parameters of the fluorescence decay curve by minimizing the difference between the observed decay and the fitted exponential function. LS fitting is computationally efficient and widely used in FLIM data analysis software [34].

Both MLE and LS fitting are iterative methods that start with an initial estimate of the decay parameters and refine them until the best fit is obtained. The resulting fitted parameters are then used to generate the true lifetime images, where each pixel represents the estimated lifetime of the fluorophore.

These conventional curve-fitting procedures, however, can be computationally intensive and time-consuming, especially when dealing with large datasets. To address this limitation, machine learning and deep learning techniques have been explored for FLIM data analysis. When using machine learning techniques for FLIM analysis, two steps are required: image segmentation to identify objects whose parameters need to be evaluated, and determination of the fluorescence decay parameters from data collected during the FLIM measurements. In general, this involves analyzing the fluorescence decay curves to extract information about the properties of the objects. The order of these steps is not strictly defined, but it depends on the specific analysis task and the data available. The conventional procedure for FLIM analysis is presented in Fig. 4.

In our pilot study, we first identified the objects of interest in the FLIM images of liver tissue, which were hepatocytes. We then extracted their parameters. To achieve this, we trained a selected convolutional neural network, modified UNet++ [35], and achieved a high level of cell detection accuracy: the F_1 metric (Dice coefficient) was ~ 0.7 . The results of the image segmentation are presented in Fig. 5.

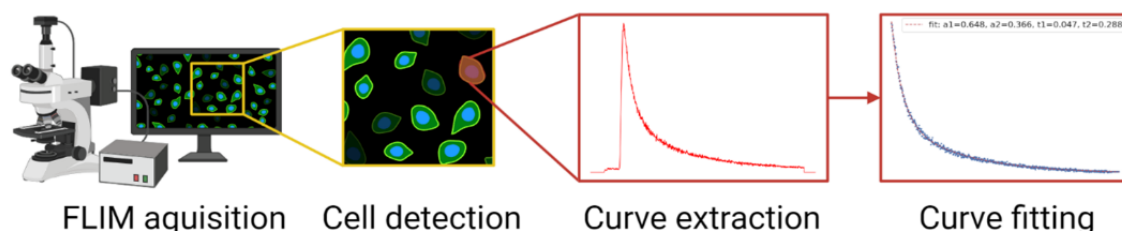


Fig. 4 The procedure for performing FLIM microscopy analysis using machine learning methods.

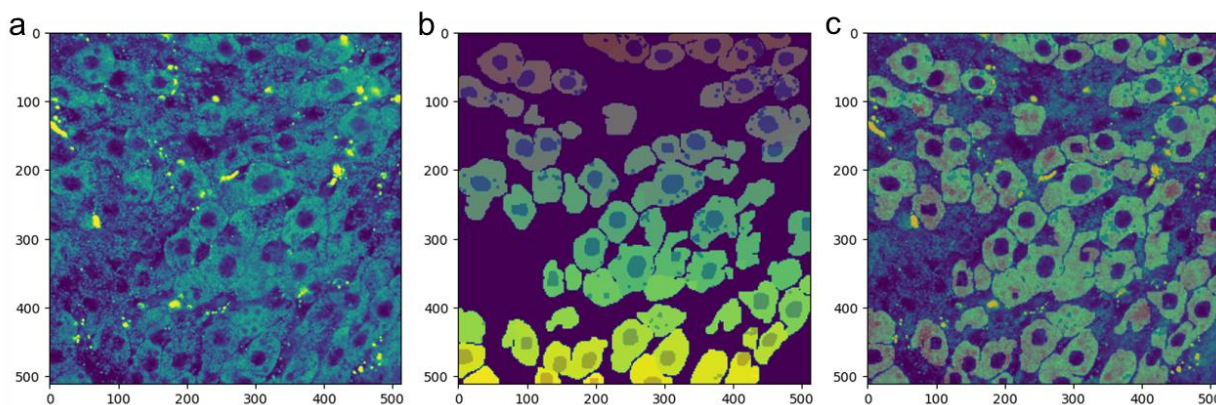


Fig. 5 The results of segmentation of the FLIM image of liver tissue: (a) an intensity image; (b) the resulting cell-by-cell mask, which includes only the cytoplasm of the studied cells; (c) the imposition of a mask on the original image.

3.1 Methods of Biomedical Image Segmentation

Biomedical image is a task that is in demand for the analysis of individual objects in various biomedical images, including the results of laser scanning microscopy, FLIM, magnetic resonance imaging and many other applications. The task of biomedical segmentation has been solved in various ways throughout the development of methods. Early methods assumed direct work with pixels, based on their intensity, color, and mutual arrangement in the image. With the development of deep learning methods, the segmentation task has moved to convolutional neural networks trained for specific segmentation tasks. Now, the most common methods are foundational networks, which are trained on a wide selection of studied images and are suitable for use in various situations. Thus, image segmentation methods are divided into three broad groups: traditional statistical methods, single-task deep learning neural networks and foundational networks.

Traditional statistical machine learning methods for cell image segmentation still are essential tools in biomedical research, allowing for the effective partitioning of cell images into distinct regions for analysis. Thresholding methods, such as Otsu's thresholding, segment images by converting them into binary images based on intensity values. The goal is to find an optimal threshold that separates foreground (cells) from the background. It is commonly used in microscopy to identify and segment cells based on their intensity differences from the background [36]. K-Means Clustering partitions the image into k clusters based on pixel intensity or color similarity, effectively grouping similar pixels together. Which is useful for segmenting images where cells exhibit distinct color or intensity profiles [37]. Random Forest is an ensemble learning method that builds multiple decision trees to classify pixels based on features extracted from the image. It is applied in semantic segmentation tasks where each pixel is classified as belonging to a specific cell type or background [38]. Markov Random Fields models spatial dependencies between pixels, allowing for segmentation that considers the relationships between neighboring pixels. It is useful in medical imaging where tissue characteristics vary spatially, helping to manage noise and improve segmentation accuracy [39]. Edge Detection methods, which identify boundaries between different regions by detecting discontinuities in pixel intensity. Common algorithms include Sobel and Canny edge detectors. It is used to delineate cell boundaries in microscopy images, particularly when cells are well-defined against their background [40].

Classical single-task trained deep learning models for cell segmentation have emerged as powerful tools in the analysis of biological images, particularly in identifying and delineating cellular structures. These models leverage deep learning architectures specifically designed for the task of segmenting cells from microscopy images. U-Net is a convolutional neural

network architecture specifically designed for biomedical image segmentation. It features a contracting path to capture context and a symmetric expanding path to enable precise localization. U-Net is widely used for segmenting nuclei and other cellular structures in fluorescence microscopy images [41]. DeepCell is a CNN model that focuses on cell segmentation using a patch-based approach. It segments images by classifying each pixel based on learned features from surrounding pixels. It is designed for segmenting diverse cell types in various imaging contexts, including high-throughput microscopy [42]. CellProfiler is not actually an ML model; it uses classical image processing algorithms and can serve as a baseline for comparison with deep learning methods. It employs various modules, including thresholding and shape-based filtering, to identify and segment cells. CellProfiler is frequently used in biological labs for high-throughput image analysis to segment nuclei and other cellular components [43]. SPPNet is designed for nuclei image segmentation and operates with significantly fewer parameters than traditional models while maintaining high performance. It is effective for real-time applications where speed is critical, such as in live-cell imaging scenarios [44]. GeneSegNet integrates gene expression data with imaging information to enhance cell segmentation accuracy. This model employs a recursive training strategy to manage noisy training labels effectively. It is particularly useful in studies involving in situ RNA detection technologies where accurate cell boundaries are crucial for analyzing gene expression data [45].

Foundational models for biomedical image segmentation have evolved significantly, particularly with the advent of deep learning techniques. These models are large-scale pre-trained deep learning models designed to generalize across a variety of tasks with high accuracy, which optimizes the segmentation process and reduces the need for task-specific models.

Many variants of deep learning networks have been developed and modified. One such example is reported in study [46] that describes a new diffusion probabilistic model (DPM) for the medical image segmentation task. Two novel components are introduced in this MedSegDiff approach: a dynamic condition encoder to enhance regional attention in the DPM, and a feature frequency parser to eliminate the negative effects of high-frequency noise components. Experiments on three different datasets demonstrated that the proposed method outperforms existing methods, achieving state-of-the-art results. The authors argue that this is the first time DPM has been used for medical image segmentation.

Research paper [47] introduces a novel neural network architecture designed to overcome the limitations of CNNs and transformers in handling long-range dependencies in medical image segmentation. U-Mamba integrates the local feature extraction capabilities of convolutional layers with the long-term dependency modeling abilities of state space sequence models (SSMs), creating a hybrid CNN-SSM architecture. U-Mamba also has a self-adaptive

mechanism that allows it to automatically adjust to different datasets without the need for manual intervention. This makes it a powerful tool for biomedical image segmentation tasks. There is, however, a paper [48] that introduces a novel approach to medical image segmentation by proposing a U-shape architecture model named Vision Mamba UNet (VM-UNet). This model leverages SSMs to address the limitations of CNN- and transformer-based models. A Visual State Space (VSS) block is introduced to capture extensive contextual information, and an asymmetrical encoder-decoder structure is utilized. VM-UNet is the first medical image segmentation model based solely on SSMs, aiming to set a baseline for more efficient and effective SSM-based segmentation systems in the future.

Another important model is SegMamba [49], which is a novel 3D medical image segmentation model. It effectively captures long-range dependencies within entire volume features at each scale. Compared with transformer-based methods, in whole volume feature modeling SegMamba outperforms these from a state-space model perspective, maintaining superior processing speed even with high-resolution volume features.

In research article [50], the authors propose LM-Net, a lightweight architecture that combines CNNs and Vision

Transformers (ViTs) to enhance segmentation accuracy. LM-Net utilizes a multi-branch module to capture multi-scale features and introduces the Local Feature Transformer (LFT) and Global Feature Transformer (GFT) to simultaneously capture local detail textures and global contextual semantics. By integrating these modules, the model achieves state-of-the-art results on various medical image segmentation tasks across different datasets, demonstrating its effectiveness and adaptability.

Paper [51] proposes a semi-supervised contrastive learning framework called ARCO for medical image segmentation. The authors argue that current contrastive learning methods may struggle with distinguishing minority tail-class samples and that this may lead to model collapse. They introduce variance-reduction techniques, which they prove, theoretically, to be universally beneficial for pixel/voxel-level segmentation tasks with limited labels. The authors experimentally validated their approach on eight benchmarks, demonstrating the consistent outperformance of such state-of-the-art semi-supervised methods.

A comparison of the different models is shown in Tables 1 and 2.

Table 1 Comparison table of biomedical image segmentation methods.

Model	Advantages	Limitations
Thresholding [36]	Simple and fast	Sensitive to noise; limited effectiveness with overlapping cells
K-Means [37]	Effective for distinct clusters	Requires predefined <i>K</i> ; struggles with complex scenes
Random Forest [38]	Robust against overfitting	May lose fine details
Markov Random Fields [39]	Captures spatial relationships	Computationally intensive
Edge Detection [40]	Provides clear boundaries	Affected by noise; may miss edges in smooth areas
U-Net [41]	High accuracy; effective with small datasets	Sensitive to overfitting; requires tuning
DeepCell [42]	Handles complex images well	Computationally intensive
CellProfiler [43]	User-friendly; highly configurable	Lacks accuracy compared to deep learning methods
SPPNet [44]	Fast inference; fewer parameters	May miss contextual details
GeneSegNet [45]	Integrates diverse data; robust against noise	Complexity in data integration
MedSegDiff [46]	State-of-the-art performance; noise reduction	Computationally intensive; sensitive tuning
U-Mamba [47]	Linear complexity; flexible	Requires extensive training data
VM-UNet [48]	Competitive performance; establishes a baseline	Basic architecture without enhancements
SegMamba [49]	Handles complex images; robust performance	High computational demands
LM-Net [50]	Fast inference; lightweight	May sacrifice accuracy in complex scenarios
ARCO [51]	Adaptive region-based segmentation	Implementation complexity; sensitivity issues

Table 2 Comparison table of the application of contemporary biomedical image segmentation methods.

Method	Application
MedSegDiff [46]	General medical image segmentation tasks, particularly where high accuracy is required; effective in segmenting complex anatomical structures across various imaging modalities
U-Mamba [47]	Long-range dependency modeling in segmentation tasks; suitable for tasks requiring detailed contextual understanding, such as organ or tumor segmentation in complex images
VM-Unet [48]	Organ and skin lesion segmentation tasks; effective in scenarios where computational efficiency is crucial without significantly sacrificing accuracy
SegMamba [49]	Complex images with overlapping structures; suitable for challenging segmentation tasks where traditional methods may struggle, such as in dense tissue imaging
LM-Net [50]	Real-time medical image segmentation tasks; suitable for applications requiring rapid processing, such as intraoperative imaging or point-of-care diagnostics
ARCO [51]	Adaptive region-based segmentation tasks; effective in segmenting heterogeneous regions, such as differentiating between various tissue types in medical scans

Table 3 Comparison table of various accuracy metrics for image segmentation methods.

Model	Mean IoU	Pixel Acc	Dice
MedSegDiff [46]	0.807		0.882
U-Mamba [47]			0.6964
VM-Unet [48]	0.8079	0.956	0.8937
SegMamba [49]	0.8859		0.6967
LM-Net [50]	0.8912		0.9409
ARCO [51]			0.852

3.2 Image Segmentation Quality Metrics

In the field of computer vision, accurate segmentation metrics are essential tools for evaluating segmentation models. These metrics provide quantitative measurements that assess how well the model separates objects in an image. They allow developers to evaluate the effectiveness of the model, identify weaknesses, and make improvements. There are several metrics available, such as Pixel Accuracy, Mean Intersection over Union (mIoU), and Dice Coefficient. Each metric provides unique insights into the performance of the model for different applications and challenges [52]:

- mAP (mean Average Precision) and AR (Average Recall) metrics to assess base accuracy.
- Dice, mIoU (Jaccard) to evaluate the mask accuracy.
- Precision and Recall to assess classification correctness.
- Confusion Matrix, per-class and per-image metrics to get more insights regarding the instance segmentation model.

A comparison of the different models based on the indicators presented is shown in Table 3.

Challenges remain, however, particularly in handling the complexity of clinical tissue samples and the large

size of whole-slide images. The difficulties of segmenting abnormal tissues in medical images, such as the heterogeneity of tumor morphology and the need to process gigapixel-scale images, remains [53, 54].

Overall, the field of tissue image segmentation has seen significant progress, driven by advances in deep learning, the creation of large annotated datasets, and innovative algorithmic approaches. However, continued research is needed to address the unique challenges posed by clinical tissue samples and to enable more accurate and efficient analysis of these critical medical images.

3.3 Curve Analysis

The FLIM decay parameters, as mentioned above, are calculated on the basis of the approximation of the decay curve, for each pixel or selected field, with further parameter estimation depending on the resulting function. Thus, the basic approach to conducting FLIM analysis is curve fitting.

Curve fitting is a fundamental tool in data analysis and modelling, used across various disciplines to describe and predict relationships between variables. Curve fitting involves finding a mathematical function that best fits a set of data points. The goal is to capture the underlying trend or pattern in the data, allowing for

interpolation, extrapolation, and prediction. Commonly used functions for curve fitting include linear, polynomial, exponential, logarithmic, and trigonometric functions.

3.4 Techniques for Curve Fitting

There are several standard curve fitting methods that include: linear regression, nonlinear regression, polynomial functions, and splines. These techniques have been used to analyze curves with great efficiency for a long time and do not require further explanation [55–58].

Machine learning offers powerful techniques for curve fitting, including neural network and deep learning approaches. These methods provide flexibility in capturing complex, nonlinear relationships compared to traditional polynomial regressions.

Indeed, neural networks excel at function approximation, making them well-suited for curve fitting problems. By using a sufficient number of hidden layers and neurons, neural networks can learn to fit arbitrary curves to data. The key is to train the network on the dataset, allowing it to learn the underlying function. One approach is to use an extreme learning machine (ELM) algorithm, which introduces an additional linear neuron to correct the output of the hidden layer neurons. The number of hidden neurons is optimized using particle swarm optimization (PSO) to achieve the best fit. This method has shown excellent performance on various benchmark datasets and real-world applications [59].

Deep learning architectures, such as deep neural networks with multiple hidden layers, offer even more flexibility in curve fitting. By stacking multiple nonlinear transformations, deep networks can capture highly complex relationships in the data. This is particularly useful when fitting curves to high-dimensional or multivariate datasets. For example, convolutional neural networks can be applied to curve fitting problems by treating the input data as an image-like structure. The convolutional layers extract relevant features, while the fully connected layers perform the actual curve fitting. This approach has been demonstrated on multidimensional curve fitting tasks [60].

Machine learning and deep learning methods provide a number of advantages for curve fitting compared to traditional methods. Neural networks and deep learning models can match a wide range of curve shapes, including highly nonlinear and complex functions. They can automatically extract relevant functions from data, without the need for manual function design, are suitable for fitting curves to large-dimensional datasets such as those used in scientific and engineering applications, and can be trained to be resistant to noise and outliers in the data, providing reliable curve fitting results. With powerful computing resources and large datasets, machine learning methods can scale to fit curves to massive datasets.

Despite the promising results of machine learning and deep learning in the field of curve fitting, there are still challenges and areas that require improvement. One such

main issue is interpreting the fitted curves produced by the deep learning models. This is particularly challenging for high-dimensional problems, where it can be difficult to understand the underlying patterns. Ensuring that the fitted curves generalize well to new data is crucial for the success of these models. Techniques such as regularization and cross-validation can help to improve such generalizations, but more research is needed in this area. Another important aspect of curve fitting is obtaining reliable estimates of uncertainty. Combining traditional curve fitting methods with machine learning techniques, such as spline interpolation or Bayesian approaches, can lead to more accurate and reliable results.

3.5 Contemporary Premade Curve Fitting Solutions

There are many premade solutions for analyzing curves, in particular, fluorescence attenuation curves, thanks to which you can quickly select the parameters of the decay curve function and obtain calculated decay parameters.

Research paper [61] introduces CurvPy, an open-source Python library designed for automated curve fitting and regression analysis. CurvPy aims to make advanced statistical and machine learning techniques more accessible by providing functionalities for optimization, smoothing, imputation, summarization, visualization, regression, evaluation, and tuning. The methodology of CurvPy involves the use of mathematical techniques such as least squares estimation, Savitzky-Golay filtering, matrix completion, gradient descent optimization, regularization, basis function regression, and standard model evaluation metrics. This highlights the contributions of CurvPy in simplifying model fitting and analysis for both non-experts and power users, emphasizing usability, flexibility, and transparency in leveraging advanced statistical methods for data analysis.

Research paper [62] describes TableCurve 2D a powerful software tool for automated curve fitting and data modeling, widely used by engineers and scientists. It offers a range of features that make it a valuable tool for analyzing data. One of the key features of TableCurve 2D is its ability to automatically fit curves to data using a library of 3,665 equations. This saves time and effort compared to manual fitting, as users can quickly find the best fit with a single click. Additionally, users can define up to 15 custom equations for more complex modeling needs. TableCurve also provides advanced estimation procedures for nonparametric models, such as Spline, Smoothing Spline, and Local Regression. These methods are useful for data that does not fit well within traditional parametric models.

Wolfram Language, introduced in Ref. [63], offers powerful tools for curve fitting, supporting both linear and nonlinear approaches. Users can quickly approximate data with linear least-squares regression using specified basis functions, while nonlinear models allow for more flexibility with constraints and initial parameter estimates. Interpolation methods are also

available to create smooth curves that pass-through data points, without assuming a specific functional form. Regularization techniques prevent overfitting by penalizing large coefficient values, and diagnostic tools help assess fit quality with statistical measures. These capabilities are useful in various fields, such as physics for modeling experimental data, biology for growth or decay curve fitting, and engineering for stress-strain analysis.

A further paper [64] introduces *ndCurveMaster*, an example of curve fitting software, providing tools for automated curve fitting, managing unlimited variables, and heuristic techniques for optimal fitting, together with advanced statistical analysis tools such as Analysis of Variance (ANOVA), regression analysis, and overfitting detection. The software's integration of machine learning techniques for equation discovery significantly enhances the efficiency of data analysis, and its extensive applicability across disciplines such as medicine, physics, ecology, and finance.

In addition to the listed options for premade methods for analyzing decay curves there another options including open-source and programming languages-based methods, which is important to mention.

ImageJ [65], a popular open-source image processing software, offers several tools for decay curve fitting, particularly useful in fluorescence lifetime imaging and other applications where analyzing decay kinetics is essential. *FILMJ* is an advanced *ImageJ* toolkit specifically designed for fluorescence lifetime data analysis. It integrates seamlessly with the *ImageJ* Ops framework and includes various fitting routines that can be extended for complex workflows. *FILMJ* supports standard nonlinear least-squares fitting using the Levenberg-Marquardt algorithm, as well as more sophisticated methods like maximum likelihood and Bayesian analysis.

Gwyddion [66] is a powerful open-source software for analyzing and processing data from scanning probe microscopy (SPM) and other imaging techniques. It includes various tools for curve fitting, particularly useful for decay curve fitting in the context of surface roughness and force-distance data analysis. *Gwyddion* employs the Levenberg-Marquardt algorithm for nonlinear curve fitting, which is suitable for statistical functions used in evaluating surface roughness parameters. The software provides a range of fitting functions, including Gaussian and exponential models, which can be applied to various datasets, including those derived from AFM measurements.

Python, alongside other programming languages-based methods, like MATLAB or R-Studio, is a powerful tool for decay curve fitting, particularly using special libraries such as SciPy and NumPy for Python. It allows users to define custom functions for fitting, enabling them to model a wide range of decay behaviors tailored to specific datasets. Libraries like SciPy provide built-in functions, which simplifies the process of fitting data to models using nonlinear least squares optimization.

4 Complex Analysis of Obtained FLIM Data Using Machine Learning

Machine learning techniques are well-suited to extracting and interpreting measurements from multi-dimensional FLIM data sets, offering significant improvements in the efficiency of data analysis. FLIM data extraction using curve-fitting techniques can be complex and computationally inefficient for real-time measurements, especially in applications like FRET that require analysis in multiple dimensions. Machine learning approaches can accelerate imaging processes, aid in data interpretation, and extract multi-dimensional data into a single exponent lifetime, addressing the challenges faced by traditional curve-fitting methods. In the context of metabolic imaging, machine learning models can be trained to map FLIM imaging data to provide target outputs, such as metabolic states or the presence of biomarkers [67].

The diverse approaches to deep learning, which are tailored to process raw FLIM data, encompass physics-based models. Physics-based deep learning (PBDL) models integrate fundamental physical principles into machine learning frameworks to enhance performance in various applications, including fluorescence lifetime imaging.

4.1 State-of-the-Art Methods of FLIM Data Analysis

Research paper [11] introduces a deep learning approach called FLI-Net which is proposed as a fit-free method for FLIM image formation that can quantify fluorescence decays simultaneously over a whole image at high speed. FLI-Net has a deep neural network architecture designed and trained for different classes of experiments, including visible FLI, near-infrared (NIR) FLIM, and NIR gated Macroscopy Fluorescent Lifetime Imaging (MFLI). FLI-Net outputs the spatially resolved lifetime-based parameters typically employed in the field, without requiring any parameter settings. The paper [68] presents AUTO-FLI, a groundbreaking deep learning framework for three-dimensional fluorescence lifetime tomography that significantly enhances accuracy and computational efficiency. By employing a two-stage convolutional neural network architecture, consisting of modified Automap and modified FLI-Net, AUTO-FLI achieves highly accurate reconstructions of three-dimensional intensity and lifetime distributions from raw two-dimensional fluorescence decays, outperforming conventional methods. The use of Monte-Carlo Xtreme for data generation reduces the time and resources required for training, making the model more robust and versatile across a wide range of biological and medical imaging applications.

Another research paper [69] introduces a deep learning framework called Few-Photon Fluorescence Lifetime Imaging (FPFLI), designed for analyzing FLIM data under low-light conditions with minimal photons per pixel. Traditional FLIM analysis methods struggle with low photon counts, requiring hundreds to thousands of

photons for accurate results. FPFLI breaks away from pixel-wise analysis, leveraging spatial correlation and intensity information to estimate lifetime images effectively with only a few photons. The framework involves two main steps: estimating local lifetime distributions using a Local Lifetime Estimator (LLE) and fusing this information with intensity images to reconstruct target lifetime images using Neural Implicit Image Interpolation (NIII). The study includes the development of a large-scale synthetic FLIM dataset for training the FPFLI algorithm, demonstrating rapid and robust lifetime estimation even with extremely low photon counts. The paper outlines the algorithm's implementation, training process, and evaluation using semi-synthetic FLIM datasets to overcome challenges in acquiring diverse FLIM images for training the deep learning algorithms.

Paper [70] introduces a compact deep learning architecture called Fluorescence Life-time AdderNet (FLAN) for high-fidelity FLIM imaging with low computing overheads. FLAN is designed for efficient implementation on field-programmable gate arrays (FPGA), aiming to improve the performance of FLIM while reducing computational complexity. The architecture uses an adder-based deep learning network that avoids multiplication-based convolutions, simplifying calculations and making it more suitable for real-time applications with limited computational resources. The network employs a log-scale (LS) merging technique to compress fluorescence decay data, achieving significant compression ratios of 0.11 and 0.23 compared to traditional convolutional neural networks, while maintaining high accuracy in lifetime retrieval. This architecture has been successfully implemented on FPGA, using post-quantization techniques to optimize computational efficiency and minimize resource usage. The implementation allows for high performance in real-time biomedical research applications. Evaluations on synthetic and experimental FLIM data show that FLAN and FLAN+LS achieve state-of-the-art accuracy and precision, outperforming traditional curve fitting methods, especially in low photon count conditions. Finally, the paper demonstrates real-time FLIM imaging by implementing FLAN+LS on an FPGA, achieving high throughput by processing multiple pixels in parallel.

Another paper [71] discusses a hardware-friendly deep learning architecture utilizing one-dimensional convolutional neural networks (1D CNN) for analyzing FLIM data. This architecture highlights the advantages of 1D CNNs over high-dimensional CNNs, emphasizing their simplicity, faster training, and quicker processing speeds. Notably, 1D CNNs excel in handling multi-exponential fluorescence decay models and have shown superior performance in fluorescence lifetime image reconstruction compared to traditional LS methods, particularly validated when using simulated data. The authors also mention the successful application of the proposed 1D CNN in analyzing two-photon FLIM images of functionalized gold nanoprobe in Hek293 and human prostate cancer cells, showcasing the speed and accuracy

of 1D CNNs in extracting lifetime parameters from fluorescence signals. Overall, the study underscores the significant potential of 1D CNNs in real-time FLIM applications.

Paper [72] introduces a deep learning architecture model, the MLP-Mixer, that can perform fast and accurate fluorescence lifetime imaging. Compared to other methods like 1D-CNN, NLSM, and VPM, the MLP-Mixer model achieves the best performance in terms of low bias and small standard deviation, together with high correlation with the reference lifetimes for both Rhodamine B and Rhodamine 6G samples across different signal-to-noise ratios. The MLP-Mixer model has a simple architecture and fewer parameters compared to the 1D-CNN model, yet it outperforms the 1D-CNN in both training time and inference time. This makes the MLP-Mixer a practical and efficient solution for real-time fluorescence lifetime imaging applications. The paper demonstrates the effectiveness of the proposed deep learning approach through both simulation and experimental results using Rhodamine 6G, Rhodamine B, and *Convallaria majalis* cell samples. The results show that such deep learning models can more accurately estimate fluorescence lifetimes compared to traditional methods like nonlinear LS fitting.

Paper [73] introduces FlimGANE, a fluorescence lifetime imaging method based on Generative Adversarial Network Estimation. It consists of a generator, a discriminator, and an estimator, with the generator using convolutional and residual blocks to generate high-quality decays from low-photon-count inputs. The discriminator is composed of fully connected layers, and the estimator involves training with mean-squared error cost functions. FlimGANE is trained in three stages: generative model training, estimative model training, and combination training. It addresses the need for a computationally efficient and high-quality method for fluorescence lifetime estimation, performing well even in ultra-low-photon-count conditions. It outperforms traditional methods like TD-MLE in barcode identification, FRET characterization, and metabolic state analysis, offering a reliable alternative for FLIM analyses, especially in clinical applications like retinal imaging and tumor margin identification.

The article [74] discusses advancements in time-domain diffuse optical tomography (TD-DOT), particularly focusing on a new reconstruction algorithm called SENSOR-NET. This technology aims to enhance the clinical application of TD-DOT for monitoring brain tissue hemodynamics by providing high spatial and temporal resolution. TD-DOT offers superior accuracy in measuring optical properties and physiological parameters compared to frequency-domain or steady-state methods.

However, it faces challenges in achieving high temporal resolution due to the computational demands of solving the inverse problem, which can lead to long reconstruction times. The authors introduce the SENSOR algorithm, which is a non-iterative approach that significantly reduces computation time while maintaining high spatial resolution (1 mm) and temporal

resolution (20–30 ms). This algorithm transforms the complex inverse problem into a more manageable format, allowing for faster processing without iterative solving. The proposed SENSOR-NET integrates deep learning with the SENSOR algorithm. By structuring the iterations of SENSOR within a neural network, SENSOR-NET eliminates the need for empirical parameter tuning prior to reconstruction. Once trained, this model can achieve accurate reconstructions in under 20 ms, independent of the number of sources or wavelengths used. The method has been validated using both numerical simulations and experimental data, demonstrating its potential for real-time brain monitoring and other applications requiring rapid imaging capabilities.

Another paper [75] introduces a Bayesian Nonparametric FLIM (BNP-FLIM) is an innovative framework designed to enhance fluorescence lifetime imaging microscopy by addressing key challenges in quantitative analysis. Traditional FLIM methods often ignore spatial correlations between pixels, rely on photon ratios instead of absolute lifetime maps, and struggle to determine the number of contributing

fluorophore species. BNP-FLIM overcomes these issues by employing a doubly nonparametric approach that uses Beta-Bernoulli process priors to [infer the number of species and Gaussian process priors for continuous lifetime mapping.

This framework leverages information from “empty pulses” – pulses that do not result in observable photons – to improve lifetime estimates. Benchmarking against synthetic and experimental data demonstrates its robustness, allowing for the recovery of lifetime differences as small as 0.3 ns with only 1000 photons. BNP-FLIM significantly enhances spatial resolution and enables rigorous uncertainty propagation.

Thus, the use of machine learning in complex FLIM imaging analysis aligns with its broader applications in medical imaging, where it plays a crucial role in tasks like disease diagnosis, characterization, and treatment response prediction [76].

A comparison of the methods of FLIM data complex analysis is shown in Table 4 and 5.

Methods available on GitHub are listed in Table 6.

Table 4 Comparison table of the application of various methods of FLIM data complex analysis.

Method	Application
FLI-Net [11]	Rapid analysis of fluorescence lifetime imaging data; effective for tasks requiring fast processing, such as monitoring dynamic biological processes or detecting changes in fluorescence lifetimes in live cells
FPFLI [69]	Advanced fluorescence lifetime imaging with enhanced sensitivity; suitable for high-resolution imaging tasks where detailed analysis of molecular interaction is required
FLAN [70]	Analyzing complex datasets with multiple parameters; suitable for studies involving microenvironmental changes or interactions among multiple fluorescent species
1D CNN [71]	Fast processing of sequential fluorescence data; effective for real-time analysis of time-series data in flim applications, particularly in simple scenarios where spatial features are less critical
MLP-Mixer [72]	General-purpose application requiring efficient processing of sequence signals; suitable for various flim applications, especially where speed and accuracy are essential, such as in embedded systems or real-time imaging
FlimGANE [73]	Enhancing image quality and resolution in fluorescence lifetime imaging; useful for improving the quality of fluorescence lifetime maps, especially under low-photon conditions or when dealing with challenging imaging scenarios
SENSOR-Net [74]	Simultaneous learning of fluorophore species and their absolute lifetime maps while leveraging spatial correlations; effective for characterizing complex biological environments by providing high-resolution lifetime maps while accounting for uncertainties inherent in experimental data
BNP-FLIM [75]	Overcoming challenges in quantitative flim analysis by learning the number of species and their absolute lifetime maps simultaneously; suitable for scenarios where distinguishing between closely spaced lifetimes is critical, such as drug discovery and monitoring biological interactions

Table 5 Comparison table of various methods of FLIM data complex analysis.

Method	Advantages	Limitations
FLI-Net [11]	High accuracy in quantifying fluorescence lifetimes; approximately 30 times faster than traditional methods like spcimage; well-suited for real-time analysis in complex biological samples	The outcome may vary depending on the unique properties of the sample being examined; to attain the best results, it may be necessary to have a substantial amount of training data
FPFLI [69]	Enhanced fluorescence lifetime imaging capabilities; provides improved sensitivity and resolution compared to traditional methods	May be complex to implement and require specialized hardware; performance can vary depending on the imaging setup and sample conditions
FLAN [70]	Designed for efficient analysis of fluorescence lifetime data; capable of handling complex datasets with multiple parameters	Implementation may be challenging without adequate computational resources; require careful calibration to ensure accurate results
1D CNN [71]	Efficient for processing sequential data, suitable for time-series analysis in fluorescence imaging; simple architecture that can be trained quickly on smaller dataset	Limited ability to capture spatial features compared to 2D or 3D cnns; may not perform as well on complex imaging tasks requiring detailed spatial context
MLP-Mixer [72]	Offers simplicity, flexibility, and scalability; performs well on large datasets while being computationally efficient	Limited interactions with neighboring neurons can result in decreased performance with smaller datasets; still relatively new with less established performance metrics in specific biomedical applications
FlimGANE [73]	Combines gans with fluorescence lifetime imaging to enhance image quality and resolution rapidly, even under low-photon conditions; up to 2800 times faster than traditional maximum likelihood estimation methods while providing accurate analyses	Training gans can be resource-intensive and time-consuming; performance may be affected by the stability of the photon-counting electronics used in imaging setup
SENSOR-Net [74]	Leverages spatial correlations and empty pulse information for robust analysis; dynamically learns the number of species and their corresponding lifetime maps simultaneously	Requires sophisticated mathematical modeling and understanding of bayesian methods; may demand significant computational power for processing large datasets
BNP-FLIM [75]	Learns the number of species and their absolute lifetime maps without prior assumptions about species count; effectively propagates uncertainties across species and lifetime estimates	The bayesian framework can be computationally intensive due to the complexity of the models used; requires high-quality data to ensure accurate modeling and predictions

Table 6 Table of various methods available on GitHub.

Model	GitHub
MedSegDiff [46]	https://github.com/SuperMedIntel/MedSegDiff.git , 21 Jan 2025
U-Mamba [47]	https://github.com/bowang-lab/U-Mamba.git , 21 Jan 2025
VM-Unet [48]	https://github.com/JCruan519/VM-UNet.git , 21 Jan 2025
SegMamba [49]	https://github.com/ge-xing/SegMamba.git , 21 Jan 2025
LM-Net [50]	https://github.com/Asunatan/LM-Net.git , 21 Jan 2025

Table 6 Cont.

ARCO [51]	https://github.com/charlesyou999648/ARCO.git , 21 Jan 2025
FLI-Net [11]	https://github.com/jasontsmith2718/DL4FLI.git , 21 Jan 2025
FPFLI [69]	https://github.com/Dooongx/FPFLI.git , 21 Jan 2025
FlimGANE [73]	https://github.com/NinaYIC/flimGANE.git , 21 Jan 2025
BNP-FLIM [75]	https://github.com/MohamadFazel/BNP2-FLIM.git , 21 Jan 2025

4.2 Advantages and Limitations of ML in FLIM

Machine learning (ML) techniques can provide more accurate estimates of fluorescence lifetimes and metabolic pathways compared to traditional analysis methods. ML algorithms can also analyze fluorescence lifetime imaging (FLIM) data faster, allowing for real-time decision-making and analysis. This enables a more intuitive understanding of the data, providing insights into cellular metabolism and disease progression. Label-free imaging combined with ML removes the need for time-consuming and expensive labeling processes, making this approach a powerful tool for researchers. ML can handle complex multi-dimensional FLIM data, extracting valuable information that can be used to improve clinical outcomes. However, implementing ML in FLIM requires large and high-quality datasets, as well as appropriate approaches to overcoming clinical, professional, and technical challenges to its successful translation to clinical practice [77]. There are many techniques for the denoising of FLIM images using machine learning and deep-learning methods.

5 Future Directions

Physics-based deep learning (PBDL) is an emerging approach that combines the principles of physics with deep learning algorithms to enhance the performance, effectiveness, and resilience of various applications. In the context of FLIM, the potential of PBDL in overcoming the constraints of conventional data-driven or physics-based methods is immense. However, there are no existing PBDL models tailored specifically for FLIM yet.

PBDL models leverage governing equations, such as partial differential equations, by embedding them into the architecture or loss functions, promoting solutions that adhere to known physics. By incorporating physical laws, PBDL models achieve high accuracy with less training data, making them particularly beneficial in scenarios where large datasets are challenging to obtain. They effectively address both forward simulations and inverse problems, allowing for a wide range of applications. The augmentation of loss functions with physical constraints ensures that predictions align with real-world behaviors, while differentiable physics simulations enable tight coupling between deep learning and physical modeling. Additionally, PBDL models can utilize virtual data points to evaluate physical

consistency, enhancing training efficiency. In FLIM and beyond, these models improve analysis by providing robust, physically consistent solutions that are resilient to noise and variability, making them valuable tools for real-time assessments and dynamic process monitoring in scientific research and practical applications [78–80].

The integration of physical constraints, such as models of fluorescence decay and photon transport, into deep learning architectures is a significant advancement in the field of FLIM. One way to effectively embed physical constraints is by integrating governing equations related to fluorescence decay and photon transport into the loss function of deep learning models. Using partial differential equations (PDEs) as regularizers guides the training process, ensuring that model predictions adhere to known physical laws. Physics-Informed Neural Networks (PINNs) are neural networks designed to solve forward and inverse problems involving nonlinear PDEs, incorporating physical laws directly into their architecture. These networks can learn from limited data while ensuring physically plausible predictions, which is particularly useful in FLIM where data can be scarce or noisy. Recent advancements have led to the introduction of filtered PDEs as surrogate constraints within PBDL frameworks. This approach enables robust solutions even in noisy environments, by calculating PDE losses based on filtered variables. It maintains the integrity of the governing equations, while being less sensitive to noise. Simulated datasets, based on known physical models, can augment real experimental data, providing additional training examples for deep learning models. This improves the model's ability to generalize across different experimental conditions, enhancing its performance [11, 81, 82].

Developing models that effectively integrate deep learning with physical simulations can be complex and requires a deep understanding of both fields. This complexity may limit accessibility for practitioners who are not experts in either domain. While PBDL reduces the need for large datasets, the quality of the available data remains critical. Poor quality or biased data can lead to inaccurate models, undermining the benefits of integrating physical knowledge. Training hybrid models that combine deep learning with physics-based simulations often requires significant computational resources, which may not be available to all researchers or institutions [78–80].

As this field continues to develop, overcoming the challenges of integration and computational demands will be crucial for unlocking the full potential of PBDL

in FLIM applications. This will ultimately advance our understanding of complex biological systems.

6 Conclusion

The review summarizes recent advances in the use of machine learning and deep learning techniques for fluorescence lifetime imaging microscopy data analysis and modeling.

An analysis of the use of deep learning methods for FLIM data analysis showed that:

- FLI-Net, 1D CNN, MLP-Mixer are suitable for real-time FLIM;
- FPFLI, FlimGANE are suitable for low-photon conditions;
- FLAN is suitable for FPGA;
- SENSOR-Net, BNP-FLIM are suitable for simultaneous FLIM estimation and classification;
- FPFLI, FlimGANE, SENSOR-Net and BNP-FLIM are suitable for high-resolution imaging;
- FLI-Net, MLP-Mixer, FlimGANE and especially BNP-FLIM are significant resistant to data noise;
- FLI-Net, FPFLI and FlimGANE are suitable for small datasets.

Traditional methods for extracting fluorescence lifetimes from FLIM data can be complex and time-consuming, but machine learning and deep learning

approaches have emerged as promising alternatives that offer faster and more accurate lifetime extraction.

Convolutional neural networks have been particularly prevalent in FLIM analysis, demonstrating effectiveness in denoising and enhancing fluorescence microscopy images, as well as in developing novel techniques.

In conclusion, machine learning and deep learning-based techniques have demonstrated great potential in FLIM data modeling, offering faster and more accurate lifetime extraction, denoising capabilities, and improved image enhancement.

Funding

Russian Science Foundation №23-15-00100.

Acknowledgment

Thanks to Svetlana Rodimova, Daria Kuznetsova and Vladislav Shcheslavsky (PRMU) for their help in writing the manuscript.

Disclosures

The authors declare no conflicts of interest.

Data Availability

No data were generated or analyzed in the presented research.

References

1. R. Datta, T. M. Heaster, J. T. Sharick, A. A. Gillette, and M. C. Skala, “[Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications](#),” *Journal of Biomedical Optics* 25(07), 1 (2020).
2. P. A. Bonilla, R. Shrestha, “[FLIM-FRET Protein-Protein Interaction Assay](#),” in KRAS, A. G. Stephen and D. Esposito (Eds.), 2797, Springer US, 261–269 (2024).
3. P. P. Provenzano, K. W. Eliceiri, and P. J. Keely, “[Multiphoton microscopy and fluorescence lifetime imaging microscopy \(FLIM\) to monitor metastasis and the tumor microenvironment](#),” *Clinical & Experimental Metastasis* 26(4), 357–370 (2009).
4. S. Karpf, C. T. Riche, D. Di Carlo, A. Goel, W. A. Zeiger, A. Suresh, C. Portera-Cailliau, and B. Jalali, “[Spectro-temporal encoded multiphoton microscopy and fluorescence lifetime imaging at kilohertz frame-rates](#),” *Nature Communications* 11(1), 2062 (2020).
5. M. Evers, N. Salma, S. Osseiran, M. Casper, R. Birngruber, C. L. Evans, and D. Manstein, “[Enhanced quantification of metabolic activity for individual adipocytes by label-free FLIM](#),” *Scientific Reports* 8(1), 8757 (2018).
6. T. Sanchez, T. Wang, M. V. Pedro, M. Zhang, E. Esencan, D. Sakkas, D. Needleman, and E. Seli, “[Metabolic imaging with the use of fluorescence lifetime imaging microscopy \(FLIM\) accurately detects mitochondrial dysfunction in mouse oocytes](#),” *Fertility and Sterility* 110(7), 1387–1397 (2018).
7. W. Becker, A. Bergmann, L. Braun, and Becker, “[Metabolic Imaging with the DCS-120 Confocal FLIM System : Simultaneous FLIM of NAD \(P \) H and FAD](#),” Application note (2018).
8. A. D. Komarova, S. D. Sinyushkina, I. D. Shchekhin, I. N. Druzhkova, S. A. Smirnova, V. M. Terekhov, A. M. Mozherov, N. I. Ignatova, E. E. Nikonova, E. A. Shirshin, L. E. Shimolina, S. V. Gamayunov, V. I. Shcheslavskiy, and M. V. Shirmanova, “[Insights into metabolic heterogeneity of colorectal cancer gained from fluorescence lifetime imaging](#),” *eLife* 13, RP94438 (2024).
9. Y. Ouyang, Y. Liu, Z. M. Wang, Z. Liu, and M. Wu, “[FLIM as a Promising Tool for Cancer Diagnosis and Treatment Monitoring](#),” *Nano-Micro Letters* 13(1), 133 (2021).
10. B. W. Weyers, M. Marsden, T. Sun, J. Bec, A. F. Bewley, R. F. Gandour-Edwards, M. G. Moore, D. G. Farwell, and L. Marcu, “[Fluorescence lifetime imaging for intraoperative cancer delineation in transoral robotic surgery](#),” *Translational Biophotonics* 1(1–2), e201900017 (2019).

11. J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, “[Fast fit-free analysis of fluorescence lifetime imaging via deep learning](#),” *Proceedings of the National Academy of Sciences* 116(48), 24019–24030 (2019).
12. A. Buades, B. Coll, and J.-M. Morel, “[A review of image denoising algorithms, with a new one](#),” *Multiscale Modeling and Simulation* 4 (2), 490–530 (2005).
13. A. Shah, J. I. Bangash, A. W. Khan, I. Ahmed, A. Khan, A. Khan, and A. Khan, “[Comparative analysis of median filter and its variants for removal of impulse noise from gray scale images](#),” *Journal of King Saud University-Computer and Information Sciences* 34(3), 505–519 (2022).
14. W. Zhou, Y. Miura, “[Denoising method using a moving-average filter for the moving-object recognition of low-illuminance video-image](#),” in 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), IEEE (2016).
15. D. Bhonsle, V. Chandra, and G. R. Sinha, “[Medical Image Denoising Using Bilateral Filter](#),” *International Journal of Image, Graphics and Signal Processing* 4(6), 36–43 (2012).
16. S. Shrestha, “[Image Denoising using New Adaptive Based Median Filters](#),” *Signal & Image Processing: An International Journal (SIPIJ)* 5(4), 1–13 (2014).
17. Z. Shan, J. Yang, M. A. F. Sanjuán, C. Wu, and H. Liu, “[A novel adaptive moving average method for signal denoising in strong noise background](#),” *The European Physical Journal Plus* 137(1), 50 (2022).
18. R. Tian, G. Sun, X. Liu, and B. Zheng, “[Sobel Edge Detection Based on Weighted Nuclear Norm Minimization Image Denoising](#),” *Electronics* 10(6), 655 (2021).
19. B. Jakisc, R. Ivokovic, B. Gara, M. Petrovic, and P. Spalevic, “[Analysis of different influence of compression algorithm on the image filtered Laplacian, Prewitt and Sobel operator](#),” *International Journal of Darshan Institute on Engineering Research & Emerging Technologies* 2(1), 68–76 (2013).
20. W. Piao, Y. Yuan, and H. Lin, “[A Digital Image Denoising Algorithm Based on Gaussian Filtering and Bilateral Filtering](#),” *ITM Web of Conferences* 17, 01006 (2018).
21. S. Paris, S. W. Hasinoff, and J. Kautz, “[Local Laplacian filters: edge-aware image processing with a Laplacian pyramid](#),” *Communications of the ACM* 58(3), 81–91 (2015).
22. Y. Raphiphan, S. Wattanakaroon, and S. Khetkeeree, “[Adaptive High Boost Filtering for Increasing Grayscale and Color Image Details](#),” in 2020 International Conference on Power, Energy and Innovations (ICPEI), IEEE, 69–72 (2020).
23. Z. Shi, Y. Chen, E. Gavves, P. Mettes, and C. G. M. Snoek, “[Unsharp Mask Guided Filtering](#),” *IEEE Transactions on Image Processing* 30, 7472–7485 (2021).
24. M. A. Digman, V. R. Caiolfa, M. Zamai, and E. Gratton, “[The phasor approach to fluorescence lifetime imaging analysis](#),” *Biophysical Journal* 94(2), L14–L16 (2008).
25. V. Mannam, Y. Zhang, X. Yuan, T. Hato, P. C. Dagher, E. L. Nichols, C. J. Smith, K. W. Dunn, and S. Howard, “[Convolutional neural network denoising in fluorescence lifetime imaging microscopy \(FLIM\)](#),” in *Multiphoton Microscopy in the Biomedical Sciences XXI*, A. Periasamy, P. T. So, and K. König (Eds.), SPIE Proceedings 11648, 116481C (2021).
26. B. Fu, X. Zhao, Y. Li, X. Wang, and Y. Ren, “[A convolutional neural networks denoising approach for salt and pepper noise](#),” *Multimedia Tools and Applications* 78, 30707–30721 (2019).
27. P. C. Tripathi, S. Bag, “[CNN-DMRI: A Convolutional Neural Network for Denoising of Magnetic Resonance Images](#),” *Pattern Recognition Letters* 135, 57–63 (2020).
28. Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, “[Transfer Learning from Synthetic to Real-Noise Denoising with Adaptive Instance Normalization](#),” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3482–3492 (2020).
29. V. Kapitany, A. Fatima, V. Zickus, J. Whitelaw, E. McGhee, R. Insall, L. Machesky, and D. Faccio, “[Single-sample image-fusion upsampling of fluorescence lifetime images](#),” *Science Advances* 10(21), eadn0139 (2024).
30. C. Callenberg, A. Lyons, D. D. Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio, and M. B. Hullin, “[Super-resolution time-resolved imaging using computational sensor fusion](#),” *Scientific Reports* 11, 1689 (2021).
31. B. Huang, J. Li, B. Yao, Z. Yang, E. Y. Lam, J. Zhang, W. Yan, and J. Qu, “[Enhancing image resolution of confocal fluorescence microscopy with deep learning](#),” *PhotoniX* 4, 2 (2023).
32. D. Xiao, Z. Zang, W. Xie, N. Sapermsap, Y. Chen, and D. D. Uei Li, “[Spatial resolution improved fluorescence lifetime imaging via deep learning](#),” *Optics Express* 30(7), 11479 (2022).
33. M. Adhikari, R. Houhou, J. Hniopek, and T. Bocklitz, “[Review of Fluorescence Lifetime Imaging Microscopy \(FLIM\) Data Analysis Using Machine Learning](#),” *Journal of Experimental and Theoretical Analyses* 1(1), 44–63 (2023).
34. W. Shi, D. E. S. Koo, M. Kitano, H. J. Chiang, L. A. Trinh, G. Turcatel, B. Steventon, C. Arnesano, D. Warburton, S. E. Fraser, and F. Cutrale, “[Pre-processing visualization of hyperspectral fluorescent data with Spectrally Encoded Enhanced Representations](#),” *Nature Communications* 11(1), 726 (2020).

35. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “[UNet++: A Nested U-Net Architecture for Medical Image Segmentation](#),” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi (Eds.), Springer International Publishing, 11045, 3–11 (2018).
36. W. Tingxi, T. Binbin, L. Yu, P. Ting, D. Yu, C. Yuping, and Z. Shanshan, “[Review of research on the instance segmentation of cell images](#),” *Computer Methods and Programs in Biomedicine* 227, 107211 (2022).
37. S. M. A. Burney, H. Tariq, “[K-Means Cluster Analysis for Image Segmentation](#),” *International Journal of Computer Applications* 96(4), 1–8 (2014).
38. W. Gao, “[Random Forest in Image Segmentation](#),” *Machine Learning Theory and Practice* 1, 11–19 (2020).
39. Z. Kato, “[Markov Random Fields in Image Segmentation](#),” *Foundations and Trends® in Signal Processing* 5(1–2), 1–155 (2011).
40. B. Karthicsonia, M. Vanitha, “[Edge Based Segmentation in Medical Images](#),” *International Journal of Engineering and Advanced Technology* 9(1), 449–451 (2019).
41. O. Ronneberger, P. Fischer, and T. Brox, “[U-Net: Convolutional Networks for Biomedical Image Segmentation](#),” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Eds.), Springer International Publishing, 9351, 234–241 (2015).
42. D. Bannon, E. Moen, M. Schwartz, E. Borba, T. Kudo, N. Greenwald, V. Vijayakumar, B. Chang, E. Pao, E. Osterman, W. Graf, and D. Van Valen, “[DeepCell Kiosk: scaling deep learning-enabled cellular image analysis with Kubernetes](#),” *Nature Methods* 18(1), 43–45 (2021).
43. A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini, “[CellProfiler: image analysis software for identifying and quantifying cell phenotypes](#),” *Genome Biology* 7(10), R100 (2006).
44. K. He, X. Zhang, S. Ren, and J. Sun, “[Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition](#),” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(9), 1904–1916 (2015).
45. Y. Wang, W. Wang, D. Liu, W. Hou, T. Zhou, and Z. Ji, “[GeneSegNet: a deep learning framework for cell segmentation by integrating gene expression and imaging](#),” *Genome Biology* 24(1), 235 (2023).
46. J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, H. Liu, and Y. Xu, “[MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model](#),” in *Medical Imaging with Deep Learning*, *Proceedings of Machine Learning Research* 227, 1623–1639 (2024).
47. J. Ma, F. Li, and B. Wang, “[U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation](#),” *arXiv preprint arXiv:2401.04722* (2024).
48. J. Ruan, S. Xiang, “[VM-UNet: Vision Mamba UNet for Medical Image Segmentation](#),” *arXiv preprint arXiv:2402.02491* (2024).
49. Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu, “[SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation](#),” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer Nature Switzerland, Cham, 578–588 (2024).
50. Z. Lu, C. She, W. Wang, and Q. Huang, “[LM-Net: A light-weight and multi-scale network for medical image segmentation](#),” *Computers in Biology and Medicine* 168, 107717 (2024).
51. C. You, W. Dai, Y. Min, F. Liu, D. Clifton, S. K. Zhou, L. Staib, and J. Duncan, “[Rethinking Semi-Supervised Medical Image Segmentation: A Variance-Reduction Perspective](#),” *Advances in neural information processing systems*, 36, 9984–10021 (2023).
52. D. Müller, I. Soto-Rey, and F. Kramer, “[Towards a guideline for evaluation metrics in medical image segmentation](#),” *BMC Research Notes* 15(1), 210 (2022).
53. H. Khastavaneh, H. Ebrahimpour-komleh, “[Automated Segmentation of Abnormal Tissues in Medical Images](#),” *Journal of Biomedical Physics & Engineering* 11(4), 415–424 (2019).
54. P. Qin, J. Chen, J. Zeng, R. Chai, and L. Wang, “[Large-scale tissue histopathology image segmentation based on feature pyramid](#),” *EURASIP Journal on Image and Video Processing* 2018, 75 (2018).
55. A. K. Kuchibhotla, L. D. Brown, A. Buja, and J. Cai, “[All of Linear Regression](#),” *arXiv preprint arXiv:1910.06386* (2019).
56. H.-H. Huang, Q. He, “[Nonlinear regression analysis](#),” in *International Encyclopedia of Education*, 4th Ed., Elsevier, 558–567 (2023).
57. R. S. Turley, “Polynomial Fitting”, Faculty Publication, 2237, 1 Jan 2018 (accessed 22 October 2024). [<https://scholarsarchive.byu.edu/facpub/2237>].
58. D. C. Vargas, E. J. Rodríguez, M. Flickner, and J. L. C. Sanz, “[Splines and Spline Fitting Revisited](#),” in *Image Technology: Advances in Image Processing, Multimedia and Machine Vision*, J. L. C. Sanz (Ed.), Springer Berlin Heidelberg, 405–437 (1996).
59. M. Li, L. D. Li, “[A Novel Method of Curve Fitting Based on Optimized Extreme Learning Machine](#),” *Applied Artificial Intelligence* 34(12), 849–865 (2020).

60. V. P. D. R. Rao, “[Machine Learning Techniques on Multidimensional Curve Fitting Data Based on R-Square and Chi-Square Methods](#),” International Journal of Electrical and Computer Engineering 6(3), 974–979 (2016).
61. S. S. Sidharth, “[Curve Fitting Simplified: Exploring the Intuitive Features of CurvPy](#),” arXiv preprint arXiv:2307.06377 (2023).
62. N. Tumbarkova, N. Deliiski, “[Applying the software package TableCurve 2D for computation of processing air medium temperature during freezing in a freezer and defrosting of logs](#),” Innovations in Woodworking Industry and Engineering Design 11, 69–76 (2017).
63. A. Dakkak, T. Wickham-Jones, and W.-m. Hwu, “[The design and implementation of the wolfram language compiler](#),” Proceedings of the 18th ACM/IEEE International Symposium on Code Generation and Optimization, San Diego, CA, USA, 212–228 (2020).
64. T. Cepowski, ndCurveMaster (8.7), SigmaLab [Computer Software], (2025) (accessed 22 October 2024). [<https://www.ndcurvemaster.com>].
65. D. Gao, P. R. Barber, J. V. Chacko, Md. A. Kader Sagar, C. T. Rueden, A. R. Grislis, M. C. Hiner, and K. W. Eliceiri, “[FLIMJ: An open-source ImageJ toolkit for fluorescence lifetime image data analysis](#),” PLoS ONE 15(12), e0238327 (2020).
66. D. Nečas, P. Klapetek, “[Gwyddion: an open-source software for SPM data analysis](#),” Central European Journal of Physics 10, 181–188 (2011).
67. L. Hu, N. Wang, J. D. Bryant, L. Liu, L. Xie, A. P. West, and A. J. Walsh, “[Machine learning prediction of cancer cell metabolism from autofluorescence lifetime images](#),” bioRxiv [preprint], 2022.2012.16.520759 (2022).
68. N. I. Nizam, V. Pandey, I. Erbas, J. T. Smith, and X. Intes, “[A Novel Technique for Fluorescence Lifetime Tomography](#),” bioRxiv [preprint], 2024.09.19.613888 (2024).
69. D. Xiao, N. Sapermsap, Y. Chen, and D. D. U. Li, “[Deep learning enhanced fast fluorescence lifetime imaging with a few photons](#),” Optica 10(7), 944 (2023).
70. Z. Zang, D. Xiao, Q. Wang, Z. Jiao, Y. Chen, and D. D. U. Li, “[Compact and robust deep learning architecture for fluorescence lifetime imaging and FPGA implementation](#),” Methods and Applications in Fluorescence 11(2), 025002 (2023).
71. D. Xiao, Y. Chen, and D. D.-U. Li, “[One-Dimensional Deep Learning Architecture for Fast Fluorescence Lifetime Imaging](#),” IEEE Journal of Selected Topics in Quantum Electronics 27(4), 1–10 (2021).
72. Q. Wang, Y. Li, D. Xiao, Z. Zang, Z. Jiao, Y. Chen, and D. D. U. Li, “[Simple and Robust Deep Learning Approach for Fast Fluorescence Lifetime Imaging](#),” Sensors 22(19), 7293 (2022).
73. Y.-I. Chen, Y.-J. Chang, S.-C. Liao, T. D. Nguyen, J. Yang, Y.-A. Kuo, S. Hong, Y.-L. Liu, H. G. Rylander, S. R. Santacruz, T. E. Yankeelov, and H.-C. Yeh, “[Deep learning enables rapid and robust analysis of fluorescence lifetime imaging in photon-starved conditions](#),” bioRxiv [preprint], 2020.2012.02.408195 (2020).
74. F. Wang, S. H. Kim, Y. Zhao, A. Raghuram, A. Veeraraghavan, J. Robinson, and A. H. Hielscher, “[High-Speed Time-Domain Diffuse Optical Tomography With a Sensitivity Equation-Based Neural Network](#),” IEEE Transactions on Computational Imaging 9, 459–474 (2023).
75. M. Fazel, S. Jazani, L. Scipioni, A. Vallmitjana, S. Zhu, E. Gratton, M. A. Digman, and S. Pressé, “[Building Fluorescence Lifetime Maps Photon-by-Photon by Leveraging Spatial Correlations](#),” ACS Photonics 10(10), 3558–3569 (2023).
76. D.-M. Koh, N. Papanikolaou, U. Bick, R. Illing, C. E. Kahn, J. Kalpathi-Cramer, C. Matos, L. Martí-Bonmatí, A. Miles, S. K. Mun, S. Napel, A. Rockall, E. Sala, N. Strickland, and F. Prior, “[Artificial intelligence and machine learning in cancer imaging](#),” Communications Medicine 2(1), 133 (2022).
77. V. Mannam, Y. Zhang, X. Yuan, C. Ravasio, and S. S. Howard, “[Machine learning for faster and smarter fluorescence lifetime imaging microscopy](#),” Journal of Physics: Photonics 2(4), 042005 (2020).
78. N. Thuerey, P. Holl, M. Mueller, P. Schnell, F. Trost, and K. Um, “[Physics-based Deep Learning](#),” arXiv preprint arXiv:2109.05237 (2021).
79. A. Arzani, L. Yuan, P. Newell, and B. Wang, “[Interpreting and generalizing deep learning in physics-based problems with functional linear models](#),” Engineering with Computers 41(1), 135–157 (2025).
80. Y. Sun, Q. Sun, and K. Qin, “[Physics-Based Deep Learning for Flow Problems](#),” Energies 14(22), 7760 (2021).
81. D. Zhang, Y. Chen, and S. Chen, “[Filtered partial differential equations: a robust surrogate constraint in physics-informed deep learning framework](#),” Journal of Fluid Mechanics 999, A40 (2024).
82. S. Kaltenbach, P.-S. Koutsourelakis, “[Incorporating physical constraints in a deep probabilistic machine learning framework for coarse-graining dynamical systems](#),” Journal of Computational Physics 419, 109673 (2020).

Design of a Photonic Integrated Circuit for Swept-Source OCT for Simultaneous Operation in Two Distinct Wavelength Ranges

Ivan V. Stepanov*, Vladimir S. Lyubopytov, Anton A. Ivanov, and Elizaveta P. Grakhova

Ufa University of Science and Technology, 32 Zaki Validi str., Ufa 450076, Russian Federation

*e-mail: stepanov.iv@ugatu.su

Abstract. The paper presents a photonic integrated circuit (PIC) concept for the swept-source optical coherence tomography (SS-OCT) systems operating within two distinct wavelength ranges: 820–880 nm and 1260–1360 nm. The PIC features two interferometric schemes connected through a joint splitter that delivers a single output for tissue scanning. The proposed PIC concept is developed for the silicon nitride platform, which offers low losses compared to another photonic integrated platform. The numerical simulations were conducted using the Ansys Lumerical software to verify the PIC concept. The simulation results indicate that the focal length of the proposed SS-OCT system is approximately 2.9 mm for both wavelength ranges, and there is no interference between the optical signals from the adjacent interferometric systems. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: optical coherence tomography; photonic integrated circuits; swept-source OCT; dual-band OCT.

Paper #9190 received 20 Nov 2024; revised manuscript received 18 Dec 2024; accepted for publication 19 Dec 2024; published online 31 Jan 2025. doi: [10.18287/JBPE25.11.010301](https://doi.org/10.18287/JBPE25.11.010301).

1 Introduction

Optical coherence tomography (OCT) prospects are highly positive, fueled by continuous technological advancements and expanding applications across various medical fields. OCT is a noninvasive method used to visualize the structure of biological tissues by analyzing the interference between an optical reference signal and the backscattered and backreflected signals from the sample being studied. Initially, OCT gained prominence in ophthalmology [1, 2], but its applications are now rapidly expanding. For instance, in cardiology, intravascular OCT images of coronary arteries provide high-resolution images of plaque morphology and stent deployment [3]. This technique aids in the diagnosis and treatment of coronary artery disease [4]. In dermatology, OCT is increasingly utilized for noninvasive skin imaging [5], assisting in the diagnosis of skin cancers [6], inflammatory skin diseases [7], and other dermatological conditions [8]. Additionally, emerging applications are being explored in gastroenterology (specifically, endoscopy), neurology, and other medical specialties.

Modern OCT systems offer significantly faster acquisition speeds and higher axial and transverse resolutions than earlier models. These advancements enable

quicker examinations, minimize motion artifacts, and allow for the visualization of finer anatomical details. One notable example is swept-source OCT (SS-OCT), which provides much faster scan rates than spectral-domain OCT (SD-OCT) [9]. Despite these improvements, OCT still faces several challenges. The development of compact OCT systems, potentially utilizing photonic integrated circuits (PICs), is facilitating the creation of more accessible and portable devices well-suited for point-of-care applications and broader use in resource-limited settings [10]. Various solutions have already been proposed for the implementation of individual components within the PIC OCT system [11, 12] and for highly compact systems that are fully integrated [13]. Modern PICs incorporate SS-OCT [14–16] and SD-OCT [17, 18] methods and encompass all relevant wavelength ranges for medical applications, specifically 800 [12, 19], and 1310 nm [17, 20].

At the same time, another trend comprises searching for techniques to advance systems imaging speed, resolution (axial and transverse), and penetration depth. Such research focuses on using dual-band imaging schemes or integrating with other imaging modalities [21, 22], that can provide a more comprehensive picture of tissue structure and function.

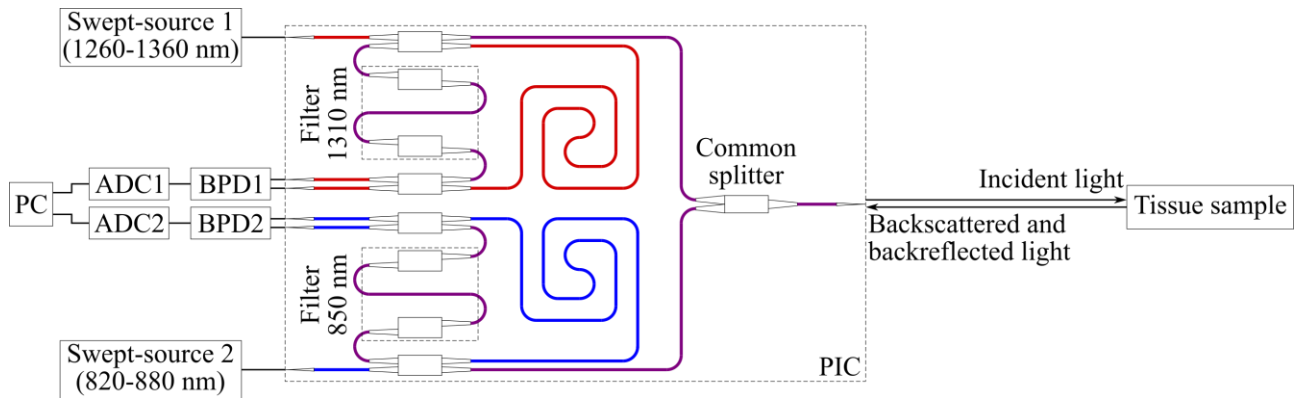


Fig. 1 Concept of a PIC-based dual-band SS-OCT system: PC – personal computer; ADC – analog-to-digital converter; BPD – balanced photodetector; PIC – photonic integrated circuit. Blue shows the waveguides in which radiation in the 820–880 nm wavelength range propagates, red – 1260–1360 nm, and violet – the waveguides in which both wavelength ranges can propagate.

The first approach seems particularly promising due to its simple integration with integrated photonics technology. Dual-band OCT presents significant benefits compared to single-band systems by utilizing different wavelengths that uniquely interact with tissue. This approach delivers complementary information, improving our tissue structure and function comprehension.

The applications of dual-band OCT are diverse and span various fields, with a significant focus on ophthalmology. In this field, dual-band OCT is particularly valuable for detailed retinal imaging, including angiography, blood flow measurement, and the assessment of retinal oxygen metabolism. It plays a crucial role in diagnosing and monitoring various eye diseases, such as glaucoma [23], where distinct differences can be observed in the OCT cross-sectional images at the two wavelengths used for the retinal nerve fiber layer. The dual-band OCT utilizes wavelengths of 840 nm and 1050 nm for visualizing the retina and the anterior segments of the eye [24].

Additionally, a second measurement in a different range enhances tissue contrast [25] and reduces image speckle, particularly in FD-OCT, while improving resolution [26]. For instance, visible light optical coherence tomography (Vis-OCT) offers superior axial resolution for imaging retinal layers, especially the retinal pigment epithelium and choroid complex. In contrast, near-infrared optical coherence tomography (NIR-OCT) provides greater penetration depth for examining deeper structures [27]. Moreover, the same research [27] has shown that Vis-OCT excels in measuring retinal blood oxygen saturation due to its increased blood optical absorption contrast in the visible spectrum.

In dermatology, researchers are exploring the use of dual-band line-field confocal optical coherence tomography (LC-OCT) systems that operate at wavelengths of 770 nm and 1250 nm. These systems aim to enhance the non-invasive diagnosis of malignant skin neoplasms, particularly basal cell carcinomas (BCCs). The dual-band approach integrates high-resolution imaging of superficial layers at 770 nm with deeper

penetration capabilities at 1250 nm, resulting in a more comprehensive view of the lesion [28]. In cardiology, different wavelengths are employed to improve the differentiation of plaque components, such as lipid-rich areas versus fibrous tissue. However, it is essential to note that specific studies demonstrating this advantage need to be directly cited in the available data.

However, the lack of readily available publications detailing fully integrated dual-band OCT on PICs suggests that this is an area of ongoing research and development. While the building blocks for a PIC-based dual-band OCT system exist (dual-band OCT systems and PIC-based OCT systems), a fully integrated solution is not clearly documented in the provided data.

Thus, the paper presents the concept of a dual-band PIC designed for the most technologically relevant method – SS-OCT. We propose integrating two data collection schemes on a single chip, operating at two optimal wavelength ranges for medical applications: around 850 nm (820–880 nm) and around 1310 nm (1260–1360 nm). Both schemes capture backscattered and backreflected signals through a single output on the chip, ensuring data acquisition at varying depths and with different resolutions from a single observation point. The paper is organized as follows. Section 2 describes the concept of the system and the basic operating principle of the SS-OCT system on PIC. In Section 3, we present the design and simulation models for the system's building blocks. Section 4 describes the simulation results of the overall system performance, while Section 5 provides a discussion and main conclusions.

2 The Dual-Band OCT Scheme

Fig. 1 illustrates the concept of the proposed dual-band SS-OCT system. The chip is equipped with two tunable-source OCT circuits linked by a splitter, providing a single scanning output across both wavelength ranges.

The operating principle for both circuits, which use two different wavelengths, is consistent across the chip. Therefore, this chapter will focus on the functioning of just one circuit. Light from a tunable laser source is

coupled to the chip through an edge coupler. A directional multimode interference coupler (MMI) splits the light beam into a sample arm and a reference arm. The light beam in the sample arm is directed to the tissue sample via a joint splitter and decoupled from the PIC through the output edge coupler. The backscattered and backreflected light from the tissue sample is then collected by the chip using the same edge coupler. The signal is divided by a joint splitter into two interference circuits. Before reaching the MMI, the signal passes through a filter in each circuit. Ultimately, the backscattered and backreflected radiation interferes with the reference signal in the subsequent MMI. A balanced photodetector then converts the interference result into an electrical signal.

The PIC design is optimized for the silicon nitride platform, which exhibits lower losses than other integrated photonics platforms [29].

3 The PIC Components Design and Simulation

Numerical methods, specifically the finite difference eigenmode (FDE) and the eigenmode decomposition (EME) methods, were employed to simulate the components of the PIC for dual-band SS-OCT.

3.1. The Waveguide Cross-Section for Both Wavelength Ranges

The essential step in any PIC design is defining the waveguide cross-section, which facilitates light transmission between different components. The proposed device utilizes strip waveguides that are 450 nm in height. To determine the appropriate width of these waveguides, graphs were created to illustrate the relationship between the effective refractive index of the modes supported by the waveguide and its width. The resulting graphs for the central wavelengths of 850 nm and 1310 nm from the light sources are displayed in Figs. 2 and 3, respectively.

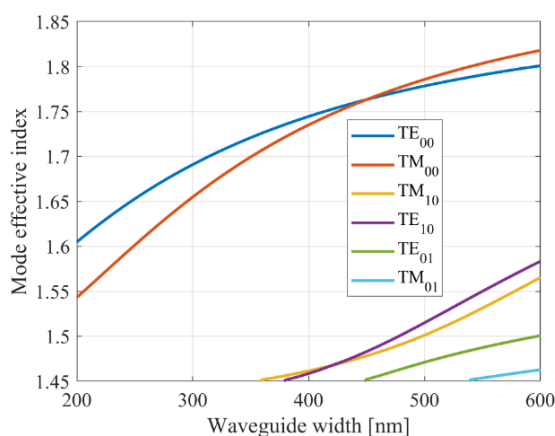


Fig. 2 The effective index of transverse modes in a waveguide versus its width for 850 nm central wavelength. TE indicates the transverse electric modes and TM – the transverse magnetic modes.

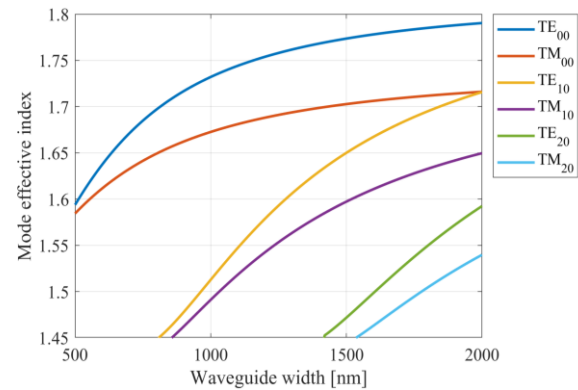


Fig. 3 The effective index of transverse modes in a waveguide versus its width for 1310 nm central wavelength. TE indicates the transverse electric modes and TM – the transverse magnetic modes.

The effective refractive index dependencies indicate that to achieve quasi-single-mode operation, the waveguide widths should be specified as follows: 500 nm for the wavelength range of 820–880 nm and 1200 nm for the wavelength range of 1260–1360 nm.

3.2 Design of the 50/50 MMI for Both Wavelength Ranges

The circuit splits the light beam using MMI, where the power distribution follows the self-imaging principle [30]. Fig. 4 illustrates the generalized structure of the MMI. The EME method was employed to determine the geometry of the MMI for 50/50 split ratios.

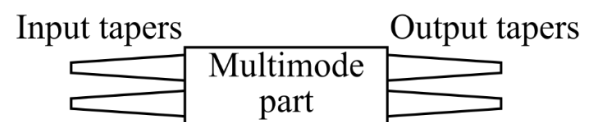


Fig. 4 Generalized structure of the MMI.

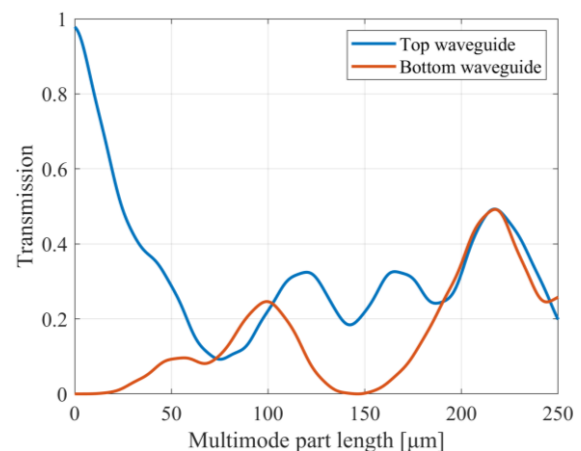


Fig. 5 The 50/50 MMI's transmission spectra relying on the length of the multimode section for a central wavelength of 850 nm.

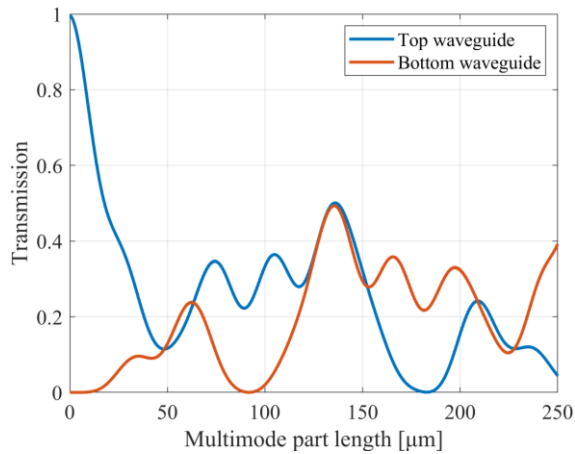


Fig. 6 The 50/50 MMI's transmission spectra relying on the length of the multimode section for a central wavelength of 1310 nm.

The width of the multimode section was set to 12 μm . The coordinates of the input waveguides were determined based on the principles outlined in Ref. [30]. Subsequently, the transmission coefficients from the upper left waveguide to the upper right and lower MMI waveguides were analyzed as a function of the length of the multimode section using the Ansys Lumerical MODE software package. The resulting data for central wavelengths of 850 nm and 1310 nm are presented in Figs. 5 and 6, respectively.

The optimal transmission point in both output waveguides determines the length of the multimode sections. At a central wavelength of 850 nm, the length was 217 μm , while at a central wavelength of 1310 nm, it was 141.2 μm . Fig. 7 illustrates the final geometry for both MMIs, while Figs. 8 and 9 show the transmission performance of the 50/50 MMIs across a broad wavelength range.

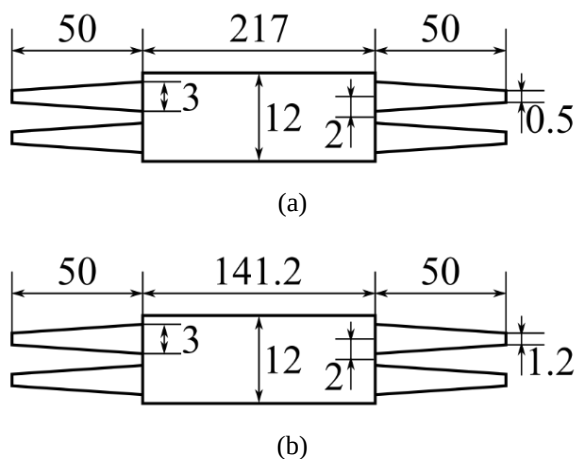


Fig. 7 50/50 MMI dimensions for (a) 850 nm and (b) 1310 nm central wavelengths. All dimensions are in μm .

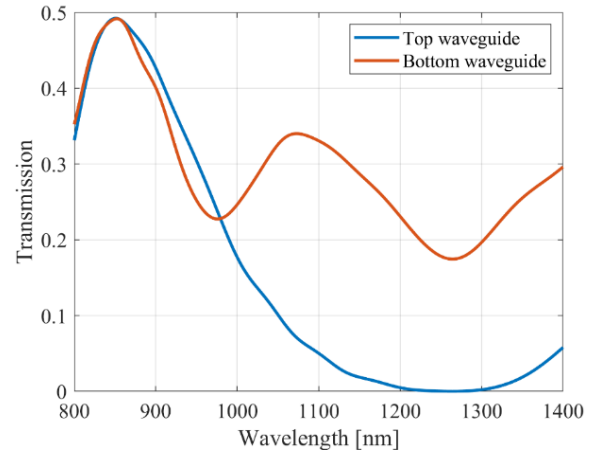


Fig. 8 Transmission spectra of the 50/50 MMI for an 850 nm central wavelength.

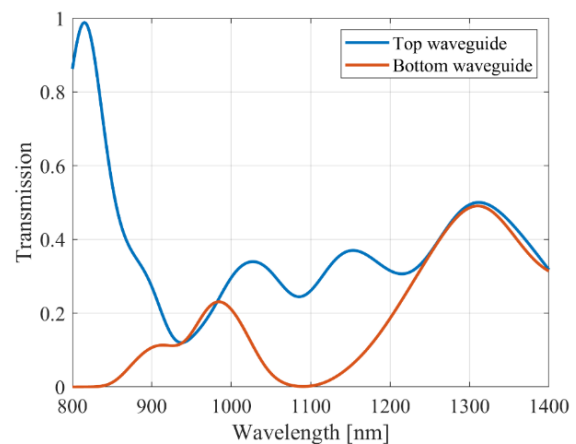


Fig. 9 Transmission spectra of the 50/50 MMI transmission spectra for a 1310 nm central wavelength.

3.3 The Splitter Design for Both Wavelength Ranges

An MMI-based joint splitter combines the interference schemes and divides the backscattered and backreflected radiation between them. The design follows the principles of a 50/50 MMI to determine its geometry. This splitter features a single input with a width set to 4 μm . As with the 50/50 MMI, we analyzed how the transmission coefficients depend on the length of the multimode section for both wavelength ranges (see Fig. 10). Consequently, the length of the multimode section was determined to be 15.1 μm by the maximum transmission for the both central wavelength. The dimensions of the joint splitter are illustrated in Fig. 11, while its transmission spectra are displayed in Fig. 12.

Joint splitter transmission spectra imbalance is conditioned by the different waveguides widths for the various wavelength ranges. Less waveguide width for the 820–880 nm wavelength range provides higher loss for the more abrupt waveguide taper. This effect does not have a significant impact on the system simulation.

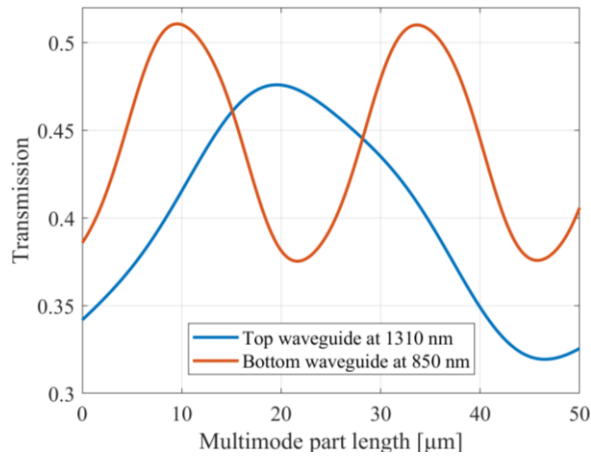


Fig. 10 Transmission spectra of the joint splitter over the length of the multimode section for central wavelengths of 850 nm and 1310 nm.

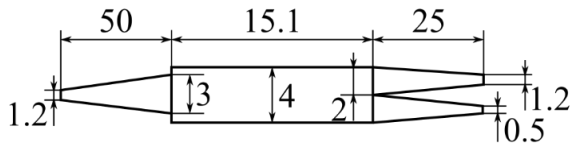


Fig. 11 The joint splitter dimensions in μm .

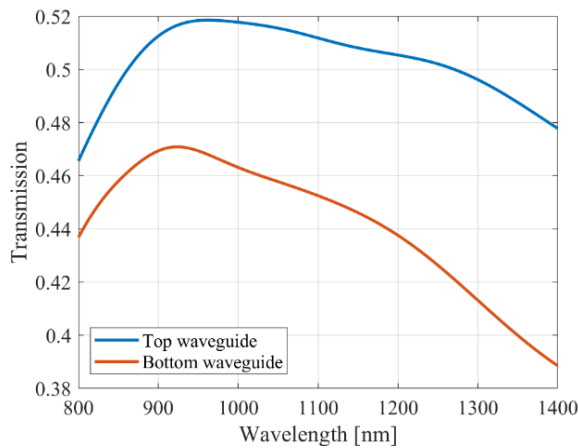


Fig. 12 Transmission spectra of a joint splitter over the broad wavelength band.

3.4 The Exponential-Shaped Edge Coupler

Light is coupled and decoupled from the PIC through an edge coupler, offering a wider bandwidth than traditional input-output gratings [31]. An exponential tapering shape was employed for the waveguide to enhance input efficiency. The length of the edge coupler is $200 \mu\text{m}$, and a specific function can represent its shape:

$$y = \exp\left(\frac{\log(1.5)}{l} \cdot x\right) - 0.9 [\mu\text{m}],$$

where l is the length of the edge coupler, x denotes the coordinates along the X axis in the segment from 0 to l . The edge coupler width spans from 200 to 1200 nm.

Fig. 13 illustrates the transmission of the edge coupler. It is important to note that the transmission was measured for a mode spot with a diameter of $2 \mu\text{m}$, positioned $5 \mu\text{m}$ from the PIC edge. For the edge coupler simulation, we used Ansys Lumerical EME software package.

Edge coupler simulation results show that average loss for the light coupling is approximately 3 dB. This value is sufficient for our simulation. But further light coupling optimization will provide higher OCT system sensitivity. Transmission spectra difference in this case is conditioned by the mode size inequality for the air and for the waveguide.

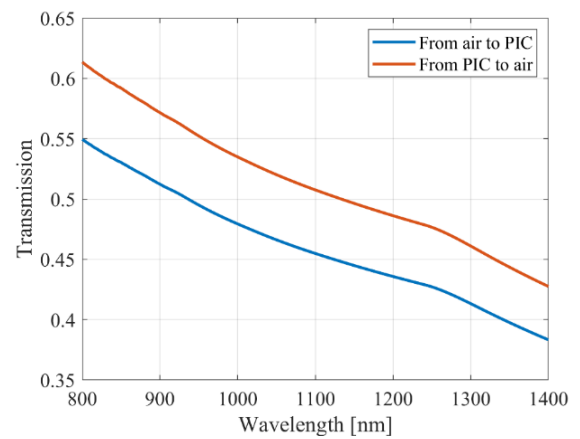


Fig. 13 Edge coupler transmission spectra.

3.5 Bandpass Filters

To effectively differentiate backscattered and backreflected radiation between two segments with distinct operating wavelength ranges, we propose utilizing a bandpass filter based on MMI within the device. These filters were developed utilizing the same principles as those used in the 50/50 MMI and the joint splitter. The key distinction in this case is that the device features only one input and one output. The geometries of the filters are illustrated in Fig. 14, while their transmission characteristics are shown in Fig. 15.

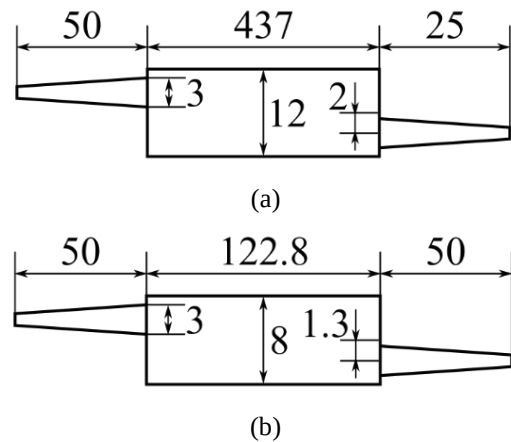


Fig. 14 Dimensions (in μm) of bandpass filters with central wavelengths of (a) 850 nm and (b) 1310 nm.

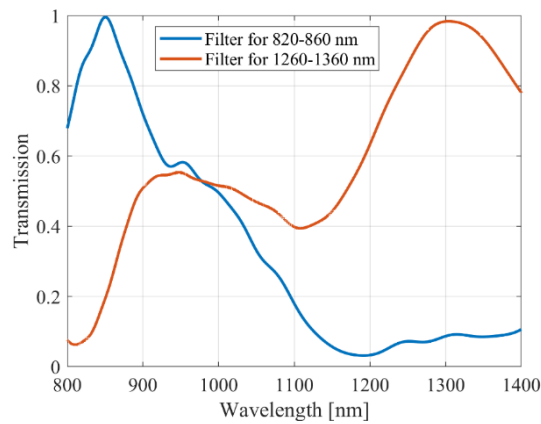


Fig. 15 MMI-based bandpass filters transmission spectra.

We chose bandpass filters lengths by the transmission spectra maximum for the needed wavelength range by analogy with the joint splitter geometry.

4 Simulation Results of the Performance of the PIC-Based Dual-Band SS-OCT System

To evaluate the performance of the SS-OCT system, we converted all the simulation results presented in the previous section of the article into S-parameters. We used the Ansys Lumerical INTERCONNECT software package to create a simulation model of the dual-band PIC. The following parameters were used in the model:

- Simulation time window: 10 ns;
- Sampling frequency: 250 THz;
- Radiation source tuning time: 10 ns;
- Output signal level at the radiation source: 0 dBm;
- Wavelength tuning ranges for the radiation sources: 820–880 nm and 1260–1360 nm;
- Signal attenuation level in waveguides: 92 dB/m [29].

A mirror with a reflectivity of 0.99 was used as a test sample to simulate biological tissue. It was connected to the circuit through a layer of air, the thickness of which was varied to determine the zero-delay point corresponding to the system's focal length.

The tuning time for the radiation sources was set to 10 ns to optimize resource usage. In commercial tunable radiation sources, this tuning time is typically 10 μ s. However, we have previously demonstrated that while tuning time influences the frequency value, it does not affect the system's overall behavior [16, 32]. This assumption does not impact the signal level or the zero delay point.

The simulation results demonstrated the relationship between the OCT system's output signal frequency and the distance from the scanned sample to the end of the PIC. To establish this relationship, we identified the power peak in the signal spectrum and recorded the corresponding frequency in a data array. Fig. 16 illustrates this dependency.

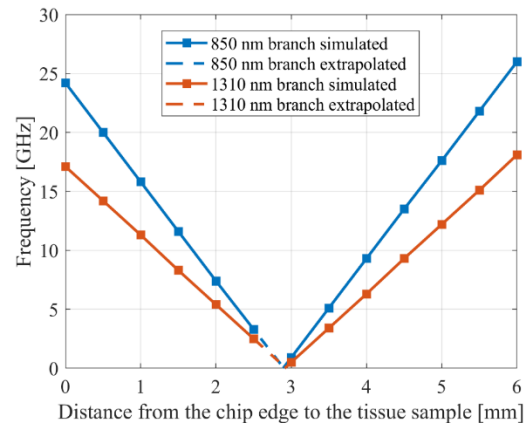


Fig. 16 The output signal frequency in relation to the distance between the scanned sample and the end of the PIC.

Based on the acquired data, extrapolation was conducted in a segment ranging from 2.5 to 3 mm to determine the precise value of the zero-delay point. As a result, the zero-delay point for the OCT branch with a central wavelength of 850 nm was found to be 2.9 mm. For the branch with a wavelength of 1310 nm, the zero-delay point value was 2.92 mm.

Figs. 17 and 18 illustrate examples of OCT signal spectra obtained from simulations at various distances between the scanned sample and the PIC edge.

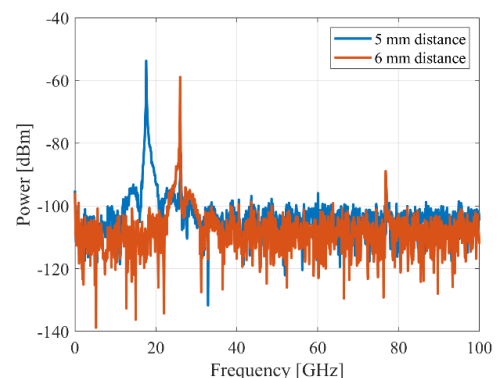


Fig. 17 OCT signal spectra for the branch with a central wavelength of 850 nm.

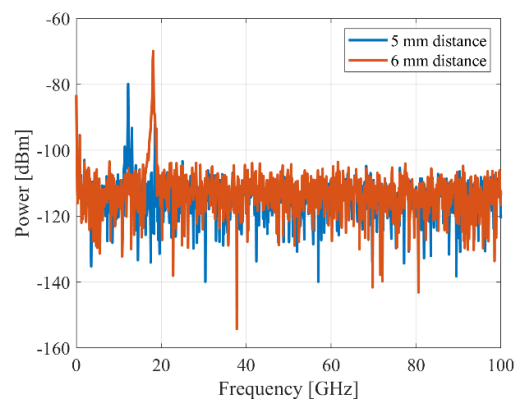


Fig. 18 OCT signal spectra for the branch with a central wavelength of 1310 nm.

The spectra show side frequency peaks. Supposedly, they arise due to the nonlinearity of the transfer characteristics of individual optical system components. It is worth noting that these peaks were also observed when using idealized bandpass optical filters that entirely suppress the adjacent frequency range in the shoulders of the sample.

5 Discussions and Conclusions

This paper presents a PIC concept for an SS-OCT system that can operate simultaneously in two different wavelength ranges: 850 nm and 1310 nm. Although initially designed for the silicon nitride platform, this PIC concept can also be adapted for other integrated photonics platforms. The chosen wavelength ranges are tailored to the most common OCT applications in medical diagnostics, including ophthalmology, dermatology, and endoscopic examinations. In each field, acquiring additional information about the biological tissue sample through measurements at a second wavelength can enhance resolution and increase visualization depth.

It is worth noting that biological tissues absorb radiation differently depending on the wavelength range [33]. This effect can be compensated with proper swept source power selection in the proposed scheme. However, medical standards must limit this power to prevent biological tissue damage.

The simulation results obtained using Ansys Lumerical software indicate that optical signals from adjacent interferometric systems do not interfere with one another. However, the distortions observed in the output signal spectrum for a single reflecting object are attributed to the nonlinear characteristics of individual components within the PIC.

To enhance the performance of dual-band SS-OCT PIC, it is crucial to improve the linearity of the transfer characteristics of its components within the specified operation bands. It can be achieved by optimizing the design of the elements based on MMI or by replacing

them with broadband directional couplers [34]. However, using broadband directional couplers may increase the PIC's sensitivity to variations in characteristics due to manufacturing errors [35]. Furthermore, optimizing the edge couplers can help reduce attenuation during the coupling and decoupling of radiation to the PIC, thereby increasing the system's overall sensitivity.

The second step to improve the performance of the SS-OCT system is to enhance the functionality of the PIC. For both interferometric schemes, the focal length can be adjusted using a tunable reference arm, as previously demonstrated by the authors in an earlier study [32]. Furthermore, OCT systems with a tunable source can integrate a *k*-clock for analog-to-digital converters. This *k*-clock can be implemented either as an unbalanced Mach-Zehnder interferometer [32] or as a microring resonator [36], utilizing dispersion regions that exceed the wavelength range of the radiation source.

Another promising solution to improve OCT image quality is spectroscopic OCT (SOCT) signal processing algorithm realization at the personal computer side. This technique is based on the short-time and short-frequency Fourier transforms to create 2-D SOCT signals indexed by wavelength and depth in the object [37]. In combination with our PIC, it will allow obtaining more information about tissue samples [38].

Disclosures

The authors declare that they have no conflict of interest.

Acknowledgments

The research is supported by the Ministry of Science and Higher Education of the Russian Federation within the state assignment for UUST (agreement № 075-03-2024-123/1 dated 15.02.2024) and conducted in the research laboratory "Sensor systems based on integrated photonics devices" of the Eurasian Scientific and Educational Center.

References

1. J. Ong, A. Zarnegar, G. Corradetti, S. R. Singh, and J. Chhablani, "Advances in Optical Coherence Tomography Imaging Technology and Techniques for Choroidal and Retinal Disorders," *Journal of Clinical Medicine* 11(17), 5139 (2022).
2. R. Shen, L. K. Y. Chan, A. C. W. Yip, and P. P. Chan, "Applications of optical coherence tomography angiography in glaucoma: current status and future directions," *Frontiers in Medicine* 11, 1428850 (2024).
3. A. Murata, D. Wallace-Bradley, A. Tellez, C. Alviar, M. Aboodi, A. Sheehy, L. Coleman, L. Perkins, G. Nakazawa, G. Mintz, G. L. Kaluza, R. Virmani, and J. F. Granada, "Accuracy of Optical Coherence Tomography in the Evaluation of Neointimal Coverage After Stent Implantation," *JACC: Cardiovascular Imaging* 3(1), 76–84 (2010).
4. A. Karanasos, J. Lighthart, K. Witberg, G. Van Soest, N. Bruining, and E. Regar, "Optical Coherence Tomography: Potential Clinical Applications," *Current Cardiovascular Imaging Reports* 5(4), 206–220 (2012).
5. B. Wan, C. Ganier, X. Du-Harpur, N. Harun, F. M. Watt, R. Patalay, and M. D. Lynch, "Applications and future directions for optical coherence tomography in dermatology," *British Journal of Dermatology* 184(6), 1014–1022 (2021).
6. J. Olsen, L. Themstrup, N. De Carvalho, M. Mogensen, G. Pellacani, and G. B. E. Jemec, "Diagnostic accuracy of optical coherence tomography in actinic keratosis and basal cell carcinoma," *Photodiagnosis and Photodynamic Therapy* 16, 44–49 (2016).

7. G. Odorici, A. Losi, S. Ciardo, G. Pellacani, and A. Conti, “Non-invasive evaluation of Secukinumab efficacy in severe plaque psoriasis with confocal microscopy and optical coherence tomography: A case report,” *Skin Research and Technology* 24(1), 160–162 (2018).
8. P. Gong, S. Es’haghian, K.-A. Harms, A. Murray, S. Rea, B. F. Kennedy, F. M. Wood, D. D. Sampson, and R. A. McLaughlin, “Optical coherence tomography for longitudinal monitoring of vasculature in scars treated with laser fractionation,” *Journal of Biophotonics* 9(6), 626–636 (2016).
9. T. Kostanyan, G. Wollstein, and J. S. Schuman, “New developments in optical coherence tomography,” *Current Opinion in Ophthalmology* 26(2), 110–115 (2015).
10. E. A. Rank, A. Agneter, T. Schmoll, R. A. Leitgeb, and W. Drexler, “Miniaturizing optical coherence tomography,” *Translational Biophotonics* 4(1–2), e202100007 (2022).
11. L. Chang, N. Weiss, T. G. Van Leeuwen, M. Pollnau, R. M. De Ridder, K. Wörhoff, V. Subramaniam, and J. S. Kanger, “Chip based common-path optical coherence tomography system with an on-chip microlens and multi-reference suppression algorithm,” *Optics Express* 24(12), 12635 (2016).
12. R. M. Ruis, A. Leinse, R. Dekker, R. G. Heideman, T. G. Van Leeuwen, and D. J. Faber, “Decreasing the Size of a Spectral Domain Optical Coherence Tomography System With Cascaded Arrayed Waveguide Gratings in a Photonic Integrated Circuit,” *IEEE Journal of Selected Topics in Quantum Electronics* 25(1), 1–9 (2019).
13. J. Sancho-Durá, K. Zinoviev, J. Lloret-Soler, J. L. Rubio-Guivernau, E. Margallo-Balbás, and W. Drexler, “Handheld multi-modal imaging for point-of-care skin diagnosis based on akinetic integrated optics optical coherence tomography,” *Journal of Biophotonics* 11(10), e201800193 (2018).
14. Y. Huang, M. Badar, A. Nitkowski, A. Weinroth, N. Tansu, and C. Zhou, “Wide-field high-speed space-division multiplexing optical coherence tomography using an integrated photonic device,” *Biomedical Optics Express* 8(8), 3856 (2017).
15. E. A. Rank, S. Nevlacsil, P. Mueller, R. Hainberger, M. Salas, S. Gloor, M. Duell, M. Sagmeister, J. Kraft, R. A. Leitgeb, and W. Drexler, “In vivo human retinal swept source optical coherence tomography and angiography at 830 nm with a CMOS compatible photonic integrated circuit,” *Scientific Reports* 11(1), 21052 (2021).
16. I. V. Stepanov, E. A. Talynev, R. V. Kutluyarov, A. A. Ivanov, and E. P. Grakhova, “4-channel SS-OCT system design based on the silicon photonics chip with a high-scale integration,” in *Optical Technologies for Telecommunications 2022*, A. V. Bourdine, O. G. Morozov, and A. H. Sultanov (Eds.), *SPIE Proceedings* 12743 (2023).
17. X. Ji, X. Yao, Y. Gan, A. Mohanty, M. A. Tadayon, C. P. Hendon, and M. Lipson, “On-chip tunable photonic delay line,” *APL Photonics* 4(9), 090803 (2019).
18. E. A. Rank, R. Sentosa, D. J. Harper, M. Salas, A. Gaugutz, D. Seyringer, S. Nevlacsil, A. Maese-Novo, M. Eggeling, P. Mueller, R. Hainberger, M. Sagmeister, J. Kraft, R. A. Leitgeb, and W. Drexler, “Toward optical coherence tomography on a chip: in vivo three-dimensional human retinal imaging using photonic integrated circuit-based arrayed waveguide gratings,” *Light: Science & Applications* 10(1), 6 (2021).
19. S. Nevlacsil, P. Mueller, A. Maese-Novo, M. Eggeling, F. Vogelbacher, M. Sagmeister, J. Kraft, E. Rank, W. Drexler, and R. Hainberger, “Multi-channel swept source optical coherence tomography concept based on photonic integrated circuits,” *Optics Express* 28(22), 32468 (2020).
20. B. I. Akca, B. Považay, A. Alex, K. Wörhoff, R. M. De Ridder, W. Drexler, and M. Pollnau, “Miniature spectrometer and beam splitter for an optical coherence tomography on a silicon chip,” *Optics Express* 21(14), 16648 (2013).
21. J. W. Evans, R. J. Zawadzki, R. Liu, J. W. Chan, S. M. Lane, and J. S. Werner, “Optical coherence tomography and Raman spectroscopy of the ex-vivo retina,” *Journal of Biophotonics* 2(6–7), 398–406 (2009).
22. M. Tamošiūnas, O. Čiževskis, D. Viškere, M. Melderis, U. Rubins, and B. Cugmas, “Multimodal Approach of Optical Coherence Tomography and Raman Spectroscopy Can Improve Differentiating Benign and Malignant Skin Tumors in Animal Patients,” *Cancers* 14(12), 2820 (2022).
23. X. Zhang, J. Hu, R. W. Knighton, X.-R. Huang, C. A. Puliafito, and S. Jiao, “Dual-band spectral-domain optical coherence tomography for in vivo imaging the spectral contrasts of the retinal nerve fiber layer,” *Optics Express* 19(20), 19653 (2011).
24. S. Fan, L. Li, Q. Li, C. Dai, Q. Ren, S. Jiao, and C. Zhou, “Dual band dual focus optical coherence tomography for imaging the whole eye segment,” *Biomedical Optics Express* 6(7), 2481 (2015).
25. F. Spöler, S. Kray, P. Grychtol, B. Hermes, J. Bornemann, M. Först, and H. Kurz, “Simultaneous dual-band ultra-high resolution optical coherence tomography,” *Optics Express* 15(17), 10832 (2007).
26. M. Zhang, L. Ma, and P. Yu, “Dual-band Fourier domain optical coherence tomography with depth-related compensations,” *Biomedical Optics Express* 5(1), 167 (2014).
27. S. Chen, X. Shu, J. Yi, A. Fawzi, and H. F. Zhang, “Dual-band optical coherence tomography using a single supercontinuum laser source,” *Journal of Biomedical Optics* 21(6), 066013 (2016).
28. A. Davis, O. Levecq, H. Azimani, D. Siret, and A. Dubois, “Simultaneous dual-band line-field confocal optical coherence tomography: application to skin imaging,” *Biomedical Optics Express* 10(2), 694 (2019).
29. R. S. El Shamy, M. A. Swillam, and X. Li, “Comparative Study of Photonic Platforms and Devices for On-Chip Sensing,” *Photonics* 10(11), 1233 (2023).

30. L. B. Soldano, E. C. M. Pennings, "[Optical multi-mode interference devices based on self-imaging: principles and applications](#)," *Journal of Lightwave Technology* 13(4), 615–627 (1995).
31. F. Amanti, G. Andrini, F. Armani, F. Barbato, V. Bellani, V. Bonaiuto, S. Cammarata, M. Campostrini, T. H. Dao, F. De Matteis, V. Demontis, S. Donati, G. Di Giuseppe, S. Ditalia Tchernij, A. Fontana, J. Forneris, L. Frontini, R. Gunnella, S. Iadanza, A. E. Kaplan, C. Lacava, V. Liberali, L. Martini, F. Marzoni, L. Morescalchi, E. Pedreschi, P. Piergentili, D. Prete, V. Rigato, C. Roncolato, F. Rossella, M. Salvato, F. Sargeni, J. Shojaii, F. Spinella, A. Stabile, A. Toncelli, and V. Vitali, "[Integrated Photonic Passive Building Blocks on Silicon-on-Insulator Platform](#)," *Photonics* 11(6), 494 (2024).
32. I. V. Stepanov, E. A. Talynev, A. A. Ivanov, R. V. Kutluyarov, and E. P. Grakhova, "[Design of a Photonic Integrated Device with an on-Chip k-Clock and Tunable Reference Arm for Swept-Source Optical Coherence Tomography](#)," *Journal of Biomedical Photonics & Engineering* 9(3), 030317 (2023).
33. A. Sabeeh, V. Tuchin, "[Recent Advances in the Laser Radiation Transport through the Head Tissues of Humans and Animals – A Review](#)," *Journal of Biomedical Photonics & Engineering* 6(4), 040201 (2020).
34. G. F. R. Chen, J. R. Ong, T. Y. L. Ang, S. T. Lim, C. E. Png, and D. T. H. Tan, "[Broadband Silicon-On-Insulator directional couplers using a combination of straight and curved waveguide sections](#)," *Scientific Reports* 7(1), 7246 (2017).
35. Y. Fei, L. Zhang, T. Cao, Y. Cao, and S. Chen, "[Ultracompact polarization splitter–rotator based on an asymmetric directional coupler](#)," *Applied Optics* 51(34), 8257 (2012).
36. S. Kumar Bag, S. K. Varshney, "[Ultrawide FSR microring racetrack resonator with an integrated Fabry–Perot cavity for refractive index sensing](#)," *Journal of the Optical Society of America B* 38(5), 1669 (2021).
37. A. L. Oldenburg, C. Xu, and S. A. Boppart, "[Spectroscopic Optical Coherence Tomography and Microscopy](#)," *IEEE Journal of Selected Topics in Quantum Electronics* 13(6), 1629–1640 (2007).
38. M. Reddikumar, K. Bose, and R. Poddar, "[Implementation of the swept source optical coherence tomography \(SSOCT\) system with a statistical method to analyze optical properties of turbid media](#)," *Optik* 127(4), 1656–1659 (2016).

ADHD Subtype Diagnosis for Multiclass Imbalanced Augmented EEG Data

Vandana Joshi^{1,2*} and Nirali Nanavati²

¹ Gujarat Technological University, Ahmedabad, Gujarat, India

² Sarvajani College of Engineering & Technology, Surat, Gujarat, India

e-mail: vandana.joshi@scet.ac.in

Abstract. Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental condition that impacts an individual's emotional state, actions, and ability to acquire knowledge. The diagnosis is challenging due to the different characteristics of its distinct subtypes: inattention, hyperactivity, and impulsivity. ADHD has been diagnosed using a range of techniques, such as fMRI and behavioural evaluations. Electroencephalography (EEG) has garnered interest due to its capacity to uncover atypical brain activity patterns. Nevertheless, using EEG data in ADHD research poses difficulties due to the limited annotated datasets and substantial inter-individual variations. To address these difficulties, we have investigated various approaches to improve EEG data: Synthetic Minority Over-sampling Technique (SMOTE) variants, Non-deterministic Conditional Tabular Generative Adversarial Network (ND-CTGAN), and EEG-GAN variants. SMOTE variants have generated synthetic samples to address class imbalances, enhancing classification performance. ND-CTGAN has generated synthetic EEG data of superior quality, preserved inter-channel correlations and spectral properties, and improved the classification model's accuracy, recall, and precision. Statistical analyses have verified that synthetic data closely reflects the distributions and connections seen in real data. This similarity allows for strong comparisons and improves the ability of models. These findings not only enhance our understanding of ADHD but also inspire hope for future studies in the field, demonstrating the potential of advanced augmentation techniques to significantly enhance classification accuracy in EEG research. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: ADHD subtypes diagnosis; EEG data augmentation; SMOTE; CTGAN; EEG-GAN variants.

Paper #9203 received 6 Jan 2025; revised manuscript received 29 Jan 2025; accepted for publication 4 Feb 2025; published online 14 Mar 2025. [doi: 10.18287/JBPE25.11.010302](https://doi.org/10.18287/JBPE25.11.010302).

1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental condition characterized by patterns of inattention, hyperactivity, and impulsivity that interfere with functioning or development [1]. The inattentive subtype is defined by trouble maintaining focus and frequent careless mistakes. Restlessness, impulsive behaviour, and fidgeting characterize the hyperactive-impulsive subtype. The mixed subtype

combines symptoms of both the hyperactive-impulsive and inattentive subtypes [1].

In recent years, various methods have been developed for the diagnosis of ADHD by using different modalities. For instance, functional magnetic resonance imaging (fMRI) [2], diagnostic and statistical manual (DSM) [3, 4], behavioural rating, i.e., child behavior checklist (CBCL), neuropsychological test with integrated visual and auditory (IVA), continuous performance test (CPT) [5]. The subtypes of ADHD present a significant challenge in clinical settings due to their pronounced

behavioural manifestations and associated neurophysiological patterns. Electroencephalography (EEG) has emerged as a valuable tool for studying these neurophysiological markers, offering insights into the atypical brain activity patterns of individuals with ADHD [6–9].

Despite its potential, the effective utilization of EEG data in ADHD research faces critical challenges. One significant issue is the limited availability of labeled EEG dataset [9, 10], which impedes the development of robust machine-learning models for diagnosis and prediction. The insufficiency of data is impaired by the variable nature of EEG signals, which require substantial training data to capture the complex dynamics associated with ADHD subtypes accurately. To address this limitation, various data augmentation techniques have been proposed as a solution to artificially expand EEG datasets, thereby enhancing the diversity and representativeness of training samples [10, 11].

The limitations associated with EEG augmentation include challenges in evaluation metrics, such as Inception Score (IS) and Fréchet Inception Distance (FID), which are prone to overfitting issues. Furthermore, Wasserstein distance (WD) introduces additional complexity, while Euclidean distance (ED) may deliver results that lack precision. It is also worth noting that the adversarial training process inherent in GANs is complex and time-consuming. Additionally, EEG augmentation is extensively utilized across various applications, including Brain-Computer Interfaces (BCI), motor imagery, the study of epilepsy, and emotion recognition. However, its potential application for ADHD subtypes EEG data has not been thoroughly investigated [11].

The usefulness of EEG applications is also constrained by significant inter-subject variability. This variability is primarily attributed to physiological distinctions among individuals, which can significantly impact the efficiency of models designed for generalization across different subjects. Generalizing from one group of individuals to another is crucial for various practical uses of EEG technology, prompting significant research efforts toward developing strategies capable of managing inter-subject variability [12]. Traditional EEG augmentation methods frequently depend on EEG image data, which can result in challenges related to the curse of dimensionality. Additionally, these methods often struggle to preserve inter-channel correlations effectively. Understanding the correlation between EEG channels is valuable in ADHD analysis [13].

Another key challenge in EEG-based ADHD research is the issue of class imbalance [14, 15]. In clinical datasets, the proportion of samples from individuals with ADHD often significantly differs from those without the disorder. This imbalance leads to biased model training, where the learning algorithms are predisposed to the majority class, resulting in poor generalization and reduced diagnostic accuracy for the minority class. Addressing class imbalance is crucial to ensure that models are equally adept at identifying both positive and

negative instances of ADHD, thereby improving their reliability and clinical applicability.

This manuscript explores advanced techniques for oversampling and augmenting data designed explicitly for EEG datasets to address limited data and imbalanced class distribution issues. The objective is to enhance the variety of training data and rectify class imbalances to improve the efficacy of machine learning models in identifying ADHD subtypes from EEG signals.

This novel approach contributes to the methodological progress in EEG analysis and shows potential for more precise and fair diagnostic procedures in ADHD research. Traditional classification methods often struggle to capture the intricate, multidimensional nature of ADHD, making it challenging to identify subtypes precisely. In this context, SMOTE and Generative Adversarial Networks (GANs) have emerged as potential tools for augmenting data and generating synthetic data, thereby improving the capability of machine learning models to handle diverse and sparse datasets.

We propose a Non-Deterministic Conditional Tabular GAN (ND-CTGAN) to improve ADHD subtype classification. By generating diverse and representative synthetic data, this approach enhances the accuracy and robustness of machine learning models in handling ADHD variability, showing promising results. We also compared our method with the EEG-GAN variant (TTS-WGAN-GP) [16], and it showed better results than it.

Our manuscript is organized into four main sections, the first of which is the introduction. The second section initiates the discussion by outlining the theoretical framework of augmentation methods. The third section delves into the methodology employed in conducting the study. The fourth section details the research's outcomes and findings, followed by the conclusion and future work.

2 Related Work

This section reviews various methodologies for generating synthetic EEG data, focusing on three principal approaches: Synthetic Minority Over-sampling Technique (SMOTE), Generative Adversarial Networks (GANs), and traditional signal processing techniques.

2.1 SMOTE Variants for EEG Data Augmentation

Traditional EEG datasets often suffer from imbalances and limited sample sizes, challenging achieving robust classification performance. To mitigate these issues, various Synthetic Minority Over-sampling Technique (SMOTE) variants have been developed. By interpolating between existing instances of the minority class, SMOTE creates synthetic samples that successfully increase the representation of the underrepresented and improve the robustness of machine learning models [17]. This technique can be helpful in ADHD studies where data might be skewed due to a smaller number of diagnosed cases compared to control

groups. In ADHD, research has been applied SMOTE to various types of ADHD datasets, including phenotypic information [18], electronic medical records (EMR) [19], FMRI [20], and EEG data [21]. This involves generating synthetic samples to balance the dataset, enhancing the machine learning model's ability to detect ADHD accurately by leveraging behavioral and neurological features. SMOTE variants have shown promise in improving training datasets in EEG data analysis in various applications. The study [22] highlights the effectiveness of the adaptive synthetic sampling (ADASYN) SMOTE variant in handling multi-class imbalances in EEG data, making it a viable technique for improving the performance of EEG signal classification systems to detect epilepsy events in the EEG signals.

In Ref. [23] the researcher introduced an effective fusion algorithm that significantly improves seizure detection in imbalanced EEG datasets. By integrating SMOTE and TomekLink (STL) synthetic sampling methods with ensemble learning, the algorithm enhances the accuracy and reliability of seizure classification. Moreover, a novel data augmentation technique, BNNSMOTE [24], generates synthetic samples by focusing on the boundary and neighborhood properties of the minority class (seizure events). It aims to improve the distribution of synthetic samples to represent the minority class better and enhance classifier training. Likewise, the Borderline-SMOTE and Radius-SMOTE [25] algorithm demonstrates their edge in efficient, emotional identification of EEG signals [26]. While Borderline-SMOTE [27], support vector machine (SVM) SMOTE, and adaptive synthetic sampling [28] were used to increase the number of samples for the class of target stimuli in P300-based brain-computer interfaces (BCIs), whereas safe-level SMOTE has enhanced classification performance for different anxiety levels from the DASPS dataset.

2.2 GAN Approaches for EEG Data Augmentation

Generative Adversarial Networks (GANs) have become a cornerstone in advanced EEG data augmentation, enabling the generation of realistic synthetic data that mirrors the complexity of real EEG signals. Variants of GAN, for instance, the Deep Convolutional GAN (DCGAN) applied on spectrogram-generated images from STFT (Short-Time Fourier Transform) of EEG, with transfer learning models, have generated high-fidelity EEG signals by minimizing divergence between real and synthetic data distributions for epileptic seizure prediction [29]. Auxiliary Classifier GAN (AC-GAN) also demonstrated that generated synthetic spikes samples improve machine learning classification tasks in the diagnosis of epilepsy [30]. Conditional Wasserstein GAN (CWGAN) enhances the performance of EEG-based emotion recognition models by conditioning the generation process on specific labels, producing more relevant synthetic data [31, 32]. Balanced Wasserstein

GAN with gradient penalty (BWGAN-GP) [33] and Common Spatial GAN (CS-GAN) [34] address specific needs in EEG data, such as balancing datasets and generating spatially consistent signals, respectively, in BCI applications. Likewise, Event-Related Potentials GAN (ERP-GAN) effectively improves the stability of the model for single-trial P300 detection [35].

Time Series GAN [36], is designed to handle temporal dependencies and generate synthetic EEG sequences that capture the dynamic nature of neural activity. These GAN-based approaches significantly enrich EEG datasets, providing diverse and high-quality synthetic data that enhances model training and performance in Seizure detection [37]. Another approach in Ref. [16] the study derives a hybrid architecture, TTS-WGAN-GP, that contains a transformer-encoder-based generator and discriminator for time series (TTS) and conditional GANs. This architecture focuses on temporal features, conditional data, and stability while generating EEG synthetic data.

2.3 Other Methods for EEG Data Augmentation

Besides SMOTE and GAN-based techniques, other EEG data augmentation methods are crucial for improving the diversity and resilience of EEG training datasets. Time domain augmentations, like smooth time masking and time reversal, enhance temporal variability. Frequency domain augmentations, such as frequency shifting and band-stop filtering, can improve training data by altering spectral properties using various materials. The spatial variety of EEG data is improved via spatial domain augmentations such as channel symmetry, channel dropout, and channel shuffling [10]. A comparative research study is presented for motor imagery EEG data in Ref. [38]. These include simple approaches such as averaging, time slice recombination, and cropping, as well as more sophisticated ones such as variational autoencoder (VAE).

3 Materials and Methods

In addressing EEG data augmentation challenges, we aim to generate synthetic EEG data specifically for the Hyperactive ADHD subtype. A major issue with existing methodologies is class imbalance, which can lead to biased outcomes and hamper the accuracy of classification models. Traditional approaches often rely on EEG image data, which leads to the curse of dimensionality and fails to maintain inter-channel correlations, thus diminishing the model's performance. Moreover, image-based techniques do not adequately capture subject-wise variability despite their effectiveness, limiting their applicability across diverse subjects. By focusing on maintaining spectral characteristics and enhancing the inter-channel correlations of synthetic EEG data, we seek to improve classification accuracy while enabling a more comprehensive exploration of the feature space.

The following subsections outline this study's materials, experimental design, and methodological approaches. We employed two methods for data augmentation and classification. Firstly, we compared augmented data using various Synthetic Minority Over-sampling Technique (SMOTE) variants. Secondly, we introduced a novel approach, the non-deterministic Conditional Tabular Generative Adversarial Network (ND-CTGAN), to improve classification performance.

3.1 Dataset

In our study, we examined EEG data across different ADHD subtypes [15], including 17 subjects with inattentive ADHD, 17 with combined ADHD, and 3 with hyperactive ADHD. The EEG recordings were captured using 64 channels, initially sampled at 2048 Hz and subsequently down-sampled to 256 Hz to facilitate efficient processing. To ensure data quality, a bandpass filter was applied to remove slow drifts below 1 Hz and high-frequency noise above 20 Hz, including line noise. The recordings were made while subjects were working and responding to auditory stimuli. The tasks involved listening to human speech streams composed of /ba/, /da/, and /ga/ syllables. Each trial had a central "Target" stream with three syllables and a "Distractor" stream with

five syllables to the right. This dataset is available openly; the link is in the data availability statement.

3.2 Data Augmentation and Classification with SMOTE Variants

To address the issue of class imbalance in our EEG dataset, we applied data augmentation using various Synthetic Minority Over-sampling Technique (SMOTE) variants. These variants included SMOTE, Borderline-SMOTE, ADASYN, SMOTE and Tomek, and SVM-SMOTE. Each is designed to generate synthetic samples for the hyperactive ADHD minority class. The augmented data were then subjected to classification using a range of machine learning algorithms. This comparative analysis aimed to evaluate the effectiveness of different SMOTE variants in enhancing classification performance by creating a more balanced dataset.

Fig. 1 shows the flow diagram of data augmentation and classification with SMOTE variants. After preprocessing raw EEG data, six non-linear features are extracted: approximate entropy, singular value decomposition entropy (SVD), spectral entropy, Petrosian Fractal dimension (FD), Katz FD, and Higuchi FD. In our previous work [38], these features yield promising results for EEG analysis.

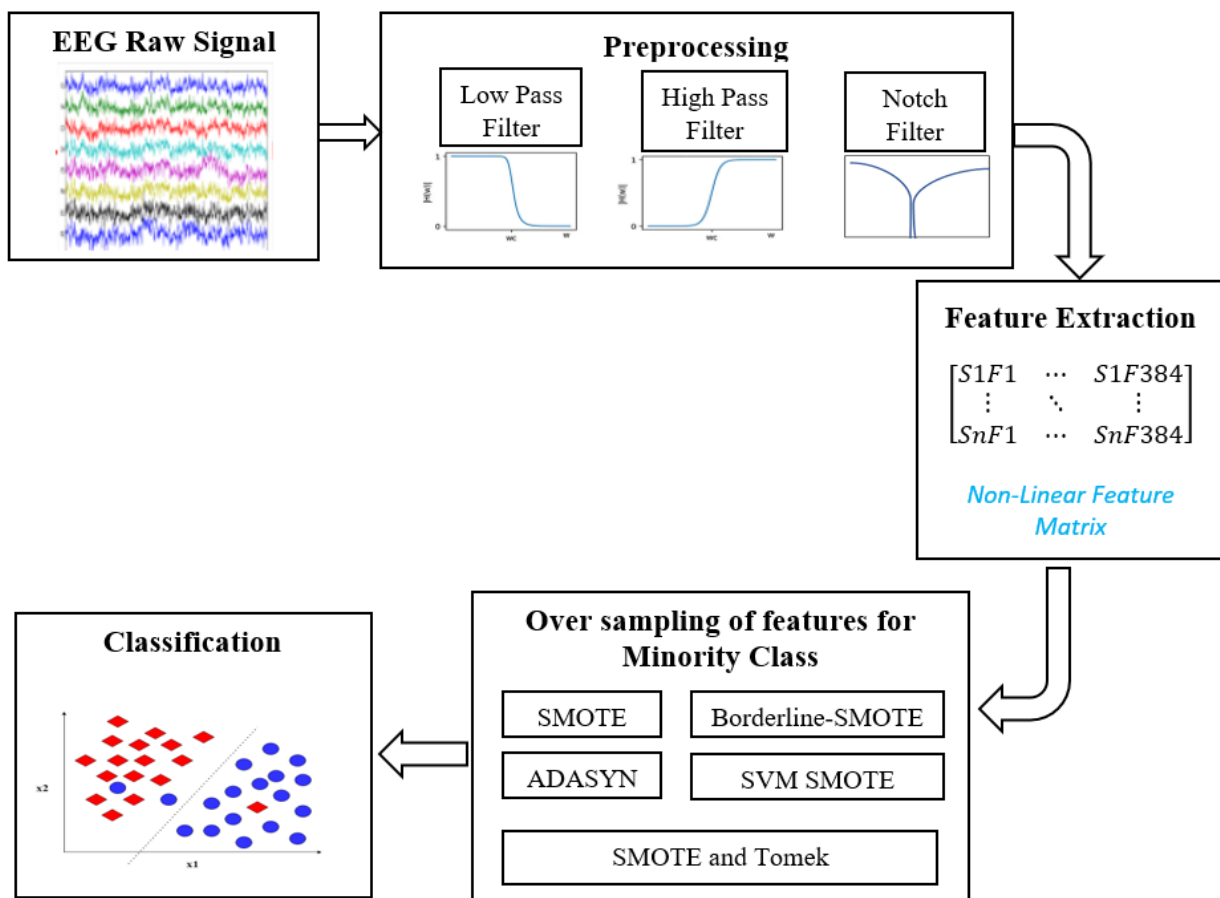


Fig. 1 Block diagram of data augmentation and classification with SMOTE variants.

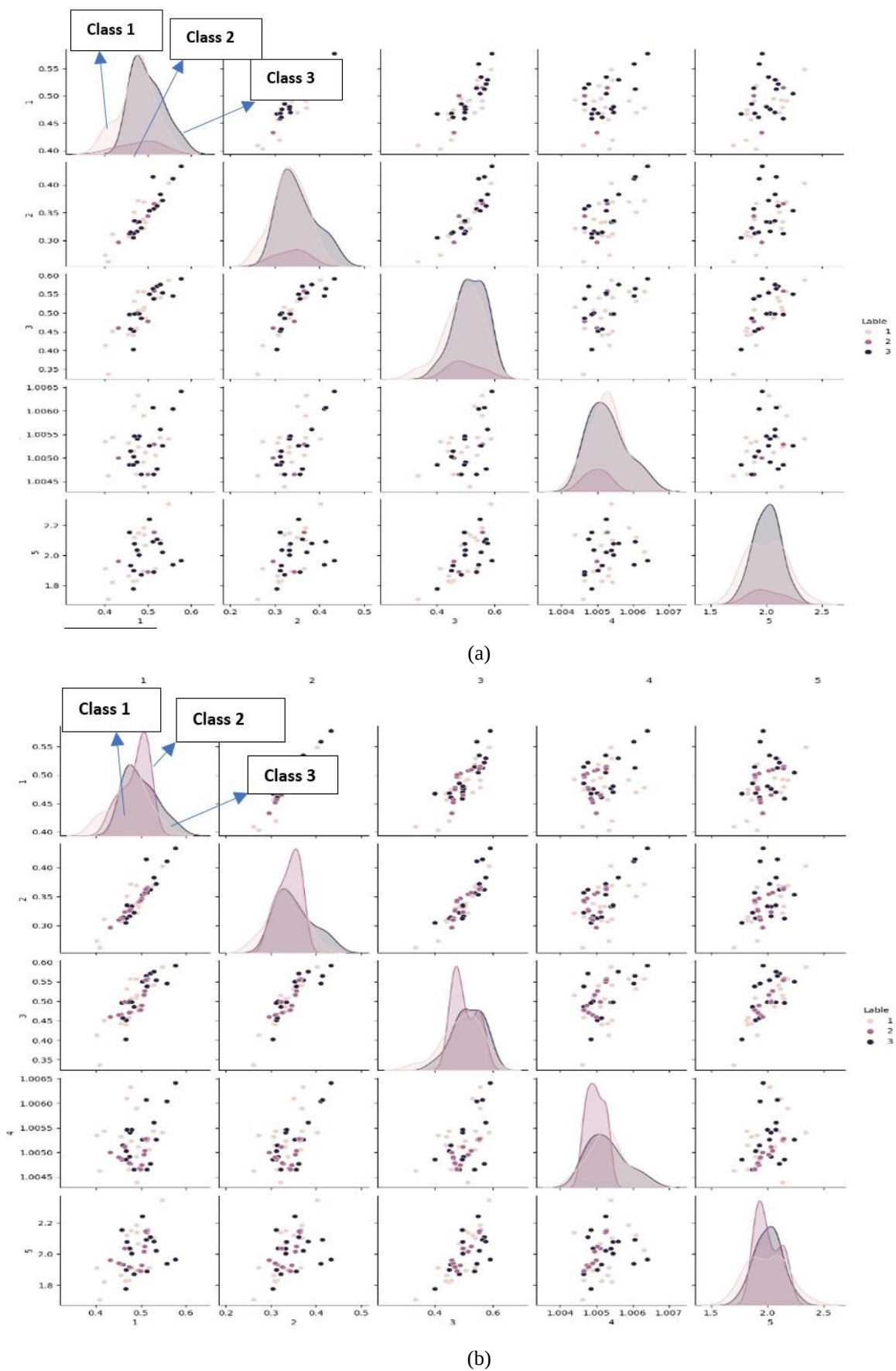


Fig. 2 Pair plot for first five features of class 1, 2, and 3 (on the X-axis and Y-axis): (a) before ADASYN, (b) after ADASYN.

A total of 37 EEG samples were initially available for oversampling, including 17 from the combined ADHD type, three from the inattentive ADHD type, and three from the hyperactive ADHD type. The dataset was balanced using oversampling techniques, resulting in an equal distribution of 51 samples across all ADHD types. These samples were analyzed using various classification methods. Fig. 2 presents a pair plot for the first five features of the feature vector, both before and after applying ADASYN. This plot illustrates the oversampling of hyperactive samples and demonstrates that the synthetic data maintains a distribution similar to the real data.

Further, five SMOTE variants including standard SMOTE, Borderline, ADASYN, SMOTE SVM, and SMOTE Tomek, were proposed by us for multiclass classification. Three classes were considered for classification: Combined (Class 1), Hyperactive (Class 2), and Inattentive (Class 3). Seven different classifier models have been applied and compared with each other. Random Forest, MLP (Multilayer Perceptron), SVM (Support Vector Machine), KNN (K-Nearest Neighbours), Logistic Regression, Decision Tree (DT), and Gaussian Naive Bayes (GaussianNB), before and after the application of several data augmentation techniques. The models were evaluated on their classification accuracy on the original dataset (indicated as “before Augmentation”) and after applying different augmentation methods. Table 1 illustrates the performance of all classifiers. The ten-cross validation (10-CV) method was used for training and testing. In the SVM and logistic regression method, three different classes were applied to the one-vs-rest (OVR) approach to evaluate the performance. The results indicate that data augmentation generally leads to an improvement in

model accuracy. For instance, Logistic Regression achieves its highest accuracy of 0.70 when Borderline SMOTE is applied.

We found that augmented EEG data using SMOTE variants cannot trace which synthetic data points are generated from which subject data, making it challenging to directly compare real and generated data for each subject. To further improve the classification of EEG data, we introduced a nondeterministic Conditional Tabular Generative Adversarial Network (CTGAN) for data augmentation (Explained in the following section).

3.3 Data Augmentation and Classification with ND-CTGAN

Unlike traditional methods, this approach generates synthetic EEG data that captures complex, non-linear relationships and retains essential inter-channel correlations and spectral characteristics. By augmenting our dataset with this advanced technique, we aimed to enhance the model’s generalization capability and robustness. In contrast, the nondeterministic CTGAN allows for the generation of data that preserves individual subject characteristics, facilitating more meaningful comparisons between real and synthetic datasets. The detailed algorithm is illustrated as Algorithm 1.

This method was applied to pre-processed EEG data of individual subjects and then used for feature extraction with the real data. The synthetic data produced by the nondeterministic CTGAN was subsequently used to extract the nonlinear features and to train various machine learning classifiers. Six nonlinear features are extracted, as stated in the 3.2 Section. Fig. 3 illustrates all the phases of the processes.

Algorithm 1 Generation of synthetic samples using CTGAN for hyperactive subject data.

Input:

- $O_{subject}$: Original EEG data for the hyperactive subject.
- $N_{desired}$: The desired number of synthetic samples

Output:

- $S_{synthetic}$: Synthetic EEG samples for the hyperactive subject.

Procedure:

- a) Load the original EEG dataset $O_{subject}$ for the hyperactive subject.
- b) Set the desired number of synthetic samples $N_{desired}$.
- c) Utilize the CTGAN model to generate an initial set of synthetic samples $S_{initial}$ based on $D_{subject}$.
- d) Compute the number of synthetic samples $N_{generated}$ in $S_{initial}$.
- e) While $N_{generated} \neq N_{desired}$:
 - i. apply CTGAN to generate additional synthetic samples $S_{additional}$;
 - ii. Combine $S_{initial}$ with $S_{additional}$ to update the synthetic dataset $S_{initial}$;
 - iii. Update $N_{generated}$ with the new count of samples in $S_{initial}$.
- f) Collect synthetic samples: Once $N_{generated} = N_{desired}$, assign $S_{synthetic} = S_{initial}$.
- g) Return the synthetic EEG samples $S_{synthetic}$ for the hyperactive subject.

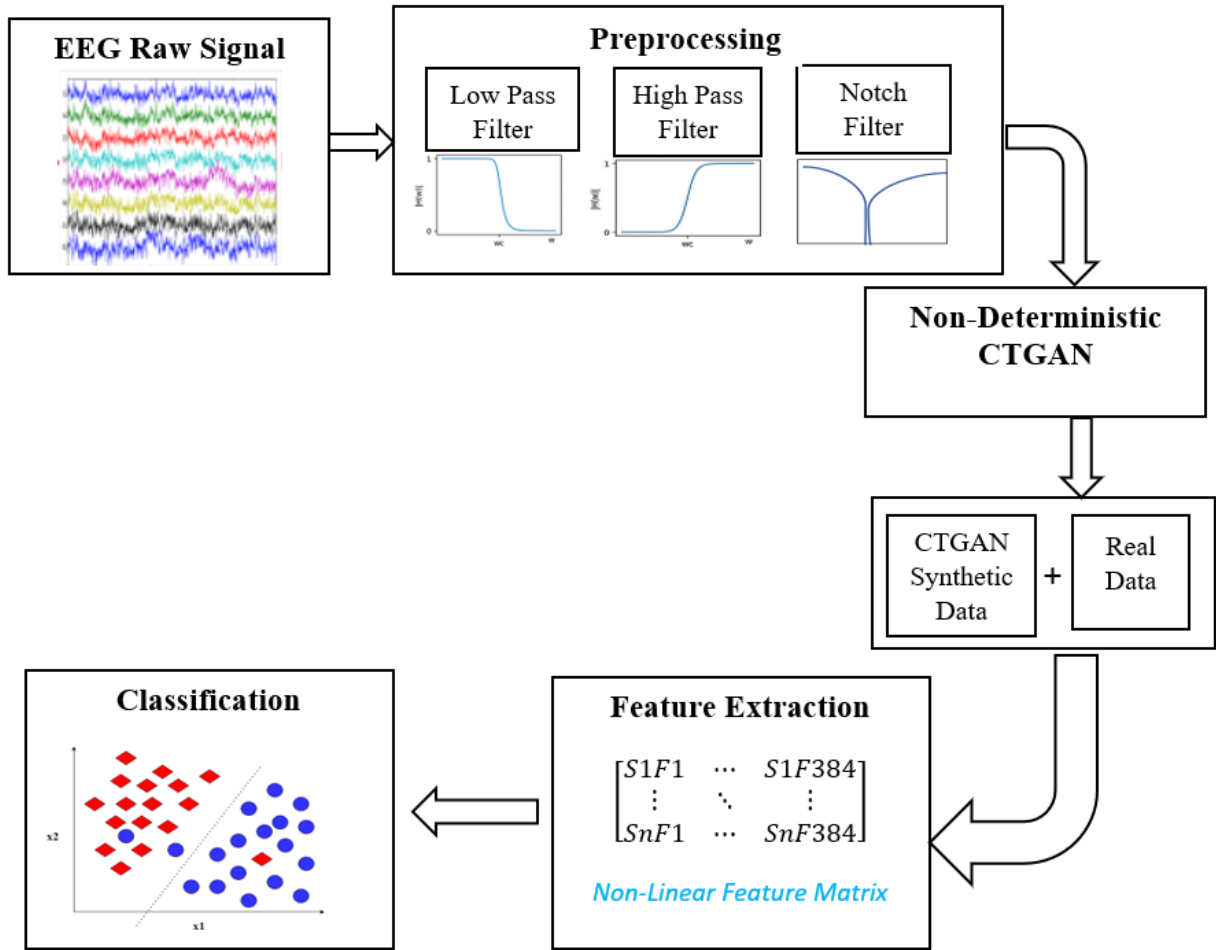


Fig. 3 Block diagram data augmentation and classification with ND-CTGAN.

CTGAN Architecture

CTGAN is a GAN-based method for modeling tabular data distribution and sampling rows. Fig. 4 illustrates the architecture. The mode-specific normalization $N(0, 1)$ is used for a continuous variable with floating point values. CTGAN employs a variational Gaussian mixture model [39]. As a unit distribution model, it estimates the number of distributions for each variable and normalizes the variable values based on the estimated distributions. During training, CTGAN uses these encoded values instead of the original values. After training, CTGAN transforms the generated data back to the original scale when generating synthetic data.

For each continuous column C_i , use the Variational Gaussian Mixture Model (VGM) to estimate the number of modes m_i and fit a Gaussian mixture. The learned Gaussian mixture is

$$PC_i(c_{i,j}) = \sum_{k=1}^3 \mu_k N(c_{i,j}; \eta_k, \Phi_k), \quad (1)$$

where μ_k and Φ_k are the weight and standard deviation of a mode, respectively.

The CTGAN model's training process is comprehensive. It does not depend on the local structure within individual columns; instead, it uses fully

connected networks in both the generator and discriminator to capture all possible correlations between columns. With its two completely connected hidden layers, the generator employs batch normalization and the ReLU activation function. The synthetic row representation blends activation functions, generating scalar values with the hyperbolic tangent function ($\tanh$) and discrete values with Gumbel Softmax. Using the leaky rectified linear unit (ReLU) function and applying dropout to every hidden layer, the discriminator operates within the PacGAN framework, using 10 samples in each pac to avoid mode collapse effectively. Additionally, the model is trained via the Wasserstein GAN (WGAN) loss function, which incorporates a gradient penalty.

The training process employs an Adam optimizer with a learning rate of $2 \cdot 10^{-4}$. Researchers in Ref. [40] It has been proven that synthetic data from the CTGAN is more suitable for EEG data. We used pre-processed raw tabular data with 1,792 samples and 64 channel characteristics per subject. We used the CTGAN algorithm to generate synthetic data for hyperactive subjects. The steps in this procedure are outlined in the following algorithm.

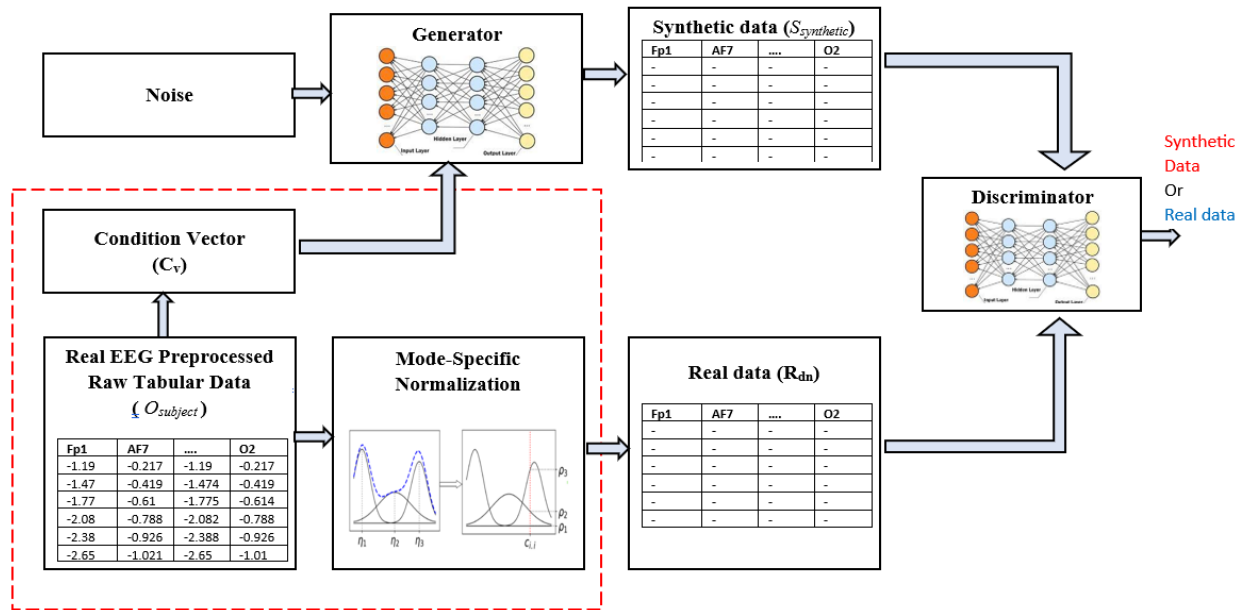


Fig. 4 The architecture of conditional tabular GAN (CTGAN).

Evaluation of Synthetic Data

Several statistical and quality metrics are used to evaluate the performance of our proposed CTGAN before using the synthetic data for classification tasks. These metrics help ensure that the generated data closely resembles the real data in terms of distribution and relationships between variables. The statistical and quality metrics comprise mean and standard deviation (STD), cumulative distribution function (CDF), quality score: Kolmogorov-Smirnov statistic (KS statistic), and correlation similarity. The CDF represents the likelihood that a random variable will have a value less than or equal to a given value. Comparing the CDFs of each characteristic in the synthetic and actual data may determine how well the distributions match throughout the whole range of values.

The KS statistic measures the maximum difference between the empirical CDFs of the real and synthetic data for each feature. It provides a non-parametric test for comparing distributions without making assumptions about their exact form. The KS statistic can be used to evaluate the similarity between actual and synthetic data distributions. A smaller KS statistic suggests a closer match between the two distributions. Correlation similarity assesses how the associations between pairs of attributes in the synthetic data look like those in the real data. It ensures that the synthetic data maintains the interconnections and associations among variables, which is essential for upholding the integrity of the fundamental structure of the data.

A total of 1793×67 samples were utilized for the KS Complement analysis. Based on the sample size, the critical values for the Kolmogorov-Smirnov test were

determined as 0.169 for a significance level of $\alpha = 0.05$ and 0.203 for $\alpha = 0.01$. The observed average KS values were recorded at 0.220 (1–0.78) and 0.31 (1–0.69). Given that both 0.220 and 0.31 exceed 0.203, the results indicate statistical significance at the 5% and 1% levels.

Analysis of Synthetic Data

We have analyzed the EEG synthetic data of ND-CTGAN and EEG-GAN variants. The visualizations in Fig. 5 show similarities between the synthetic and actual data. The mean aligns well, and the standard deviation shows a small change for ND-CTGAN, but it shows a difference for EEG-GAN variants.

Fig. 6 shows the distribution of channels for real and synthetic data of channels like FP1, PC1, and PC4. This indicates that the distribution of synthetic data is well matched with the real for ND-CTGAN compared to EEG-GAN. Likewise, by considering the cumulative sums per channel, as in Fig. 7, most of the channel data in the synthetic data match closely with the actual data for ND-CTGAN but not for EEG-GAN.

Fig. 8 illustrates the “Column Shapes” property of the quality score for each EEG channel computed using the KS-metric. The average KSComplement values obtained were 0.78 for ND-CTGAN and 0.69 for EEG GAN. In the SDMetrics library, the KSComplement is derived by inverting the KS statistic using the formula “ $1 - (\text{KS statistic})$ ”. A higher score signifies greater data quality, with a score of 1.0 representing the optimal scenario in which the synthetic data is indistinguishable from the real data, while a score of 0.0 indicates complete dissimilarity between the two. In our analysis, a score of 0.7 approaches 1.0, suggesting a reasonable similarity between the real and synthetic data.

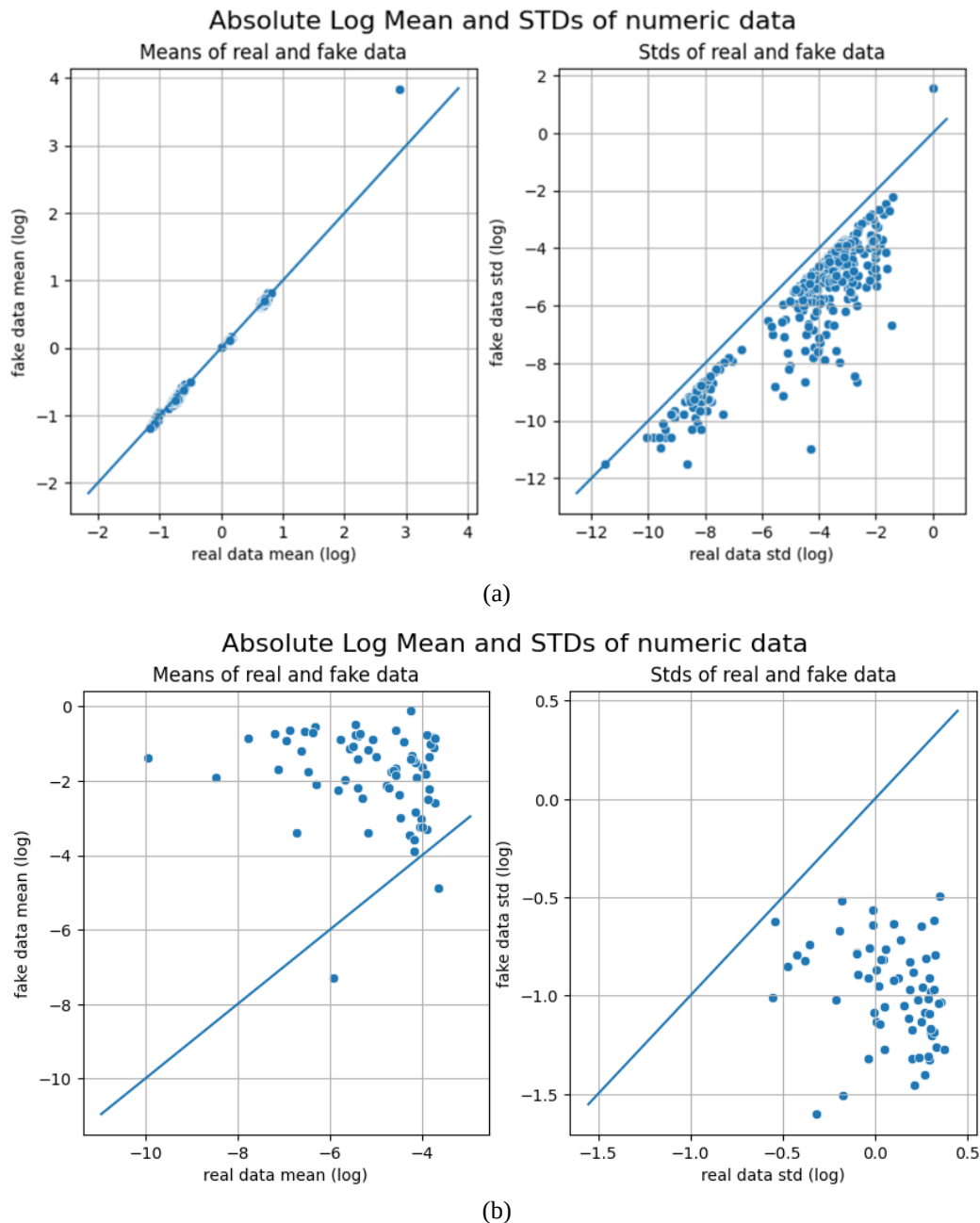


Fig. 5 Mean and STD (Real vs Synthetic) of (a) ND-CTGAN, (b) EEG-GAN.

The correlation similarity plots of samples in Fig. 9 illustrate a comparison of 64 channel properties using both real and generated data. This analysis helps us understand the relationships between channels in real and synthetic datasets. For instance, the similarity score between the FT8 and PO3 pair is 0.95. The correlation coefficient between the Fz and O1 channels is 0.38 for real data and 0.3 for synthetic data, suggesting a strong resemblance. These findings indicate that the synthetic data generated possess qualities that closely resemble real data.

Performance Evaluation of Classifiers

Following a thorough evaluation of synthetic data generated by ND-CTGAN and EEG-GAN, both

generated and real datasets are employed for classification purposes. We examine seven distinct classifiers: Random Forest, Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Logistic Regression, Decision Trees (DT), and Gaussian Naive Bayes (NB), which are widely recognized in ADHD classification research [41]. The classifiers are assessed using key performance metrics, including accuracy, recall, precision, and the Receiver Operating Characteristic (ROC) curve. To ensure a reliable evaluation of the models, we implemented a cross-validation strategy with 10 folds utilizing stratified K-Fold. This approach enables a robust assessment of model performance by testing it on various subsets of data, resulting in an averaged measure of effectiveness across multiple data splits.

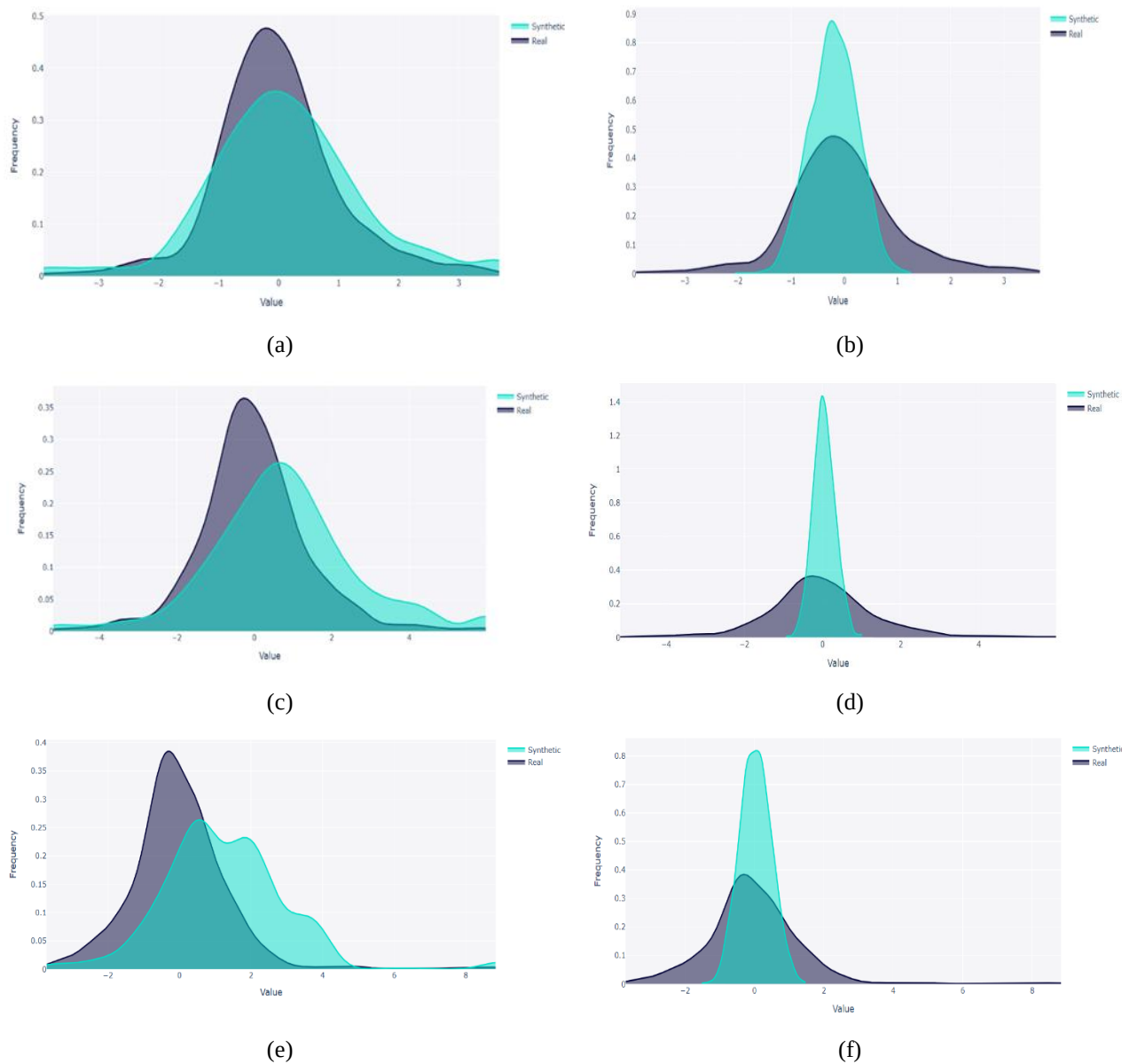
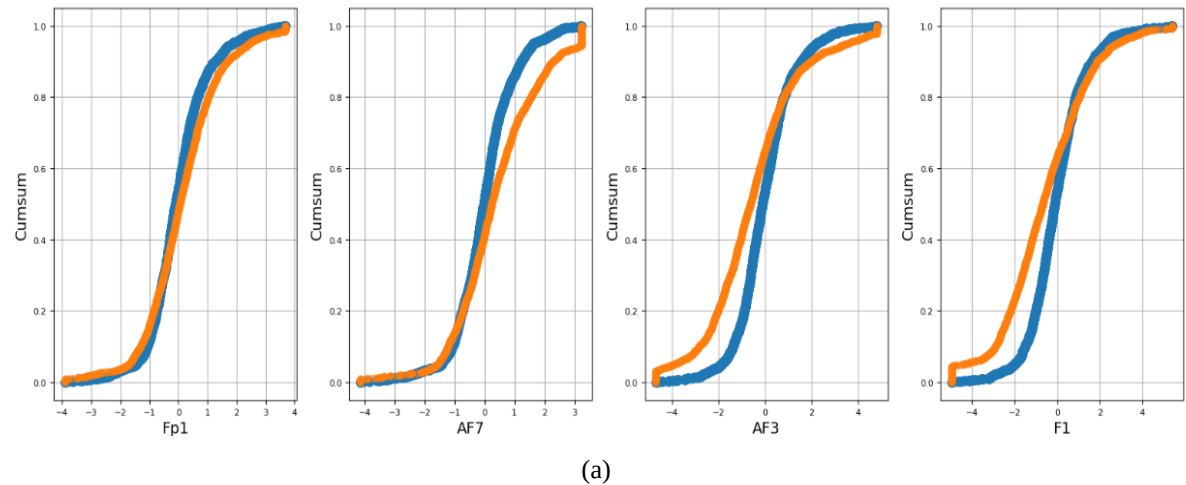


Fig. 6 Comparison of the distribution of channel data (Real vs Synthetic) between (a, c, e) ND-CTGAN and (b, d, f) EEG-GAN: FP1 channel of (a) ND-CTGAN and (b) EEG-GAN, FC1 channel of (c) ND-CTGAN and (d) EEG-GAN, PO4 channel of (e) ND-CTGAN and (f) EEG-GAN.



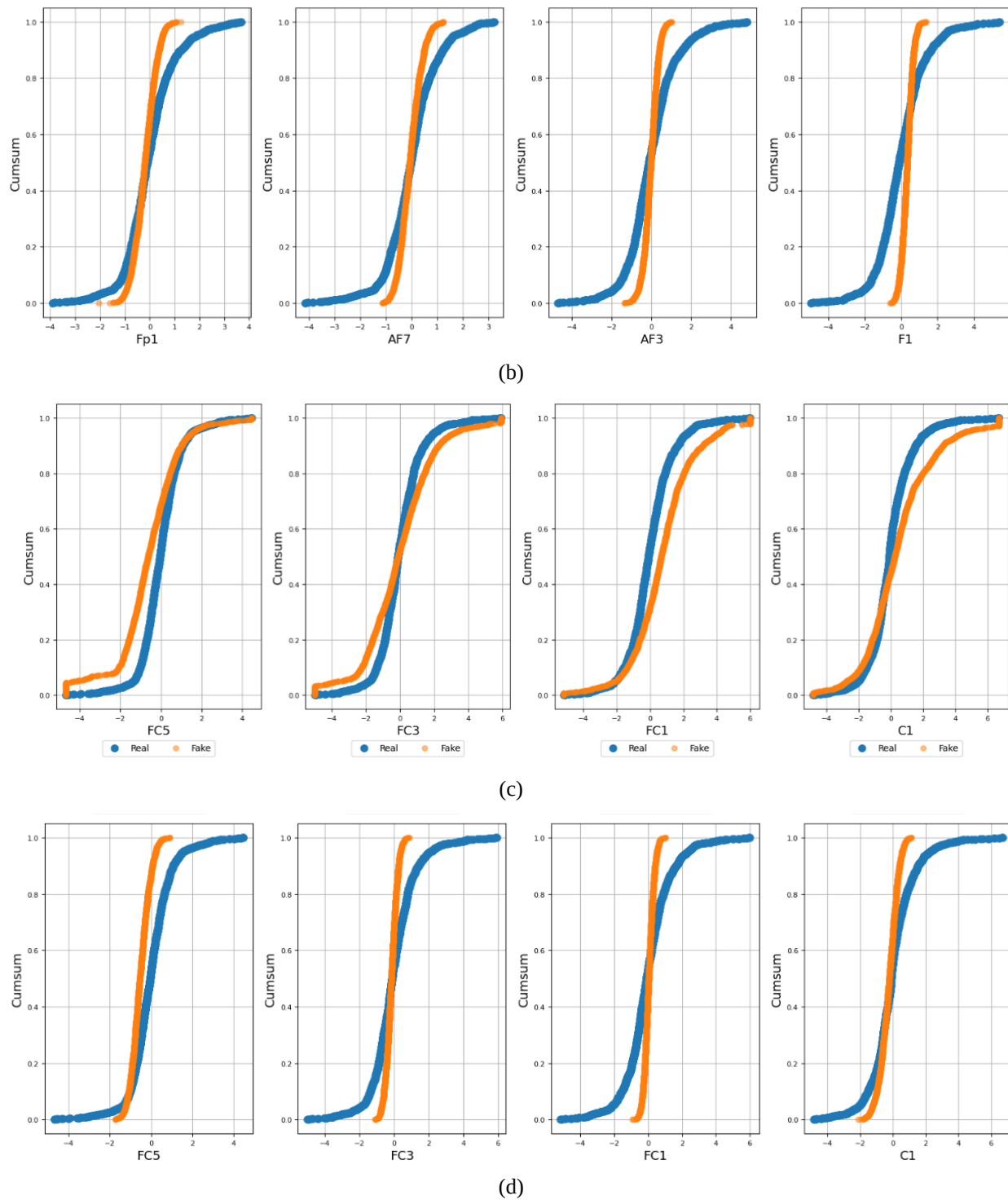


Fig. 7 Comparison of cumulative sum per feature of channel data (Real (blue) vs. Synthetic (red)) between (a, c) ND-CTGAN and (b, d) EEG-GAN.

Table 1 shows the comparative analysis of classification accuracy for the SMOTE variants and GAN methods. In SMOTE, augmentation was applied to the features matrix, whereas EEG-GAN and ND-CTGAN were applied to pre-processed data. We tested the performance of seven classifier models, and the result shows a 16% to 18% improvement in accuracy after

augmentation. SMOTE borderline and ADASYN perform well among other SMOTE variants. Likewise, MLP reaches its peak accuracy of 0.69 with ND-CTGAN pre-processed data. Random Forest, SVM, KNN, and DT models also exhibit their best performance after using ND-CTGAN on pre-processed data compared to EEG-GAN.

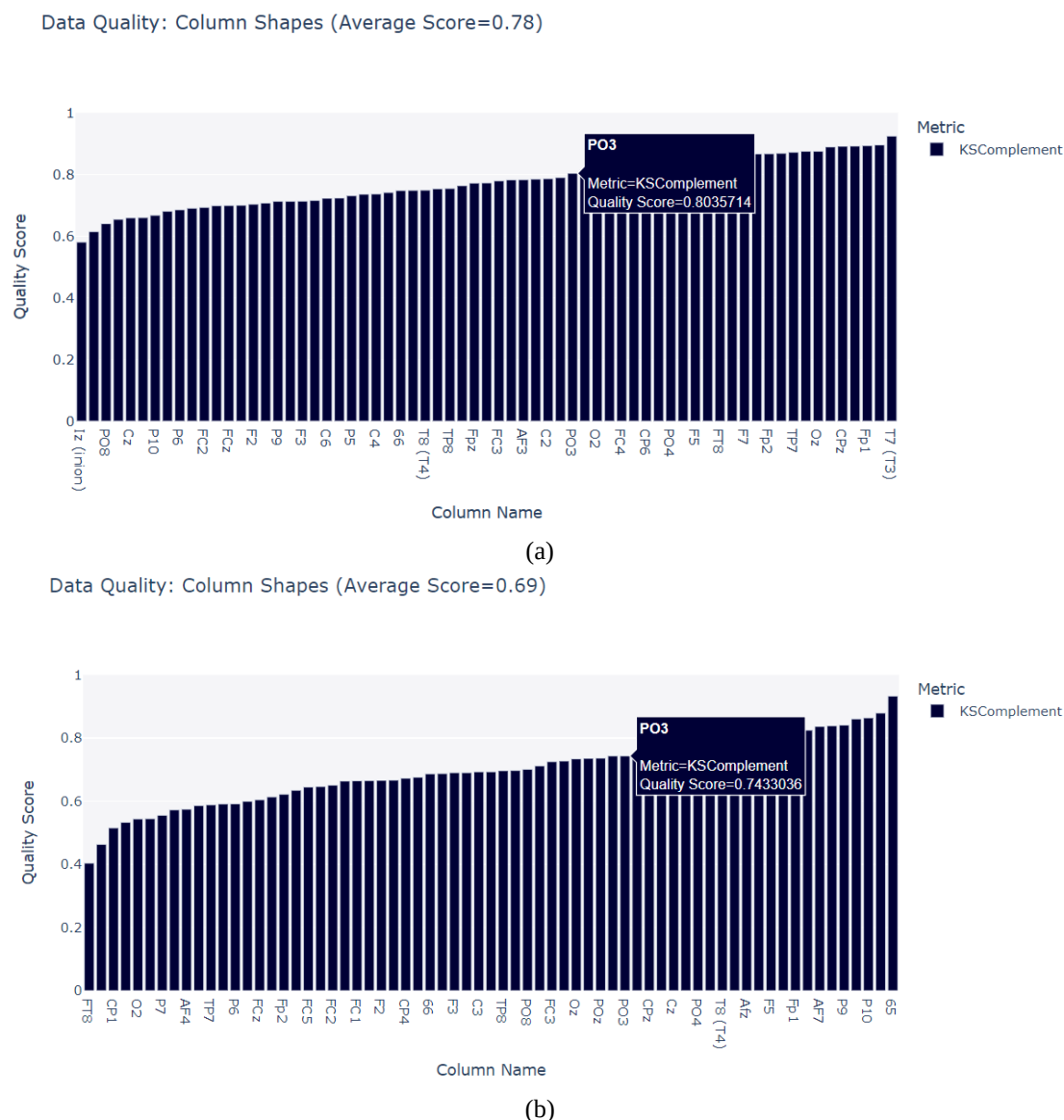


Fig. 8 Quality Score of all channels using the KSComplement metric of synthetic data:
(a) ND-CTGAN, (b) EEG-GAN.

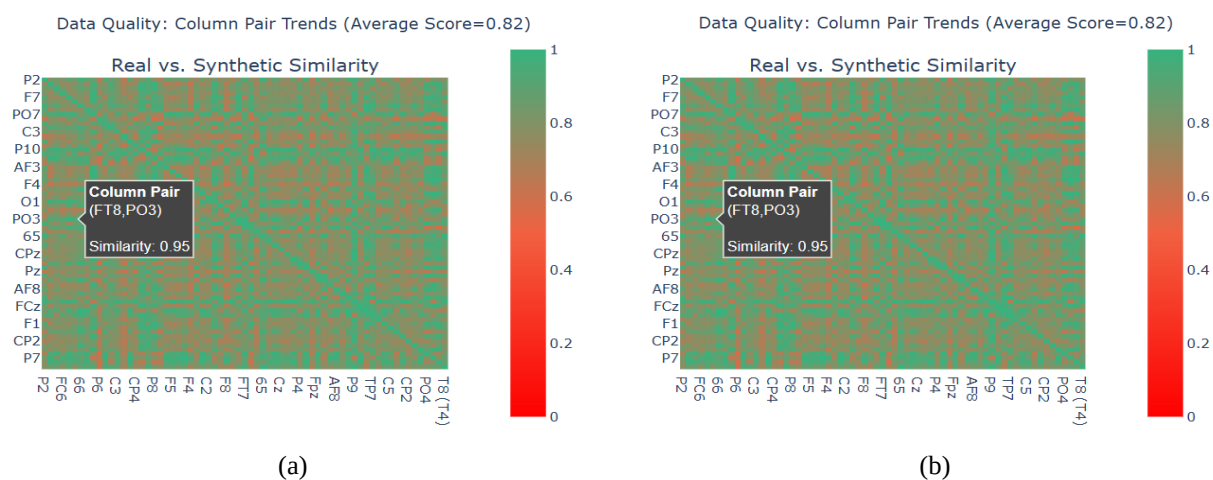


Fig. 9 Correlation similarity (column pair trends) of channel data (real vs synthetic): (a) ND-CTGAN, (b) EEG-GAN.

Table 1 Performance evaluation of SMOTE variants, EEG-GAN, and ND-CTGAN.

			Random Forest	MLP	SVM	KNN	Logistic	DT	Gaussian NB
before augmentation		Accuracy	0.32	0.37	0.38	0.4	0.46	0.38	0.35
		Recall	0.23	0.27	0.37	0.3	0.33	0.38	0.35
		Precision	0.16	0.23	0.24	0.3	0.31	0.27	0.35
after augmentation (augmentation applied on)	feature matrix	SMOTE	Accuracy	0.59	0.55	0.51	0.49	0.64	0.61
			Recall	0.583	0.57	0.5	0.48	0.66	0.54
			Precision	0.533	0.49	0.51	0.35	0.59	0.53
		Borderline SMOTE	Accuracy	0.58	0.63	0.59	0.51	0.7	0.64
			Recall	0.63	0.65	0.5	0.48	0.7	0.66
			Precision	0.62	0.57	0.43	0.39	0.66	0.59
		ADASYN	Accuracy	0.59	0.62	0.4	0.53	0.67	0.61
			Recall	0.55	0.64	0.4	0.53	0.66	0.63
			Precision	0.49	0.57	0.4	0.53	0.67	0.54
	pre-processed data	SMOTE SVM	Accuracy	0.47	0.54	0.32	0.5	0.63	0.51
			Recall	0.48	0.63	0.3	0.5	0.65	0.53
			Precision	0.5	0.59	0.3	0.5	0.62	0.52
		SMOTE Tomek	Accuracy	0.57	0.61	0.51	0.47	0.57	0.59
			Recall	0.56	0.6	0.44	0.48	0.58	0.65
			Precision	0.48	0.5	0.32	0.32	0.48	0.58
		ND-CTGAN	Accuracy	0.6	0.69	0.67	0.59	0.65	0.61
			Recall	0.58	0.6	0.66	0.58	0.65	0.55
			Precision	0.43	0.52	0.58	0.53	0.64	0.55
		TTS-WGAN-GP (EEG-GAN)	Accuracy	0.56	0.54	0.6	0.56	0.6	0.54
			Recall	0.51	0.54	0.61	0.58	0.61	0.51
			Precision	0.51	0.47	0.62	0.55	0.58	0.51

The ROC curve was generated for three classes: Combine (Class 0), Hyperactive (Class 1), and Inattentive (Class 2). Fig. 10 illustrates that the 0.83 to 0.85 AUC for the hyperactive class (class 1) shows the good performance of classifiers compared to the other two classes.

4 Discussion and Conclusion

This study focuses on the issues associated with less EEG data availability and imbalances within classes that frequently hamper ADHD diagnosis and classification accuracy. Conventional methods that depend on EEG image data have difficulties with the number of dimensions and maintaining the inter-channel correlations. To address these problems, the study focuses on preserving the spectral properties and improving the inter-channel relationships of synthetic

EEG data to enhance classification accuracy, explicitly emphasizing the Hyperactive ADHD subtype.

The study employed three primary techniques for data augmentation: several versions of the Synthetic Minority Over-sampling technique (SMOTE), EEG-GAN variants (TTS-WGAN-GP), and a novel method called Non-Deterministic Conditional Tabular Generative Adversarial Network (ND-CTGAN). EEG data from subjects with various subtypes of ADHD was used, and multiple SMOTE modifications were employed to address class imbalances. The expanded datasets were classified using seven various classifiers, resulting in notable improvements in accuracy. ND-CTGAN generates high-quality synthetic data that maintains complex, non-linear relationships and essential inter-channel correlations compared to EEG-GAN. This strategy is effective, as the data generated resembles real data features similar to the individual's.

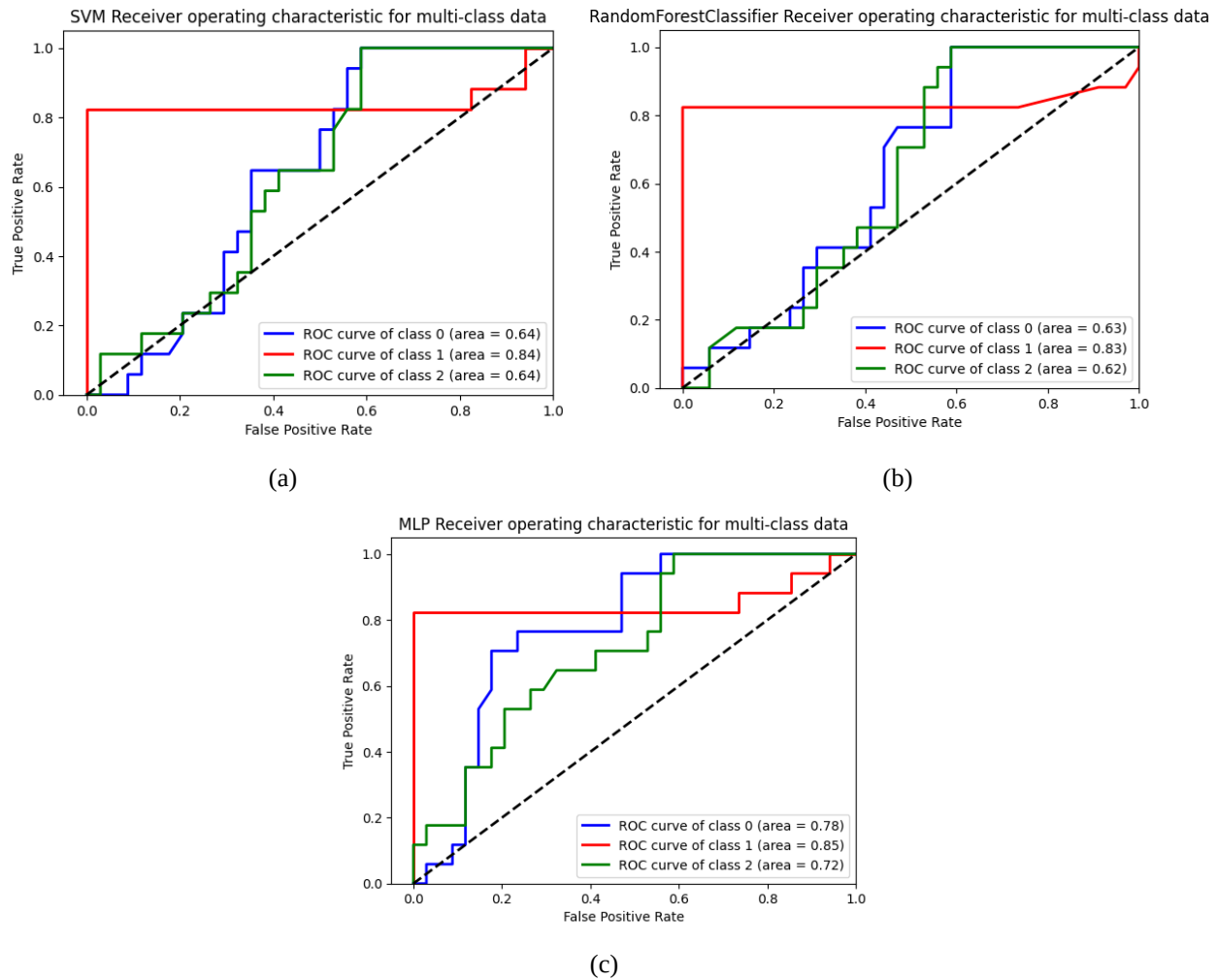


Fig. 10 ROC for Class 0 (Combine), Class 1 (Hyperactive), Class 2 (Inattentive) data: (a) for SVM, (b) for Random Forest, (c) for MLP using ND-CTGAN classifiers.

The synthetic and real data analysis using statistical and quality metrics comprise mean and STD, CDF, quality score: KS statistic, correlation similarity indicates that implementing sophisticated data augmentation techniques, ND-CTGAN can significantly enhance the classification accuracy of EEG data for individuals with Hyperactive ADHD. It primarily maintains the unique characteristics of each participant, facilitating meaningful comparisons between actual and synthetic datasets. This research emphasizes the potential of sophisticated generative models in dealing with class imbalances and improving the ability of machine learning models to generalize when applied to EEG data. Further research can enhance and explore the applicability of these methods to different forms of EEG and relevant neurological disorders.

Data Availability Statement

The data which was used in this research is publicly available on the following website: <https://osf.io/4285y/>.

Conflict of Interest

The authors declare no conflicts of interest.

Ethical Standards

Not applicable (Datasets are mentioned in the dataset availability statement).

Clinical Trials

This research did not undertake any clinical trials.

Funding Statement

The authors did not receive support from any organization for the submitted work.

References

1. R. J. Barry, A. R. Clarke, and S. J. Johnstone, "A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography," *Clinical Neurophysiology* 114(2), 171–183 (2003).
2. M. N. I. Qureshi, B. Min, H. J. Jo, and B. Lee, "Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural mri study," *PLoS ONE* 11(8), e0160697 (2016).
3. S. L. Karalunas, J. T. Nigg, "Heterogeneity and subtyping in attention-deficit/hyperactivity disorder—considerations for emerging research using person-centered computational approaches," *Biological Psychiatry* 88(1), 103–110 (2020).
4. D. Katsuki, H. Yamashita, K. Yamane, S. Kanba, and K. Yoshida, "Clinical subtypes in children with attention-deficit hyperactivity disorder according to their child behavior checklist profile," *Child Psychiatry & Human Development* 51(6), 969–977 (2020).
5. M. Rostami, S. Farashi, R. Khosrowabadi, and H. Pouretmad, "Discrimination of ADHD subtypes using decision tree on behavioral, neuropsychological and neural markers," *Basic and Clinical Neuroscience Journal* 11(3), 359–368 (2019).
6. A. Kamida, K. Shimabayashi, M. Oguri, T. Takamori, N. Ueda, Y. Koyanagi, N. Sannomiya, H. Nagira, S. Ikunishi, Y. Hattori, K. Sato, C. Fukuda, Y. Hirooka, and Y. Maegaki, "EEG power spectrum analysis in children with ADHD," *Yonago Acta Medica* 59(2), 169–173 (2016).
7. M. Rostami, R. Khosrowabadi, B. Albrecht, A. Rothenberger, and H. Pouretmad, "Classifying ADHD subtypes/presentations considering the joint effect of three levels of investigation," *Nordic Journal of Psychiatry* 75(1), 31–37 (2021).
8. R. Aldemir, E. Demirci, H. Per, M. Canpolat, S. Özmen, and M. Tokmakçı, "Investigation of attention deficit hyperactivity disorder (ADHD) sub-types in children via EEG frequency domain analysis," *International Journal of Neuroscience* 128(4), 349–360 (2018).
9. A. R. Clarke, R. J. Barry, R. McCarthy, and M. Selikowitz, "EEG-defined subtypes of children with attention-deficit/hyperactivity disorder," *Clinical Neurophysiology* 112(11), 2098–2105 (2001).
10. C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on EEG: a systematic comparison," *Journal of Neural Engineering* 19(6), 066020 (2022).
11. A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: an overview," *Journal of NeuroEngineering and Rehabilitation* 20(1), 40 (2023).
12. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering* 16(5), 051001 (2019).
13. V. Joshi, N. Nanavati, "Inter-class correlation-based EEG channel selection for ADHD classification," *IAENG International Journal of Computer Science* 51(8), 961–975 (2024).
14. J. J. Newson, T. C. Thiagarajan, "EEG frequency bands in psychiatric disorders: a review of resting state studies," *Frontiers in Human Neuroscience* 12, 521 (2019).
15. J. Kwas, A. Noyce, and B. Shinn-Cunningham, "Top-down attention modulates neural responses in neurotypical, but not ADHD, young adults," *The Journal of the Acoustical Society of America* 150(4_Supplement), A64–A64 (2021).
16. C. C. Williams, D. Weinhardt, M. Wirzberger, and S. Musslick, "Augmenting EEG with generative adversarial networks enhances brain decoding across classifiers and sample sizes," *Proceedings of the Annual Meeting of the Cognitive Science Society* 45(45) (2023).
17. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research* 16, 321–357 (2002).
18. N. Alsharif, M. H. Al-Adhaileh, and M. Al-Yaari, "Accurate identification of attention-deficit/hyperactivity disorder using machine learning approaches," *Journal of Disability Research* 3(1), (2024).
19. L. Shahmoradi, Z. Liraki, M. Karami, B. Savareh, and M. Nosratabadi, "Development of decision support system to predict neurofeedback response in adhd: an artificial neural network approach," *Acta Informatica Medica* 27(3), 186 (2019).
20. T. V. Padmavathy, M. Vinothkumar, and M. N. Vimal Kumar, "Modified cuckoo search- cascade forest (Mcs-cf) for attention deficit hyperactivity disorder (adhd) diagnosis," *NeuroQuantology* 18(7), 83–94(2020).
21. H. Alkahtani, T. Aldhyani, Z. Ahmed, and A. Alqarni, "Developing system-based artificial intelligence models for detecting the attention deficit hyperactivity disorder," *Mathematics* 11(22), 4698 (2023).
22. A. Alhudhaif, "A novel multi-class imbalanced EEG signals classification based on the adaptive synthetic sampling (ADASYN) approach," *PeerJ Computer Science* 7, e523 (2021).
23. Z. Jiang, W. Zhao, "Fusion algorithm for imbalanced EEG," *Seizure* 91, 207–211 (2021).
24. P. Zhang, X. Zhang, and A. Liu, "Effects of data augmentation with the bnnsmote algorithm in seizure detection using 1D-MobileNet," *Journal of Healthcare Engineering* 2022, 1–13 (2022).

25. R. Wardoyo, I. M. A. Wirawan, and I. G. A. Pradipta, “[Data processing in seizure detection oversampling approach using radius- SMOTE for imbalance electroencephalography datasets](#),” *Emerging Science Journal* 6(2), 382–398 (2022).
26. Y. Chen, R. Chang, and J. Guo, “[Effects of data augmentation method borderline-smote on emotion recognition of eeg signals based on convolutional neural network](#),” *IEEE Access* 9, 47491–47502 (2021).
27. T. Lee, M. Kim, and S. P. Kim, “[Data augmentation effects using borderline-SMOTE on classification of a P300-based BCI](#),” 8th International Winter Conference on Brain-Computer Interface, 9–12 (2020).
28. T. Lee, M. Kim, and S.-P. Kim, “[Improvement of P300-based brain-computer interfaces for home appliances control by data balancing techniques](#),” *Sensors* 20(19), 5576 (2020).
29. K. Rasheed, J. Qadir, T. J. O’Brien, L. Kuhlmann, and A. Razi, “[A generative model to synthesize EEG data for epileptic seizure prediction](#),” *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29, 2322–2332 (2021).
30. D. Geng, Z. S. Chen, “[Auxiliary classifier generative adversarial network for interictal epileptiform discharge modeling and EEG data augmentation](#),” 10th International IEEE/EMBS Conference on Neural Engineering, 1130–1133 (2021).
31. M. P. Kalashami, M. M. Pedram, and H. Sadr, “[EEG feature extraction and data augmentation in emotion recognition](#),” *Computational Intelligence and Neuroscience* 2022, 1–16 (2022).
32. A. Zhang, L. Su, Y. Zhang, Y. Fu, L. Wu, and S. Liang, “[EEG data augmentation for emotion recognition with a multiple generator conditional Wasserstein GAN](#),” *Complex & Intelligent Systems* 8(4), 3059–3071 (2022).
33. M. Xu, Y. Chen, Y. Wang, D. Wang, Z. Liu, and L. Zhang, “[BWGAN-GP: an EEG data generation method for class imbalance problem in RSVP tasks](#),” *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30, 251–263 (2022).
34. Y. Song, L. Yang, X. Jia, and L. Xie, “[Common spatial generative adversarial networks based EEG data augmentation for cross-subject brain-computer interface](#),” arXiv preprint arXiv:2102.04456 (2021).
35. R. Zhang, Y. Zeng, L. Tong, J. Shu, R. Lu, K. Yang, Z. Li, and B. Yan, “[ERP-WGAN: A data augmentation method for EEG single-trial detection](#),” *Journal of Neuroscience Methods* 376, 109621 (2022).
36. J. Yoon, D. Jarrett, and M. van der Schaar, “[Time-series generative adversarial networks](#),” *Advances in Neural Information Processing Systems* 32, 1–11 (2019).
37. B. Ganti, G. Chaitanya, R. S. Balamurugan, N. Nagaraj, K. Balasubramanian, and S. Pati, “[Time-series generative adversarial network approach of deep learning improves seizure detection from the human thalamic SEEG](#),” *Frontiers in Neurology* 13, 755094 (2022).
38. O. George, R. Smith, P. Madiraju, N. Yahyasoltani, and S. I. Ahamed, “[Data augmentation strategies for EEG-based motor imagery decoding](#),” *Heliyon* 8(8), e10240 (2022).
39. L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, “[Modeling tabular data using conditional GAN](#),” *Advances in Neural Information Processing Systems* 32, (2019).
40. M. J. Cheon, D. H. Lee, J. W. Park, H. J. Choi, J. S. Lee, and O. Lee, “[CTGAN VS TGAN? Which one is more suitable for generating synthetic EEG data](#),” *Journal of Theoretical and Applied Information Technology* 99(10), 2359–2372 (2021).
41. V. Joshi, N. Nanavati, “[A review of EEG signal analysis for diagnosis of neurological disorders using machine learning](#),” *Journal of Biomedical Photonics and Engineering* 7(4), (2021).

Seismocardiograph Monitoring Using SMS Fiber Structure with PDMS Enclosure

Frans Rizal Agustiyanto^{1,2}, Agus M. Hatta^{1*}, Dhany Arifianto¹, Mahenda Radityo¹, Maulana Putra Santoso¹, Selfi Stendafity¹, and Budi Susetio Pikir³

¹ Department of Engineering Physics, Faculty of Industrial Technology and Systems Engineering, Institut Teknologi Sepuluh Nopember (ITS), 10 ITS str., Surabaya 60111, East Java, Indonesia

² Department of Physics Education, State Islamic University of Mahmud Yunus Batusangkar, 137 Sudirman General str., Tanah Datar 27212, West Sumatera, Indonesia

³ Faculty of Medicine, Airlangga University, 47 Mayjen. Prof. Dr. Moestopo str., Surabaya 60132, East Java, Indonesia
e-mail: amh@its.ac.id

Abstract. This study presents a novel seismocardiography (SCG) monitoring system utilizing a singlemode-multimode-singlemode (SMS) fiber structure encased within a polydimethylsiloxane (PDMS) enclosure. The system employs a simple intensity-based interrogation method. Experimental characterization of the SMS fiber sensor demonstrated its sensitivity to vibrations, achieving an average calibration error of 1.63%. Validation against a gold-standard electrocardiogram (ECG) yielded an average error of 2.29%, representing a 1.27% accuracy improvement over existing fiber-optic SCG sensors (97.71% vs. 96.48%). This design offers a simplified and more robust alternative to current fiber-optic SCG sensors while maintaining high accuracy.
© 2025 Journal of Biomedical Photonics & Engineering.

Keywords: fiber optic sensor; heart monitoring; polydimethylsiloxane (PDMS); seismocardiography (SCG); singlemode-multimode-singlemode (SMS).

Paper #9167 received 24 Sept 2024; revised manuscript received 27 Dec 2024; accepted for publication 9 Jan 2025; published online 20 Mar 2025. doi: [10.18287/JBPE25.11.010303](https://doi.org/10.18287/JBPE25.11.010303).

1 Introduction

Cardiovascular disease (CVD) remains a leading cause of global mortality, accounting for an estimated 17.9 million deaths in 2019, representing 32% of all deaths worldwide [1]. This burden disproportionately affects low- and middle-income countries, where over 75% of CVD deaths occur. Heart attacks and strokes are the primary contributors to this alarming mortality rate [1]. The rise in CVD prevalence is closely linked to modern lifestyles, emphasizing the urgent need for early prevention strategies to mitigate the associated risks [2].

Current clinical assessment of cardiovascular health relies on various diagnostic tools, including electrocardiography (ECG), magnetic resonance imaging (MRI), Computed Tomography (CT) scans, echocardiography, and nuclear myocardial perfusion scans. ECG, the most widely used method, records the heart's electrical activity, providing crucial insights into its function [3]. However, traditional ECG systems are often confined to hospital settings due to their size, complexity, and the need for specialized personnel.

To address these limitations, research has explored innovative solutions, such as wireless ECG sensors [4], smartphone-based ECG systems [5], and the application of convolutional neural networks (CNNs) for automated ECG analysis on wearable devices [6]. ECG data offers valuable information on various aspects of the cardiac cycle, including atrial and ventricular contraction and ventricular relaxation [7].

Seismocardiography (SCG) offers an alternative approach to cardiac monitoring. SCG measures mechanical vibrations generated by the heart's contraction and relaxation, providing insights into cardiac activity, including valve motion and ventricular filling time [8]. While SCG remains primarily a research tool with limited clinical application, its potential is being actively explored. Ongoing projects include the development of wearable SCG devices incorporating accelerometers and gyroscopes, with applications ranging from heart rate monitoring using machine learning algorithms [9] to the assessment of pre-ejection period (PEP) [10] and stroke volume (SV) in perioperative settings [11].

Both SCG and ECG capture key physiological events like isovolumetric contraction, valve opening and closing, ejection, and heart filling [12]. Moreover, SCG has shown promise in predicting cardiac intervals such as left ventricular ejection time (LVET) and electromechanical systolic pre-ejection period (PEP) [8]. Consequently, ECG data can serve as a valuable reference for validating SCG signals from newly developed devices.

Limitations of accelerometer-based SCG and ECG during magnetic resonance imaging (MRI) necessitate alternative approaches for monitoring cardiac activity in this critical context [13]. Such monitoring is vital to ensure participant safety and mitigate anxiety-related complications [14]. Fiber optic sensors offer a promising solution due to their immunity to electromagnetic interference, potential for miniaturization, and cost-effective fabrication [14, 15].

Previous research has demonstrated the utility of fiber optic sensors for cardiac monitoring. Fiber Bragg Gratings (FBGs) have been employed to measure heart rates through vibrations transmitted to various surfaces [16], and also to assess respiration and heart rates simultaneously [17]. Singlemode-multimode-singlemode (SMS) fiber structures have emerged as an alternative to FBGs, offering advantages in fabrication simplicity, interrogation system complexity, and cost [18–20]. Notably, SMS fiber structures have been successfully utilized for apex cardiography [21] and integrated into smart beds for concurrent heart and respiration rate monitoring [22].

Prior research has demonstrated that encapsulating Fiber Bragg Gratings (FBGs) within a polydimethylsiloxane (PDMS) enclosure can substantially enhance optical fiber reflectivity, reaching up to 95.7% [14]. Building upon this concept, the present study investigates the application of a PDMS-enclosed SMS fiber structure sensor for SCG signal detection. The PDMS enclosure serves as a protective packaging for the sensor, aiming to maintain or improve SCG signal detection performance while providing crucial mechanical protection.

2 Materials and Methods

2.1 Fabrication and Calibration of SMS Structure Fiber Structure

The SMS fiber structure, depicted in Fig. 1, comprises a multimode fiber (MMF) section spliced between two single-mode fibers (SMFs). This SMS fiber construction utilizes a step-index MMF with a numerical aperture (NA) of 0.22 and MMF length of 45 mm as well as a singlemode optical fiber known as SMF28. The MMF section in the SMS fiber structure plays important role for the SMS fiber structure-based devices. The length of MMF section in the SMS fiber structure plays important role for the SMS fiber structure-based devices. A variation of MMF length can shift a spectral response [23] and a baseline intensity level [20] for a

spectral wavelength measurement scheme using a broadband light source and an intensity measurement scheme at a fixed wavelength light source, respectively. This paper utilizes the intensity measurement scheme, which offers a simpler configuration compared to spectral wavelength measurement. While variations in the MMF length affect the baseline intensity for the SMS fiber structure, changes in the measurand (e.g., vibration, temperature, or strain) induce corresponding intensity variations. These variations, after suitable calibration, can be effectively employed for measurement or monitoring purposes.



Fig. 1 SMS structure fiber optic sensor configuration.

SMS fiber structures were fabricated using standard fiber optic equipment, including a fiber cleaver and a fusion splicer. To ensure precise MMF lengths in the SMS fiber structures, a digital caliper and a mechanical translation stage were employed during the cleaving and splicing process. This setup, as detailed in Ref. [23], allowed for precise MMF length control with an error margin of ± 0.5 mm. Several SMS structures with identical MMF specifications were fabricated and then encased in a protective polydimethylsiloxane (PDMS) enclosure.

Fig. 2 illustrates the experimental setup employed for vibration testing of the SMS fiber structure, a principle underpinning SCG monitoring. The sensor operates based on the acousto-optic effect: acoustic waves (generated by cardiac vibrations) induce strain in the fiber, altering its electrical permittivity through the photoelastic effect [24]. This strain, in turn, modifies the refractive index within the SMS fiber structure, leading to changes in the propagation constant and ultimately affecting the optical power output. Consequently, the heart's vibrational pattern modulates the transmitted optical power, allowing for its detection and analysis [25].

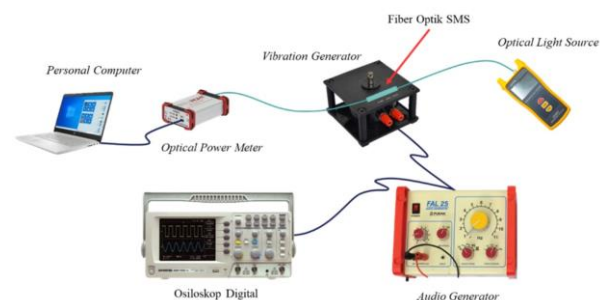


Fig. 2 PDMS Fiber-Optic Seismocardiograph calibration setup.

In the experimental setup, the SMS fiber structure is placed on a vibration generator (Pudak FAL29) connected to an oscilloscope (GW Instek GDS-1042) and a signal generator (Pudak FAL25). This configuration facilitates the acquisition of power fluctuation data and corresponding graphs resulting from vibrations in proximity to the optical fiber. Infrared light at a wavelength of 1550 nm and an optical power meter PM100USB from Thorlab was employed were utilized [21]. Fast Fourier Transform (FFT) is employed to convert the time-domain optical signal into the frequency domain, enabling the determination of its frequency components.

The SMS fiber structure is encased within a PDMS film, applied specifically to the multimode fiber (MMF) region. This encapsulation aims to mitigate the impact of environmental vibrations and disturbances that could potentially damage the delicate SMS fiber structure. The PDMS utilized in this study is in sheet form, with a thickness of 1 mm and dimensions of 80 mm (length) and 40 mm (width) [14]. Two PDMS sections are cut according to these dimensions, and the SMS fiber optic sensor is positioned between them. To investigate the influence of PDMS layer thickness on the sensor's output, varying thicknesses of 2 mm, 4 mm, and 6 mm are employed.

Calibration of the fabricated sensors, both unencapsulated and with varying enclosure thicknesses, was conducted using a vibration generator. Low-frequency vibrations within the range of 5–20 Hz, in 5 Hz increments, were applied. This frequency range was chosen to align with the typical low-frequency band of SCG signals, while avoiding higher frequencies prone to motion artifacts. The results of this calibration process are presented in Table 1.

As evident in Table 1, increasing enclosure thickness correlates with a higher average error. This observation can be attributed to the photoelastic effect: thicker enclosures dampen the vibrations reaching the SMS optical fiber, particularly within the sensitive MMF region, thereby reducing the sensor's response. Consequently, it is crucial to determine an optimal enclosure thickness that balances effective capture of cardiac vibrations with adequate protection for the optical fiber. In this study, a 2 mm thick enclosure was selected for subsequent data collection.

Table 1 Calibration results.

Reference frequency	Detected frequency			
	No enclosure	with 2 mm enclosure	with 4 mm enclosure	with 6 mm enclosure
5	5.021	5.1	5.14	5.22
10	10.198	10.2	10.44	10.31
15	15.237	15.19	14.89	15.37
20	20.4	20.29	20.22	20.4
Average error	1.01%	1.63%	2.25%	2.99%

3 Results

3.1 SCG Monitoring Using SMS Structure Fiber Optic Sensor with PDMS Enclosure

The SCG signal manifests as vibrations in the superficial layers of the body, particularly overlying the heart [8]. These vibrations induce strain in the MMF within the SMS fiber structure, leading to changes in multimode interference and, consequently, the output optical power. This photoelastic effect forms the basis of SCG signal transduction using the fiber optic sensor.

Fig. 3 illustrates the five potential sensor placements on the chest and thorax, each correlating to a specific cardiac region: position 1 (left ventricle), position 2 (left atrium), position 3 (right atrium), and positions 4 and 5 (right ventricle) [26]. For this study, the left ventricle (position 1) was chosen.

A healthy individual was selected as the participant for data collection. The sensor with PDMS enclosure was placed to the ventricle as in Fig. 4. A three-lead Heal Force Handheld Electrocardiogram ECG Monitor PC-80B with a single channel ECG was also placed to the participant allowing simultaneous SCG and ECG recordings.



Fig. 3 Seismocardiogram data collection positions.



Fig. 4 Data acquisition documentation of SMS fiber optic sensor with PDMS enclosure.

Figs. 5(a) and 5(b) show the measured optical power from the SMS fiber structure, and the ECG signal, respectively. It is noted that for SCG signal extraction, the optical power meter of PM100USB was set at a sampling frequency of 350 Hz. Raw data underwent pre-processing using a bandpass filter (0.5–25 Hz) to eliminate physiological motion artifacts. Subsequently, a smoothing technique was applied to segment the SCG signal into cardiac cycles, as detailed in Ref. [27]. To further reduce background noise arising from blood flow [28], an adjacent-average smoothing method was also employed. Fig. 5(c) presents the form of SCG signal.

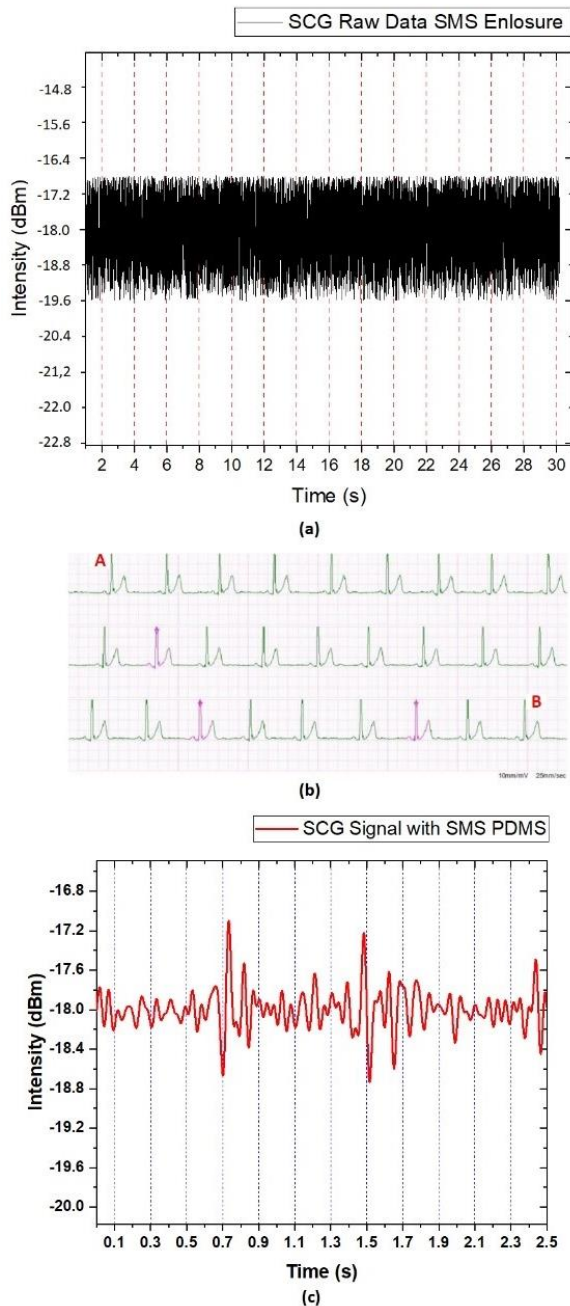


Fig. 5 (a) Raw data, (b) simultaneous ECG signal, and (c) filtered SCG signal.

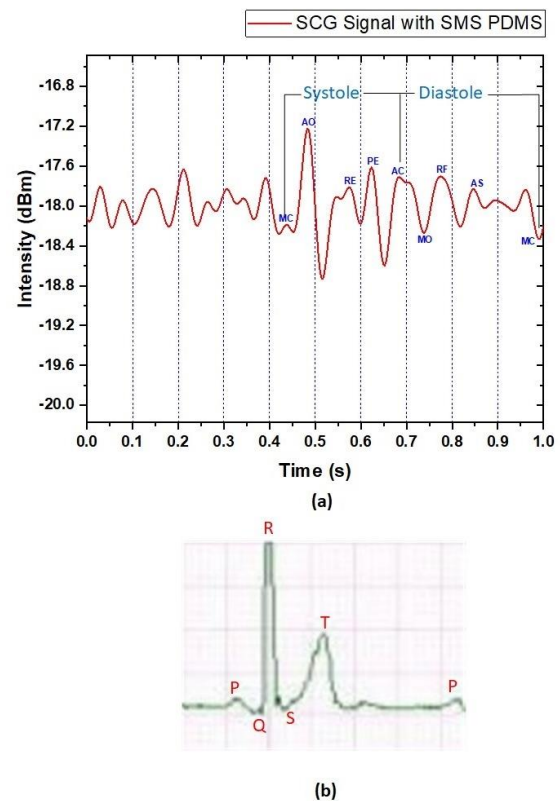


Fig. 6 Comparison of SCG and ECG signal, (a) one cycle of SCG signal with PDMS enclosure, (b) one cycle of simultaneous ECG signal.

The processed SCG signal was then compared to the ECG signal, as depicted in Fig. 6. Systolic and diastolic phases were identified from both signals: AO-AC and AC-MC peaks for SCG, and QRST and T-P intervals for ECG, respectively [8], as shown in Figs. 6(a) and 6(b). This work used a manual method to ascertain the location of each peak in both the SCG signals. By comparing it to a reference [13], the tallest peak was identified as the AO peak. The other peaks (MC, RE, PE, AC, MO, RF, AS) were identified following the acquisition of the AO peak.

As a comparison to the measured SCG signal, an ECG recording using a three-lead Heal Force Handheld Electrocardiogram ECG Monitor PC-80B with a single channel ECG was conducted. The recorded ECG signal data were extracted to identify PQRST peaks and to determine T-P time interval as in Refs. [8, 13]. The ECG monitor can provide about $5 \text{ mm.mV} \pm 10\%$ scale with a wave form sweeping speed of $20 \text{ mm/s} \pm 10\%$.

Tables 2 and 3 present a comparison of systolic and diastolic intervals derived from SCG and ECG signals, for the SMS fiber optic sensor without and with PDMS enclosure, respectively. Data validation involved 10 repetitions of SCG and ECG measurements. To assess the suitability of using the mean error value, the standard deviation of the overall error was calculated. For the unencapsulated sensor, the standard deviation was 0.0306, indicating acceptable homogeneity and justifying the use of the mean error [29]. The corresponding accuracy, derived from this mean error, was 96.48%.

Table 2 Comparison of systole and diastole periods (s) of SCG and ECG signals without PDMS enclosure.

No	ECG		SCG		Error	
	Systole	Diastole	Systole	Diastole	Systole	Diastole
1	0.49	0.72	0.55	0.69	9.86%	3.88%
2	0.49	0.50	0.52	0.48	6.12%	4.97%
3	0.33	0.67	0.32	0.67	1.54%	0.30%
4	0.50	0.72	0.52	0.72	3.58%	0.97%
5	0.32	0.63	0.31	0.63	1.89%	0.32%
6	0.41	0.52	0.42	0.52	0.97%	0.58%
7	0.33	0.47	0.30	0.49	9.37%	3.19%
8	0.40	0.46	0.40	0.47	0.99%	1.29%
9	0.46	0.53	0.44	0.49	4.32%	7.55%
10	0.41	0.72	0.41	0.78	1.48%	7.18%
Average Error					3.52%	

Table 3 Comparison of systole and diastole periods (s) of SCG and ECG signals with PDMS enclosure.

No	ECG		SCG		Error	
	Systole	Diastole	Systole	Diastole	Systole	Diastole
1	0.34	0.45	0.37	0.48	8.89%	5.26%
2	0.38	0.57	0.38	0.58	1.05%	1.38%
3	0.33	0.65	0.30	0.63	9.43%	2.53%
4	0.36	0.64	0.35	0.66	2.29%	3.34%
5	0.34	0.64	0.34	0.64	1.17%	0.16%
6	0.34	0.68	0.33	0.68	2.99%	0.00%
7	0.44	0.62	0.44	0.61	0.23%	0.49%
8	0.33	0.58	0.33	0.58	0.61%	0.17%
9	0.33	0.64	0.34	0.65	4.13%	1.08%
10	0.46	0.60	0.46	0.60	0.43%	0.17%
Average Error					2.29%	

For the PDMS-encapsulated sensor, the standard deviation of the overall error was 0.27729, again permitting the use of the mean error. The average error value, as shown in Table 3, demonstrates a marked improvement compared to the unencapsulated sensor, with a 53.7% reduction in error magnitude. The resulting accuracy for the encapsulated sensor was 97.71%.

The results highlight the efficacy of the PDMS enclosure in enhancing the accuracy and reliability of the SMS fiber optic sensor for SCG signal acquisition. The observed reduction in error magnitude and improved accuracy underscore the potential of this approach for

developing robust and practical SCG monitoring systems.

To further evaluate the impact of the PDMS enclosure, we conducted a statistical analysis comparing measurements with and without the enclosure (Tables 2 and 3). Our analysis revealed no statistically significant difference between the two datasets for both systolic and diastolic measurements (*p*-values of 0.536 and 0.173, respectively). These *p*-values exceed the 0.05 significance level, indicating that observed variations are likely due to random fluctuations and the limited sample size. While a larger study might yield more definitive conclusions, our findings suggest that the PDMS

enclosure does not negatively affect sensor performance. This observation aligns with previous research [30] where PDMS was successfully employed in flexible mechanical sensors. PDMS, with its Poisson's ratio near 0.5, exhibits incompressibility, making it highly sensitive to even small deformations, which is beneficial for sensor applications.

4 Conclusions

This study investigated the feasibility of a novel SCG monitoring system utilizing a SMS fiber structure encased in a protective PDMS enclosure. Positioned on the anterior thorax to capture cardiac vibrations, the sensor's optical power output was processed to generate the SCG signal. Calibration using vibration and audio signals (5–20 Hz) demonstrated high accuracy, with an average error of 1.63%. Validation against a gold-standard electrocardiogram (ECG) confirmed the sensor's reliability, achieving an average error of 2.29% and an overall accuracy of 97.71%. This represents a 1.27% improvement compared to existing fiber-optic SCG sensors. Statistical analysis further confirmed that the PDMS enclosure did not negatively impact sensor performance while providing essential mechanical protection. This novel SCG system offers a promising

pathway for developing reliable and cost-effective cardiac monitoring solutions.

Acknowledgments

This work was supported by “Hibah Penelitian Doktorat” from Ministry of Education and Research Technology, Indonesia.

Funding

This research received funding from the “Hibah Penelitian Doktorat” under contract number 1689/PKS/ITS/2024 from the Ministry of Education and Research Technology, Indonesia.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Disclosures

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. World Health Organization, “[World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals](#),” Geneva, 19 May 2023 (accessed 22 June 2023).
2. M. Y. Desai, S. Windecker, P. Lancellotti, J. J. Bax, B. P. Griffin, O. Cahlon, and D. R. Johnston, “[Prevention, diagnosis, and management of radiation-associated cardiac disease](#),” *Journal of the American College of Cardiology* 74(7), 905–927 (2019).
3. P. K. Jain, A. K. Tiwari, “[Heart monitoring systems—A review](#),” *Computers in Biology and Medicine* 54, 1–13 (2014).
4. S. Majumder, L. Chen, O. Marinov, C.-H. Chen, T. Mondal, and M. J. Deen, “[Noncontact wearable wireless ecg systems for long-term monitoring](#),” *IEEE Reviews in Biomedical Engineering* 11, 306–321 (2018).
5. H. T. Haverkamp, S. O. Fosse, and P. Schuster, “[Accuracy and usability of single-lead ECG from smartphones - A clinical study](#),” *Indian Pacing and Electrophysiology Journal* 19(4), 145–149 (2019).
6. L. Meng, K. Ge, Y. Song, D. Yang, and Z. Lin, “[Long-term wearable electrocardiogram signal monitoring and analysis based on convolutional neural network](#),” *IEEE Transactions on Instrumentation and Measurement* 70, 1–11 (2021).
7. B. Ngai, K. Tavakolian, A. Akhbardeh, A. P. Blaber, B. Kaminska, and A. Noordergraaf, “[Comparative analysis of seismocardiogram waves with the ultra-low frequency ballistocardiogram](#),” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2851–2854 (2009).
8. A. Taebi, B. E. Solar, A. J. Bomar, R. H. Sandler, and H. A. Mansy, “[Recent advances in seismocardiography](#),” *Vibration* 2(1), 64–86 (2019).
9. O. T. Inan, M. Baran Pouyan, A. Q. Javaid, S. Dowling, M. Etemadi, A. Dorier, J. A. Heller, A. O. Bicen, S. Roy, T. De Marco, and L. Klein, “[Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients](#),” *Circ: Heart Failure* 11(1), e004313 (2018).
10. M. M. H. Shandhi, B. Semiz, S. Hersek, N. Goller, F. Ayazi, and O. T. Inan, “[Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation](#),” *IEEE Journal of Biomedical and Health Informatics* 23(6), 2365–2374 (2019).
11. B. Semiz, A. M. Carek, J. C. Johnson, S. Ahmad, J. A. Heller, F. G. Vicente, S. Caron, C. W. Hogue, M. Etemadi, and O. T. Inan, “[Non-invasive wearable patch utilizing seismocardiography for peri-operative use in surgical patients](#),” *IEEE Journal of Biomedical and Health Informatics* 25(5), 1572–1582 (2021).
12. P. K. Sahoo, H. K. Thakkar, W.-Y. Lin, P.-C. Chang, and M.-Y. Lee, “[On the design of an efficient cardiac health monitoring system through combined analysis of ecg and scg signals](#),” *Sensors* 18(2), 379 (2018).

13. J. Kolarik, R. Kahankova, J. Brablik, and R. Martinek, “[Comparison of SCG and ECG based cardiac activity monitoring in laboratory conditions](#),” IFAC-PapersOnLine 52(27), 550–555 (2019).
14. J. Nedoma, M. Fajkus, P. Siska, R. Martinek, and V. Vasinek, “[Non-invasive fiber optic probe encapsulated into polydimethylsiloxane for measuring respiratory and heart rate of the human body](#),” Advances in Electrical and Electronic Engineering 15(1), 93–100 (2017).
15. C. Perezcampos Mayoral, J. Gutiérrez Gutiérrez, J. L. Cano Pérez, M. Vargas Treviño, I. B. Gallegos Velasco, P. A. Hernández Cruz, R. Torres Rosas, L. Tepech Carrillo, J. Arnaud Ríos, E. L. Apreza, and R. Rojas Laguna, “[Fiber optic sensors for vital signs monitoring, a review of its practicality in the health field](#),” Biosensors 11(2), 58 (2021).
16. Y. S. Yaras, O. Kocaturk, and F. L. Degertekin, “[FBG based electric field sensor for MRI safety](#),” in Proceedings of Optical Sensors and Sensing Congress, STu4D.7 (2020).
17. C. Tavares, C. Leitão, D. Lo Presti, M. F. Domingues, N. Alberto, H. Silva, and P. Antunes, “[Respiratory and heart rate monitoring using an FBG 3D-printed wearable system](#),” Biomedical Optics Express 13(4), 2299 (2022).
18. N. Irawati, A. M. Hatta, Y. G. Y. Yhuwana, and Sekartedjo, “[Heart rate monitoring sensor based on singlemode-multimode-singlemode fiber](#),” Photonic Sensors 10(2), 186–193 (2020).
19. Q. Wu, A. M. Hatta, P. Wang, Y. Semenova, and G. Farrell, “[Use of a bent single sms fiber structure for simultaneous measurement of displacement and temperature sensing](#),” IEEE Photonics Technology Letters 23(2), 130–132 (2011).
20. A. M. Hatta, Y. Semenova, Q. Wu, and G. Farrell, “[Strain sensor based on a pair of single-mode-multimode-single-mode fiber structures in a ratiometric power measurement scheme](#),” Applied Optics 49(3), 536–541 (2010).
21. F. R. Agustiyanto, A. M. Hatta, D. Arifianto, and B. S. Pikir, “[Use of a singlemode multimode singlemode fiber structure for apex cardiography monitoring](#),” in Fourth International Seminar on Photonics, Optics, and Its Applications, Proceedings of SPIE 117892, 117890E (2021).
22. N. R. Amalia, A. M. Hatta, S. Koentjoro, A. M. T. Nasution, and I. Puspita, “[Fiber sensor for simultaneous measurement of heart and breath rate on smartbed](#),” in IEEE 8th International Conference on Photonics, IEEE, 68–69 (2020).
23. A. M. Hatta, G. Farrell, Q. Wang, G. Rajan, P. Wang, and Y. Semenova, “[Ratiometric wavelength monitor based on singlemode-multimode-singlemode fiber structure](#),” Micro & Optical Tech Letters 50(12), 3036–3039 (2008).
24. A. Yariv, P. Yeh, “[Optical waves in crystal propagation and control of laser radiation](#),” (1983).
25. Y. Zhao, X. Li, F. Meng, and Z. Zhao, “[A vibration-sensing system based on SMS fiber structure](#),” Sensors and Actuators A: Physical 214, 163–167 (2014).
26. T. Allsop, G. Lloyd, R. S. Bhamber, L. Hadzievski, M. Halliday, D. J. Webb, and I. Bennion, “[Cardiac-induced localized thoracic motion detected by a fiber optic sensing scheme](#),” Journal of Biomedical Optics 19(11), 117006 (2014).
27. P. K. Jain, A. K. Tiwari, “[An algorithm for automatic segmentation of heart sound signal acquired using seismocardiography](#),” in 2016 International Conference on Systems in Medicine and Biology, 157–161 (2016).
28. S. Koyama, T. Hayase, S. Miyauchi, A. Shirai, S. Chino, Y. Haseda, and H. Ishizawa, “[Influence on measurement signal by pressure and viscosity changes of fluid and installation Condition of FBG sensor using blood flow simulation model](#),” IEEE Sensors Journal 19(24), 11946–11954 (2019).
29. S. El Omda, S. R. Sergeant, “[Standard deviation](#),” StatPearls Publishing, Treasure Island (FL) (2021).
30. Q. She, P. Wu, W. Lin, Y. Huang, F. Wang, and Z. Ye, “[Design of flexible mechanical sensor based on PDMS grating](#),” Advanced Fiber Laser Conference 12595, 560–566 (2023).

Dynamic Conditional Encoding and Feature Frequency Parsing in Diffusion Probabilistic Models for Diabetic Foot Ulcer Detection Using Thermographic Imaging

Kousar Nikhath A.^{1*}, Sireesha C.², Swapna M.³, Vijayetha Thoday⁴, and Esther Rani D.⁵

¹VNR Vignana Jyothi Institute of Engineering and Technology, Vignana Jyothi Nagar, Pragathi Nagar, Nizampet, Hyderabad 500090, Telangana, India

²Vasavi College of Engineering, Ibrahimbagh, Hyderabad 500031, Telangana, India

³Stanley College of Engineering Technology for Women, Chapel Road, Abids, Hyderabad 500001, Telangana, India

⁴Mukesh Patel School of Technology, Management and Engineering, NMIMS, Mumbai, Maharashtra 400056, India

⁵SV College of Engineering, Karakambadi Road, Tirupati 517507, India

e-mail: kousarnikhath@vnrvijet.in

Abstract. Diabetic patients' odds of survival can be significantly increased by early identification of diabetic foot ulcers (DFU). Thermography is utilised to identify changes in planter temperature because diabetic foot causes planter ulcers. This study uses publicly accessible thermographic imaging data from patients in the diabetes group as to ll as the control group. For DFU detection, both image-level and patch-level thermograms are used. The work suggests an optimizer-based Diffusion Probabilistic Segmentation model (Op-DPS) for DFU identification. The work suggests Dynamic Conditional Encoding, which creates state-adaptive conditions for every sampling step, to progress the step-wise regional attention in DPS for segmentation tasks. In moderate the detrimental impact of high-frequency noise components in this approach, to also suggest the Feature Frequency Parser (FF-Parser). The paper presents a Football Optimisation Algorithm (FbOA) that uses high dimensionality, nonlinearity, and many local optima to fine-tune the parameters of the suggested model. FbOA is based on the strategic balance between exploration and corruption that is seen in football gameplay. The proposed Op-DPS model achieved exceptional performance with a precision of 89.55%, recall of 85.53%, F_1 score of 87.50%, Intersection-over-Union (IoU) of 88.96%, and an overall accuracy of 99.27% on diabetic foot ulcer segmentation tasks. These results demonstrate significant improvements over existing methods, highlighting the model's capability for accurate and reliable DFU detection. In order to efficiently explore and take advantage of the solution space, the algorithm incorporates short passes, long passes, and positional modifications to replicate players' tactical placement and movement. Image-level, patch-level, and image-patch combination data are used to generate the findings. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: diabetic foot ulcer; diffusion probabilistic segmentation model; football optimization algorithm; patch level images; thermographic image data; dynamic conditional encoding.

Paper #9193 received 4 Dec 2024; revised manuscript received 10 Jan 2024; accepted for publication 22 Jan 2025; published online 26 Mar 2025. [doi: 10.18287/JBPE25.11.010304](https://doi.org/10.18287/JBPE25.11.010304).

1 Introduction

Amputations related to diabetic foot ulcers (DFUs) pose a global health and financial burden [1]. They make up 10% of the diabetic health services expenditure in the UK [2]. The UK would save more than £250 GBP million annually if DFU cases to be reduced by one-third [3]. In addition to the costs of diabetes, the projected cost of DFU in the United States is between \$9 and \$13 billion [4]. “More than half of all DFUs result in infection, and patients with diabetes have a lifetime risk of up to 34% [5]. DFU is thought to be a clinical indicator of a higher risk of death and amputation [6]. Diabetes mellitus (DM), which is brought on by the body’s inability to produce enough insulin, produces persistently elevated blood glucose levels, or hyperglycemia [7]. Long-term uncontrolled diabetes can result in problems such retinopathy, nephropathy, and the development of Charcot feet, amputation, or even death [8]. Uncontrolled diabetes mellitus affects nerves; sensory diabetic neuropathy is a loss of sensation that results from damage to the nerves in the legs or feet. When neuropathy prevents a patient from feeling a cut or ache in his foot, the cut leads to infection and exacerbates the foot disease [9]. The decreased blood flow is the other circumstance. Low blood flow in the arms and legs is a symptom of peripheral vascular disease. Ulcers might form if the injury is not healing because of inadequate blood supply [10]. Individuals with diabetes are most likely to develop diabetic foot ulcers (DFUs); around 15% of individuals have this condition [11].

Problems with diabetic feet can be costly and negatively affect a person’s quality of life. By doing an early risk assessment and examination of diabetic patients’ foot health, this may often be avoided or significantly postponed [12]. Temperature may be important for that reason. Neuropathy, ischaemia, or infection can cause changes in the temperature of diabetics’ plantar feet [13–14]. While the normal difference is usually less than 1 °C, temperature disparities between the right and left feet of more than 2.2 °C (4 F) are deemed abnormal [15]. By using a thermal-imaging camera, issues may be found early on, thus saving time and money. If the temperature is higher than absolute zero, infrared thermography can be used to provide a clear image of the thermal energy generated by the site under real-time monitoring [16]. A non-invasive, non-contact, quick, painless, and economical method of taking a patient’s skin temperature is thermography. Temperature variations on human skin may be detected using this imaging method [17].

A qualified specialist must evaluate the thermogram in order to identify the ulcer. It can be difficult to find these specialists, especially in isolated places [18]. Consequently, the provision of an automatic DFU identification system has been the focus of several research investigations. While several of them took into account visible-band (RGB camera) pictures [19], only a small number of them used foot thermograms [20]. Diabetic foot thermogram pictures are taken into consideration for DFU identification in this investigation.

Three levels of thermogram data are used in the tests for DFU recognition: image-level, patch-level, and a combination of image-patch thermogram data. Using transfer learning, an optimizer-based deep-learning model is used to identify the DFU feet. To observe that lesions and organs are frequently unclear and challenging to distinguish from the background in thermos picture segmentation tasks. To get correct findings in this situation, an adaptive calibration procedure is essential. In order to create the suggested model, to suggest using Dynamic Conditional Encoding instead of vanilla conditional DPS. The suggested model uses the picture to condition each stage of the iterative sampling process before learning the segmentation map. At each stage, to include the segmentation map of the current phase into the picture pre encoding to accomplish adaptive regional attention. In particular, to adopt a multi-scale approach to fuse the picture prior on the feature level with the current step segmentation mask. As a result, the reconstruction accuracy is increased and the condition characteristics are dynamically enhanced by the corrupted current-step mask. We propose to use the Feature Frequency Parser (FF-Parser) to filter the features in the Fourier space in order to remove high-frequency noise from the damaged mask. Every skip connection path uses FF-Parsers for multi-scale integration.

The remainder of the paper’s details are as follows. The background information and relevant research are presented in Section 2, and the dataset and specifics of the suggested model are included in Section 3. Section 4 presents the findings and comments. Section 5 now includes the conclusion.

2 Related Works

In order to precisely identify ulcer areas in images of patients’ feet, Sarmun et al. [21] suggested a deep learning-based approach. Using innovative model ensemble techniques including to highlighted the bounding box fusion (WBF), Soft-NMS, and non-maximum suppression (NMS), our method integrates predictions from state-of-the-art object recognition models. The results obtained from the many state-of-the-art model architectures used in this study are both improved and more generally applicable when they are used in concert with one another. With an impressive 12.4% improvement over the previous benchmark, our WBF-based technique utilising YOLOv8m and FRCNN-ResNet101 achieves a mAP score of 86.4% at the Intersection-over-Union (IoU) threshold of 0.5 on the DFUC2020 dataset. Our models have consistently outperformed the competition in qualitative analyses using the IEEE DataPort Diabetic Foot dataset, which is another dataset that to test externally. Finally, a new ensemble model for DFU detection was developed and tested in our study. Medical professionals benefit from early screening and an improved diagnosis process thanks to an AI-driven tool that increases sensitivity to potential DFU cases. While we acknowledge that false positives can happen, our study shows that by integrating

human medical knowledge with AI-based DFU management strategies, patient care can be improved.

Medical image analysis has been the primary focus of El-Kady et al. [22] in their efforts to improve the accuracy of DFU diagnosis. The famous ResNet50 model and a hybrid model combining ResNet50 with Generative Adversarial Networks to be tested. With an F_1 -score of 0.75 and an average accuracy and precision of 0.76, the ResNet50 worked fantastically, yielding outstanding results. The hybrid model accomplished better than these metrics with an average F_1 -Score, precision, and accuracy of 0.84. More precise and effective DFU detection in clinical settings is within reach, thanks to our study's contribution to medical image processing literature. An important step forward in the treatment and management of DFU has been the hybrid model's substantial increase in diagnostic accuracy.

Basiri et al. [23] has created a standardised framework for DFU data gathering by developing a case report form, a comprehensive dataset called Zivot that includes clinical feature breakdowns of the patient population, and a baseline for DFU identification using this dataset and a UNet architecture. Following this protocol, to create the Zivot dataset featuring 269 patients with active DFUs, 3700 RGB images of the DFUs with corresponding temperature and depth maps were collected. An efficient method for collecting uniform and clean datasets was demonstrated using a deep learning network for bounding box prediction that was constructed using the UNet architecture with EfficientNet as the feature extractor. The network was trained using the Zivot dataset, and the evaluation metrics shows to be promising results: a mAP segmentation metric of 0.79 and an F_1 -score metric of 0.86. This study and the Zivot database lay the groundwork for additional research on holistic and multimodal methods of DFU.

In order to assess the interpretability of deep learning (DL) models that are driven by explainability, Biswas et al. [24] introduced the DFU XAI framework. The goal of DFU XAI is to build an open DL framework by evaluating five DL models (Xception, DenseNet121, ResNet50, InceptionV3, and MobileNetV2) with three state-of-the-art explanation methods: Grad-CAM, SHAP, and LIME-local interpretable model-agnostic explanations. Outperforming the other four models, ResNet50 achieved remarkable results with 98.75% accuracy, 99.2% precision, 97.6% recall, 98.4% F_1 -score, and 98.5% AUC. Using the DFU dataset, it is able to differentiate between healthy feet and diabetic foot ulcers, and to localise diabetic foot ulcers to specific locations on diabetic feet. Using a heat map, the location of the ulcer can be accurately determined and appropriate treatment can be provided.

In their presentation of an automated system for detecting and categorising DFU, Almufadi et al. [25] rely on transfer learning. In most cases, two physically distinct states – ischaemic and infection – are used to classify DFU to evaluate the autonomous DFU detection performance of trained Deep Convolutional Neural

Network (DCNN) models. Seven models are compared: EfficientNetB0, DenseNet121, ResNet101, VGG16, MobileNetV2, InceptionV3, and InceptionResNetV2. One other model that to propose, E-DFu-Net, is built on pre-existing architectures and can help decrease overfitting. Experimental results show that E-DFu-Net does an excellent job, correctly identifying ischaemia with 97% accuracy and infections with 92%. Practitioners may now more effectively detect DFU cases because to this advancement, which improves upon previous methods.

An approach to image segmentation analysis of DFU through the internet of things using the technique of virtual sensing for semantically similar objects was proposed by Thota et al. [26] for deeper image segmentation. This method takes into account four levels of range segmentation: region-based, edge-based, image-based, and computer-aided design-based range segmentation. To simplify the multimodal process for semantic segmentation, object co-segmentation is employed in this study. It is anticipated that this will lead to improved validity and reliability evaluations. The experimental results demonstrate that the proposed model can perform segmentation analysis more efficiently and with less error than the existing methods. Results on the multiple-image dataset show that DFU produces an average segmentation score of 90.85% and 89.03% in two types of labelled ratios, i.e., 25% and 30%, respectively, after DFU using virtual sensing. Compared to the prior best performance, this is an improvement of 10.91% and 12.22%. In actual DFU tests, our proposed method achieved a 59.1% improvement over existing deep segmentation-based approaches; in comparison, its average picture smart segmentation improvements are 15.06%, 23.94%, and 45.41%. Interobserver reliability on the positive test, namely the likelihood ratio test set with just 0.25 million parameters, is 73.9% when using the specified range-based segmentation at the pace of labelled data.

Das et al. [27] described a novel Convolutional Neural Network (CNN) architecture called HCNNet, which combines several hybridised blocks from inception, residual, and dense modules with appropriately placed Squeeze-and-Excitation (SE) blocks and intermediate transition layers. Training the proposed HCNNet multiple times with varied optimiser and learning rate settings maximises its performance. Its 0.999 area under the curve (AUC) for ischaemia detection is quite encouraging. The experimental results show that the proposed HCNNet outperforms the existing SOTA methods.

Dhar et al. [28] introduced FUSegNet, a new model for segmenting diabetic foot ulcers that uses the pre-trained EfficientNet-b7 as its foundation, to circumvent the issue of insufficient training data. An enhanced spatial and channel squeeze-and-excitation (scSE) module that merges additive and max-out scSE is called P-scSE, or parallel scSE. At the middle of each decoder stage, the module is fused, resulting in a new configuration. In order to avoid max-out scSE at the top

decoder stage – which has a limited number of feature mappings – a shorted P-scSE is built. A number of augmentations, including morphological, intensity-based, and geometric augmentations, are applied to the data before it is fed into the network. The proposed model is initially evaluated using a publicly accessible dataset pertaining to chronic wounds. Its data-based dice score of 92.70% is the greatest of any method that has been documented. The model outperforms earlier scSE-based U-Net models in most categories according to Pratt's Figure of merits (PFOM) evaluations, which evaluate the accuracy of edge localisation. The model is subsequently entered into the MICCAI 2021 FUSeg challenge, where it will compete against x-FUSegNet, a variation of FUSegNet. In the FUSeg Challenge, the model that averages the outputs produced by FUSegNet using 5-fold cross-validation – x-FUSegNet – topped the leaderboard with a dice score of 89.23%.

According to Biswas et al. [29], a novel multi-scale feature fusion network called FusionNet can accurately differentiate between healthy and DFU skin by utilising multiple pre-trained convolutional neural network (CNN) approaches. Six hundred and eighty-one training images, 672 validation images, and 211 testing images were generated using a database including 6963 skin images (3,574 normal and 3,389 ulcer) taken from various patients. Beginning with three image preprocessing techniques – motion blur estimation, median filter, and Gaussian filter – fuzzy, irrelevant, and noisy data was first eliminated. After that, DenseNet201, VGG19, and NASNetMobile – three pre-trained convolutional neural network (CNN) algorithms – are used to extract high-frequency features from the input images. In order to predict DFU, a meta-tuner module was then fed the most discriminative features. Statistically significant differences between FusionNet and other sub-networks are found using analysis of variance (ANOVA) and Friedman tests. Finally, three eXplainable Artificial Intelligence (XAI) algorithms – Grad-CAM, LIME, and SHAP – are integrated into FusionNet to better explainability and transparency. The classification results provided by the FusionNet classifier are exceptional, with an accuracy of 99.05%, recall of 98.18%, precision of 100%, AUC of 99.09, and F_1 -score of 99.08%. The medical field could benefit from our proposed FusionNet since it can distinguish between DFU and healthy skin.

Zhao et al. [30] developed the Diabetic Foot Smart tool, an intelligent measurement tool for diabetic foot ulcers. In order to confirm its usefulness in assessing DFU, this study set out to quantify the ulcer area. The sizes of the ulcers on 150 DFU images are measured using three different assessment tools: the ruler method, the Smart APP software suite, and the gold standard application, Image J to compared the three approaches' measuring times and outcomes. To characterise the intra- and inter-rater reliability, the Pearson correlation coefficient, intra-group correlation coefficient, and coefficient of variation were used. The Image J program shows that the median ulcer size was 4.02 cm² and that

the mean measurement duration was 66.37 ± 7.95 s. The ruler method, which had an average measuring time of 171.47 ± 46.43 s, shows that the median ulcer area was 5.14 cm². The median ulcer size was 3.70 cm², and the mean measurement duration was 38.25 ± 6.81 s, as reported by the APP software. Although there was no notable distinction between the APP and Image J software ($Z = 1.103$, $p > 0.05$), there were notable differences between the ruler technique and the industry standard Image J software ($Z = -4.123$, $p < 0.05$). Furthermore, with intra- and inter-rater reliability both reaching 0.99, the APP software proved to be quite reliable.

Finding consensus among stakeholders involved in DFU care about relevant and viable quality indicators (QIs) was the goal of Lusendi et al.'s study [31], which was the second part of a Belgian QI selection research that aimed to find QIs for DFU therapy. Stakeholders in diabetic foot care included a patient organisation representative and healthcare professionals from both primary care and speciality fields. Using the RAND/UCLA Appropriateness Method, stakeholders are asked to rate the appropriateness of 42 prospective evidence-based indicators on a 9-point Likert scale. To classify QIs, the median scores is used and the disagreement index, which is derived from the inter-percentile range corrected for symmetry. Ultimately, 17 QIs are determined to be appropriate after a three-stage process. Five of them described previously unseen subjects: protocolised care (the implementation of care and prevention management protocols), administration of low-density lipoprotein-cholesterol of ring medication, systematic evaluation of the patient's nutritional status, and integration of wound care speciality in the interdisciplinary team.

By evaluating the likelihood of DFU recurrence in elderly diabetic patients, Hong et al. [32] hoped to enhance both therapeutic and prophylactic measures. The goal is to construct accurate prediction models that assess the probability of DFU recurrence in older diabetic patients by taking into account high-risk factors such age, blood sugar management, alcohol usage, and smoking. Machine learning models are constructed utilising data collected from 138 elderly diabetic patients after data cleansing, screening for outliers, and feature integration. Using Support Vector Machines (SVM) are able to achieve an accuracy rate of 93%. Experimental research demonstrating the efficacy of SVM in predicting the recurrence risk of DFUs in elderly diabetic patients has provided clinical practitioners with a more trustworthy assessment tool. In particular, the study highlights the significance of machine learning for predicting the probability of recurrence in the treatment of diabetic foot ulcers in older persons. This technique offers medical solutions supported by computers for the management of diabetes in the elderly, allows for rapid intervention, and reduces the likelihood of recurrence.

In their attempt to assess the efficacy of ChatGPT and similar AI chatbots in providing data regarding DFUs, Shiraishi et al. [33] relied on previously established

criteria. Seven AI chatbots are evaluated by asking clinical questions (CQs) that are based on the DFU criteria. Additionally, the chatbots' reference accuracy was verified. Although there are variations in the level of evidence and recommendation grade, the average accuracy rate of the AI chatbots in ring CQs was 91.2%. Claude-2 outperformed the other chatbots with 99.6% of its references confirmed, while ChatGPT had the proportion of reference authenticity at 66.3%. This study demonstrates the potential of artificial intelligence chatbots as tools for medical knowledge dissemination and their accuracy in addressing clinical question types relevant to DFU. Careful use and further refinement are required for medical applications due to concerns such as AI hallucinations and the unpredictable accuracy of these chatbots. This study highlights the importance of enhancing these tools for efficient usage in patient education and clinical decision-making, as well as the ways in which AI is transforming the healthcare business.

Using a combination of a convolutional neural network (CNN) and a Swin Transformer architecture, MG et al. [34] presented a deep learning model for infection classification in DFU images. The objective is to improve detection accuracy using end-to-end mapping without prior information by integrating global and local feature extraction. The proposed model makes use of a combination of a Convolutional Neural Network (CNN) and a Swin Transformer. With the use of the Grad CAM method, we can see how the CNN and Transformer blocks make decisions. The DFUC Challenge dataset is used for both training and evaluation, showcasing the model's accuracy in classifying DFU images as infected or not. Accuracy (96.52%), specificity (97.08%), sensitivity (95.08%), and Matthews Correlation Coefficient (0.93) are excellent performance measures achieved by the model. The results show that the model can quickly identify DFU infections, which is promising for its future usage in healthcare. In difficult cases, the hybrid CNN and Swin Transformer architecture effectively combines the strengths of both models to correctly identify infected or non-infected DFU images. With the insights it provides into the model's decision-making process, Grad CAM aids in the localisation of sick areas in DFU photos. Clinical evaluation and management of DFU infections could be enhanced utilising this technique.

After utilising medical digital image processing to reconstruct the 3D model of blood vessels, Wu et al. [35] utilised this data, together with the original data's non-calcified blood vessel part, as training material for deep learning (DL). Their idea for a better model to identify small targets in vascular stenosis was based on YOLOv5. In order to enhance the identification of three distinct stenosis targets (I: <50%, II: 51%–99%, and III: entirely occluded), they substituted the Path Aggregation Network (PANET) with the Bidirectional Feature Pyramid Network (BiFPN), enhanced the backbone with the Convolutional Block Attention Module (CBAM), and swapped out C3 in the neck for GhostC3. For faster reasoning and convergence, they upgraded the Complete-

IoU (CIoU) loss function to Alpha-Scylla-IoU (ASIoU) loss. Additionally, they opted for K-Means++ over K-Means to enhance algorithm convergence performance. Lastly, they benchmarked approach against five other popular models. With a mean Average Precision (mAP) of 85.40% and 74.60 FPS, this method was the most effective in detecting three different kinds of stenosis in diabetic feet, demonstrating the viability of utilising DL for this purpose.

3 Proposed Methodology

In this section, the proposed methodology is used to segment the DFU from the input images that is graphically shown in Fig. 1 and each of the block is explained in detailed.

The dataset was captured using two infrared cameras, the FLIR E60 and FLIR E6. Both cameras were configured to a standard resolution for capturing plantar thermograms, ensuring consistency across the dataset.

The cameras used identical operational conditions, including placement 1 m from the subject's feet and operation within a controlled temperature environment ($20 \pm 1^\circ\text{C}$). This ensured uniform data acquisition.

The setup included an IR-obstructing material applied to areas outside the plantar region to prevent thermal interference from other body parts. The data collection focused on capturing natural thermal emission from the plantar surface under a resting condition, avoiding external heating or cooling sources.

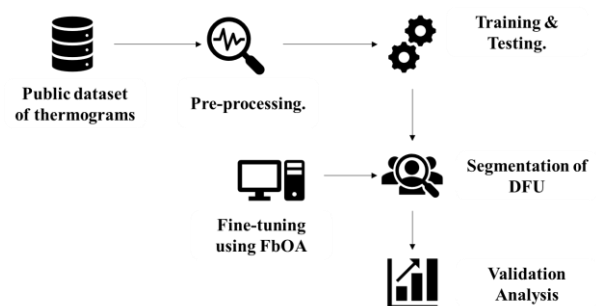


Fig. 1 Workflow of the proposed model.

3.1 Dataset Description and Augmentation

This study utilises a publicly available dataset consisting of thermograms [36]. Included in this data set are 334 plantar thermograms taken from 122 people with a diabetes mellitus (DM) diagnosis and 45 healthy controls. Sixteen women and twenty-nine men, ranging in age from 25 to 35, made up the DM group. There are 89 females and 33 males in the control group, and their ages ranged from 45 to 65. The participants are scouted from Puebla, Mexico, and their thermograms are taken over the course of 3 years (2012–2014). See Fig. 2 for an example of a pair of feet from each category [37].

The dataset comprises 334 plantar thermograms, with 122 thermograms from diabetic patients and 45 from healthy controls. The diabetic group includes individuals aged 25–35, with a gender distribution of 16 women and

29 men. The healthy control group comprises participants aged 45–65, with 89 females and 33 males. Exclusion criteria were applied to ensure data quality, eliminating patients with unrelated foot conditions or insufficient thermographic records. While comorbidities were not explicitly recorded, this remains a limitation of the study and a focus for future enhancement. These rigorous inclusion and exclusion criteria ensure the dataset's reliability for analyzing diabetic foot conditions. Four more pictures depicting plantar angiosomes are included in every thermogram. In this study, those are referred to as patches. Fig. 3 shows an example of both full-image and matching patch images (plantar angiosomes) from both groups. An infrared (IR) camera was placed 1m away from the subject's feet as they slept on a bed in order to acquire these images [37]. An IR obstructing substance was applied so that the temperature could not be detected from other parts of the body. At a temperature of 20 ± 1 °C, 2 infrared cameras, the FLIR E60 and the FLIR E6, are employed [37]. The segmented foot and patch RGB images are utilised without pre-processing since they are already part of the database. Nevertheless, data augmentation is employed to both enlarge the

dataset and achieve class parity. The process of augmentation involves flipping the photos horizontally or vertically, or both at the same time, as well as rotating them at 90°, 180°, or 270°. For image-level augmentation, 500 samples are made for every class, and for patch-level augmentation, 1500 samples are made for every class. Tabulated in Table 1 is the thermogram data broken down by class.

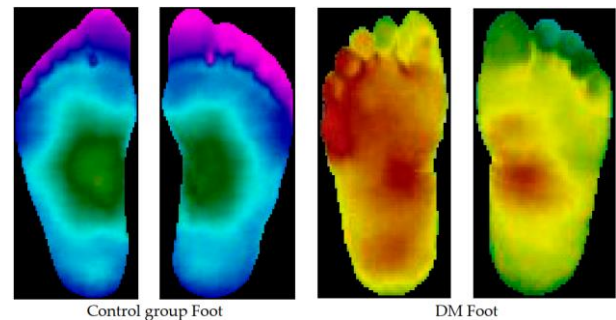


Fig. 2 Sample images of healthy foot and DM foot [37]. Reprinted under the Creative Common CC BY license from Ref. [37].

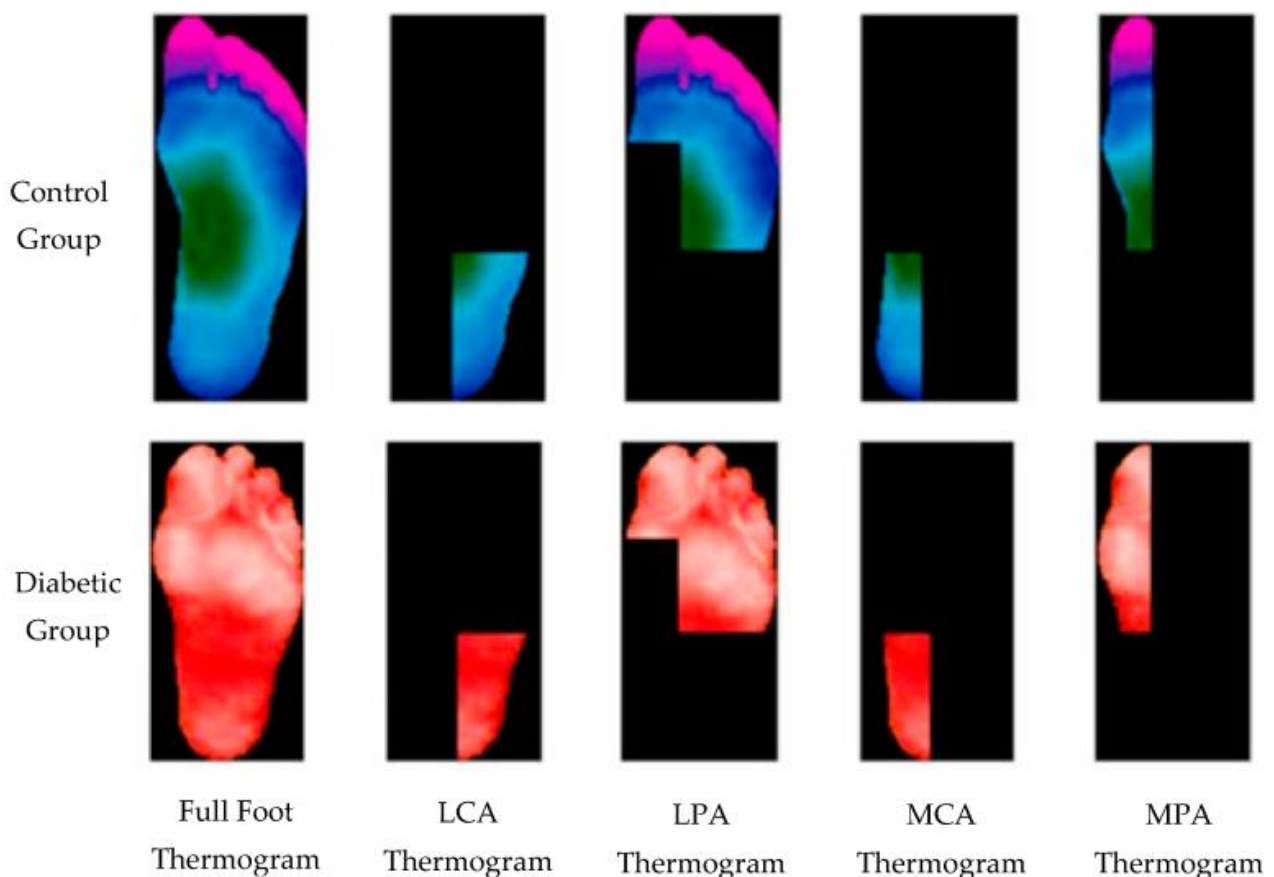


Fig. 3 Patch pictures (plantar angiosomes) and thermograms (for both the diabetes and control groups) are shown below. LCA stands for lateral calcaneal artery, LPA for lateral plantar artery, MCA for medial calcaneal artery, MPA for medial plantar artery, and so on [37]. Reprinted under the Creative Common CC BY license from Ref. [37].

Table 1 Thermogram image data detail: original and after augmentation.

Category	Diabetic Group	Control Group	Total
No. of cases	122	45	167
Original full images	244	90	334
Images after augmentation	500	500	1000
Original patches	976	360	1336
Patches after augmentation	1500	1500	3000

3.2 Segmentation Method

The research builds its model on diffusion models, which are generative models with two stages – a forward diffusion stage and a reverse diffusion stage – to extract the illness region from the input photos. Over the course of T iterations, the forward process corrupts a segmentation label x_0 with Gaussian noise. By turning the noising process backwards, the initial data can be recovered by training a neural network. One way to depict this is as:

$$p_\theta(x_{0:T-1}|x_T) = \prod_{t=1}^T p_\theta(x_{t-1}|x_t), \quad (1)$$

where θ is parameters for the reverse process. Starting from a Gaussian noise distribution, $p_\theta(x_T) = N(x_T; 0, I_{n \times n})$, where I is the original duplicate, the reverse process transforms the latent variable distribution $p_\theta(x_T)$ to the data distribution $p_\theta(x_0)$. The reverse process is symmetrical with respect to the forward process since it recovers the noisy image in stages to achieve the final, clear segmentation.

We use a UNet as our learning network to follow the typical DPM implementation. We used the raw image previously to condition the step estimate function ϵ in order to do the segmentation:

$$\epsilon_\theta(x_t, I, t) = D((E_t^I + E_t^x, t), t), \quad (2)$$

where E_t^I is the conditional feature embedding of the raw image, E_t^x at this stage is the segmentation map's feature embedding. In order to reconstruct, these 2 embeddings are combined and sent through a UNet decoder D . Along with the new embedding and decoding capabilities, the step index t is also included.

3.2.1 Dynamic Conditional Encoding

The conditional prior in the majority of conditional DPM implementations is an embedding of unique features. Nevertheless, ambiguous items are notoriously difficult to handle in medical picture segmentation, making it impossible to differentiate between the backdrop and lesions or tissues. Image modalities with low contrast increase this problem. Our proposed solution involves

using a Dynamic Conditional Encoding at each stage to fix this issue. While the raw image does include precise segmentation target information, it has a difficult time differentiating itself from the background. The current-step segmentation map, however, does include improved target areas, though it is far from accurate. We are therefore inspired to incorporate the segmentation information from the current phase into the conditional raw picture encoding in order to achieve mutual complement. Integration is implemented at the feature level in particular. To improve the raw image encoder's intermediate feature, it is necessary to enable the features of the current-step encoding. The conditional feature map scales encoding features (m_l^k, m_x^k) , where k is the layer index, are fused together to form a common shape. An attentive-like mechanism, A is used to accomplish the fusion. To create an affinity map using this method, it is need to normalise both feature maps using a layer and then multiply them. To further improve the attentive region, it is need to multiply the affinity map with the condition encoding features. Here is how it works:

$$\mathcal{A}(m_l^k, m_x^k) = (LN(m_l^k) \otimes LN(m_x^k)) \otimes m_l^k, \quad (3)$$

where $\otimes$ denotes element-wise multiplication, and LN denotes layer normalization. Both of the intermediate stages, which are convolutional stages built according to ResNet34, undergo this process.

The model is able to calibrate and localise itself in real-time with the use of this Dynamic Conditional Encoding technique. The incorporation of the noise-corrupted x_t embedding, however, might bring in extra high-frequency noise. In order to tackle this problem, the study suggests using the FF-Parser to limit the features' high-frequency components.

3.2.2 FF-Parser

The study uses FF-Parser as part of the feature integration pathway to fix the problem of high frequency noise caused by x_t embedding integration. The purpose of FF-Parser is to limit the x_t features that are associated with noise. Acquiring a parameterised attentive map for use with Fourier-space characteristics is the central concept. To initiate the process, we apply a 2D fast Fourier transform (FFT) to each spatial dimension of a decoder feature map $F[m] \in R^{H \times W \times C}$, which can be expressed as:

$$M = F[m] \in R^{H \times W \times C}, \quad (4)$$

where $F[\cdot]$ denotes the 2D FFT. Next, we modulate the spectrum of m by multiplying a parameterized attentive map $A \in C^{H \times W \times C}$ to M :

$$M' = A \otimes M, \quad (5)$$

where $\otimes$ signifies the product of elements. The last step is to use inverse FFT to return M' to the spatial domain:

$$m' = \mathcal{F}^{-1}[M']. \quad (6)$$

A frequency filter is a common tool in digital image processing; FF-Parser is like a learnable version of it. While spatial attention focusses on individual frequencies, FF-Parser modifies their components on a global scale. As a result, it is trainable to limit adaptive integration's high-frequency components.

3.2.3 Training and Architecture

The training procedure for the proposed model adheres to the DPM standard. To be more precise, the loss can be expressed as:

$$\mathcal{L} = E_{x_0, \epsilon, t} \left[\left\| \epsilon - \epsilon_0 (\sqrt{\hat{x}_t x_0} + \sqrt{1 - \hat{x}_t} \epsilon, I_t, t) \right\|^2 \right]. \quad (7)$$

In each iteration, a random pair of raw images I_i and segmentation label S_i is sampled for training. The iteration sum is sampled from a uniform distribution, besides ϵ is sampled from a Gaussian distribution.

Using a ResNet encoder and a UNet decoder, we can construct a modified ResUNet, which forms the core architecture of the suggested model. Two separate encoders are used to encode I and x_t . There are several residual blocks in each encoder's convolution stages. Each stage's number of residual blocks is based on ResNet34's. The two convolutional blocks that make up each residual block are as follows: a group-normalization layer, a SiLU activation layer, and finally, a convolutional layer. After passing through a linear layer, a SiLU activation, and still another linear layer, the time embedding reaches the residual block. After that, the output of the initial convolutional block is supplemented with the result. The final step of the encoding process involves adding the E^I and E^{x_t} that have been obtained. In order to forecast the end segmentation outcome, a conventional convolutional decoder is linked.

3.2.4 Fine-Tuning Using Football Optimizer

In this research, the learning parameters of the proposed segmentation model are optimally tuned using the Football Optimizer (FbOA) [38], as elaborated here. A comprehensive experimental setup is designed to evaluate the performance and effectiveness of the proposed FbOA. The objective is to assess the algorithm's exploration and exploitation capabilities and its ability to avoid local optima. The implementation includes an in-depth analysis of the algorithm's exploration and exploitation benchmarks, incorporating elements such as dynamic decision-making and ball-passing strategies inspired by football tactics. This process employs various mathematical equations and update procedures that simulate the strategic approaches used by football players, ensuring a balanced exploration and exploitation of productive zones within the solution space. The following sub-sections provide a detailed discussion of these aspects, beginning with the algorithm's exploration performance.

3.2.4.1 Exploration Performance

In the FbOA, exploration performance is based on the assumption that in every football pass, one of the players performs his pass with a certain speed velocity. This speed variation is important to balance the exploration of the search space and exploitation of the existing regions to avoid getting trapped in the local optimum.

To model this dynamic process, the next stage of the algorithm, represented by $Fb(S(t + 1))$, is defined as:

$$Fb(S(t + 1)) = i, \quad (8)$$

where $Fb(S(t + 1))$ denotes the updated state or solution at the next iteration $t + 1$, i represents the player as the agent involved in the optimization process, contributing to the variation in exploration.

Football Velocity

The Eq. (9) represents the velocity of each player (or agent):

$$V_n = F_{max} \left(b_x \cdot a_i [F_{ext} - F_{min}] + r \cdot b_y \cdot a_j [F_{best} - F_{min}] \times \cos\left(\frac{\pi}{Iteration}\right) \right), \quad (9)$$

where V_n is the velocity of the player (agent) at iteration n , F_{max} is the maximum force applied, representing the upper limit of velocity, b_x and b_y are coefficients determining the influence of different directions in the search space, a_i and a_j are acceleration factors impacting the speed adjustments, F_{ext} represents the external force or influence in the optimization context, F_{min} is the minimum force, indicating the lower bound of velocity, F_{best} is the best force or position found so far in the optimization process, r is a random factor introducing variability and enhancing exploration, $\cos\left(\frac{\pi}{Iteration}\right)$ modulates the velocity based on the number of iterations to balance exploitation.

The equation captures the dynamic nature of exploration by allowing agents to adapt their velocities, thus effectively covering a broader search space.

Update for Best Force

To further refine the exploration process, the best force is updated using the following Eq. (10):

$$F_{best} = \frac{1}{K} \sum_{n=0}^K \left(\frac{F_{max}^n}{(2n+1)^2} \right), \quad (10)$$

where F_{best} represents the best position or force in the current optimization context, K is an exponential factor that increases from 0 to 1 throughout the iterations, balancing between exploration and exploitation, F_{max} is raised to the power of n^2 , which intensifies the search focus on promising regions as the iterations proceed,

$(2n + 1)^2$ is a normalization term to control the update rate.

By continuously adjusting the velocities and updating the best force, FbOA enhances its ability to explore the solution space thoroughly while adapting to the iterations.

3.2.4.2 Exploitation Performance

The tuning aspect of the FbOA in exploitation is centered on enhancing the solutions within the specified search space to get the best or nearly the best solutions. The exploitation mechanism is meant to further focus on the more rewarding areas during the exploration. The update rule for adjusting the position of an agent in the exploitation phase is given by:

$$Fb(S(t + 1)) = F_i + z_3 \cdot Fb(S(t)) + K \cdot \sin\left(\frac{\pi}{Iteration}\right), \quad (11)$$

where $Fb(S(t + 1))$ is the updated solution or position at the next iteration $t + 1$, F_i represents the current position or force influencing the aggrandisement, z_3 is a control parameter that adjusts the balance between the current and the new position, $Fb(S(t))$ is the current state or position at iteration t , K is an exponential factor that increases throughout the iterations, gradually shifting focus from exploration to exploitation, $\sin\left(\frac{\pi}{Iteration}\right)$ introduces a sinusoidal modulation to enhance the convergence speed toward the optimal solution.

This update mechanism helps FbOA take advantage of the solution space and ensures that FbOA converges on precise solutions during the iterations.

Mutation

To avoid this, the Football Optimization Algorithm used in this study incorporates a mutation strategy that introduces variability into the search process, in which the algorithm initially converges with local optima. Through the mutation mechanism, the site of the agents changes constantly in the search space, thus making a transition and avoiding being trapped in an area with poor solutions. The mutation is defined by:

$$S(t) = \left(K \cdot a_q \left(\frac{2n+1}{x}\right) + K \cdot \cos\left(\frac{\pi}{Iteration}\right)\right), \quad (12)$$

where $S(t)$ represents the mutated solution at iteration t , K is an exponential factor influencing the extent of mutation, a_q is a factor that scales the mutation magnitude based on the current solution, $\frac{2n+1}{x}$ acts as a normalization term to control the mutation effect, $\cos\left(\frac{\pi}{Iteration}\right)$ modulates the mutation with a cosine function to introduce randomness and diversity in the search, iteration represents the current iteration number.

By incorporating this mutation strategy, FbOA can search distant solution areas and thus avoid local optima in the optimization framework, resulting in higher

performance and robustness of the algorithm. Table 2 displays the initial parameter values used in this study for all the algorithms: population size, sum of iterations and independent runs. All these general parameters are set for all the algorithms with no exception. Additionally, specific parameters for the Feedback-based Optimization Algorithm (FbOA) are detailed, such as the ranges for constants a_1, a_2, b_1, b_2 , and other control variables like r_1, z, θ , and a , which influence the behaviour of the algorithm.

Table 2 Initial parameters of algorithms.

Algorithm	Parameter	Value
All algorithms	Population size	30
	Sum of iterations	500
	Number of runs	30
FbOA	a_1	[0, 1]
	a_2	[0, 1]
	b_1	[0, 1]
	b_2	[0, 1]
	r_1	[0, 2]
	z	[0, 2]
	Θ	[0, 12π]
	a	[-8, 8]

4 Results and Discussion

The PyTorch platform was used to implement all experiments, and four Tesla P40 GPUs with 24GB of RAM are used for training and testing [39]. The dimensions of all the photographs are adjusted to 256×256 pixels. With the help of the FbOA optimiser, the networks are trained end-to-end. The rate was set to 10^{-4} besides the batch size was 32 when the model was trained [40].

4.1 Validation Analysis of Input Data

Here, 3 set of input data is used to test the effectiveness of the projected model that is shown in Table 3.

4.2 Validation Analysis of Training and Testing Ratio Data

Fig. 4 mentions the efficiency of the proposed prototypical based on training and testing data variations in terms of different metrics.

The proposed model's performance was assessed across different data split ratios (80–20%, 70–30%, and 60–40%) to examine its effectiveness in training and testing scenarios. For an 80–20% split, the model accomplished a high training accuracy of 0.98 and a test accuracy of 0.92, with an AUC of 0.92, precision at 0.93, recall and F_1 -score both at 0.92, and a specificity of 0.98, indicating balanced sensitivity and specificity.

Table 3 Validation analysis based on input types.

Parameters	Input data	MPA	Mean Intersection over Union (MioU)	Running Time
86 M	Image level	0.8331	0.7058	2.67 s/item
307 M	Patch level	0.8313	0.7151	4.07 s/item
632 M	Combined	0.8376	0.7185	4.2 s/item

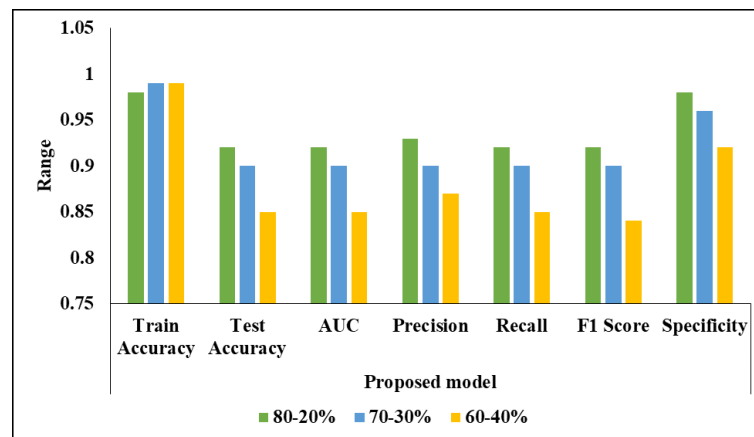


Fig. 4 Graphical description of proposed model.

In the 70–30% split, the model demonstrated a slightly higher train accuracy of 0.99 and a test accuracy of 0.90, with AUC, precision, recall, and F_1 -score all at 0.90, and a specificity of 0.96, showcasing consistent performance despite a larger test set. For the 60–40% split, the model maintained a high train accuracy of 0.99 but experienced a slight drop in test accuracy to 0.85. Its AUC, recall, and F_1 -score stood at 0.85, while precision was 0.87 and specificity was 0.92, reflecting the impact of reduced training data. Overall, the proposed model displayed robust accuracy and balanced metrics across varying data splits, demonstrating its adaptability and effectiveness in different training and testing conditions.

4.3 Validation Analysis of Input Batch Size

Table 4 presents the validation analysis of proposed model by changing its batch sizes in terms of different metrics.

The experimental analysis of the proposed model across different batch sizes (4, 8, 16, and 32) highlights its performance in training besides testing accuracy, AUC, precision, recall, F_1 -score, and specificity. For batch sizes of 4 and 8, the model achieved perfect test accuracy of 1.00, with all evaluative metrics (AUC, precision, recall, F_1 -score, and specificity) at 1.00, indicating optimal model performance and consistency in both training and testing phases. At a batch size of 16, although the model reached full training accuracy (1.00), the test accuracy and all other metrics dropped significantly to 0.50, reflecting underperformance likely due to batch size inefficiencies. Similarly, for a batch size of 32, while training accuracy remained perfect (1.00), the test accuracy and other evaluative metrics are moderate at 0.75. This comparison indicates that smaller

batch sizes (4 and 8) provided the best performance consistency, while larger batch sizes resulted in reduced generalization, as observed by lower test accuracy and evaluative scores.

4.4 Comparative Analysis: Segmentation Results

Table 5 and Fig. 4 provide the experimental analysis of projected model with existing techniques that are taken from the related work section. However, the existing models use different datasets for validation, the basic model is implemented by our considered dataset and results are averaged in this work.

The performance comparison of various methods, with Support Vector Machine (SVM), DNN, Deep Convolutional Neural Network (DCNN), ResNet50, MobileNetV2, HCNNet, FusionNet, and the proposed Op-DPS model, is summarized across key metrics: precision, recall, F_1 -score, overall accuracy, besides IoU. The Op-DPS model demonstrates the highest precision at 0.8955 and recall at 0.8553, resulting in an F_1 -score of 0.8750, indicating its superior capability to balance precision and recall effectively. Additionally, it achieves an overall accuracy of 0.9927 and an IoU of 0.8896, reflecting exceptional performance in accurately classifying instances. In comparison, FusionNet follows closely with an F_1 -score of 0.8678 and an overall accuracy of 0.9807, while the remaining methods show lower metrics across the board, with DNN recording the lowest F_1 -score of 0.7995 besides an IoU of 0.8476. Overall, the results highlight the Op-DPS model as the most effective approach among those evaluated, excelling in all critical performance indicators.

Table 4 Experimental Analysis of proposed model on various batch size.

Batch Size	Test accuracy	AUC	Precision	Recall	F ₁ -score	Train accuracy	Specificity
4	1.00	1.00	1.00	1.00	1.00	0.98	1.00
8	1.00	1.00	1.00	1.00	1.00	0.98	1.00
16	0.50	0.50	0.50	0.50	0.50	1.00	0.50
32	0.75	0.75	0.75	0.75	0.75	1.00	0.75

Table 5 Comparative analysis of various algorithms.

Method	Precision	Recall	F ₁ -score	Overall Accuracy	IoU
SVM [32]	0.8639	0.8178	0.8402	0.9780	0.8505
DNN [21]	0.8891	0.7264	0.7995	0.9742	0.8476
DCNN [25]	0.8343	0.7367	0.7825	0.9710	0.8061
ResNet50 [22]	0.8827	0.7944	0.8363	0.9780	0.8477
MobileNetV2 [24]	0.8848	0.8139	0.8479	0.9793	0.8570
HCNNNet [27]	0.8758	0.8367	0.8558	0.9801	0.8634
FusionNet [29]	0.8856	0.8231	0.8678	0.9807	0.8652
Op-DPS	0.8955	0.8553	0.8750	0.9927	0.8896

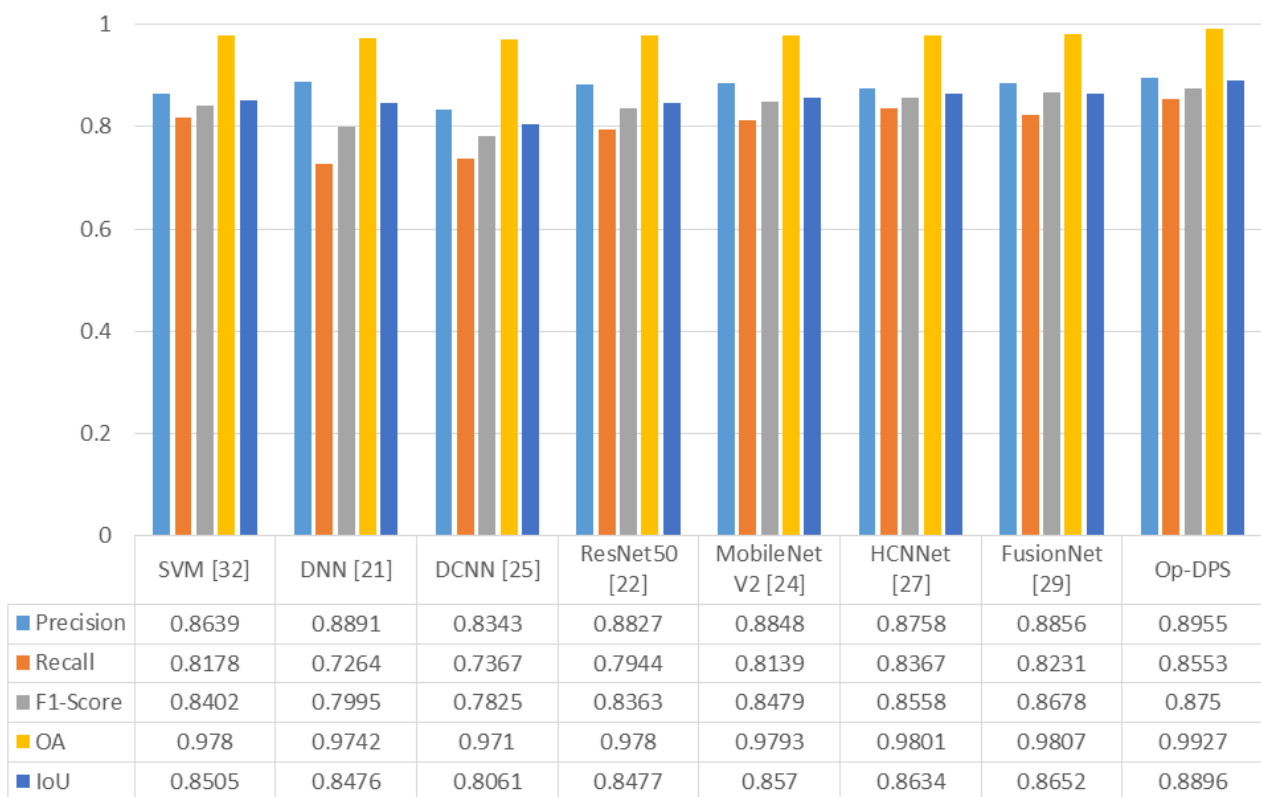


Fig. 5 Graphical Description of proposed model with existing models.

Baseline classification algorithms such as SVM, DNN, and ResNet50 were optimized using conventional techniques like grid search and random search. These techniques explored the parameter space by systematically evaluating combinations of hyperparameters, such as learning rate, regularization factors, and activation functions. In comparison, the proposed approach utilized the Football Optimization Algorithm (FbOA), which mimics football dynamics to optimize model parameters efficiently. FbOA's strategies, such as velocity adjustment and mutation, enable it to explore a larger solution space and avoid local optima, leading to superior parameter tuning. This difference in optimization techniques ensures the proposed model's performance is not underestimated relative to the baseline methods.

5 Conclusion

The segmentation of DFU foot is demonstrated in this paper by means of a thorough and comparative examination of deep-learning and machine-learning procedures. At the image, patch, and combined image-patch levels, the thermogram data of both DFU besides non-DFU foot are utilised. Image segmentation system Op-DPS is introduced in this paper. It improves segmentation performance by combining 2 unique techniques: FF-Parser and Dynamic Conditional Encoding. Results showed that FbOA converged more quickly, solved the problem accurately, and was more robust when used to fine-tune the parameters of the suggested model. An algorithm that optimises a search space by minimising the chances of getting stuck by local

optimality through high exploration and low exploitation incorporates many football techniques. To show the outcomes of the machine-learning methods we used a number of thermogram characteristics. From the various machine-learning approaches, the suggested model emerged victorious. The suggested deep-learning model outperformed both the pre-trained and competing models. Using the thermogram data, comparing the suggested model to an existing solution was not easy. Nevertheless, when looking at the 2 models at a more abstract level, the suggested model outperformed the others considering the limitations.

Data Availability Statement

The dataset considered in this work are taken from publicly available site, which is referred in Refs. [36–37].

Conflict of Interest

Author declares there is no conflict of interest.

Acknowledgment

We thank all the staff of our institution for their cooperation and kind support throughout the course of our research.

Funding Statement

Author declares there is no funding are provided for this work.

References

1. C. C. Kendrick, B. Cassidy, J. M. Pappachan, C. O'Shea, C. J. Fernandez, E. Chacko, K. Jacob, N. D. Reeves, and M. H. Yap, "Translating Clinical Delineation of Diabetic Foot Ulcers into Machine Interpretable Segmentation," in Diabetic Foot Ulcers Grand Challenge, M. H. Yap, C. Kendrick, and R. Brüngel (Eds.), Springer Nature Switzerland, 15335 (2025).
2. A. Mahbod, G. Schaefer, R. Ecker, and I. Ellinger, "Automatic foot ulcer segmentation using an ensemble of convolutional neural networks," in 26th International Conference on Pattern Recognition (ICPR), IEEE, 4358–4364 (2021).
3. R. V. Prakash, K. S. Kumar, "Development of automatic segmentation techniques using convolutional neural networks to differentiate diabetic foot ulcers," International Journal of Advanced Computer Science and Applications 13(11), (2022).
4. J. Reyes-Luévano, J. A. Guerrero-Viramontes, J. Rubén Romo-Andrade, and M. Funes-Gallanzi, "DFU_VIRNet: A novel Visible-InfraRed CNN to improve diabetic foot ulcer classification and early detection of ulcer risk zones," Biomedical Signal Processing and Control 86, 105341 (2023).
5. D. Kucharski, A. Kostuch, F. Noworolnik, A. Brodzicki, and J. Jaworek-Korjakowska, "DFU-Ens: End-to-End diabetic foot ulcer segmentation framework with vision transformer based detection," in Diabetic Foot Ulcers Grand Challenge, M. H. Yap, C. Kendrick, and B. Cassidy (Eds.), Springer International Publishing, 13797, 101–112 (2023).
6. T.-Y. Liao, C.-H. Yang, Y.-W. Lo, K.-Y. Lai, P.-H. Shen, and Y.-L. Lin, "HardNet-DFUS: enhancing backbone and decoder of HardNet-MSEG for diabetic foot ulcer image segmentation," in Diabetic Foot Ulcers Grand Challenge, M. H. Yap, C. Kendrick, and B. Cassidy (Eds.), Springer International Publishing, 13797, 21–30 (2023).
7. H.-N. Huang, T. Zhang, C.-T. Yang, Y.-J. Sheen, H.-M. Chen, C.-J. Chen, and M.-W. Tseng, "Image segmentation using transfer learning and Fast R-CNN for diabetic foot wound treatments," Frontiers in Public Health 10, 969846 (2022).

8. A. Kairys, R. Pauliukienė, V. Raudonis, and J. Ceponis, “Towards home-based diabetic foot ulcer monitoring: a systematic review,” *Sensors* 23(7), 3618 (2023).
9. B. Gunapriya, T. Rajesh, A. Thirumalraj, and B. Manjunatha, “LW-CNN-based extraction with optimized encoder-decoder model for detection of diabetic retinopathy,” *Journal of Autonomous Intelligence* 7(3), (2023).
10. H. J. Chae, S. Lee, H. Son, S. Han, and T. Lim, “Generating 3D bio-printable patches using wound segmentation and reconstruction to treat diabetic foot ulcers,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2539–2549 (2022).
11. D. Bouallal, H. Douzi, and R. Harba, “Diabetic foot thermal image segmentation using Double Encoder-ResUnet (DE-ResUnet),” *Journal of Medical Engineering & Technology* 46(5), 378–392 (2022).
12. H. Yi, W. Xu, Z. Jiang, J. Gao, Q. Kang, Q. Lao, and K. Li, “OCRNet for diabetic foot ulcer segmentation combined with edge loss,” *Diabetic Foot Ulcers Grand Challenge* 13797, 31–39 (2023).
13. M. Hassib, M. Ali, A. Mohamed, M. Torki, and M. Hussein, “Diabetic foot ulcer segmentation using convolutional and transformer-based models,” in *Diabetic Foot Ulcers Grand Challenge*, M. H. Yap, C. Kendrick, and B. Cassidy (Eds.), Springer International Publishing, 13797, 83–91 (2023).
14. A. Kairys, V. Raudonis, “Analysis of training data augmentation for diabetic foot ulcer semantic segmentation,” *Electronics* 12(22), 4624 (2023).
15. P. Xu, X. Wu, Y. Li, E. U. Haq, J. Yin, and K. Li, “Ensemble learning for diabetic foot ulcer segmentation based on DFUC2022 dataset,” *Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering*, 1750–1754 (2022).
16. R. Brüngel, S. Koitka, and C. M. Friedrich, “Unconditionally generated and pseudo-labeled synthetic images for diabetic foot ulcer segmentation dataset extension,” in *Diabetic Foot Ulcers Grand Challenge*, M. H. Yap, C. Kendrick, and B. Cassidy (Eds.), Springer International Publishing, 13797, 65–79 (2023).
17. J. Zhang, Y. Qiu, L. Peng, Q. Zhou, Z. Wang, and M. Qi, “A comprehensive review of methods based on deep learning for diabetes-related foot ulcers,” *Frontiers in Endocrinology* 13, 945020 (2022).
18. G. Scebbba, J. Zhang, S. Catanzaro, C. Mihai, O. Distler, M. Berli, and W. Karlen, “Detect-and-segment: A deep learning approach to automate wound image segmentation,” *Informatics in Medicine Unlocked* 29, 100884 (2022).
19. C. Cao, Y. Qiu, Z. Wang, J. Ou, J. Wang, A. H. Hounye, M. Hou, Q. Zhou, and J. Zhang, “Nested segmentation and multi-level classification of diabetic foot ulcer based on mask R-CNN,” *Multimedia Tools and Applications* 82(12), 18887–18906 (2023).
20. J. P., S. K. B. K., and S. Jayaraman, “Automatic foot ulcer segmentation using conditional generative adversarial network (AFSegGAN): A wound management system,” *PLOS Digit Health* 2(11), e0000344 (2023).
21. R. Sarman, M. E. H. Chowdhury, M. Murugappan, A. Aqel, M. Ezzuddin, S. M. Rahman, A. Khandakar, S. Akter, R. Alfkey, and A. Hasan, “Diabetic foot ulcer detection: combining deep learning models for improved localization,” *Cognitive Computation* 16(3), 1413–1431 (2024).
22. A. M. El-Kady, M. M. Abbassy, H. H. Ali, and M. F. Ali, “Advancing diabetic foot ulcer detection based on resnet and gan integration,” *Journal of Theoretical and Applied Information Technology* 102(6), 2258–2268 (2024).
23. R. Basiri, K. Manji, P. M. LeLievre, J. Toole, F. Kim, S. S. Khan, and M. R. Popovic, “Protocol for metadata and image collection at diabetic foot ulcer clinics: enabling research in wound analytics and deep learning,” *BioMedical Engineering OnLine* 23(1), 12 (2024).
24. S. Biswas, R. Mostafiz, B. K. Paul, K. M. M. Uddin, Md. A. Hadi, and F. Khanom, “DFU_XAI: A deep learning-based approach to diabetic foot ulcer detection using feature explainability,” *Biomedical Materials & Devices* 2(2), 1225–1245 (2024).
25. N. F. Almufadi, H. F. Alhasson, and S. S. Alharbi, “E-DFu-Net: An efficient deep convolutional neural network models for diabetic foot ulcer classification,” *Biomolecules and Biomedicine* 25(2), 445–460 (2025).
26. C. Thota, D. Jackson Samuel, M. Musa Jaber, M. M. Kamruzzaman, R. V. Ravi, L. J. Gnanasigamani, and R. Premalatha, “Image smart segmentation analysis against diabetic foot ulcer using internet of things with virtual sensing,” *Big Data* 12(2), 155–172 (2024).
27. S. K. Das, S. Namasudra, and A. K. Sangaiah, “HCNNet: hybrid convolution neural network for automatic identification of ischaemia in diabetic foot ulcer wounds,” *Multimedia Systems* 30(1), 36 (2024).
28. M. K. Dhar, T. Zhang, Y. Patel, S. Gopalakrishnan, and Z. Yu, “FUSegNet: A deep convolutional neural network for foot ulcer segmentation,” *Biomedical Signal Processing and Control* 92, 106057 (2024).
29. S. Biswas, R. Mostafiz, M. S. Uddin, and B. K. Paul, “XAI-FusionNet: Diabetic foot ulcer detection based on multi-scale feature fusion with explainable artificial intelligence,” *Heliyon* 10(10), e31228 (2024).
30. N. Zhao, L. Yu, X. Fu, W. Dai, H. Han, J. Bai, J. Xu, J. Hu, and Q. Zhou, “Application of a Diabetic Foot Smart APP in the measurement of diabetic foot ulcers,” *International Journal of Orthopaedic and Trauma Nursing* 54, 101095 (2024).
31. F. M. Lusendi, A.-S. Vanherwegen, F. Nobels, and G. A. Matricali, “A multidisciplinary Delphi consensus to define evidence-based quality indicators for diabetic foot ulcer care,” *European Journal of Public Health* 34(2), 253–259 (2024).

32. S. Hong, Y. Chen, Y. Lin, X. Xie, G. Chen, H. Xie, and W. Lu, “[Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms](#),” *Technology and Health Care* 32, 265–276 (2024).
33. M. Shiraishi, H. Lee, K. Kanayama, Y. Moriwaki, and M. Okazaki, “[Appropriateness of artificial intelligence chatbots in diabetic foot ulcer management](#),” *The International Journal of Lower Extremity Wounds* 15347346241236811 (2024).
34. M.G Sumithra, C. Venkatesan, “[SwinDFU-Net: Deep learning transformer network for infection identification in diabetic foot ulcer](#),” *Technology and Health Care* 33(1), 601–618 (2025).
35. C. Wu, C. Xu, S. Ou, X. Wu, J. Guo, Y. Qi, and S. Cai, “[A novel approach for diabetic foot diagnosis: Deep learning-based detection of lower extremity arterial stenosis](#),” *Diabetes Research and Clinical Practice* 207, 111032 (2024).
36. D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. de J. Rangel-Magdaleno, and F. J. Renero-Carrillo, “[Plantar thermogram database for the study of diabetic foot complications](#),” *IEEE Access* 7, 161296–161307 (2019).
37. I. Khosa, A. Raza, M. Anjum, W. Ahmad, and S. Shahab, “[Automatic diabetic foot ulcer recognition using multi-level thermographic image data](#),” *Diagnostics* 13(16), 2637 (2023).
38. E.-S. M. El-Kenawy, F. H. Rizk, A. M. Zaki, M. E. Mohamed, A. Ibrahim, A. A. Abdelhamid, N. Khodadadi, E. M. Almetwally, and M. M. Eid, “[Football Optimization Algorithm \(FbOA\): A novel metaheuristic inspired by team strategy dynamics](#),” *Journal of Artificial Intelligence and Metaheuristics* 8(1), 21–38 (2024).
39. R. Aluvalu, T. Sharma, U. M. Viswanadhula, A. D. Thirumalraj, M. V. V. Prasad Kantipudi, and S. Mudrakola, “[Komodo dragon mlpir algorithm-based CNN model for detection of illegal tree cutting in smart iot forest area](#),” *Recent Advances in Computer Science and Communications* 17(6), e260124226370 (2024)
40. V. Uma Maheswari, S. Stephe, R. Aluvalu, A. Thirumalraj, and S. N. Mohanty, “[Chaotic satin bowerbird optimizer based advanced ai techniques for detection of COVID-19 diseases from CT scans images](#),” *New Generation Computing* 42(5), 1065–1087 (2024).

Non-Contact Neonatal Heart Rate Monitoring Using Blue Filter Technique with Photoplethysmography Imaging under UV Illumination

Huda Ali Hashim

Middle Technical University, Electrical Engineering Technical College, Baghdad, Iraq

e-mail: huda.ali@mtu.edu.iq

Abstract. Non-contact heart rate (HR) monitoring is essential for neonates, as traditional adhesive sensors can damage their delicate skin. The rise in contagious viral infections further emphasizes the need for contactless neonatal healthcare monitoring systems to reduce physical interaction and the risk of disease transmission. This study proposes a novel non-contact HR monitoring system for neonates, combining a blue-filter technique with skin color enhancement under UV illumination. The system utilizes digital camera-based photoplethysmography imaging and a tile-based region of interest (ROI) selection method. Also, a neonatal graphical user interface (GUI) was developed for real-time HR monitoring and clear visualization of findings. We compared the findings of the proposed technique with standard UV photoplethysmography imaging without the blue filter technique to validate its effectiveness. Experimental findings demonstrate the proposed system's high performance, with a Pearson correlation coefficient (PCC) of 0.998 and a mean absolute error (MAE) of 1.12 bpm, closely matched to ECG monitor readings. These findings highlight the system's potential as a reliable, non-contact solution for neonatal HR monitoring in the NICU, offering a promising advancement in neonatal care and broader biomedical applications. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: heart rate; neonatal; digital camera; biomedical.

Paper #9219 received 14 Jan 2025; revised manuscript received 8 Feb 2025; accepted for publication 13 Mar 2025; published online 29 Mar 2025. doi: [10.18287/JBPE25.11.010305](https://doi.org/10.18287/JBPE25.11.010305).

1 Introduction

Monitoring neonatal heart rate is critical for accurately and continuously assessing their health status [1, 2]. This process facilitates the early detection of changes in temperature, heart rate, or respiratory patterns, enabling timely interventions to enhance safety and mitigate potential health risks [3, 4]. In emergency situations, traditional heart rate monitoring devices primarily rely on contact-based sensors, such as electrodes, probes, or straps [5]. These methods often involve adhesive biosensors and electrodes [6], such as those utilized for electrocardiograms (ECG), spirometry, and pulse oximetry [7]. However, the application and removal of these sensors can damage the delicate skin of neonates and restrict their mobility, posing additional challenges in clinical settings [8].

In contrast, non-contact approaches to vital sign monitoring have emerged as a crucial area of research in biomedical and clinical applications [3, 4, 9]. These methods aim to minimize physical interaction and reduce the risk of disease transmission, particularly in the context of highly contagious environments. Such advancements hold significant promise for improving neonatal care while addressing the limitations of traditional contact-based systems [10].

Nevertheless, noise and motion artifacts mostly influence the previous non-contact approaches; these techniques are vulnerable to noise and motion artifacts, which restrict mobility [11]. Moreover, they pose risks to human health and may have various biological effects on physiological tissues [12, 13]. Moreover, they possess a limited distance range and need a region area of interest (ROI) for analytical measurements [14]. Consequently,

limiting the patient's distant mobility poses a significant challenge owing to the expense of the sensor, which hinders the collection of sufficient bodily samples. Their poor resolution limits the detection range. Moreover, the emitted radar radiation is likely detrimental to human health and may induce various biological consequences on bodily tissues in the future [15, 16].

These imperceptible cardiac signals may be very beneficial, particularly in biological and therapeutic contexts. Blood circulation induces small variations in human skin color that may be amplified to evaluate and extract important information on cardiac activity when epidermal tissue is lit, since hemoglobin absorbs light more effectively than the adjacent epidermal tissue [17, 18]. Consequently, many studies have enhanced a computer vision approach using video cameras as non-contact sensors [19].

Vital signs may be extracted from certain regions of the human epidermis using video camera imaging, which is an appropriate non-contact method. This innovative approach may be classified into 2 primary categories: color-based method, often known as imaging photoplethysmography (PPG) [20] and motion-based method [9]. Several studies [21–23] used color-based methods for non-contact monitoring of cardio signals. However, they obtained PPG signals at small distances of 0.07–0.25 m [24, 25], 0.5 m [26], 0.35 m [27], and 1–1.5 m [28]. Gibson et al. [29] introduced a non-contact infant HR value monitoring method. The proposed method uses 2 Nikon cameras to detect face color variations resulting from blood perfusion. The ROI was manually designated by delineating a box around the face. Poh et al. [30] presented a PPG-based video technique that used RGB data from the face area to measure HR values. The ROI was manually delineated in the early frames, whereas later frames used optical flow tracking. However, the proposed technique of optical flow tracking encountered considerable motion artifacts, resulting in inaccuracies in heart rate estimates. Alnaggar et al. [31] proposed new non-contact techniques for RR and HR estimation from the subjects' faces. The recorded videos are magnified using the magnification technique and then analyzed. The HR estimate was achieved by employing face mesh to annotate the boundaries by additionally annotating for the boundaries of the most ROI from the face image in both HSV and RGB and color images.

According to a review of the literature, the manual ROI selection method was mostly used for non-contact biological signal monitoring and had issues, such as the fact that adding the bounding boxes by hand was time-consuming and prone to error, which made the extracted HR signals less consistent and reliable. The previous methods were also inadequate in managing subject or patient motion, which might result in the face moving beyond the designated ROI. Moreover, the findings illustrate the constraints of the remote sensing technique in terms of range and field of vision. It has a restricted operational range of less than 1.5 m from the body. Furthermore, the existing studies reveal a substantial and

important gap owing to the inadequate investigation of non-contact neonatal vital sign monitoring in the NICU under UV illumination. The aforementioned limitations serve as the impetus for developing a monitoring system for neonatal HR values under UV illumination.

The primary goal of this study is to implement a neonatal healthcare system for non-contact monitoring. This work investigates the effectiveness of a proposed blue filter technique combined with a proposed neonatal ROI selection method to accurately extract HR values under UV illumination in the NICU settings. The contributions of this work are as follows:

- A non-contact neonatal HR monitoring system utilizing digital camera technology under UV illumination was developed.
- A new integration of the blue filter technique and EVM method was proposed to enhance neonatal cardiac signal extraction under UV illumination by reducing noise and amplifying subtle skin color variations.
- A robust tile-based ROI selection method for neonatal cardiac signal monitoring was introduced, ensuring accurate heart rate calculation despite infant movements, and eliminating the need for manual bounding box adjustments.
- A neonatal Graphical User Interface (GUI) system was designed and implemented to facilitate real-time monitoring and visualization of outcomes from the proposed non-contact heart rate monitoring system.

2 Methods and Materials

2.1 Subjects

We conducted this work at the Child Central Teaching Hospital (CCTH) in Baghdad, Iraq. We obtained consent from guardians to image videos using non-contact technology at suitable times when neonates are in quiet positions. We implemented this work after explaining its nature and the experiment's aim to collect samples without any contact with babies. We kept the neonate's identities by covering their faces. In this study, we utilized an ECG device monitor as a reference to evaluate and compare the performance of the 2 suggested experiments for evaluating the HR of neonates.

2.2 Experiment Setup

In this study, we implemented 2 experiments to evaluate a noncontact technique for calculating neonates' heart rates. We conducted both experiments under UV illumination. The first experiment investigates the ability of the proposed noncontact technique to calculate neonatal heart rates using a blue filtering approach applied to phototherapy imaging; thus, it will be referred to as the HRBF experiment, while the second experiment calculates neonate heart rates without employing the blue filtering technique under UV illumination, termed the HRWB experiment.

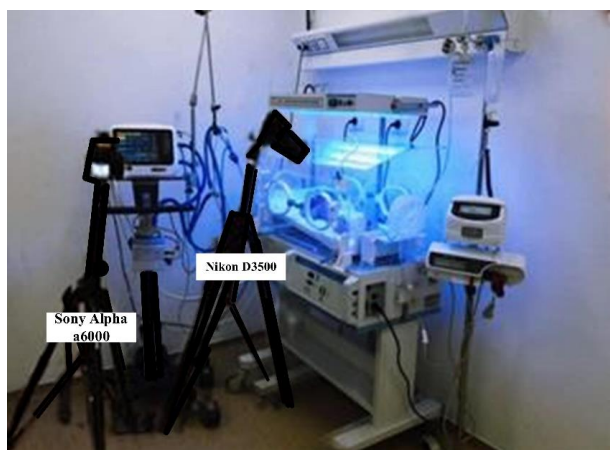


Fig. 1 Experimental setup of neonate's recorded.

We divided 10 neonates into 2 groups: 5 neonates participated in the HRBF experiment, while the remaining 5 were involved in the HRWB experiment. The neonates' gestational ages ranged from 25 to 39 weeks, and their weights varied between 1450 and 3300 g. Fig. 1 depicts the experimental setup, utilizing 2 digital cameras to record video clips. A Nikon D3500 camera, capable of recording at a resolution of 1920×1080 px at 30 fps (frames per second), was tripod-mounted at a distance of 0.55–1.5 m to capture videos of the neonates. A Sony Alpha a6000 camera, also tripod-

mounted, recorded vital signals displayed on the ECG monitor.

We synchronized both cameras to start recording simultaneously, ensuring consistent timing between the ECG data and the contactless measurements. We stored all videos in MOV format. Data collection in a neonatal intensive care unit (NICU) is inherently challenging due to the delicate environment and the need for minimal interference with medical devices. To validate our study, we created a small database by recording neonates when they were stationary and under continuous ECG monitoring. We maintained the camera-to-subject distance at 1.5–2.5 m. We recorded 20-s clips for each neonate, capturing five clips per neonate for analysis and comparison. In total, we collected 50 clips for each experimental condition (HRBF and HRWB), resulting in a robust dataset for evaluating the proposed technique.

3 System Framework

This section will present the developed healthcare system proposed based on a non-contact HR values monitoring system for neonates. It will include details on the blue filter technique proposed, a tile-based ROI selection technique, a neonatal heart rate calculation technique, and a neonatal graphical interface system proposed. Fig. 2 illustrates the framework of the proposed non-contact neonatal HR monitoring.

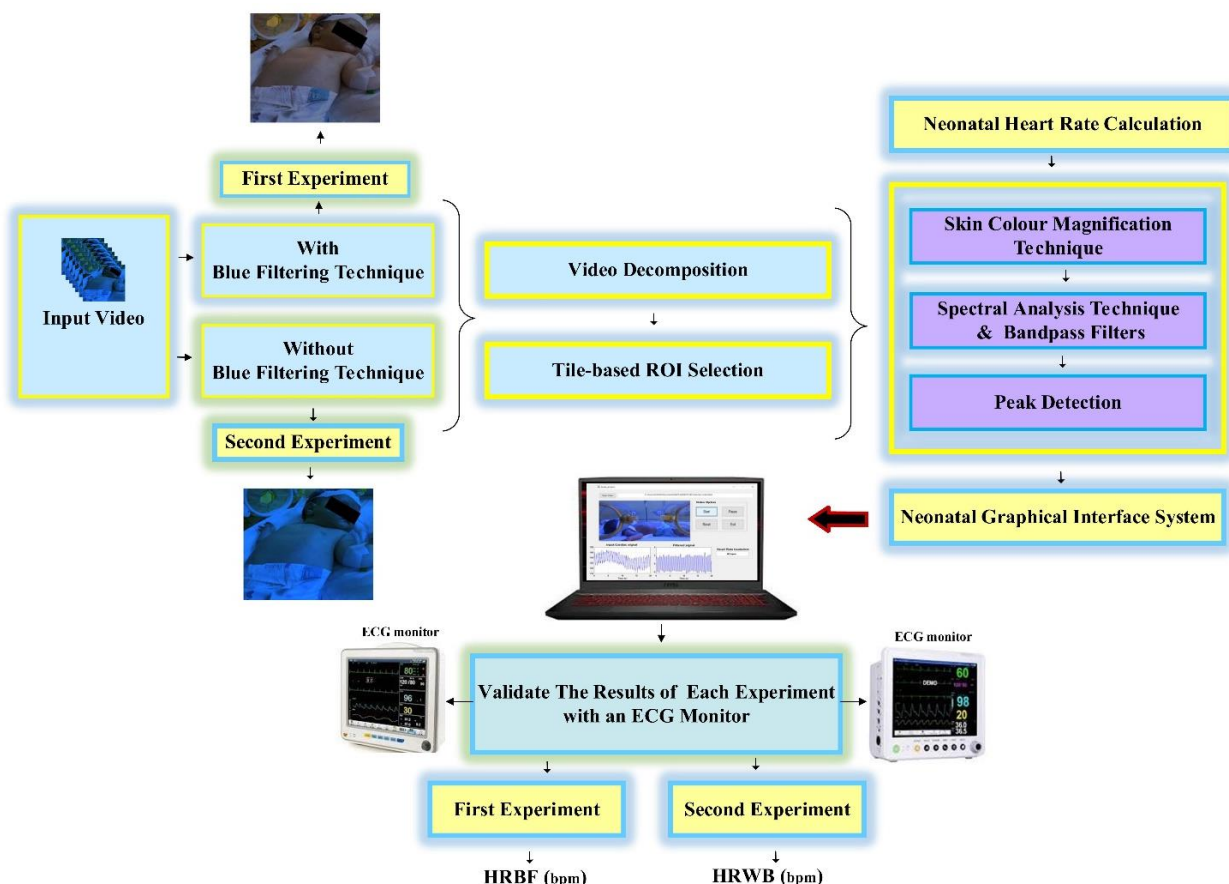


Fig. 2 The system framework of the proposed study.

3.1 Blue Filtering Technique

The basic principle that relies on the PPG is that blood absorbs light more effectively than adjacent tissues. This absorption induces slight alterations in skin pigmentation as a result of cardiovascular activity. The reflectance characteristics of human skin, obtained from video recordings using PPG, provide significant insights into cardiovascular health, including heart rate and blood oxygen saturation [32]. In PPG-based analysis, cyclical alterations in face skin color are shown as changes in brightness (intensity) values within picture sequences, facilitating the detection of heartbeats.

The second principle addresses the challenge of phototherapy images captured under UV light, where the blue channel is dominant compared to the red and green channels. In contrast, natural photographs exhibit a more balanced distribution among the 3 color channels (RGB). To address this disparity, we propose a blue filter technique using the color space transforming method for phototherapy photos in the HRBF experiment. This technique equalizes the image by calibrating the intensity of the red (Adj_R) and blue (Adj_B) channels to align with the brightness of the green channel.

We compute the adjustments for the red and blue channels using the following Eqs.:

$$Adj_R = \frac{1}{w \times h} [\sum_{i=1}^w \sum_{j=1}^h G_{ij} - \sum_{i=1}^w \sum_{j=1}^h R_{ij}], \quad (1)$$

$$Adj_B = \frac{1}{w \times h} [\sum_{i=1}^w \sum_{j=1}^h B_{ij} - \sum_{i=1}^w \sum_{j=1}^h G_{ij}]. \quad (2)$$

Next, we calculate the updated pixel intensities for the red (R_{ij}) and blue (B_{ij}) channels by applying the following Eqs.:

$$R_{ij} = \min(255, R_{ij} + Adj_R), \quad (3)$$

$$B_{ij} = \max(0, B_{ij} + Adj_B). \quad (4)$$

The introduced blue-filtering technique will enhance the performance and reliability of heart rate measurements during UV phototherapy by reducing inter-channel discrepancies to be similar relatively to obtained images framed from natural illumination.

3.2 Video Decomposition

The raw video is initially decomposed into tiles on the first frame, with each frame being divided into smaller segments corresponding to a grid. The tile size, denoted as $D \times D$, is selected based on heuristic methods. Tile creation begins at the (1,1) coordinate of the raw image frame. In cases where the image width (w) or height (h) is not an exact multiple of D , the tiles do not fully cover the entire raw image frame. Each grid cell represents a distinct tile video. This tile-based division is consistent across all frames in the video, so each frame is divided into a similar yellow grid structure for 2 proposed

experiments (HRBF and HRWB), as shown in Fig. 3(a) and (b). The total number of tiles is calculated as follows:

$$T_n = (w/D \times h/D), \quad (5)$$

where, w and h refer to the width and height of the raw image frame, respectively. For each ROI, the tile video is constructed to match the raw video in terms of the number of frames and frame rate.

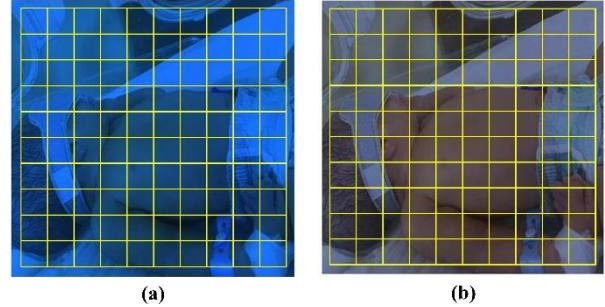


Fig. 3 Image decomposed into tiles: (a) HRBF, (b) HRWF.

3.3 Tile-Based ROI Selection

In this study, we employed a tile-based ROI selection approach for extract cardiac signals from PPG. Each ROI was extracted from the decomposed UV photograph, with the tile grid encompassing the neonatal chest region, which was designated as the ROI for heart rate calculation in two proposed. This procedure was repeated across all frames.

Traditionally, manual ROI selection techniques delineate the chest or head region by enclosing it within a bounding box, from which vital indicators are retrieved. In contrast, our proposed technique selects the tiles containing the region of interest through visual inspection rather than relying on manual bounding boxes.

The primary advantage of this tile-based ROI selection for neonates is that it eliminates the need for constant adjustments when the infant changes posture. As the neonate typically moves laterally, the chest region remains consistently within the initially designated tiles. The adjustable tile size enables us to establish a safe dimension that accommodates the neonate's dynamic movements.

3.4 Neonatal Heart Rate Calculation

This study used a color-based technique to extract the cardiac signal as the first step, utilizing the principle of light absorption by hemoglobin in blood. Hemoglobin absorbs more light than the surrounding tissue, resulting in subtle color variations in the infant's skin that are invisible to the naked eye but detectable with a digital camera. The changes in blood volume within the microvascular tissue during each cardiac cycle reflected in these color variations serve as a means to measure the heart rate. To measure the cardiac signal, the green channel of the RGB color space was selected, as it

provides the strongest cardiac frequency signal due to higher absorption in this wavelength. The spatial averaging from the brightness of the pixels within the ROI was then calculated to extract the raw cardiac signal, as described by Eq. (6). This step minimizes the influence of camera noise associated with individual pixels, thereby improving the signal-to-noise ratio. The resulting process effectively reduces noise and enhances the clarity of the cardiac signal representation.

$$I_G(t) = \frac{\sum_{x,y \in ROI} I(x_i, y_i, t)}{|ROI|}, \quad (5)$$

where $I(x_i, y_i, t)$ is indicated as the illuminance values at the pixel image position (x_i, y_i) at time phase (t) , and $|ROI|$ specifies the size of the recorded video selected by ROI. The neonatal cardio signal obtained from the G channel information was selected for calculate the neonatal HR values of 2 experiments (HRBF and HRWB).

In the second step, we applied a magnification technique to further improve the cardiac signal's extraction and clarity. Based on the EVM technique [31], this technique of magnification brings out small changes in color over time that are caused by the heart's pulsed blood flow. We then add this enhanced signal back to the original video frame to create a clearer representation of the heart rate-related changes in skin color. This magnification approach enhances the cardiac signal extraction and amplification, enabling improved detection of tiny fluctuations, especially when the baby is far from the camera or in low signal-to-noise ratio conditions. This enhanced methodology considerably elevates the dependability and sensitivity of non-contact heart rate monitoring devices.

In the third step, the raw time series cardio signals were converted from the time domain to the frequency domain using the spectrum analysis method using FFT. Subsequently, to get the desired frequency band, 2 ideal band-pass filters were used at 1 Hz to 3 Hz [33], equivalent to the neonatal pulse range of 60 to 180 bpm. Following this, the inverse FFT was applied to the filtering signals to derive the time series of the neonatal cardiac signals.

In the fourth step, temporal sequence neonatal cardiac signals are analyzed by determining the peaks and their respective positions (*locs*) between consecutive peaks. The MATLAB function "findpeaks" was used to identify peaks. The total cycle length between 2 peaks is determined using the following Eq. (7):

$$T_c = \text{mean}(\text{diff}(\text{locs})). \quad (7)$$

The number of peaks pk can be calculated by the following Eq. (8):

$$pk = \frac{T}{T_c}, \quad (8)$$

where T is the duration time in seconds of the recording video.

Finally, the measured values HR in bpm (beats per minute) is calculated using Eq. (9):

$$HR = \frac{60 \cdot pk}{T}. \quad (9)$$

3.5 Neonatal Graphical Interface System

We implemented a neonatal GUI system for non-contact HR values monitoring using MATLAB 2023b. The proposed system features a user-friendly interface that includes neonatal cardio signal processing and displays the neonatal HR value findings based on integrating a blue-filter technique with skin color magnification under UV illumination. It also offers an efficacious environment for neonatal video uploads.

The proposed system includes text fields for displaying HR values. Additionally, it includes a variety of uploaded neonatal video controls, such as start, pause, extract, and reset options. Moreover, it includes analytical tools, including 3 axes for neonatal video upload, input cardio signal display, and filtered signal output.

4 Experimental Results

4.1 Blue Filter and ROI Selection Proposed

This section presents the performance of the blue filter technique and tile-based ROI selection of 2 experiments. Fig. 4 compares the original image with phototherapy images, both before and after applying the blue filtering method. Fig. 4 shows that the proposed blue-filtering method works to improve the accuracy and reliability of HR measurements during UV phototherapy by reducing differences between channels and reducing the blue tint significantly, especially in areas of skin that can be seen. Other areas, such as neonatal eye masks, bedding, and the incubator bed, also exhibit reduced blue hues, simulating conditions of natural light illumination. This approach plays a crucial role in achieving consistent and clinically meaningful outcomes in neonatal NICU settings under UV illumination.

Fig. 5(a) and (b) show the tile-based ROI selection method utilized in the HRWB and HRBF experiments. As demonstrated in Fig. 5(a) yellow tile encapsulates the neonatal chest region, which was identified as the ROI for heart rate (HR) evaluation. This ensures accurate heart rate calculation despite infant movements and eliminates the need for manual bounding box adjustments. This targeted selection ensures the accurate extraction of HR values by focusing on a relevant anatomical area, minimizing interference from non-relevant regions.

4.2 Heart Rate Extraction

In our works, we obtained raw cardio signals for 600 frames by applying spatial averaging to detected ROIs using a color -based technique. Fig. 6(a) and (b) illustrates the cardio signal from the HRWB and HRB experiments, respectively.

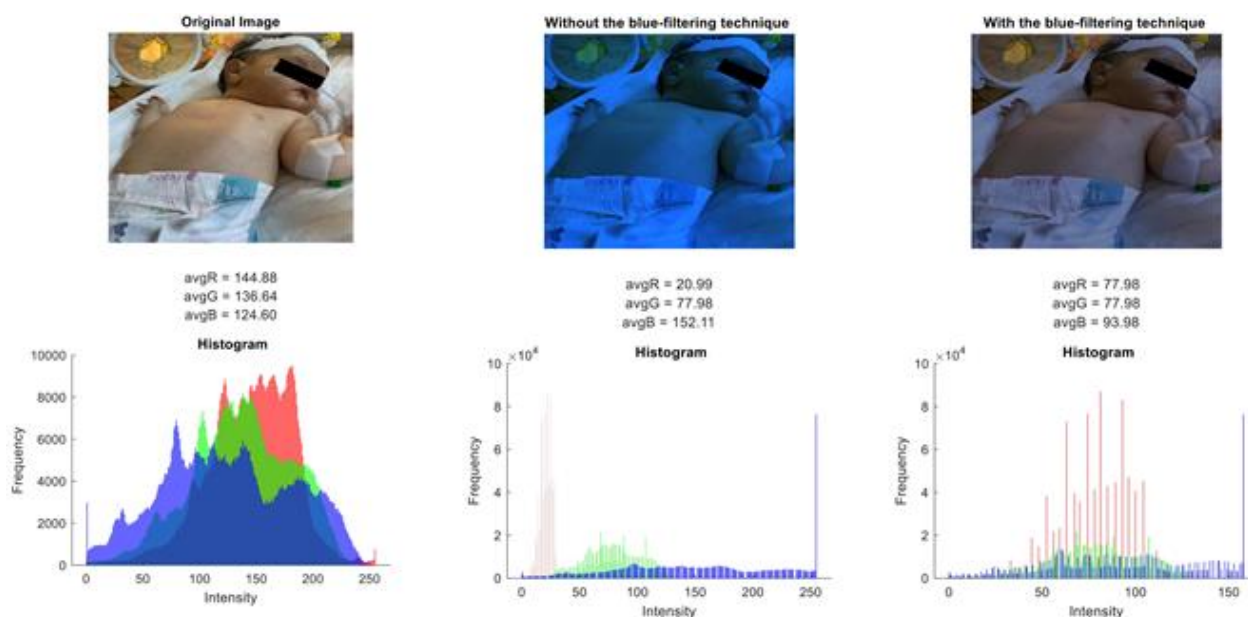


Fig. 4 Comparison of phototherapy images captured with and without the blue-filtering technique.

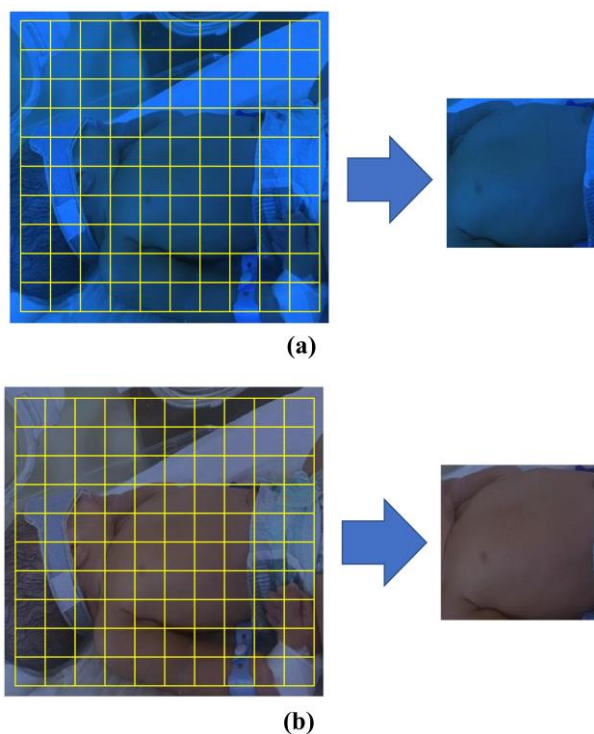


Fig. 5 Tile-based ROI selection of two experiments. (a) HRWB experiment; (b) HRBF experiment.

Fig. 7(a) and (b) displays the cardio signal extraction after applying the magnification technique for two experiments, respectively. The magnification technique extracted the cardiac signal more accurately and amplified it, enabling better detection of subtle variations. This magnified approach significantly improves the reliability and sensitivity of non-contact heart rate monitoring systems.

Fig. 8 (a) and (b) demonstrate the results of filtered with detection peaks cardio signals obtained after

applying FFT, band-pass filtering, inverse FFT, and Eq. (7), respectively. In this work, the window duration of FFT was twenty seconds. The green color points indicate the peak positions of the filtered signals, as displayed in Fig. 8(a) (HRWB experiment) and Fig. 8(b) (HRBF experiment).

Fig. 9 shows the neonate's GUI monitoring proposed, which contains the HR values (bpm) calculation after applying Eq. (8), the cardio input signal, and the filtered cardio signal. Additionally, this suggested healthcare system enables real-time remote monitoring of neonates' states and automatically stores cardio signals and HR measurement outputs on the computer as accessible medical data to assist in future diagnoses of neonatal cardio disease.

4.3 Statistical Analysis

This section compares the performance of two experiments (HRWB and HRBF) by analyzing the measured values (HR) by Eq. (9) and reference readings (ECG) using statistical analysis to validate the proposed technique. The statistical analysis methods will be applied, including linear regression, Pearson's correlation coefficient (PCC), Bland-Altman plots, mean bias, Spearman's correlation coefficient (SCC).

The total sample size collected from neonatal captured video was set at $n = 25$. The statistical analyses of the samples measured readings collected from the HRWB experiment, as shown in Fig. 10a. The Bland-Altman analysis discloses a mean bias of -2.9 bpm, with 95% limits of agreement ranging from -11 to $+5$. Also, the PCC and SCC results between the measured and reference readings were 0.981 and 0.948, respectively. Fig. 10(b) displays the linear regression analysis of the relationship between the experimental measure readings and the linear fit obtained, with an R^2 value of 0.963.

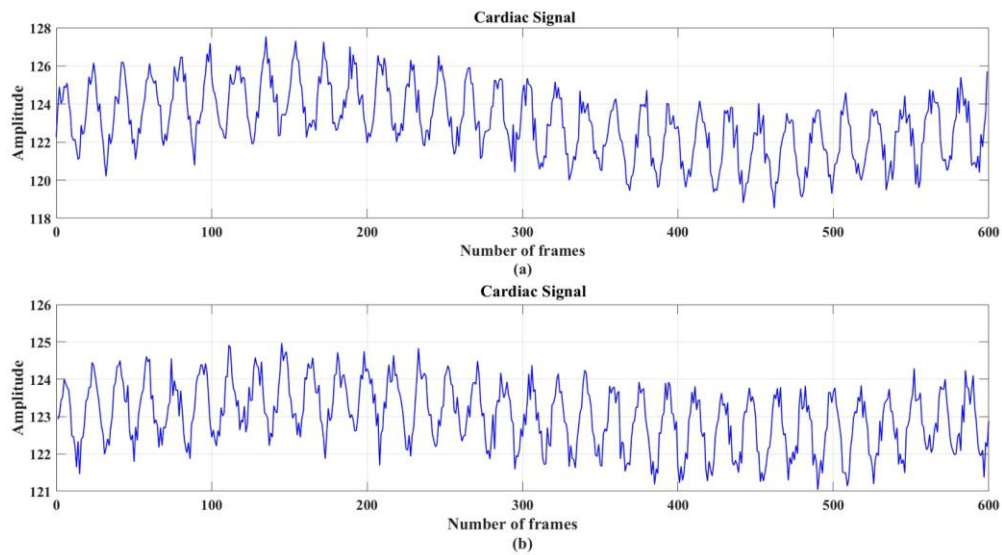


Fig. 6 Raw cardio signals of two experiments. (a) HRWB experiment, (b) HRBF experiment.

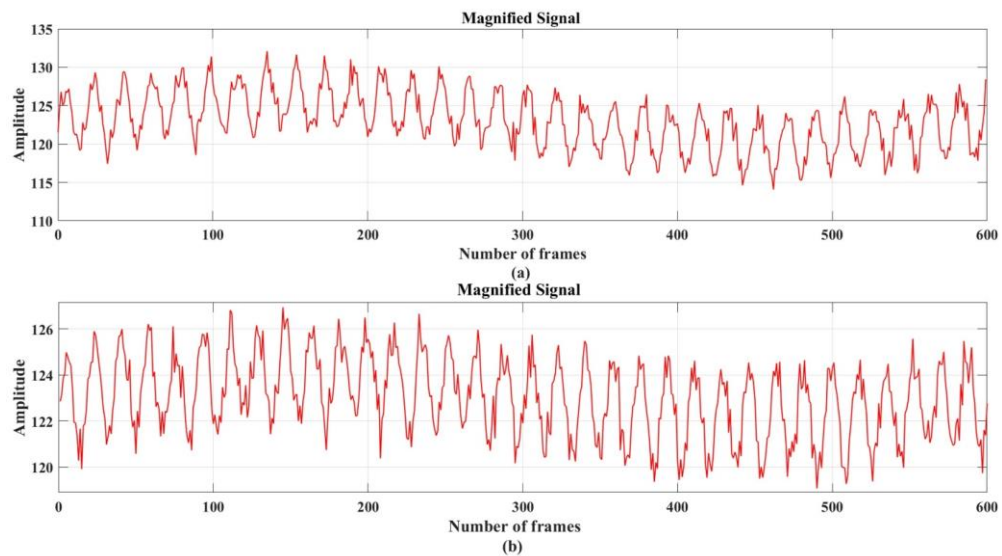


Fig. 7 Magnification results of the raw cardio signals. (a) HRWB experiment, (b) HRBF experiment.

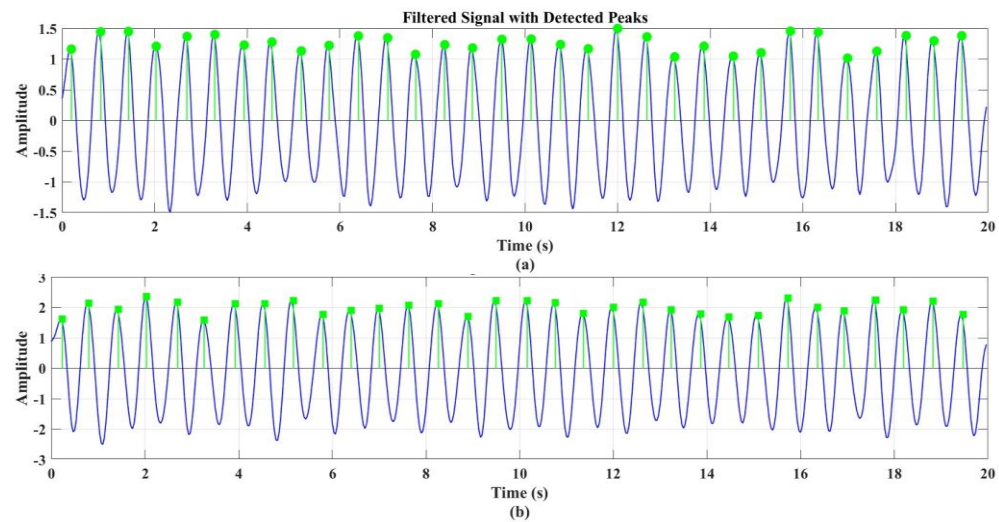


Fig. 8 Filtered and detection peak of cardio signals. (a) HRWB experiment, (b) HRBF experiment.

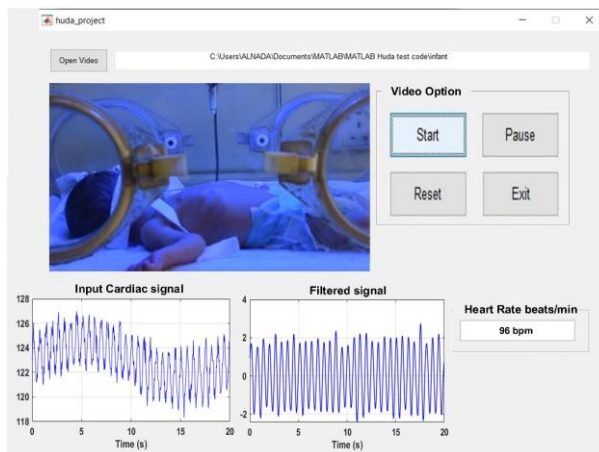
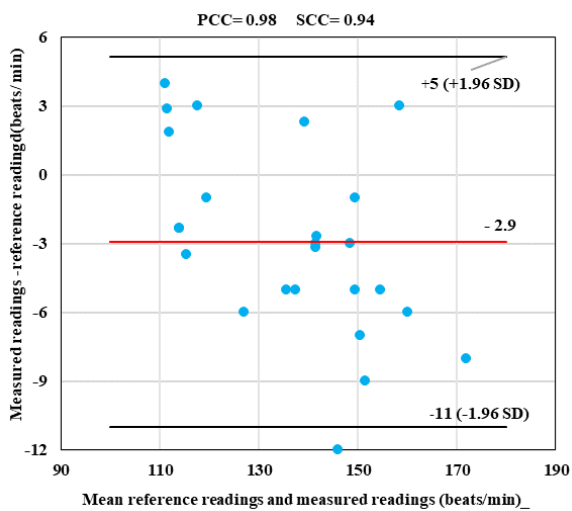
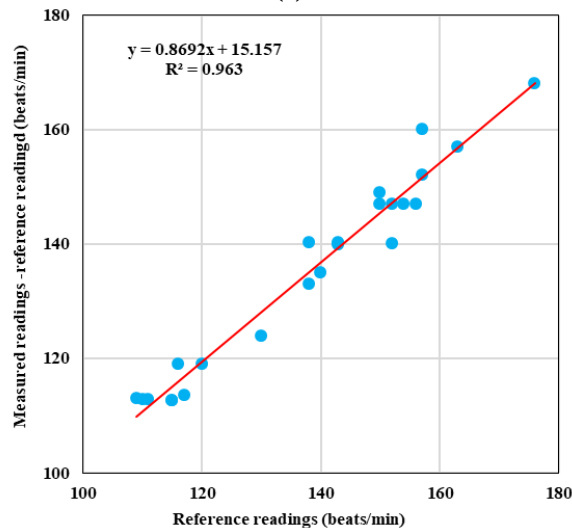


Fig. 9 Screenshot of proposed neonatal system GUI.



(a)

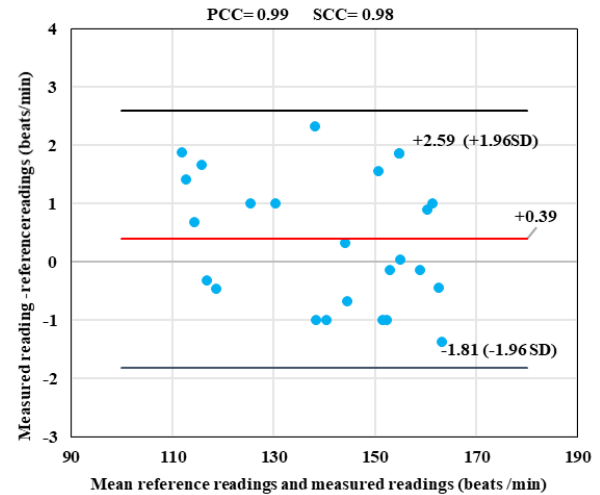


(b)

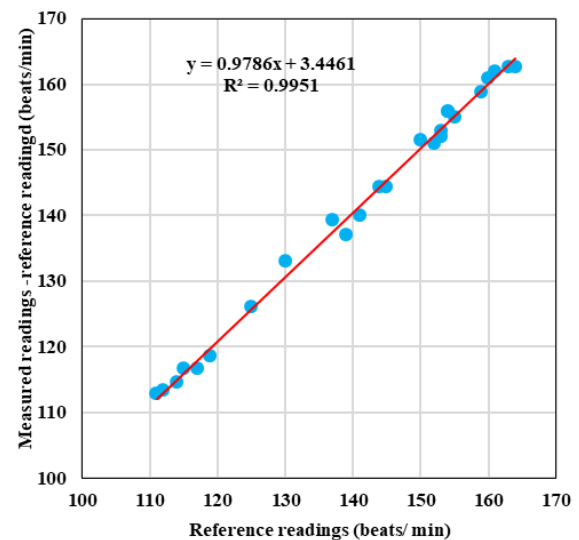
Fig. 10 Statistical analyses of the HRWB experiment: (a) Bland-Altman plot, (b) correlation plot.

Fig. 11 shows a statistical approach related to measured readings from the HRBF experiment. As demonstrated in Fig. 11(a), the Bland-Altman statistics display a mean bias of 0.39 bpm, with 95% confidence

intervals ranging from -1.81 to $+2.59$. Also, the highest obtained results of the PCC and SCC between readings and reference readings were 0.998 and 0.982, respectively. Also, the linear regression coefficients demonstrate a strong match output between the values of measured readings and the linear fit, with $R^2 = 0.995$, as shown in in Fig. 11(b).



(a)



(b)

Fig. 11 Statistical analyses of the HRBF experiment: (a) Bland-Altman plot, (b) correlation plot.

To validate the proposed technique, the performance results from the HRBF experiment were compared with those from the HRWB experiment. The findings demonstrated that the HRBF experiment achieved the lowest average MAE of 1.12 bpm, whereas the HRWB experiment recorded the highest average MAE of 4.2 bpm.

Therefore, the proposed technique is crucial for accurately extracting neonatal cardiac signals under UV illumination, compared to the results obtained without using the blue filter method in the HRWB experiment. This methodology is considered novel and more effective in the neonatal healthcare system, particularly in the rise in contagious viral infections such as COVID-19.

5 Discussion

This study focuses on the urgent need for a neonatal healthcare noncontact system to monitor HR values under UV illumination during the spread of transitional epidemics like COVID-19 and other new diseases. The main goal of the research was to develop a neonatal healthcare noncontact system that combines noncontact technology and image processing to accurately evaluate the heart rate of neonates using noncontact camera technology during an emergency. The proposed blue filter technique, combined with the EVM method, was utilized to extract cardio signals effectively to enhance the accuracy and reliability of signal detection by reducing the interference caused by UV illumination and amplifying the subtle temporal color variations in the skin. The blue filter was specifically employed to balance the intensity of the color channels under UV light, mitigating the impact of excess blue or red light reflections and isolating the green channel, which is most sensitive to pulsatile blood flow.

The proposed approach was critical for ensuring precise cardiac signal extraction in neonates exposed to UV illumination compared with obtained results without using a blue filter method in the HRWB experiment.

Also, this approach for skin color magnification offers several advantages. Enhancing skin color variations in the temporal domain enhances signal clarity without requiring physical contact, making it ideal for neonatal applications. Furthermore, the reduction in computational overhead ensures that the method is deployable in real-time monitoring systems, an essential criterion for practical implementation in NICU settings.

The effectiveness of this technique was demonstrated through a comparative analysis of video sequences with

and without magnification. This advancement not only accelerates execution time but also increases the robustness of non-contact physiological monitoring solutions.

On the other hand, this study highlights the effectiveness of the ROI tile-based selection method in monitoring vital signs, demonstrating improved adaptability to neonatal movements and potentially reducing the need for manual adjustments. The primary challenges encountered in this study are outlined as follows:

1. The difficulty of acquiring a sufficiently large dataset of infants under UV illumination, compounded by their critical health conditions.

2. The presence of complex medical apparatus obstructs the capture of the neonate's body posture, thereby limiting the effectiveness of the monitoring system.

3. The presence of NICU personnel, such as nurses and doctors, or items like toys that resemble infants, particularly when they are situated on bedsheets, can lead to false detection incidents.

4. The data acquisition process was time-consuming and required effort, especially given the requirement for each neonate to be recorded multiple times in their most appropriate posture. UV illumination and lighting variability.

5. One of the significant challenges was ensuring accurate performance validation by recording neonatal data under both UV and non-UV illumination as the same subjects. This challenge necessitated additional work to mitigate disparities across channels in the equalized phototherapy image to simulate natural lighting conditions.

Table 1 Comparison the proposed system (HRBF) with the previous studies.

Authors	Sensor	ROI	Technique	Region	MAE (bpm)	PCC, SCC
Poh et al. [34]	Video camera	Bounding box	Chrominance	Face	3	0.98, -
Wei et al. [35]	Video camera	Bounding box	ICA	Face	2.65	0.92, -
Wang et al. [36]	RBG camera	Tile	Spatial processing	Wrist	-, -	-, -
Revanur et al. [37]	Video Footage	-	Video Transformer	Face	13	-, -
Molinaro et al. [38]	Digital cam	-	PPG	Face	5	-, -
Cheng et al. [39]	RGB Camera	Bounding box	EEMD	Face	-	0.96, -
Sharma et al. [32]	Low-cost video camera	Bounding box	FFT	Face	-	-, -
Stockwell et al. [40]	Forehead-based PPG sensor	Bounding box	Reflection mode PPG	Forehead	8.44	-, 0.981
Al-Naji et al. [41]	Digital camera	Bounding box	Noise artifact removal	Face	2.91	0.97, -
Khanam et al. [42]	Digital camera	Image decomposition	Color based method	Face	3.56	0.979, -
Alnaggar et al. [31]	RGB Camera	Tile	EVM	chest	1.62	-, -
HRBF (Our proposed)	Nikon D3500	Tile	Blue filter, EVM	Face	1.12	0.998, 0.982

To evaluate the efficacy of our findings from the HRBF experiment. We compared them with 11 reliable studies using non-contact technology-based monitoring methods, as shown in Table 1. The findings indicated that the suggested system achieved the lowest average MAE of 1.12 bpm and exhibited high PCC and SCC metrics of 0.997 and 0.982, respectively. The error rate study indicates that the proposed blue filter technique combined with the proposed tile-based ROI selection method significantly enhances the accuracy of HR values in comparison to previous studies. The suggested technique is appropriate for use in healthcare or indoor home monitoring.

6 Conclusions

In conclusion, this study introduced a novel non-contact neonatal heart rate monitoring for neonates under UV illumination, addressing the limitations of traditional contact-based monitoring devices. A robust tile-based ROI selection method for neonatal cardiac signal monitoring was introduced, ensuring accurate heart rate calculation despite infant movements and eliminating the need for manual bounding box adjustments. A blue filter technique combined with the EVM method was utilized to enhance the extraction of cardiac signals by mitigating interference from UV illumination and amplifying subtle temporal skin color variations. Furthermore, a specialized neonatal GUI was developed, offering a user-

friendly platform for biomedical engineers to monitor vital signs remotely in the neonatal NICU. Experimental results demonstrated the system's efficacy, achieving high correlation coefficients (PCC = 0.998) and low error rates (MAE = 1.12 bpm) in the HRBF experiment, outperforming results from the HRWF experiment (PCC = 0.981; MAE = 4.2 bpm). The system's low cost and ease of deployment position it as a viable solution for real-time neonatal healthcare monitoring system, particularly during transitional epidemics like COVID-19 and other scenarios requiring reduced physical contact. Future work will focus on enhancing system reliability by integrating advanced signal processing techniques and exploring machine learning approaches to optimize ROI selection and improve overall performance.

Acknowledgments

The author would like to thank the staff of the Department of Medical Instrumentation Techniques Engineering, Electrical Engineering Technical College, Middle Technical University, for their support during this work.

Disclosures

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera," *IEEE Transactions on Biomedical Engineering* 61(10), 2593–2601 (2014).
2. A. G. Shabeeb, H. A. Hashim, and S. K. Gharghan, "Heart disease classification based on combination of PCA /ANFIS model," *Research on Biomedical Engineering* 40(3), 609–625 (2024).
3. X. He, R. A. Goubran, and X. P. Liu, "Wrist pulse measurement and analysis using Eulerian video magnification," *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 41–44 (2016).
4. N. Kumar, G. Akangire, B. Sullivan, K. Fairchild, and V. Sampath, "Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront," *Pediatric Research* 87(2), 210–220 (2020).
5. H. Ali Hashim, N. Mahmood Ahmed, and A. G. Shabeeb, "Infant heart rate estimation based on non-contact UV photoplethysmography," *Indonesian Journal of Electrical Engineering and Computer Science* 31(1), 180 (2023).
6. D. Kolosov, V. Kelefouras, P. Kourtessis, and I. Mporas, "Contactless camera-based heart rate and respiratory rate monitoring using ai on hardware," *Sensors* 23(9), 4550 (2023).
7. L. Scalise, N. Bernacchia, I. Ercoli, and P. Marchionni, "Heart rate measurement in neonatal patients using a webcamera," *IEEE International Symposium on Medical Measurements and Applications Proceedings*, IEEE, Budapest, Hungary (2012).
8. J. Brieve, E. Moya-Albor, O. Rivas-Scott, and H. Ponce, "Non-contact breathing rate monitoring system based on a Hermite video magnification technique," in *14th International Symposium on Medical Information Processing and Analysis*, SPIE Proceedings 10975, 1097504 (2018).
9. M. Chen, Q. Zhu, H. Zhang, M. Wu, and Q. Wang, "Respiratory rate estimation from face videos," *2019 IEEE EMBS International Conference on Biomedical & Health Informatics*, Chicago, IL, USA (2019).
10. H. A. Hashim, "Non-contact automatic respiratory rate monitoring for newborns using digital camera technology and deep learning," *Journal of Biomedical Photonics & Engineering* 10(4), 040317 (2024).
11. A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: principles of remote measurements and review of methods," *IEEE Access* 5, 15776–15790 (2017).
12. A. Al-Naji, A. G. Perera, and J. Chahl, "Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle," *Biomedical Engineering Online* 16(1), 101 (2017).

13. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "[Remote plethysmographic imaging using ambient light](#)," *Optics Express* 16(26), 21434 (2008).
14. L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "[Non-contact video-based vital sign monitoring using ambient light and auto-regressive models](#)," *Physiological Measurement* 35(5), 807–831 (2014).
15. Y. Zeng, D. Yu, X. Song, Q. Wang, L. Pan, H. Lu, and W. Wang, "[Camera-based cardiorespiratory monitoring of preterm infants in NICU](#)," *IEEE Transactions on Instrumentation and Measurement* 73, 1–13 (2024).
16. G. O. Ganfure, "[Using video stream for continuous monitoring of breathing rate for general setting](#)," *Signal, Image and Video Processing* 13(7), 1395–1403 (2019).
17. H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. Jane Wang, "[Video-based human heart rate measurement using joint blind source separation](#)," *Biomedical Signal Processing and Control* 31, 309–320 (2017).
18. H. Hashim, S. Mohammed, and S. Gharghan, "[Accurate localization of elderly people based on neural and wireless sensor networks](#)," *Journal of Engineering and Applied Science* 14(11), 3777–3789 (2019).
19. C. Gonzalez Viejo, S. Fuentes, D. D. Torrico, and F. R. Dunshea, "[Non-contact heart rate and blood pressure estimations from video analysis and machine learning modelling applied to food sensory responses: a case study for chocolate](#)," *Sensors* 18(6), 1802 (2018).
20. Y. Sun, C. Papin, V. Azorin-Peris, R. Kalawsky, S. Greenwald, and S. Hu, "[Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam](#)," *Journal of Biomedical Optics* 17(3), 037005 (2012).
21. A. Al-Naji, J. Chahl, "[Remote respiratory monitoring system based on developing motion magnification technique](#)," *Biomedical Signal Processing and Control* 29, 1–10 (2016).
22. K. S. Tan, R. Saatchi, H. Elphick, and D. Burke, "[Real-time vision based respiration monitoring system](#)," in 7th International Symposium on Communication Systems, Networks & Digital Signal Processing, 770–774 (2010).
23. M. Bartula, T. Tigges, and J. Muehlsteff, "[Camera-based system for contactless monitoring of respiration](#)," 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2672–2675 (2013).
24. K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "[Face-based heart rate signal decomposition and evaluation using multiple linear regression](#)," *IEEE Sensors Journal* 16(5), 1351–1360 (2016).
25. D.-Y. Chen, J.-J. Wang, K.-Y. Lin, H.-H. Chang, H.-K. Wu, Y.-S. Chen, and S.-Y. Lee, "[Image sensor-based heart rate evaluation from face reflectance using hilbert–huang transform](#)," *IEEE Sensors Journal* 15(1), 618–627 (2015).
26. W. Wang, S. Stuijk, and G. de Haan, "[A novel algorithm for remote photoplethysmography: Spatial subspace rotation](#)," *IEEE Transactions on Biomedical Engineering* 63(9), 1974–1984 (2016).
27. S. Yu, S. Hu, V. Azorin-Peris, J. A. Chambers, Y. Zhu, and S. E. Greenwald, "[Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise](#)," *Journal of Biomedical Optics* 16(7), 077010 (2011).
28. L. A. M. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. Bambang Oetomo, and W. Verkruysse, "[Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit — A pilot study](#)," *Early Human Development* 89(12), 943–948 (2013).
29. K. Gibson, A. Al-Naji, J. Fleet, M. Steen, A. Esterman, J. Chahl, J. Huynh, and S. Morris, "[Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study](#)," *Pediatr Research* 86(6), 738–741 (2019).
30. M.-Z. Poh, D. J. McDuff, and R. W. Picard, "[Advancements in noncontact, multiparameter physiological measurements using a webcam](#)," *IEEE Transactions on Biomedical Engineering* 58(1), 7–11 (2011).
31. M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Z. Rashad, "[Video-based real-time monitoring for heart rate and respiration rate](#)," *Expert Systems with Applications* 225, 120135 (2023).
32. N. Sharma, S. Kaman, and P. K. Mahapatra, "[Non-contact measurement of human heart rate using low cost video camera](#)," 2019 Fifth International Conference on Image Information Processing, 58–62 (2019).
33. Z. Wu, N. E. Huang, "[Ensemble empirical mode decomposition: a noise-assisted data analysis method](#)," *Advances in Adaptive Data Analysis* 01(01), 1–41 (2009).
34. M.-Z. Poh, D. J. McDuff, and R. W. Picard, "[Non-contact, automated cardiac pulse measurements using video imaging and blind source separation](#)," *Optics Express* 18(10), 10762 (2010).
35. W. Wei, K. Vatanparvar, L. Zhu, J. Kuang, and A. Gao, "[Remote photoplethysmography and heart rate estimation by dynamic region of interest tracking](#)," in 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, 3243–3248 (2022).
36. H. Wang, S. Zhang, "[Non-contact heart rate measurement using low-cost RGB camera under complex light conditions](#)," *Multimedia Tools and Applications* 84(3), 1561–1575 (2025).
37. A. Revanur, A. Dasari, C. S. Tucker, and L. A. Jeni, "[Instantaneous physiological estimation using video transformers](#)," in *Multimodal AI in Healthcare*, A. Shaban-Nejad, M. Michalowski, and S. Bianco (Eds.), Springer International Publishing 1060, 307–319 (2023).
38. N. Molinaro, F. Zangarelli, E. Schena, S. Silvestri, and C. Massaroni, "[Cardiorespiratory parameters monitoring through a single digital camera in real scenarios: ROI tracking and motion influence](#)," *IEEE Sensors Journal* 23(17), 20097–20106 (2023).

39. J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "[Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition](#)," IEEE Journal of Biomedical and Health Informatics 21(5), 1422–1433 (2017).
40. S. J. Stockwell, T. C. Kwok, S. P. Morgan, D. Sharkey, and B. R. Hayes-Gill, "[Forehead monitoring of heart rate in neonatal intensive care](#)," Frontiers in Physiology 14, 1127419 (2023).
41. A. Al-Naji, J. Chahl, "[Remote optical cardiopulmonary signal extraction with noise artifact removal, multiple subject detection & long-distance](#)," IEEE Access 6, 11573–11595 (2018).
42. F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, K. Gibson, and J. Chahl, "[Remote vital signs monitoring in neonatal intensive care unit using a digital camera](#)," International Journal of Biomedical and Biological Engineering 16(10), 138–144 (2022).

Optimized Classification with CBGRU-CapNet on Thermographic Analysis to Detect Diabetic Foot Ulcers with Precise Diagnosis

Ashok Babu Ane^{1*}, Gireesh Babu C N², Karthik Govindan Manoharan³, Rafeeq Md.⁴, Bhupathi Rajarao⁵, and Kadiyala Vijaya Kumar⁶

¹ Department of Electronics & Communication Engineering, P.V.P Siddhartha Institute of Technology, Vijayawada 520007, Andhra Pradesh, India

² Department of Computer Science and Engineering, BMS Institute of Technology & Management, Bengaluru 560064, Karnataka, India

³ Department of Information Security, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

⁴ Department of Computer Science and Engineering, CMR Engineering College, Hyderabad 501401, Telangana, India

⁵ Department of Electronics and Communication Engineering, Eluru College of Engineering and Technology, Eluru 534004, Andhra Pradesh, India

⁶ Department of Computer Science and Engineering – Artificial Intelligence, G. Pullaiah College of Engineering and Technology, Kurnool 518002, Andhra Pradesh, India

e-mail: aneashokbabu@gmail.com

Abstract. Diabetic foot ulcers (DFUs) are a severe complication of diabetes mellitus, often leading to lower limb or foot amputation due to infections and a high recurrence rate. Recent research focuses on early detection of underlying conditions that cause skin and tissue damage before visible lesions appear by analysing infrared thermograms of the plantar aspect of both feet. Most existing studies rely on the publicly available INAOE dataset, which includes thermographic images of both healthy and diabetic subjects. However, advancements in the field have been supported by the newly released STANDUP dataset, which offers a more extensive and diverse collection of thermograms. This study utilizes a broader dataset to improve DFU classification. Features are extracted from preprocessed input data using both statistical methods and deep feature extraction techniques. A novel deep learning model, CBGRU-CapNet – combining convolutional layers, bidirectional gated recurrent units (BiGRU), and capsule networks – is employed for classification. To enhance accuracy and reduce computational complexity, the study introduces a fine-tuning approach using the Modified Directional Transport Metaheuristic Algorithm (mDTMA). Experimental results demonstrate that the proposed model achieves superior predictive performance and robustness, making it highly suitable for accurate and consistent DFU detection across multiple evaluation metrics. The results demonstrate that the proposed model provides superior predictive power and robustness, indicating its suitability for tasks requiring high accuracy besides consistency across multiple estimation metrics. © 2025 Journal of Biomedical Photonics & Engineering.

Keywords: Diabetic foot ulcer; thermography; directional transport metaheuristic algorithm; bidirectional gated recurrent units; deep features; Diabetes mellitus.

Paper #9198 received 12 Dec 2024; revised manuscript received 18 Feb 2025; accepted for publication 25 Feb 2025; published online DD Mar 2025. doi: [10.18287/JBPE25.11.010306](https://doi.org/10.18287/JBPE25.11.010306).

1 Introduction

Diabetes mellitus (DM) is another name for diabetes. Low blood sugar, or hyperglycemia, is the root cause of this chronic condition that can cause amputation of lower limbs, heart disease, blindness, and renal failure, among other devastating complications [1]. Diabetes foot ulcers (DFUs), commonly known as sores, are open sores that can develop on the foot of a diabetic. More than 422 million people had diabetes in 2014, up from 108 million in 1980, according to statistics from around the world. Among adults, for example, the total frequency increased from 4.7% in 1980 to 8.5% in 2014 [2]. In addition, 600 million individuals will have diabetes globally by the end of 2035, according to predictions [3]. The fact that most of these people are from impoverished countries with inadequate healthcare systems and generalized ignorance is also important to consider. Amputation of the lower limbs may be necessary in 15–25% of diabetic individuals who develop DFU later in the course of the disease if adequate care is not accessible [4].

Due to insufficient diagnosis and treatment, almost one million diabetics with “high-risk feet” lose a limb every year [5]. Extra care is required of those with diabetes, who must maintain good personal cleanliness, take their prescriptions as prescribed, and see their doctors frequently for exams [6]. Consequently, the financial burden will be particularly heavy for families dealing with diabetes, particularly in underdeveloped nations where the cost of treatment is approximately 5.7 periods the yearly income [7]. The prevalence of diabetes is expected to increase, putting an increasing number of entities at risk for diabetic foot ulcers [8]. The already-scarce resources and expertise will be even more stretched thin as a result of this. Inadequate treatment causes almost one million diabetics to lose a vital body part every year, such as a foot [9].

To get to the correct diagnosis, doctors have to go through mountains of medical records. Manual diagnostic procedures are more prone to error and can be costly, while computer-aided diagnostic techniques have the potential to be more efficient and cost-effective [10]. Technological advancements in the field of wearable health and mobile health have alleviate complications associated with diabetes by tracking damaging foot stress and inflammation, enhancing patients’ well-being, and extending the time it takes for remission to set in Ref. [11]. There is an immediate need for a cost-effective, non-invasive, remotely deployable, and highly reliable automated DFU diagnostic method. Deep learning (DL), intelligent systems, and the proliferation of applications have all contributed to the emergence of DFU disease detection as a vibrant and rapidly evolving area of research [12]. The DFU test now includes numerous important mechanisms for early diagnosis, thanks to current clinical practice. A diabetic foot specialist or evaluates the patient’s medical history and does an evaluation of the DFU. Additional tests, including an X-ray, CT scan, or MRI, administered as needed.

Technological advancements in the last several years have allowed the biomedical field to make great strides, particularly in areas such as the explanation of protein structures, the classification of biomedical images, and the alignment of genetic sequences [13]. In order to comprehend, evaluate, and preserve the growing body of biomedical data, dependable and effective computer systems are essential. DL-based methods are capable of addressing these intricate problems. It is possible to efficiently acquire hierarchical DL features from raw data using DL techniques, and neural networks in particular [14]. Because of this, these algorithms are able to decipher complex biomedical data, including its representations and linkages. Due to the increased focus and investment in DL recently, it has grown and is now widely used in many industries. The biomedical industry has started to prioritize the use of DL techniques because of their ability to assess massive datasets, find patterns that are relevant to tasks, and generate predictions. In the modern, digital crucial for the diagnosis of numerous health issues [15].

When it comes to medical imaging, traditional ML and DL algorithms succeed by picking and extracting features that are more shape, colour, and size sensitive. Using ML and DL approaches, DFUs have been successfully discovered and detected in previous studies. Numerous scholars have looked into these challenging difficulties, but thus yet, no practical answers have been identified [16]. Unfortunately, a lot of ML and DL algorithms are complicated and opaque, so it is hard to figure out how the model is making judgments or predictions and to spot possible biases or mistakes. Medicine, economics, and law enforcement are just a few fields where model errors can have far-reaching consequences. The difficulty in tuning and generalizing black box models to fresh data is further compounded by their sensitivity to hyperparameter choice [17]. Another problem that can arise with black box models is overfitting, which can cause them to perform poorly when tested with new data. Despite black box models’ success in many domains, their opaqueness and lack of interpretability raise serious concerns.

If possible, ML and DL models should be able to estimate the severity of the lesion and categorize people at risk of emerging ulcers using a single thermogram that includes the plantar feet. Thorough data interpretation is of the utmost importance in the healthcare setting. While numerous indicators have been suggested for use in thermography-based foot disease identification, it is difficult to ascertain which ones are most indicative of DFUs [18]. It could be difficult to understand data when there are a lot of features. Experts’ differing interpretations of the data can cause more variation in clinical decision-making due to discrepancies in illness diagnoses. Consequently, in order to improve the overall cost-performance for classification and decrease decision variability and data misinterpretation, it is necessary to develop a subset of significant variables for thermography-based foot disorder detection [19]. Because lowering the amount of features can memory

resources, classifiers with improved cost-performance ratios are attained by using a subset of features containing relevant information [20]. This appropriate subset of characteristics cannot be established due to the imbalanced datasets cases and the absence of consistency among thermograms.

To find out what properties are important for early DFU detection, ML and DL models have been investigated. The research primarily used the INAOE dataset, which is comprised of thermograms showing the plantar aspect of both feet, and was the only publicly available dataset, with a few exceptions. Just recently, a comparable dataset called STANDUP was made public. It offers ways to improve upon the existing state of the art by just adding more samples to train the ML and DL representations. Additionally, the supplementary dataset allows for the assessment of the set of state-of-the-art features' generalizability, which was previously determined using classical and DL methods.

How the residue of the paper is structured is as shadows: In Section 2, the relevant literature is reviewed; in Section 3, the suggested classical is fleetingly described; in Section 4, results are analysed; and in Section 5, conclusion is drawn.

2 Related Works

The research was carried out by Özgür et al. [21] from November 2023 to February 2024. The information was gathered from three Turkish hospitals' endocrinology and metabolic diseases departments. Data was collected using an identification form besides the Foot Care Scale. The significance of the variables was predicted using gradient boosting methods. The research included XGBoost, LightGBM, and Random Forest as its machine learning methods. Patients whose scores were either higher than the mean FCS-OD score or lower than it was predicted using the procedures. The best result was achieved by XGBoost, with an area under the curve (AUC) of 0.7469. In all of the models, gender (0.0038), perceived health condition (0.0218), and treatment regimen (0.0027) were the most common components. The best area under the curve (AUC) model, XGBoost, also found that income level (0.0055) and A1c (0.0020) were predictors of score.

The goal of the study by Chen et al. [22] was to find potential treatment targets by detecting genes associated with macrophages in DFU. This work utilised datasets from the gene RNA-seq datasets linked to DFU. Macrophage subpopulation distribution in the DFU was shown by analysis of the given single-cell data. For additional bioinformatics analysis, four separate RNA-seq datasets were combined into a single DFU cohort. Diverse machine learning techniques for biomarker identification, enrichment analysis, and differential expression (DEG) analysis were all a part of it. Western transcription-quantitative polymerase chain reaction (RT-qPCR) were used to authorize the important findings. Lastly, western blotting and RT-qPCR were used to confirm the results. Using single-cell analysis, to were able to recover 802 genes associated with macrophages. There were 743 differentially expressed

genes (DEGs) found following the study of expression. Through the process of cross-analysis, 37 differentially expressed genes (DEGs) related with macrophages were discovered. Four machine learning techniques were used to screen and cross-analyse 37 junctions. Lastly, HMOX1 was recognised as a biomarker that could be very useful. Several metabolic pathways, including the insulin signalling system, were substantially connected with HMOX1. In DFU samples, HMOX1 was shown to be considerably overexpressed. Finally, this study's analytical results showed that HMOX1 is a macrophage-associated biomarker in DFU that could be useful. Our work sheds light on how macrophages contribute to this condition and could lead to more effective targeted treatments for diffuse focal ulceration (DFU).

Using machine learning, Hong et al. [23] assessed the likelihood of diabetic foot ulcers (DFUs) returning in older diabetic patients with the goal of improving DFU prevention and care. The objective is to develop reliable models that can predict the likelihood of DFU recurrence in older diabetic patients by taking into account risk factors such age, blood sugar control, alcohol intake, and smoking. Machine learning models were built using data obtained from 138 older diabetic patients following data cleansing, screening for outliers, and feature integration. A 93% success rate was achieved using Support Vector Machine (SVM). The experimental consequences demonstration that SVM is a good tool for forecasting the probability of DFU recurrence in older diabetic patients, giving doctors a better gauge to work with. This research shows that machine learning is very useful for forecasting the likelihood of recurrence of foot ulcers in older diabetes patients, which is crucial for their management. This strategy utilises computer-assisted medical procedures to control diabetes in the elderly in a way that allows for timely intervention, which in turn reduces the likelihood of patient recurrence.

In their study on semantic segmentation of photographs of diabetic foot ulcers (DFUs), Saeedi et al. [24] looked at how the U-Net model was applied inside a federated learning framework. Training in both a centralised and a federated learning environment is the goal, with a set of randomly generated U-Net model parameters serving as the basis. We will use an open-source dataset of diabetic foot ulcers supplied by Medetec because medical photos are quite sensitive. There is no longer any need to centralise raw data thanks to federated learning, which allows us to decentralise our approach to machine learning. One of the methods utilised is to compare and contrast centralised and federated machine learning systems through simulations. According to the findings, federated learning in conjunction with the U-Net architecture can obtain a noteworthy dice score of 0.9, which is comparable to the performance of centralised models, while also eliminating the need for centralised data collection. Attaining a balance between privacy issues and analytical accuracy, this conclusion highlights the potential of federated learning to improve detection approaches for DFUs. The federated learning codebases provided by this study allow interested researchers to

replicate and build upon the findings, making a significant contribution to the biomedical imaging field. On GitHub, you can find the source code.

Verma [25] has used a deep learning model to analyse thermal photographs of the foot in diabetic patients in order to diagnose foot ulcers. The model's efficacy was estimated by comparing it to previous research. The study utilised thermal pictures that are available for free online. There are two categories of diabetic foot photos in the dataset: normal and pathological. Out of a total of 1055 photographs in the collection, 543 depict the patient's normal foot and the remaining images show the patient's aberrant foot. After the study's dataset was pre-processed using watershed segmentation and clever edge detection, it was transformed into a new dataset. The pre-processed dataset was subsequently enhanced and balanced by data augmentation. Following this, learning model was utilised for prediction purposes in order to diagnose foot ulcers. The pre-processed dataset, improved by clever edge detection and segmentation, can lower computing cost and improve model performance for accurate predictions. Both the original datasets were used to evaluate our proposed model, which utilised EfficientNetB0. The pre-dataset had edge detection and segmentation applied to it. Impressively, ResNet50 achieved 89% and 89.1% accuracy for the two datasets, correspondingly, whereas EfficientNetB0 outperformed with 96.1% and 99.4% accuracy, respectively. The results were incredibly encouraging.

Wu et al. [26] proposed a developed YOLOv5-based model for vascular stenosis small target detection. Different types were identified by using GhostC3 and tested on the publicly available dataset for detection of diabetic feet.

Using hybrid deep learning models, Chee et al. [27] have studied the use of gait diabetes detection. Hybrid data, like ECG data and the Pima Database, are used in current research. This study uses a hybrid deep layers make up the suggested CNN-LSTM model. CNN-LSTM can learn patterns for classification and extract significant features by merging the two models that are used to compare the CNN-LSTM model's performance to that of the CNN and LSTM models. The accuracy of CNN-LSTM is 91.25%, which is higher than that of current techniques. Therefore, this study shows that non-invasive diabetes detection methods have the potential to displace more conventional invasive approaches. Future research could examine muscle activation and forces in conjunction with gait acceleration to enhance the model's ability to detect diabetes.

The objective of Castillo-Morquecho et al. [28] is to use thermography to examine the foot sole investigate any relationships with clinical variables. A machine learning method was also created for classification. There were twenty-seven age- and sex-matched controls and twenty-three diabetic patients. The plantar foot surface's thermograms were obtained and divided into areas of interest. Using predetermined regions of interest, the mean foot temperature and temperature change index were computed. Glycated haemoglobin, body mass

index, and temperature measurements were all correlation analysis. Principal component investigation and support vector machines were used in a two-layered cross-validation model for classification. Mean foot temperature was found to have significant positive correlations with both body mass index ($\rho = 0.35$, $p = 0.013$) and glycated haemoglobin ($\rho = 0.44$, $p = 0.0015$). There were no discernible relationships among the temperature change index and clinical variables. Based on thermal data, the machine learning model classified the diabetic and control groups with a moderate sensitivity (78.3%) and high overall accuracy (90%) and specificity (100%).

The main goal of Agrawal et al. [29] is to use insoles with wireless plantar pressure sensors to identify and categorise diabetic feet from healthy ones using dynamic plantar pressure. Comparing machine learning (ML) and statistical-based classification techniques is the secondary goal. In line with earlier studies, data from 150 participants who wore the insoles while walking showed that diabetic feet have higher plantar pressure than healthy feet. With an accuracy of 0.85, the Adaptive Boosting (AdaBoost) machine learning model outperformed the statistical approach, which had an accuracy of 0.67. These results imply that plantar pressure features can be used to efficiently classify diabetic and healthy feet when paired with machine learning models and insoles with built-in pressure sensors. Future studies will concentrate on classifying different stages of diabetic neuropathy using these insoles and machine learning, with the goal of predicting foot ulcers early in the home environment.

Using bioinformatics and machine learning analysis, Shi et al. [30] seek to find and validate putative biomarkers of Gln metabolism linked to DFU. Two microarray datasets were obtained from the Gene Expression Omnibus (GEO) database to identify the related genes. To examine the relationship between immune cell infiltration status and deGlnMRGs, the connections were investigated. Notably, WGCNA can be used to pinpoint genes that are differentially expressed (DEGs) within particular clusters. The best machine learning model was then created and vetted. Additionally, two clusters were found in DFU. Immune heterogeneity was found in these two clusters, according to immunological infiltration analysis.

In improve the accuracy of diagnosing Diabetic Foot Ulcers, El-Kady et al. [31] has concentrated heavily on medical image analysis. The hybrid models were evaluated by using the publicly available dataset. Notable results were obtained; the ResNet50 scored 0.75 on the F_1 -score and averaged 0.76 on accuracy and precision. With an F_1 -score of 0.84, a regular accuracy of 0.84, besides precision of 0.85, the hybrid model outperformed these measures. More accurate and efficient DFU identification in clinical settings may be possible thanks to this study, which adds to the growing in medical image processing. The diagnostic accuracy was gradually increased due to the implementation of hybrid proposed model.

Building a convolutional neural network to categorise thermal imageries of the feet as “normal” or “diabetic” has been the primary focus of Sulayman et al. [32]. The authors have performed multiple experiments on various parameters, including optimisers, activation functions, and transfer learning base models, in an effort to increase the model’s accuracy. This work uses the TensorFlow library, and the Keras package for scripting the code. Matplotlib and Excel are used to simulate the experiments and plot the results as training and validation accuracy and loss graphs. For the sake of comparison, only one parameter is varied in each experiment; all other parameters stay the same. So that the findings are solely impacted by the optimiser change, other parameters are kept constant for all trials in an experiment to find an effective optimiser, and the same process is used for all the other experiments. The built convolutional neural network makes advantage of transfer learning to enhance training efficiency and speed while compensating for the short dataset. A comparison is provided between ResNet-50, EfficientNetB3, and VGG16, three of the top transfer learning models. Adam and SGD were evaluated in terms of optimisers. The sigmoid function and the ReLU activation function were compared. The prevailing belief was that the optimal outcome could be achieved by integrating the most effective parameters from each experiment. With a validation accuracy of 90.5%, it was clearly an improvement above previous results. Afterwards, an accuracy of 86.7% was achieved when the prototypical was pictures that were not included in the training dataset.

Illuminate®, a point-of-care platform developed by Sridhar et al. [33], combines autofluorescence intensity imaging with deep learning in less than 2 min. The DenseNet with UNet model was used to detect the gram-type DFU model.

The purpose of the study by Cakir et al. [34] is to assess how well a DL algorithm can distinguish between bone marrow signals from osteomyelitis. The study’s local ethics committee gave their assent to the retrospective design. Partitioning the 148 patients into their respective T1- and T2-weighted images yielded 679 labelled regions 714 labelled regions (160 CNO, 272 OM, and 282 TR), respectively. For T1-weighted images, the ResNet-50 accuracy values were calculated to be 96.2% and 97.1%, correspondingly. Also, for T2-weighted images, the ResNet-50 achieved an accuracy of 95.6% and the EfficientNet-b0 96.8%. In addition, ResNet-50 has a sensitivity of 91.1% for CNO, 92.4% for OM, and 96.6% for TR, as reported by BCC. EfficientNet-b0, on the other hand, has a sensitivity of 93.2% for T1, 97.6% for OM, and 98.1% for TR. For TR, ResNet-50 achieves specificity scores of 98.1%, 97.9%, and 94.7% in T1-weighted pictures, and 98.6%, 97.5%, weighted images, respectively. In a similar vein, EfficientNet-specificity b0’s for T1-weighted pictures are 99.1%, 98.5%, and 98.7%, while for T2-weighted images it is 98.7%, 98.7%, and 98.4%. Deep learning techniques accurately distinguish CNO, OM, and TR in diabetic feet without invasive procedures.

By combining state-of-the-art technology with traditional algorithms, smart sensors linked to an Internet of Things infrastructure and a Raspberry Pi, Murugan and Mohankumar [35] suggest a method for early diagnosis of diabetic complications. It greatly enhances the treatment that is given to people with diabetes. An example of real-time data is the utilisation of a Convolutional Neural Network (CNN) model to interpret things like heel pressure, temperatures, and mobility habits. By distinguishing between normal foot health and the first signs of chronic difficulties, the CNN allows for remote monitoring and rapid treatments to lessen the likelihood of diabetes’s complications. Individuals are able to take charge of their health thanks to a user-friendly interface and the rapid notifications provided by a cloud-based infrastructure. The application of this novel tactic signifies an important advancement in the management of foot complications associated with diabetes. It provides a preventative measure to guarantee patients’ health and welfare and enhance systemic management. The system’s potential to revolutionise diabetes treatment methods is demonstrated by how easily it can be integrated into daily life. This paves the way for a thorough and effective method of reducing diabetic foot complications.

3 Materials Besides Methods

A state-of-the-art set of features may be extracted thermograms using the deep learning methodology presented in this study. The feature extraction process took into account three input datasets that were created by combining separate datasets. Using both traditional and deep learning methods, the study extract a subset of features from the input datasets. The CBGRU-CapNet classifier was fed a subset of features that were common to all the methods used. The study work employs the mDTMA to fine-tune the CBGRU-CapNet settings. Each set of retrieved characteristics from the STANDUP besides expanded databases was evaluated and compared using the suggested classifier as a reference. Furthermore, features that were more relevant and robust were retrieved in this work and compared to features that were extracted using only the INAOE dataset. In this section, brief clarification of projected model for DFU is mentioned in Fig. 1 and each of its block is explained in detailed.

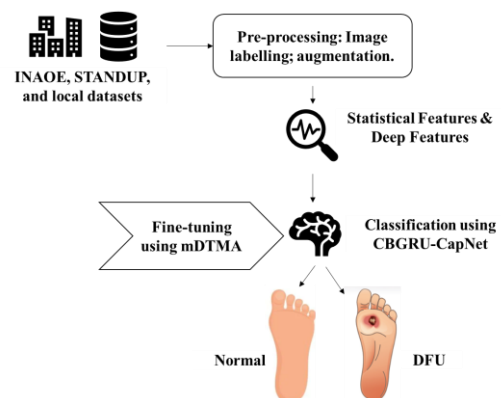


Fig. 1 Workflow of the research model.

3.1 Datasets Description

Throughout this work, infrared (IR) photos captured at various locations and with various sensors are used as part of the datasets. For every dataset, you can also find RGB versions of the same scene. The sole of the foot can be more easily separated from its surroundings with the help of these photos. Also, a different demographic sample was used to launch the matching acquisition campaigns at different times. You may find more information on the infrared acquisition procedure used for the INAOE, STANDUP, and local datasets elsewhere.

3.1.1 STANDUP Dataset

For DFU finding using infrared thermography, the STANDUP dataset was created and released in June 2023 [36]. Among the 227 participants in the study, 145 had diabetes and 82 were considered to be in good health. After 10 min of resting in a position to achieve thermodynamic equilibrium, photographs were taken at T0 for each participant. Then, after a cold stress test, photographs were taken at T10. R0 was for people without neuropathy or ischaemia; R1 was for people with neuropathy but no ischaemia; and R2 was for people with ischaemia in the diabetic group. This study did not differentiate between diabetic and healthy patients when determining pathology grades; all subjects were included for classification. In addition, T0 photos are the main focus of the analysis, which allows for useful comparisons with other INAOE and local. In other words, thermograms from people who were cold-stress tested were not taken into account. As a consequence, 145 individuals with diabetes and 38 healthy controls were removed from STANDUP dataset. Take note that this dataset is severely skewed towards cases of diabetes, just like the INAOE dataset.

3.1.2 INAOE Dataset

Among the 167 volunteers included in the INAOE thermogram dataset, which was made public in December 2019, 122 had diabetes and 45 did not [37]. The INAOE dataset has an imbalance in the number of cases with diabetes. It was initially planned to use this dataset to compare the distribution of plantar area temperatures in participants with and without diabetes. Nonetheless, thermograms have found extensive application in the early diagnosis of DFU.

3.1.3 Local Dataset

This effort is associated with the public release of the local dataset, which was acquired in 2021 [38–39]. The dataset includes samples taken from 22 healthy unpaid helpers at four separate time points; however, metaphors from the 15-min resting site before acquisition were used. This dataset was combined INAOE dataset to create a larger dataset that aims to make up for the INAOE dataset's bias towards diabetes cases.

3.2 Pre-Processing

In order to recover the accuracy of the classification, it is necessary to clean the input data before processing it.

3.2.1 Image Labeling Approach

On the other hand, a large quantity of labelled data is necessary for a deep learning model to perform optimally, particularly one that is CNN-based. However, it can be time-consuming and expensive to acquire a mountain of medical data. Consequently, methods such as regularisation, data picture labelling can be utilised to improve the performance of DL models and address overfitting. This problem is solved by expanding the dataset using a patch labelling strategy. This method of sample labelling involves selecting large portions of a sample, cropping them, and then assigning the appropriate class to the cropped samples. Initially, to extract the Region of Interest (ROI) from the samples, the research work uses the 224-pixel-wide and 224-pixel-high windows value. Classification of the patches into ulcer or healthy categories is based on the areas that encompass both ulcer skin and normal skin. Instead of focussing on the full sample, the DFU_ XAI framework labelled areas that were relevant (i.e., areas connected to ulcers) within bigger samples. This allowed the DL models to be more efficient. We used this method to extract relevant DL data for classification. Overfitting can be mitigated in part by training DL models with relevant DL data, which decreases computational load and memory allocations.

3.2.2 Data Augmentation

Data augmentation approaches are utilised to augment data in order to make it more varied and larger. These methods include random scaling besides rotation, besides contrast enhancement using new colour spaces. We used the rotation method in our research, which involves spinning each of the input skin patches through two different angles of 90° and 180°. The “train_test_split” function makes short work of dividing the DFU datasets into a “train” set consisting of 80% of the data besides a “test” set containing of 20%. This method is imported using the “sklearn.model_selection” package in Python. Ten percent of the training set's data is separated once more for validation using the same splitting procedure.

3.3 Features Extraction

This subsection outlines the two distinct feature extraction methodologies that to compare: the statistical method and the deep features-based method.

3.3.1 Statistical Features Extraction

Extracting features from raw data using various statistical procedures is what statistical features extraction is all about. Following a method that is commonly used in literature, we calculated 9 statistical indicators for each variable and time window: median, standard deviation,

skewness, kurtosis, quartile. In order to facilitate the ML algorithm's eventual discovery of patterns in the data, the acquired values are then normalised so that all of the values are on the same scale. Ultimately, we obtained a matrix in $\mathbb{R}^{N \times 9 \times V}$, where V is the sum of variables and N is the number of time windows.

3.3.2 Convolutional Neural Network

One category of neural network developed to detect input spatial patterns is the CNN [40]. It is based on the principle of weight sharing, wherein a single weight is utilised by multiple neurons in a layer. Hence, a layer's output, also known as the feature map, can be determined by taking the layer's convolutional product with a specific matrix, also known as a convolutional filter. For this reason, the feature m in the formula for a 1D-CNN with a single filter is mentioned as $M \in \mathbb{R}^K$ can be written as linear elements in the input layer $I = \{I_{j,h}\} \in \mathbb{R}^{N \times H}$ weighted by the filters $W = \{W\} \in \mathbb{R}^L$:

$$M_k = f\left(\sum_{l=1}^L \sum_{h=1}^H W_l I_{(k-1)*s+l,h}\right), \quad (1)$$

where s is the stride of the convolution and the dimension of the feature map, K is equaled to $\left\lfloor \frac{N+1-L}{s} \right\rfloor$.

The transposed convolutional layer, also called the deconvolutional layer, is another variety of convolutional layer. For the convolutional layer, it is a common way to go back to the input's original dimensions while still keeping the spatial relationships between data intact. The feature map is thus derived from the aforementioned formulation. $M \in \mathbb{R}^K$ has the dimension equal to $K = N + (L - 1) * s$.

3.3.3 AutoEncoder

Reducing the feature space's dimensionality is the goal of an AutoEncoder (AE) neural network. Projecting the raw data into a subset called latent space allows us to complete this task. By filtering out irrelevant data and background noise, the projection is engineered to retain key details.

The encoder and decoder are the two main parts of an AE. The first one maps the inputs onto the latent space, and the second one uses the encoded input to rebuild the features. The decoder is required to train the complete AE, even if our aim only requires the encoder. By using the l2 norm of reconstruction provided by the AE as a loss function, the network is trained to extract a meaningful latent space.

The AE is built using the CNN in our application. Specifically, there are layers in the encoder and two deconvolutional layers in the decoder. To get a latent space with dimensions comparable to the statistical method, the kernel size and stride of the encoder layers are selected. On the flip side, deconvolutional layer hyperparameters are selected with the goal of achieving Decoder symmetry with respect to the Encoder.

The AE is built using the CNN in our application. Specifically, there are two convolutional layers in the

encoder and two deconvolutional layers in the Decoder. The goal of selecting the encoder layers' kernel sizes and strides is twofold: first, to achieve a latent space with dimensions comparable to those of the statistical method; and second, to maintain a smooth dimensionality drop between them. Deconvolutional layered feature maps, on the other hand, are created with an effort to maintain encoder-decoder symmetry. For activation function, the study uses the ReLU. Other options, like sigmoid and tanh, have also been investigated. But they do not seem to boost ReLU's performance much at first look, so to stick with this activation function.

3.4 Classification Using CBGRU-CapNet

In this work, the CBGRU-CapNet is used to detect the DFU from the extracted statistical and deep features from CNN-AE. The multi-layered architecture of the model is examined in the following subsections.

3.4.1 Convolutional Layer

Three convolutional layers – conv1, conv2, and conv3 – in a convolutional neural network (CNN) are used by the model to extract regional patterns input data. These layers use kernel sizes of 2, 3, and 4, sliding filters over token embeddings to create feature maps. These maps capture interactions among neighboring tokens, extracting spatial, and temporal highlights associated with DFU. With 128 channels of varying sizes, the convolutional layers perform operations and derive a feature sequence f_s from the input data. This sequence undergoes maximum pooling and concatenation, ultimately contributing to the subsequent layer's final feature vector.

$$fn = f(wtxt + b). \quad (2)$$

3.4.2 BiGRU Layer

A bidirectional Gated Recurrent Unit (BiGRU), a specific kind of recurrent patterns and relationships within the data, is used in this layer. The GRU's bidirectional capability allows it to analyse input sequences both forward and backward, at the same time collecting data from the past and the future. In the BiGRU setup, a forward GRU processes succeeding sequences (f_1 to f_{128}), while a backward GRU handles preceding sequences (f_{128} to f_1).

The combination of CNN and BiGRU in a proposed architecture is a key to boosting accuracy. The integration of CNN and BiGRU not only ensures the extraction of features but also finds unique trends and patterns in the contextual features of the various ROI of three datasets. Hence, the above-discussed combination provides extraordinary mapping between the input images and extracted features in the DFU detection process.

$$ForwardHiddenState(h_f) = GUR(L_{fs}), \quad n \in [1, 128], \quad (3)$$

$$\text{BackwardHiddenState}(h_b) = \text{GUR}(L_{fs}), n \in [128,1], \quad (4)$$

$$\text{TotalHiddenState}(h_t) = [h_f, h_b]. \quad (5)$$

3.4.3 GRU Cell

The cell is equipped with gate (i_t), the forget gate (f_t), and the output gate (o_t). These gates serve the purpose of preserving and modifying information related to the data preceding time t . The cell states and updating the parameters is governed by Eq. (6) to Eq. (7).

The Gated Recurrent Unit (GRU) is a particular kind of cell created to detect long-range relationships in sequential input while minimising vanishing gradient issues. It is made up of a candidate activation ($\tilde{h}_t$) that computes the intermediary hidden state, an update gate (z_t), and a reset gate (r_t), which manage the movement of information, and the last hidden state to be computed (h_t), which acts as the cell's output. The equations for the GRU cell are as shadows:

$$(\text{updategate})z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \quad (6)$$

$$(\text{Resetgate})r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \quad (7)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t]), \quad (8)$$

$$(\text{newhiddenstate})h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t. \quad (9)$$

Here, σ stands for the sigmoid activation function, $\tanh$ for the hyperbolic tangent activation function, W_z , W_r , and W_h respectively, for weight matrices, h_{t-1} for the prior hidden state, and x_t for the input at time step t . The update gate z_t regulates how much of the preceding hidden state and candidate activation are combined to create the new hidden state h_t , while the reset gate r_t regulates which portions of the previous hidden state are disregarded when calculating the candidate activation.

To capture context-rich sequences in both directions (forward as well as backward), the model applies a BiGRU layer to the result of the convolutional layer. The outputs of the BiGRU for forward and backward directions are represented by Eqs. (3) and (4), respectively. The contextual representation of the input text is combined with information from both directions in the BiGRU output. This representation based on BiGRU combines the forward (h_f) and backward (h_b) hidden states, aligning them to reconstruct comprehensive, contextually-informed sequences. Eq. (5) ultimately expresses the combined contextual sequence as the final hidden state h_t , which is then passed to the capsule network layer.

3.4.4 Capsule Layer

Furthermore, the precision of DFU detection is ameliorated by using the capsule layer. Traditional CNNs lack distinct feature extraction and lose essential information due to pooling methods. Capsule networks address this by capturing linguistically supplemented

features that consider order and local context. They excel in tasks like classification and retrieval, outperforming CNNs. Capsules comprise smaller capsules and use vectors to convey classification likelihood and aspect orientation, resulting in more efficient representation. Unlike CNNs, capsules generate vectors instead of scalars, and the dynamic routing algorithm adjusts weights, enhancing feature extraction.

The Capsule Network (CapsNet) is handling hierarchical relationships and spatial hierarchies in data. Capsules are a fundamental building block in CapsNets, and they aim to represent specific patterns or entities in a more robust manner. The key equations for Capsules and routing by-agreement are given. Capsule C_i in layer l is defined as a vector s_i representing the instantiation parameters of an entity, such as the pose (position, orientation, etc.). In order to guarantee that the length of s_i is from 0 to 1, it also includes a squash function connected to it.

$$s_i = W_i \cdot x_i, \quad (10)$$

$$v_i = \text{squash}(s_i). \quad (11)$$

Routing-by-Agreement (Dynamic Routing).

Information is sent from capsules to higher-level capsules via dynamic routing to guarantee agreement among each capsule on the presence of entities.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} (\text{Softmax to compute routing weights}), \quad (12)$$

$$s_j = \sum_i c_{ij} \cdot u_{ji} (\text{Weighted sum of predictions}), \quad (13)$$

$$v_j = \text{squash}(s_j) (\text{Squash function}), \quad (14)$$

where C_i represents a capsule in the lower layer, v_i is the output of that capsule, W_i represents the weight matrix, x_i is the input to the capsule, c_{ij} are the routing weights, b_{ij} are the log prior probabilities, u_{ji} is the prediction vector from capsule i to capsule j , v_j is the output of the higher-level capsule.

CapsNets are designed to hierarchical dealings by routing information between capsules and have shown promise in various tasks, especially in vision.

Our model utilizes capsules to improve DFU classification. As a result, the final hidden state, denoting the output from the BiGRU layer, is transmitted to the capsule network layer. Dynamic routing is used for building coupling coefficients c_{ij} , which enables the model to ignore trivial terms in the input data. The importance of DFU-related characteristics is correlated with the coupling coefficient c_{ij} ; larger values denote more significant features, which capture crucial semantic representations of ROI in several directions. The coupling coefficient c_{ij} is computed using the Softmax function, and the output of the capsule, s_j , is obtained by adding up all of the prediction vectors.

3.4.5 Dense and Output Layers

The connected layer receives the output vector, v_j , from capsule network. Finally, using the sigmoid function on the dense layer output, the output layer classifies DFU. The proposed CBGRU-CapNet model utilizes a binary cross-entropy loss function.

3.5 Fine-Tuning of CBGRU-CapNet Using mDTMA

Here, a metaheuristic manifold algorithm [41] is used to fine-tune the CBGRU-CapNet parameters. Its architecture is comparable to the classic metaheuristic approach on Euclidean space, but its implementation of operations is different. Each operation must be performed in this implementation to maintain the individual's position on the manifolds. To be more precise, mDTMA, the anticipated algorithm, begins with a randomly generated population, with each person drawn from the manifold's points. The next step is to optimise the population towards better locations using a unique manifold version of the update and selection technique. This can be done by selecting individuals with superior fitness values for a goal, which is a real-valued function. Throughout the process, the population stays on the manifold, ensuring that workable answers will be found. In an effort to make the proposed algorithm more accessible, we have tried to name it after well-known operations from classic metaheuristic algorithms, like selection; however, these operations have been modified or improved to meet the needs of manifolds.

Optimisation issues involving real-valued functions on matrix manifolds are taken into account in this study. Each element of the solution set can be naturally characterized in matrix arrays, since the set is located on a matrix manifold. This allows for the progress of differential geometry through the formulation and control of matrix algebra. Positive definite simplex manifolds, Grassman manifolds, oblique spaces, product manifolds (rotation with Euclidean space), sphere manifolds, and matrices manifolds are among the manifolds to which our approaches are applied.

3.5.1 Main Framework of mDTMA

Initialising the population, selecting the tournament, crossing, mutation of polynomials, directional retraction are the fundamental steps of mDTMA. Directional are valuable because they facilitate mutation and crossover processes.

Let us consider a real valued function $f: \mathcal{M} \rightarrow \mathcal{R}$, in which a submanifold $\mathcal{R}^{n \times p}$ is embedded, which is explicit limit set in $\mathcal{R}^{n \times p}$. For instance, Stiefel manifold $St(p, n)$ is a collection of submanifold of $\mathcal{R}^{n \times p}$. In particular, when $p = 1$, it sphere; when $p = n$, it is orthogonal matrices $O(n)$.

Let us represent $\mathcal{P}^t \in \mathcal{M}$ as size of N population for t group, $\mathcal{M}$ as manifold, and $\mathcal{P}^0$ as the initial population, where $\mathcal{P}^0$ is defined as $\mathcal{P}^0 = \{x_1^0, x_2^0, \dots, x_N^0\}$, each agent (chromosome) x_n^0 appears as $x_n^0 \rightarrow \{R^{a \times b}, \dots, R^{c \times d}\}_k$

and k denotes the element and the issue definition determines the sum of matrices (or solution components) in each agent.

The mDTMA procedure starts populace $\mathcal{P}^0$, where each agent is located on the manifold M . Initially, we used tournament selection to return the indices of N solutions by 2-tournament preference (random list of candidates) based on their fitness values. The contender with the lowest fitness score $f(x)$ will be designated selection. When compared to the conventional crossover, in which the offspring are not given any parental figures, this crossover is unique. The two parents only partially operate in this stage; the remaining work is done by the retraction step, which to call pre-crossover. In order to increase the variety of targets and their exploitability, mutation is a necessary step. Two goals from the previous two steps will typically depart from the manifold in the case that the provided M is not linear, according to directional transport. The directed transport operator takes the objective and breaks it down into its component parts, keeping the direction and length. In contrast to parallel transport, which uses a geodesic path while maintaining length but not angle, this operator for directional transport takes both into account. Finally, two possibilities are generated by this stage. Using the retraction are retrieved to the manifold as a new generation.

3.5.2 Implementation of Operators

In agents, pre-cross over as $x_i^t, x_j^t \rightarrow \{R^{a \times b}, \dots, R^{c \times d}\}_k$, k iterations are executed, one element at a time. Let us assume that for every element in both agents, parent1 and parent2, β are randomly generated where β belongs to $\mathcal{R}^{c \times d}$, $\beta \in \mathcal{R}^{c \times d}$

To obtain the goal, let us compute the crossover using β generated d -length vector:

$$v_1(target1) \leftarrow [\beta \odot (Q_1^t - Q_2^t) \odot \frac{1}{2}], \quad (15)$$

$$v_2(target2) \leftarrow [-\beta \odot (Q_1^t - Q_2^t) \odot \frac{1}{2}], \quad (16)$$

where $\odot$ achieves element-wise multiplication.

If you want to describe an isometric movement along a curve from tangent vectors to tangent space, you can use parallel transports. Assuming the existence and uniqueness of a minimising geodesic, the curve is typically left out when to implicitly use it as the curve. On the other hand, solving equation is often expensive, and it is necessary for algorithmic parallel transport. Consequently, the pragmatic option pertains to computationally efficient methodologies, transporters, and a stand-in for equivalent transport that is not constrained by connection limitations.

Orthogonal projector: Orthogonal map $Proj_x: \varepsilon \rightarrow \varepsilon$ categorized by three properties: (1) $im(Proj_x) = T_x \mathcal{M}$; (2) $Proj_x \circ Proj_x = Proj_x$; (3) $\langle u - Proj_x(u), v \rangle = 0$ for all $v \in T_x \mathcal{M}$ and $u \in \varepsilon$.

Denote $Proj_x$ the orthogonal from ε to $T_x\mathcal{M}$ and deduce $Proj_x^\perp = Id - Proj_x$ which is the orthogonal projector to $N_x\mathcal{M}$. Operation $T_x\mathcal{M}$ and $N_x\mathcal{M}$ are complements in ε , where Id represents the individuality operator.

To calculate the orthogonal projection, we need to choose a Riemannian manifold, M . $T_x\mathcal{M} = \{v \in R^d : x^T v = 0\}$ is well defined by orthogonal accompaniment of x in Euclidean space ε . Therefore, any component aligned with x is removed by embedded manifold (the underlying Euclidean metric) to a tangent space of x . Evaluate the orthogonal projectors' features, for all $u, v \in \varepsilon$, there is $\langle Proj_x(u), v - Proj_x(v) \rangle = 0$. This movement $z \in \varepsilon$ from a point $x \in \mathcal{M}$ how tangential and normal motion are composed. To suggest an operator, a directed transport operator, move the candidate isometrically informally.

Let us equip $\mathcal{M}$ with metric. Then, there is a map $\eta_x(z) \cdot \eta_x(z) : \varepsilon \rightarrow T_x\mathcal{M} \oplus N_x\mathcal{M} \rightarrow (Proj_x(z) \oplus Proj_x^\perp(z))$, where $\eta_x(z)$ describes the procedure to $z \in \varepsilon$ of the point x on the manifold $\mathcal{M}$.

Given $Proj_x : \varepsilon \rightarrow T_x\mathcal{M} : u \mapsto Proj_x(u) = u - \langle x, u \rangle x$, which describes a map of $T\mathcal{M} \rightarrow \mathcal{M}$. The projection $Proj_x^\perp$ constituent satisfies $R_x(Proj_x^\perp) = x$, only when manifold $\mathcal{M}^{d-1}$ with embedded space R^d , for the purpose of defining retraction. There are some areas where to could theoretically avoid doing the same searches twice.

4 Results and Discussion

Our work is executed on Dell computers powered by Intel Core i7 processors. The utilisation of a robust GPU, and more especially the P100 prototypical, was critical in expediting the training process. The computational load was efficiently handled by our system with 32 GB of system RAM, thanks to the Kaggle kernel environment. Python 3.9.7 and TensorFlow 2.12.0 were used for the implementation. Table 1 discusses the training parameters.

Table 1 Training hyperparameters.

Setting	Hyperparameter
32	Batch size
90%	Training besides validation split
10%	Test split
mDTMA	Optimizer
150 × 150 px	Input size
Categorical cross-entropy	Loss function
0.001	Learning rate
10	Epochs

4.1 Validation Study of Proposed Model on First Dataset

Fig. 2 provides the graphical description of projected model on different training and testing data ratio on first dataset in terms of different metrics. The comparative analysis of various algorithms in terms of different metrics is shown in Fig. 3.

The proposed model demonstrates notable improvements in performance as the training data percentage increases. With a 60–40 training-testing split, the model accomplishes an accuracy of 69.04%, sensitivity of 90%, specificity of 43%, and an F_1 -score of 73%, indicating a moderate detection.

The comparative analysis across various models (Table 2) highlights that the projected model achieves the highest presentation in all metrics, with an accuracy of 97.38%, F_1 -score of 97.05%, precision of 96.98%, besides recall of 97.12%. ResNet50 performs notably well with an accuracy of 95.02% and F_1 -score of 95.61%, proving its capability in handling complex classification tasks. CNN-LSTM and CNN also maintain competitive performance, achieving accuracies of 94.36% and 94.12%, respectively. Meanwhile, SVM, and XGBoost yield slightly lower metrics but still demonstrate strong classification accuracy and robustness. Overall, the proposed model's superior metrics underscore its effectiveness and reliability for the intended application, outperforming other approaches in the study.

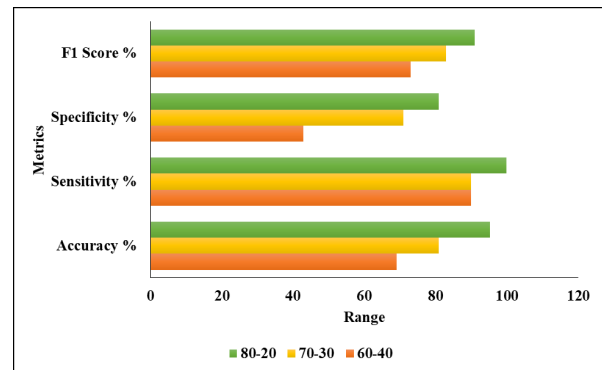


Fig. 2 Visual study of proposed prototypical on first dataset.

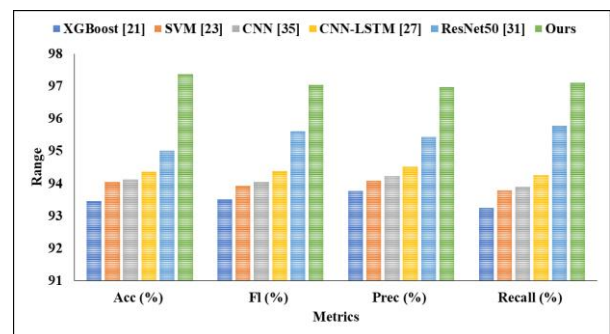


Fig. 3 Graphical description of projected model.

Table 2 Validation analysis of the desired model using existing techniques.

Methods	Accuracy (%)	F_1 -score (%)	Precision (%)	Recall (%)
XGBoost [21]	93.46	93.51	93.78	93.25
SVM [23]	94.06	93.94	94.09	93.79
CNN [35]	94.12	94.05	94.22	93.89
CNN-LSTM [27]	94.36	94.39	94.53	94.26
ResNet50 [31]	95.02	95.61	95.44	95.79
Ours	97.38	97.05	96.98	97.12

4.2 Validation Analysis of Proposed Classical on Second Dataset

Fig. 4 provides the graphical description of predictable model on second dataset by using different training ratio.

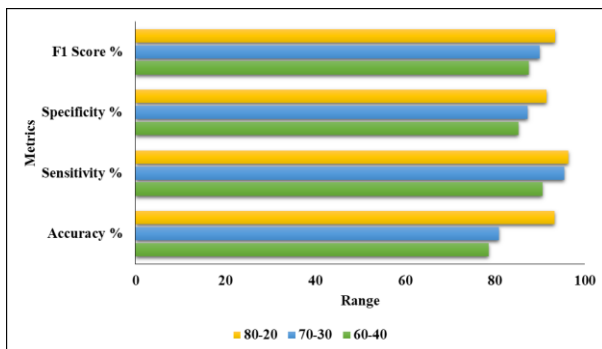


Fig. 4 Visual analysis of proposed model by modifying the second data training and testing ratio.

The proposed model exhibits a clear performance improvement with an increase in the training data percentage, reflecting the model's robustness with more training data. For the 60–40 training-testing split, the model accomplishes an accuracy of 78.57%, sensitivity of 90.65%, specificity of 85.23%, besides an F_1 -score of 87.58%, showing solid performance. With a 70–30 split, the model's accuracy rises to 80.95%, sensitivity to 95.52%, specificity to 87.25%, and F_1 -score to 89.95%, indicating enhanced detection capabilities. The highest metrics are observed with an 80–20 split, where the model achieves 93.26% accuracy, 96.36% sensitivity, 91.52% specificity, and an F_1 -score of 93.36%. This upward trend underscores the model's improved reliability and effectiveness in correctly identifying relevant patterns as more data becomes available for training.

The comparative analysis of various methods highlights the superior performance of the proposed model and it is visually shown in Fig. 5. XGBoost achieves an accuracy of 82.99% score of 83.74%, precision of 81.03%, and recall of 85.45%. SVM performs slightly better with an accuracy of 84.14%, F_1 -score of 85.43%, precision of 83.48%, and recall of

86.39%. CNN achieves 86.18% accuracy, F_1 -score of 86.05%, precision of 86.31%, and recall of 89.79%. The CNN-LSTM model further improves, with an accuracy of 90.18%, F_1 -score of 89.91%, precision of 89.69%, and recall of 90.13%. ResNet50 outperforms the previous methods, achieving an accuracy of 91.33%, F_1 -score of 91.03%, precision of 90.48%, and recall of 91.59%. The proposed model surpasses all others, with an accuracy of 93.26%, F_1 -score of 92.36%, precision of 91.29%, and recall of 93.46%, demonstrating its effectiveness and robustness in classification tasks.

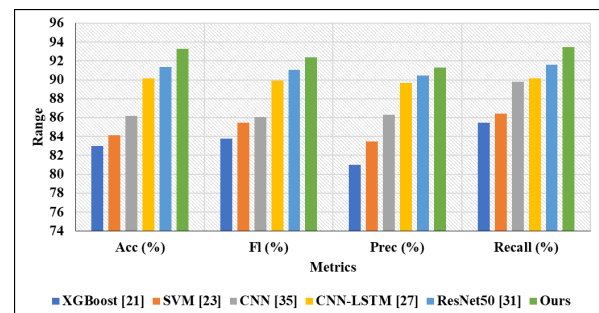


Fig. 5 Graphical comparison of projected model with existing works on second dataset.

4.3 Validation Analysis of Projected Model on Third Dataset

Figs. 6 and 7 presents the visual analysis of various models with proposed model on third dataset in terms of various metrics.

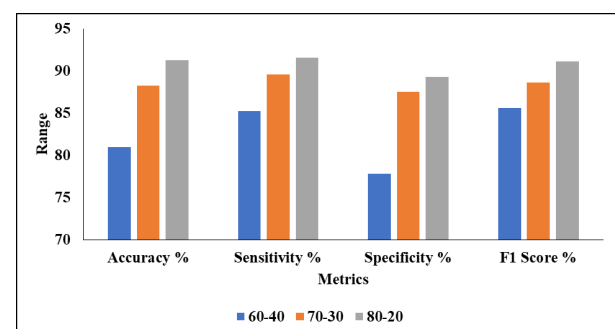


Fig. 6 Training and testing ratio of proposed model.

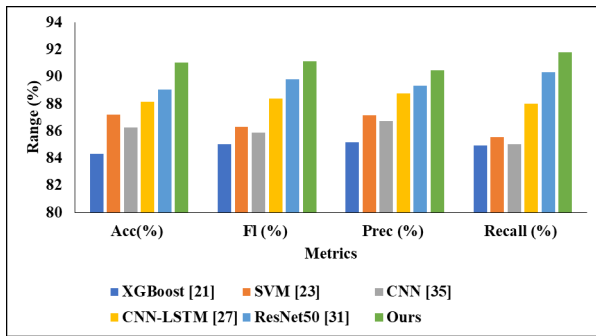


Fig. 7 Visual comparison of proposed model with existing techniques.

In this performance analysis of the proposed model across different training/testing splits, the accuracy, sensitivity, specificity, and F_1 -score show a positive trend as the training percentage increases. With a 60–40 split, the model accomplishes an accuracy of 80.95%, sensitivity of 85.25%, specificity of 77.85%, besides an F_1 -score of 85.64%. Moving to a 70–30 split, the accuracy rises to 88.25%, sensitivity to 89.58%, specificity to 87.52%, and F_1 -score to 88.64%, indicating improved performance. With an 80–20 split, the model achieves the highest metrics, accuracy of 91.23%, sensitivity of 91.56%, specificity of 89.25%, and F_1 -score of 91.12%. This analysis recommends that increasing the training data proportion enhances the model's predictive capabilities, particularly in achieving higher accuracy and F_1 -scores.

This comparative analysis of classification methods highlights the improved performance of the proposed model (“Ours”) over established techniques such as XGBoost, SVM, CNN, CNN-LSTM, and ResNet50. The

proposed model attained the highest accuracy of 91.02%, F_1 -score of 91.12%, precision of 90.46%, and recall of 91.79%. Among existing models, ResNet50 performed best, with 89.02% accuracy and a 90.31% recall, while CNN-LSTM achieved a strong F_1 -score of 88.36%. The results designate that the proposed model consistently outperforms other methods across altogether metrics, demonstrating its efficacy and robustness for classification tasks, particularly in precision and recall metrics, which are crucial for balanced model performance.

4.4 Analysis of Our Model on Merged Data

Tables 3 and 4 delivers the comparative study of various models on merged data in terms of loss and accuracy.

In the analysis of model performance based on parameters, training accuracy, and loss, the proposed model exhibits exceptional results with a training accuracy of 100% and a minimal loss of 0.0021, indicating highly effective training. CNN-LSTM follows closely with an accuracy of 99.97% and a loss of 0.0027, and CNN achieves 99.91% accuracy with a slightly higher loss of 0.0042. SVM also demonstrates strong performance with 99.86% accuracy and 0.0057 loss, despite having significantly fewer parameters at 8.1M. In contrast, ResNet50, with the highest parameter count (86M), shows relatively lower training accuracy (74.84%) and a higher loss of 0.5115, suggesting overfitting or inefficiency with the given dataset. This analysis highlights the proposed model's balance between parameter efficiency and training effectiveness, outperforming in accuracy and loss across the evaluated models.

Table 3 Various Models' investigation on merged dataset.

Model name	Parameters	Training Accuracy	Loss
XGBoost [21]	61.9 M	0.8763	0.3233
SVM [23]	8.1 M	0.9986	0.0057
CNN [35]	5.3 M	0.9991	0.0042
CNN-LSTM [27]	11.2 M	0.9997	0.0027
ResNet50 [31]	86 M	0.7484	0.5115
Ours	22.9 M	1.0000	0.0021

Table 4 Comparative analysis of different models on merged data.

Model name	Accuracy %	Precision %	Recall %	F_1 -score %
XGBoost [21]	87	87	89	87
SVM [23]	89	88	90	89
CNN [35]	91	90	92	92
CNN-LSTM [27]	93	92	93	93
ResNet50 [31]	95	94	94	94
Ours	96	96	96	96

In the comparative analysis of different models on merged data, the projected model outperforms other approaches with score of 96% across all metrics. ResNet50 follows closely, achieving 95% accuracy, and F_1 -score, while CNN-LSTM achieves 93% across these metrics, indicating strong performance but slightly lower than ResNet50 and the proposed model. CNN achieves a balanced 91% in all categories, with SVM and XGBoost at 89% and 87%, respectively. The results demonstrate that the proposed model provides superior predictive power and robustness, indicating its suitability for tasks requiring high accuracy besides consistency across multiple estimation metrics.

5 Conclusion

To enhance classification accuracy and minimize decision variability and data misinterpretation, it is crucial to identify a subset of significant variables for early thermographic detection of foot illnesses. However, due to the imbalance in datasets concerning diabetic cases and inconsistencies among thermograms, defining an optimal subset of features remains challenging. In this study, lightweight deep neural network models for DFU recognition were developed, integrating convolutional networks, BiGRU, and capsule networks. The proposed approach effectively captures features from multiple orientations through the capsule network, while the CNN layer identifies semantic and contextual details. Additionally, the BiGRU model leverages its hidden layers to extract valuable patterns from contextual information. To further refine classification performance, the mDTMA was employed for optimization. The results indicate that the proposed model improves classification accuracy and efficiently localizes ulcers, thereby

increasing diagnostic confidence among medical professionals. Beyond enhancing patient care and outcomes, this approach can aid in accurately diagnosing foot ulcers caused by conditions beyond diabetes.

Future research could focus on developing a multiscale transfer learning framework that integrates multiple pre-trained CNN models to enhance efficiency and classification accuracy. This approach aims to reduce training time while improving specificity in skin lesion categorization. Furthermore, expanding the proposed framework to detect a broader range of dermatological diseases and various skin lesions can be explored by leveraging multiple pre-trained CNN networks instead of a single model. Additionally, incorporating explainable AI techniques could provide deeper interpretability, aiding clinicians in decision-making and ensuring reliable diagnosis.

6 Data Availability Statement

The INAOE and STANDUP datasets employed in this study are publicly obtainable at <https://ieee-dataport.org/open-access/plantar-thermogram-databasestudy-diabetic-foot-complications> and https://www.standupproject.eu/manager/?conf=default&route=/STANDUP_Database (accessed 2 Feb 2024), respectively. The local dataset has been made available at <https://www.iac.es/en/science-and-technology/technology-transfer-iactec/forms> (accessed 17 Feb 2024). Notice that, due to privacy restrictions, only 19 of the 22 images sets originally acquired were released.

Disclosures

Author declares there is no conflict of interest.

References

1. J. Reyes-Luévano, J. A. Guerrero-Viramontes, J. Rubén Romo-Andrade, and M. Funes-Gallanzi, “DFU_VIRNet: A novel Visible-InfraRed CNN to improve diabetic foot ulcer classification and early detection of ulcer risk zones,” *Biomedical Signal Processing and Control* 86, 105341 (2023).
2. S. K. Das, P. Roy, P. Singh, M. Diwakar, V. Singh, A. Maurya, S. Kumar, S. Kadry, and J. Kim, “Diabetic foot ulcer identification: A review,” *Diagnostics* 13(12), 1998 (2023).
3. A. Alqahtani, S. Alsubai, M. P. Rahamathulla, A. Gumaiei, M. Sha, Y.-D. Zhang, and M. A. Khan, “Empowering foot health: Harnessing the adaptive weighted sub-gradient convolutional neural network for diabetic foot ulcer classification,” *Diagnostics* 13(17), 2831 (2023).
4. D. G. Armstrong, T.-W. Tan, A. J. M. Boulton, and S. A. Bus, “Diabetic foot ulcers: A review,” *JAMA* 330(1), 62 (2023).
5. S. Nagaraju, K. V. Kumar, B. P. Rani, E. L. Lydia, M. K. Ishak, I. Filali, F. K. Karim, and S. M. Mostafa, “Automated diabetic foot ulcer detection and classification using deep learning,” *IEEE Access* 11, 127578–127588 (2023)..
6. C. Cao, Y. Qiu, Z. Wang, J. Ou, J. Wang, A. H. Hounye, M. Hou, Q. Zhou, and J. Zhang, “Nested segmentation and multi-level classification of diabetic foot ulcer based on mask R-CNN,” *Multimedia Tools and Applications* 82(12), 18887–18906 (2023).
7. O. Niță, L. I. Arhire, L. Mihalache, A. D. Popa, G. Niță, A. Gherasim, and M. Graur, “Evaluating classification systems of diabetic foot ulcer severity: A 12-Year retrospective study on factors impacting survival,” *Healthcare* 11(14), 2077 (2023).
8. E. Tamir, O. Rabau, Y. Beer, Y. Smorgick, H. Kaufman, and A. S. Finestone, “A novel classification for diabetic foot ulcers of the first ray,” *Advances in Skin & Wound Care* 36(1), 30–34 (2023).

9. M. S. A. Toofanee, M. Hamroun, S. Dowlut, K. Tamine, V. Petit, A. K. Duong, and D. Sauveron, “[Federated learning: Centralized and P2P for a siamese deep learning model for diabetes foot ulcer classification](#),” *Applied Sciences* 13(23), 12776 (2023).
10. P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, “[Diabetic foot ulcer detection using deep learning approaches](#),” *Sensors International* 4, 100210 (2023).
11. S. K. Das, S. Namasudra, A. Kumar, and N. R. Moparthi, “[AESPNet: Attention enhanced stacked parallel network to improve automatic diabetic foot ulcer identification](#),” *Image and Vision Computing* 138, 104809 (2023).
12. B. Gunapriya, T. Rajesh, A. Thirumalraj, and B. Manjunatha, “[LW-CNN-based extraction with optimized encoder-decoder model for detection of diabetic retinopathy](#),” *Journal of Autonomous Intelligence* 7(3), (2023).
13. M. H. Alshayeji, S. ChandraBhasi Sindhu, and S. Abed, “[Early detection of diabetic foot ulcers from thermal images using the bag of features technique](#),” *Biomedical Signal Processing and Control* 79, 104143 (2023).
14. K. McDermott, M. Fang, A. J. M. Boulton, E. Selvin, and C. W. Hicks, “[Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers](#),” *Diabetes Care* 46(1), 209–221 (2023).
15. V. Sathya Preiya, V. D. A. Kumar, “[Deep learning-based classification and feature extraction for predicting pathogenesis of foot ulcers in patients with diabetes](#),” *Diagnostics* 13(12), 1983 (2023).
16. A. Hernandez-Guedes, N. Arteaga-Marrero, E. Villa, G. M. Callico, and J. Ruiz-Alzola, “[Feature ranking by variational dropout for classification using thermograms from diabetic foot ulcers](#),” *Sensors* 23(2), 757 (2023).
17. I. Khosa, A. Raza, M. Anjum, W. Ahmad, and S. Shahab, “[Automatic diabetic foot ulcer recognition using multi-level thermographic image data](#),” *Diagnostics* 13(16), 2637 (2023).
18. S. Biswas, R. Mostafiz, B. K. Paul, K. M. Mohi Uddin, M. M. Rahman, and F. N. U. Shariful, “[DFU_MultiNet: A deep neural network approach for detecting diabetic foot ulcers through multi-scale feature fusion using the DFU dataset](#),” *Intelligence-Based Medicine* 8, 100128 (2023).
19. M. Bundó, B. Vlachos, J. Llusa, I. Bobé, M. Aivar, C. Ciria, A. Martínez-Sánchez, J. Real, M. Mata-Cases, X. Cos, M. Dòria, J. Viade, J. Franch-Nadal, and D. Mauricio, “[Prediction of outcomes in subjects with type 2 diabetes and diabetic foot ulcers in Catalan primary care centers: a multicenter observational study](#),” *Journal of Foot and Ankle Research* 16(1), 8 (2023).
20. T. Lan, Z. Li, and J. Chen, “[FusionSegNet: Fusing global foot features and local wound features to diagnose diabetic foot](#),” *Computers in Biology and Medicine* 152, 106456 (2023).
21. S. Özgür, S. Mum, H. Benzer, M. K. Toran, and İ. Toygar, “[A machine learning approach to predict foot care self-management in older adults with diabetes](#),” *Diabetology & Metabolic Syndrome* 16(1), 244 (2024).
22. Y. Chen, Y. Zhang, M. Jiang, H. Ma, and Y. Cai, “[HMOX1 as a therapeutic target associated with diabetic foot ulcers based on single-cell analysis and machine learning](#),” *International Wound Journal* 21(3), e14815 (2024).
23. S. Hong, Y. Chen, Y. Lin, X. Xie, G. Chen, H. Xie, and W. Lu, “[Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms](#),” *Technology and Health Care* 32, 265–276 (2024).
24. M. Saeedi, H. T. Gorji, F. Vasefi, and K. Tavakolian, “[Federated versus central machine learning on diabetic foot ulcer images: Comparative simulations](#),” *IEEE Access* 12, 58960–58971 (2024).
25. G. Verma, “[Leveraging smart image processing techniques for early detection of foot ulcers using a deep learning network](#),” *Polish Journal of Radiology* 89, 368–377 (2024).
26. C. Wu, C. Xu, S. Ou, X. Wu, J. Guo, Y. Qi, and S. Cai, “[A novel approach for diabetic foot diagnosis: Deep learning-based detection of lower extremity arterial stenosis](#),” *Diabetes Research and Clinical Practice* 207, 111032 (2024).
27. L. Z. Chee, S. Sivakumar, K. H. Lim, and A. A. Gopalai, “[Gait acceleration-based diabetes detection using hybrid deep learning](#),” *Biomedical Signal Processing and Control* 92, 105998 (2024).
28. R. Castillo-Morquecho, E. Guevara, J. L. Ramirez-GarciaLuna, M. A. Martínez-Jiménez, M. G. Medina-Rangel, and E. S. Kolosovas-Machuca, “[Digital infrared thermography and machine learning for diabetic foot assessment: thermal patterns and classification](#),” *Journal of Diabetes & Metabolic Disorders* 23(2), 1967–1976 (2024).
29. D. K. Agrawal, W. Jongpinit, S. Pojprapai, W. Usaha, P. Wattanapan, P. Tangkanjanavelukul, and T. Vitoonpong, “[Smart insole-based plantar pressure analysis for healthy and diabetic feet classification: statistical vs. machine learning approaches](#),” *Technologies* 12(11), 231 (2024).
30. H. Shi, X. Yuan, X. Yang, R. Huang, W. Fan, and G. Liu, “[A novel diabetic foot ulcer diagnostic model: identification and analysis of genes related to glutamine metabolism and immune infiltration](#),” *BMC Genomics* 25(1), 125 (2024).
31. A. M. El-Kady, M. M. Abbassy, H. Hamdy, A. Farid, A. Moussa, “[Advancing diabetic foot ulcer detection based on resnet and gan integration](#),” *Journal of Theoretical and Applied Information Technology* 102(6), 2258–2268 (2024).
32. S. A. Sulayman, A. R. Qishqish, and G. F. Dayhoom, “[Early detection of diabetic foot using thermal imaging and deep learning techniques](#),” *African Journal of Advanced Pure and Applied Sciences*, 564–567 (2024).
33. A. Sridhar, R. Balasubramanian, J. A. Krishnaswamy, G. Radhakrishnan, and B. Pesala, “[Deep-learning aided multispectral autofluorescence imaging for rapid bioburden detection and gram-type classification on diabetic foot ulcers](#),” *Optics and Biophotonics in Low-Resource Settings* X 14 (2024).

34. M. Cakir, G. Tulum, F. Cuce, K. B. Yilmaz, A. Aralasmak, M. İ. Isik, and H. Canbolat, “[Differential diagnosis of diabetic foot osteomyelitis and charcot neuropathic osteoarthropathy with deep learning methods](#),” *Journal of Imaging Informatics in Medicine* 37(5), 2454–2465 (2024).
35. S. Murugan, N. Mohankumar, “[Smart foot monitoring: A cutting-edge solution for early detection of diabetic complications using CNN model](#),” *International Journal of Advances in Signal and Image Sciences* 10(1), 45–55.
36. D. Bouallal, A. Bougrine, R. Harba, R. Canals, H. Douzi, L. Vilcahuaman, and H. Arbanil, “[STANDUP database of plantar foot thermal and RGB images for early ulcer detection](#),” *Open Research Europe* 2, 77 (2022).
37. D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. de J. Rangel-Magdaleno, and F. J. Renero-Carrillo, “[Plantar thermogram database for the study of diabetic foot complications](#),” *IEEE Access* 7, 161296–161307 (2019).
38. N. Arteaga-Marrero, L. C. Bodson, A. Hernández, E. Villa, and J. Ruiz-Alzola, “[Morphological foot model for temperature pattern analysis proposed for diabetic foot disorders](#),” *Applied Sciences* 11(16), 7396 (2021).
39. E. Villa, N. Arteaga-Marrero, and J. Ruiz-Alzola, “[Performance assessment of low-cost thermal cameras for medical applications](#),” *Sensors* 20(5), 1321 (2020).
40. R. Aluvalu, T. Sharma, U. M. Viswanadhula, A. D. Thirumalraj, M. V. V. P. Kantipudi, and S. Mudrakola, “[Komodo dragon mlpir algorithm-based CNN model for detection of illegal tree cutting in smart IoT forest area](#),” *Recent Advances in Computer Science and Communications* 17(6), 1–12 (2024).
41. V. Snášel, L. Kong, and S. Das, “[From constraints fusion to manifold optimization: A new directional transport manifold metaheuristic algorithm](#),” *Information Fusion* 113, 102596 (2025).