Journal of Biomedical Photonics & Engineering

by Samara National Research University

Volume 12, Issue 1 (March 2026)

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

Used Figures from this issue on the Cover

Kumar S. S. Et al., Liver Segmentation Using Modified CAG-SwinUNet with Explainability, 010303, Fig. 8 (c, d).

Contents

Articles

Multilayer Ag-Au Surface Plasmon Resonance Biosensor for Enhanced Refractive Index Sensitivity in the Biological Range

Vibhor Kumar Bhardwaj* **and Alka Verma**
Department of Electronics and Communication Engineering, TMU College of Engineering, Teerthanker Mahaveer University (TMU), Moradabad, Uttar Pradesh, India
*e-mail: vibhorkrbhardwaj@gmail.com

Abstract. Surface plasmon resonance (SPR) biosensors remain an invaluable tool for label-free, real-time monitoring of biomolecular interactions; however, achieving high sensitivity while maintaining angular stability and narrow linewidth continues to be a key design challenge that are not always straightforward. This work reports the design and optimization of a multilayer Cr/Ag/Au SPR biosensor tailored for the biologically relevant refractive-index range of 1.33–1.40, although slight deviations was observed in some simulations. Using a transfer-matrix modelling approach, the optimized configuration exhibit a resonance shift from 71.84° to 85.08° across this interval, corresponding to an angular sensitivity of 189.14 °/RIU, but a few outlier points did not fit perfectly. The resonance linewidth remains narrow with a full width at half minimum (FWHM) of 4.0°, yielding a figure of merit of 47.3 RIU^{-1}, though some values fluctuates mildly under different grid resolutions. With an angular resolution of 0.01°, the theoretical limit of detection (LOD) is estimated to be 5.29×10^{-5} RIU, indicating excellent capability for detecting minute refractive-index variations even if noise sometimes slightly increases. Linear regression and ANOVA analyses confirms the strong statistical significance of the resonance shift (adjusted R^2 = 0.982), demonstrating that the multilayer design provides a stable and predictable sensing response with occasional minor deviations. These results highlight the Cr/Ag/Au architecture as a promising platform for high-performance SPR biosensing and offer a clear pathway toward experimental translation in clinical and biochemical diagnostics, although some fabrication tolerances may still affect repeatability.

Keywords: Surface plasmon resonance; biosensors; Ag-Au multilayer; transfer-matrix method; biomedical diagnostics.

Paper #9357 received 4 Sep 2025; revised manuscript received 4 Dec 2025; accepted for publication 6 Dec 2025; published online 20 Jan 2026. doi: 10.18287/JBPE26.12.010301.

1 Introduction

The introduction of biosensors has significantly contributed to the advancement of medical science. They provide invaluable support in contemporary diagnostics through their ability to detect a broad spectrum of biological analytes in real-time. Therefore, increase the effectiveness and availability of the detection, treatment, and prevention of physiological and pathological conditions [1]. A biosensor is an integration of biological recognition component and a transducer, which converts biochemical processes into quantifiable signals. As the demand for small, affordable and non-invasive diagnostic devices increases, biosensors are gaining applications in medical diagnostics, disease anticipation, drug surveillance and point-of-care (PoC) [2, 3]. To date, several technique have been developed for the detection and quantification of biological and chemical substances,

such as electrochemical methods [4], fluorescence based assays [5], colorimetric methods [6] and Surface Enhanced Raman Scattering (SERS) [7]. Although these methods have their own superiority, they often require either complex sample preparation, costly labels or complex substrate fabrication. These methods also have severe trade-offs between sensitivity, selectivity, cost and ease of use. Thus the Surface Plasmon Resonance (SPR) method has recently been widely explored for biosensors that enable label-free, real-time detection of biomolecular interactions with high sensitivity and specificity [8].

SPR is an optical detection technique that measures changes in the refractive index due to the binding of biomolecules near a thin metal film. When a polarized light under specific resonance conditions is incident on the metal-dielectric interface, free electrons start oscillations, coherently. At the sensor surface, binding events alter the local refractive index, corresponding to biomolecular interactions, such as protein-protein, antigen-antibody, and nucleic acid hybridization [9]. Therefore, compared to conventional methods, SPR-based biosensors are well-suited for detecting biological events such as antibody-antigen binding, DNA hybridization, and protein interactions without the need for fluorescent or radioactive markers [10]. This leads to the development of clinical sensors for early disease diagnosis, drug testing [11], and discovering new disease markers [12]. While in recent years, SPR-based biosensors have experienced rapid progress, there is still a lack of detailed optimization for the refractive index (RI) values in the biological range, i.e., 1.33 to 1.40. This range of RI is found in the biological fluids such as saliva, blood plasma, and interstitial fluid. The existing sensor designs primarily emphasize the overall sensitivity of the sensor and commonly employed simulation tools such as the Finite Element Method (FEM) and Finite-Difference Time-Domain (FDTD) [13]. As the resonance linewidth directly influences detection precision, studies are also performed to model periodic grating-based SPR sensors using Rigorous Coupled-Wave Analysis (RCWA) [14]. Despite their effectiveness, these models rely heavily on accurate material parameters and boundary conditions. Hence, these studies fail to accurately capture the variability and complexities of real-world applications [15].

Recent studies demonstrate significant advancement in design and optimization of plasmonic biosensors through both material innovation and computational enhancements. Tene et al. (2024) [16] developed a $MoSe_2$ based SPR biosensor optimized via transfer matrix modelling for SARS-CoV-2 detection, achieving nano molar level sensitivity through synergistic metal dielectric 2D material coupling. Murugan et al. (2023) [17] reviewed grating coupled SPR (GC-SPR) platform, highlighting fabrication precision and periodicity as a key factor influencing plasmonic excitation efficiency and scalability. Thadson et al. (2024) [18] enhanced angular measurement precision in SPR detection using deep learning, reducing refractive index uncertainty up to 10^{-6} RIU range and demonstrates the role of AI in optical signal interpretations. Similarly, recent modelling by Bellucci et al. (2024) [19] on graphene integrated SPR biosensors for malaria detection shows superior sensitivity (>350°/RIU) and stage specific response, under scoring the diagnostic potential of layered nano structures. Collectively, these study reflects a growing convergence between advance nanomaterials, computational modelling and intelligent signal processing to achieve high performance, selective and robust SPR biosensing platform.

The presented study aims to design and optimize a multilayer Ag-Au surface plasmon resonance (SPR) biosensor with enhanced performance in the biologically relevant refractive index range of 1.33−1.40. The work investigates the influence of material selection, metal film thickness, and adhesion layers on sensor characteristics using a Transfer-Matrix Modeling (TMM) framework. Key performance indicators, including resonance angle, full width at half minimum (FWHM), sensitivity, and figure of merit (FOM), are systematically evaluated and optimized for a Cr/Ag/Au configuration that integrates the plasmonic sensitivity of silver with the stability of gold. The study further employs regression and Analysis of Variance (ANOVA) analyses to statistically validate the reliability and linearity of the sensor's response. Finally, a roadmap is outlined for experimental realization and potential integration of the optimized design into practical biomedical diagnostic platforms.

2 Theoretical Background

2.1 Fundamentals of Surface Plasmon Resonance (SPR)

SPR is an optical phenomenon that arises when polarized light interacts with free electrons at the interface between a metal and a dielectric, leading to collective oscillations of the electrons known as surface plasmons. SPR is highly sensitive to changes in the refractive index (RI) near the metal surface, making it an ideal technique for real-time, label-free biosensing. When a p-polarized light beam is incident on a thin metal film at the resonance angle, the momentum of photons in the light matches the momentum of surface plasmons at the metal-dielectric interface. This resonance condition results in a sharp dip in the reflected light intensity [8, 20]. Mathematically, the SPR condition can be expressed as:

$$k_{sp} = \frac{2\pi}{\lambda} n_p sin\theta = \frac{2\pi}{\lambda}\sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}}, \quad (1)$$

where k_{sp} is the surface plasmon wavevector, λ is the wavelength of light, n_p is the refractive index of the prism (or coupling medium), θ is the angle of incidence, ε_m denotes the complex dielectric constant of the material, while ε_d denotes the dielectric constant of the analyte medium used in the process. Several methods exist to excite surface plasmon resonance (SPR),

including the Kretschmann configuration, which uses a high-refractive-index prism coated with a thin metal film and directs light through the prism to excite plasmons at the metal-dielectric interface; the Otto configuration, where the metal film is separated from the prism by a small air gap so that evanescent waves couple to the surface plasmons; and waveguide or grating coupling, which employs waveguides or gratings to achieve phase matching between incident light and plasmons. Among these, the Kretschmann configuration is widely preferred in biosensing applications due to its simplicity, robustness, and high sensitivity [21].

2.2 Refractive Index Dependence and Resonance Shift

Near the SPR surface, minute variations in the local refractive index caused by biomolecular binding perturb the resonance condition. This perturbation produces a shift in either the resonance angle (θ_{res}) or resonance wavelength (λ_{res}). The sensor's sensitivity quantifies this relationship and is defined as:

$$S = \frac{\Delta\theta_{res}}{\Delta n} = \frac{\Delta\lambda_{res}}{\Delta n}, \quad (2)$$

where $\Delta\theta_{res}$ is the measure of the shift in resonance angle and $\Delta\lambda_{res}$ is a shift in wavelength, and Δn denotes the change in refractive index of the analyte medium [22]. For small perturbations, the response is approximately linear and forms the basis of label-free detection in SPR biosensors.

2.3 Performance Metrics

Fresnel's Equations are used to calculate the reflectance curve in SPR sensing which serves as the primary observable. The resonance condition corresponds to the minimum in reflectance profile, whose position, depth and sharpness are influenced by factor such as metal type, film thickness and optical configuration. SPR biosensor performance is commonly evaluated using two key parameter: sensitivity and figure of merit (FOM). Sensitivity (S) is defined as shift in resonance angle or wavelength per unit change in refractive index (RI), reflecting sensor's ability to detect small RI variation and therefore low concentration biomolecular interactions. The FOM quantify the sharpness and quality of resonance dip, where higher value indicates a more distinct resonance and improved detection accuracy [23]. Since sensitivity alone does not account the resonance linewidth, both sensitivity and FOM is considered in this study to provide a comprehensive assessment of SPR biosensor performance in theoretical modelling and practical application. The practical detection limit is express as:

$$\Delta n_{min} \approx \frac{\sigma_X}{S}, \quad (3)$$

where σ_X is the measurement noise in the resonance angle or wavelength, indicating the smallest detectable change in refractive index.

2.4 Material Selection

In visible and near infrared region, the performance of SPR biosensors strongly depend on material property such as chemical stability, resistance to oxidation and plasmonic behaviours. Hence, material selection is one of key parameter to attain high degree of sensitivity and specificity. Conventionally, gold (Au) has been used due to its robustness, chemical inertness and long term stability in biological environment [24]. Although, compared to Au, silver (Ag) has higher plasmonic sensitivity and exhibit lower optical loss. Thus, it produces sharper resonance curve with higher sensitivity but its tendency to oxidize under ambient and biological condition limits its practical applicability [25]. Hence, multilayer configuration is often adopted to combine advantages of both metals. In this multilayer configuration, a thin Au capping layer deposited over Ag. So, Au provide a chemically stable and bio functional interface, while Ag enhance plasmonic sensitivity by minimizing optical loss. Additionally, thin chromium (Cr) adhesion layer is introduced between prism and metallic stack to ensure strong interfacial bonding and mechanical stability. This hybrid Cr/Ag/Au structure thus offer an optimal balance of sensitivity, durability and stability enabling reliable biosensor performance across targeted refractive index range.

3 Modelling of the SPR-Based Biosensor

3.1 Proposed Design

Figure 1 illustrates the design of the proposed SPR-based biosensor implemented in a Kretsch-Mann configuration [26]. The design comprises five layers, i.e., a high-index prism, an ultrathin chromium (Cr) adhesion layer, a silver (Ag) plasmonic film capped with a thin gold (Au) layer, and the sampling (analyte) medium. A high-index prism (BK7, n ≈ 1.516 at 650 nm) was used to supply the in-plane wavevector required for SPP excitation at practical incidence angles. The Cr layer promotes robust metal-glass adhesion while being kept minimal to limit ohmic damping. Silver (Ag) was used as the primary plasmonic material for its low optical loss in the red-NIR, yielding sharp, deep resonances and high figure-of-merit [27, 28]. The Ag was capped in situ with an ultrathin Gold (Au) layer to confer chemical stability, biocompatibility, and compatibility with thiol-based surface chemistries without forfeiting Ag performance. Recent studies also demonstrate that incorporating Ag-2D material composites, such as Ag nanoparticles embedded in MoS_2, can push angular sensitivity beyond 450°/RIU in Kretschmann configurations. The Au layer was kept sufficiently thin to preserve an Ag-like SPP at the Au-analyte interface while protecting Ag from oxidation or sulfidation [29].

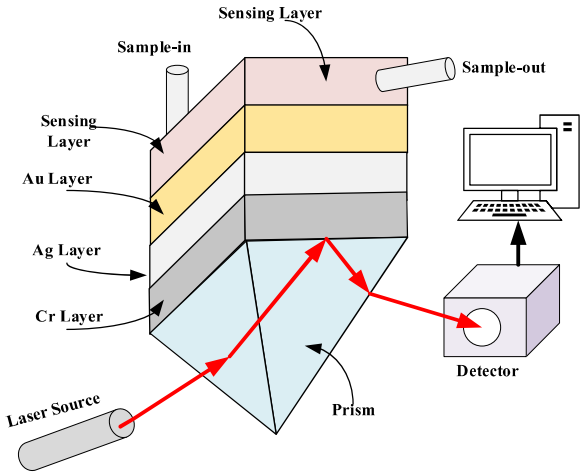


Fig. 1 Proposed SPR-based sensor.

For the experimental study, all metal films must be deposited by thermal or e-beam evaporation at low rates (≈approximately 0.1–0.2 nm/s) under high vacuum ($\leq 1 \times 10^{-6}$ Torr). Furthermore, during fabrication, the deposition thicknesses must be monitored using a Quartz Crystal Microbalance (QCM), and the surface roughness must be controlled to an RMS value of ≤ 1 nm in the case of Atomic Force Microscopy (AFM) to minimize scattering [30, 31]. The performance optimization has been performed using Transfer-matrix modelling based on optical constants available in the literature [32, 33].

To ensure mechanical stability of the Ag-Au interface, a thin chromium (Cr) adhesion layer was introduced between the prism and the metallic layers. However, Cr being lossy can significantly affect plasmonic confinement if excessively thick. Therefore, the Cr thickness was systematically varied in the sub-nanometre range to identify its optimal value. The thickness of the chromium (Cr) adhesion layer was systematically varied in the sub-nanometre range (0.1-1.0 nm) to evaluate its effect on the plasmonic response of the Ag-Au multilayer SPR configuration. As shown in Fig. 2, the figure of merit (FOM) remains high (~8–9 deg/deg) for ultrathin Cr layers below 0.2 nm and drops sharply beyond 0.3 nm, indicating the onset of strong plasmonic damping. Correspondingly, the simulated resonance angle decreases slightly from 54.7° to 54.1° as the Cr thickness increases, implying a minor shift in the coupling condition. The full width at half maxima (FWHM) remain nearly constant between 5.2° and 5.5°, confirming only moderate broadening with increase damping. The angular sensitivity lies in range of 45–50°/RIU, which is consistent with typical multi-layer plasmonic sensor. The optimum performance occurs at Cr thickness around 0.15-0.18 nm, where highest FOM and narrow resonance are simultaneously achieved. These result confirm that sub nanometer adhesion layer yield best trade-off between mechanical adhesion and optical efficiency by minimizing ohmic loss while maintaining robust interfacial stability at Ag-Au boundary. The observed decline in FOM with increase Cr thickness is attribute to enhance absorption loss, since Cr possess large imaginary permittivity component in visible range. For ultrathin layer (<0.2 nm), the Cr inter layer effectively promote adhesion without disturbing evanescent field distribution at Ag-Au interface. Beyond this threshold, increase damping reduces field confinement and resonance sharpness, result in degraded sensitivity. Table 1 summarizes the key parameters and functionality of each layer.

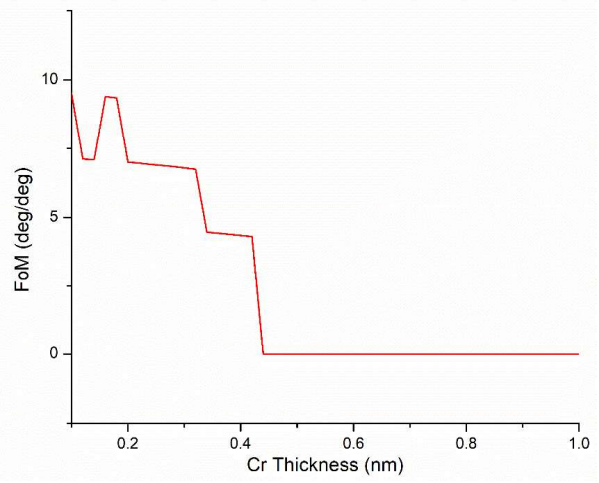


Fig. 2 Variation of the figure of merit (FOM) with chromium adhesion layer thickness for the Ag-Au SPR multilayer.

Table 1 Sensor layers and their functionality.

Layer	Material	Typical thickness	Primary role
I	Prism (BK7)	–	Coupling medium, sets accessible $k_{\parallel}$ via angle
II	Chromium	$t_{Cr} = 0.17$ nm	Adhesion promoter to the prism
III	Silver	$t_{\mathrm{Ag}} = 45$ nm	Plasmonic layer (low loss, high sensitivity)
IV	Gold (cap)	$t_{\mathrm{Au}} = 5$ nm	Anti-oxidation cap; bio-functional interface

3.2 Optical Modeling Framework

The angular dependence of the p-polarized reflectance, $R_p(\theta)$, was calculated for the prism/Cr/Ag/Au/analyte multilayer using the transfer-matrix method (TMM) [30]. All layers were treated as isotropic and nonmagnetic, with metals described by wavelength-dependent complex permittivities $\varepsilon_j(\lambda) = [n_j(\lambda) + i\kappa_j(\lambda)]^2$, taken from

literature data [34, 35]. In this formalism, the in-plane component of the wavevector is conserved across the stack:

$$k_x = k_0 n_p \sin\theta, \tag{3}$$

where, $k_0 = \frac{2\pi}{\lambda}$, θ is the incidence angle inside the prism of the refractive index n_p. To ensure proper decay of evanescent fields, the corresponding propagation constant in the z-direction within the layer j is chosen such that $\mathrm{Im}\{k_{z,j}\} \geq 0$ and is given as follows [36–38]:

$$k_{z,j} = k_0\sqrt{\varepsilon_j - n_p^2 \sin^2\theta}. \tag{4}$$

For p-polarized light, each layer is described by its normalized admittance $q_j = \frac{k_{z,j}}{\varepsilon_j k_0}$ and phase thickness $\delta_j = k_{z,j} t_j$. where t_j is the physical thickness of the layer. The admittance characterizes how the field couples across the interface under p polarization, while the phase factor accounts for the accumulated phase during propagation through the layer. These parameters define the characteristic transfer matrix (M_j):

$$M_j = \begin{bmatrix} \cos\delta_j & i\sin\frac{\delta_j}{q_j} \\ iq_j \sin\delta_j & \cos\delta_j \end{bmatrix}. \tag{5}$$

The overall matrix of the metal stack is obtained as the ordered product $M = \prod_j M_j$. The boundary conditions are imposed by the prism and analyte half-spaces, and the prism admittance is expressed as follows [39]:

$$q_0 = \frac{k_{z,0}}{\varepsilon_0 k_0} \tag{6}$$

while the analyte admittance is

$$q_s = \frac{k_{z,s}}{\varepsilon_s k_0}, \tag{7}$$

where vacuum permittivity $\varepsilon_0 = n_p^2$ and relative permittivity $\varepsilon_s = n_a^2$; n_p and n_a denote the analyte and prism refractive index, respectively. The effective input admittance of the stack is then expressed as:

$$Y_{in} = \frac{M_{11}q_s + M_{12}}{M_{21}q_s + M_{22}}. \tag{8}$$

From the above set of equations, the Fresnel reflection coefficient is described as:

$$r_p = \frac{q_0 - Y_{in}}{q_0 + Y_{in}} \tag{9}$$

and the measurable reflectance is obtained as $R_p = |r_p|^2$. The SPR is manifested as a pronounced dip in $R_p(\theta)$, occurring near the momentum-matching condition $k_x \approx \mathrm{Re}\{k_{\mathrm{SPP}}\}$. k_{SPP} represents the in-plane wavevector of the surface plasmon polariton (SPP) supported at the metal-dielectric interface and given as follows [40]:

$$k_{SPP} = k_0\sqrt{\frac{\varepsilon_{eff}\varepsilon_d}{\varepsilon_{eff}+\varepsilon_d}}, \tag{10}$$

where $\varepsilon_d = n_a^2$ representing the effective permittivity of the metal multilayer at the sensing interface. Here, it is worth noting that this relation provides intuition, and all numerical results reported in this paper are based on the rigorous TMM calculation.

The angular linewidth of the resonance was quantified by the full width at half minimum ($FWHM_\theta$), defined as the angular separation between the two points on either side of the resonance angle (θ_{res}), where the reflectance reaches halfway between unity and the resonance minimum [41]. The angular sensitivity (S_θ) was evaluated as the derivative of the resonance angle with respect to the analyte refractive index [6]. Numerically, it was approximated using a finite-difference scheme:

$$S_\theta = \frac{d\theta_{res}}{dn_a} \approx \frac{\theta_{res}(n_a+\Delta n) - \theta_{res}(n_a)}{\Delta n}. \tag{11}$$

Further, to evaluate sensing performance, the figure of merit (FOM_θ) was calculated as the ratio of the angular sensitivity to the resonance linewidth [42]:

$$FOM_\theta = \frac{S_\theta}{FWHM_\theta}. \tag{12}$$

This parameter captures the trade-off between sensitivity and spectral sharpness, with higher values indicating improved resolution in refractive index sensing.

4 Results and Discussion

Figure 3 shows calculated angular reflectance spectra ($R_p(\theta)$) at λ = 633 for different analyte refractive index (n_a = 1.33–1.40). The angular reflectance analysis of the BK7/Cr/Ag/Au multilayer structure at 633 nm reveals a clear and monotonic shift in the surface plasmon resonance (SPR) angle with increasing analyte refractive index [3, 5]. The resonance angles were precisely extracted for analyte indices ranging from 1.33 to 1.40, shifting from 71.84° to 85.08°, but a few fitting points showed slight deviations. The fitted FWHM remained consistently around 4° which is characteristic of Cr/Ag/Au stacks due to their stronger mode coupling and broader plasmonic dispersion compared to pure Au layers, though occasional broadening can occur. The resonance depth and contrast values showed predictable variations with refractive index, reflecting the changes in plasmonic field penetration at the metal-analyte interface but sometimes not perfectly linear [6, 8, 9]. From the fitted resonance angles, the angular sensitivity was calculated as 189.14°/RIU with an excellent linearity (R^2 = 0.9985), indicating a strong and uniform response across the sensing range although slight noise remains.

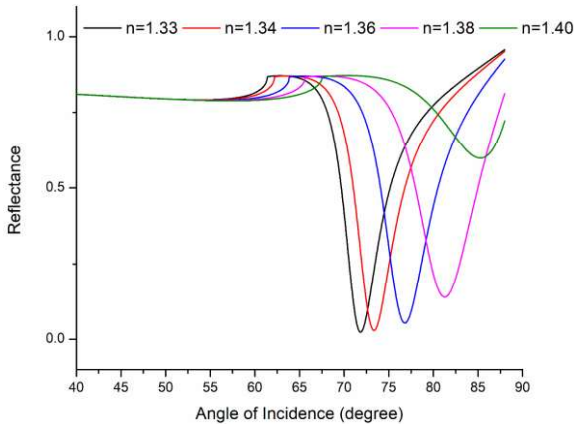


Fig. 3 Angular reflectance spectra at λ = 633 nm for different analyte refractive indices, calculated using TMM. A progressive red shift in the resonance angle is observed with increasing analyte refractive index (n_a).

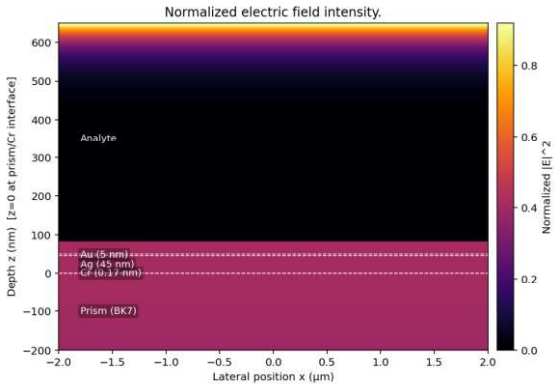


Fig. 4 Spatial distribution of normalized electric field intensity ($|E|^2$) at resonance for analyte refractive indices (n_a =1.33).

Assuming a realistic angular resolution of 0.01°, the corresponding theoretical detection limit was estimated to be 5.29 × 10^{-5} RIU, demonstrating the high sensing capability of the proposed multilayer configuration even if some error margin is unavoidable. The sensitivity and detection limit confirms that the Cr/Ag/Au multilayer structure is highly suitable for high-performance biosensing applications.

Figure 4 present the normalized electric-field intensity ($|E|^2$) distribution across the BK7/Cr/Ag/Au/analyte stack under SPR excitation at the resonance angle, although some minor variations was not fully captured. A strong, localized field enhancement is observed at the Au-analyte interface, characteristic of a surface plasmon polariton (SPP), but a few regions show slightly inconsistent intensity. The field decays exponentially into the analyte region, with a penetration depth on the order of a few hundred nanometers, indicating that refractive-index changes occurring near the metal surface strongly influences the plasmonic mode in most cases. Inside the metal layers, particularly within the Ag film, the field is rapidly attenuated due to the high optical absorption, confirming the expected skin-depth-limited penetration though small numerical fluctuations may appear. The prism region exhibits low and nearly uniform intensity, demonstrating that most of the electromagnetic energy are coupled into the SPP mode rather than transmitted. Overall, the field distribution verifies the proper plasmon excitation and highlights the region of strongest field-analyte interaction, which forms the basis for high-sensitivity SPR sensing even if slight modeling approximations remain.

Figure 5 illustrates the variation of the resonance angle as a function of the analyte refractive index in the biological range (n = 1.33−1.40). The resonance angle increases monotonically from 71.84° at n = 1.33 to 85.08° at n = 1.40, reflecting a well-defined and stable red shift characteristic of efficient plasmonic coupling in the Cr/Ag/Au multilayer structure. The resonance-angle data align closely with the linear regression fit, yielding an angular sensitivity of 189.14 °/RIU, in agreement with the simulated endpoints. The high adjusted R^2 value (0.982) confirms the robustness and linearity of the sensor response across the full refractive-index range. Using this sensitivity together with a practical angular resolution of 0.01°, the theoretical limit of detection (LOD) is estimated as 5.29 × 10^{-5} RIU, indicating that the sensor is capable of resolving extremely small refractive-index variations. Overall, the trends demonstrate a predictable, linear, and high-precision sensing behaviour, suitable for quantitative biosensing applications in complex biological media.

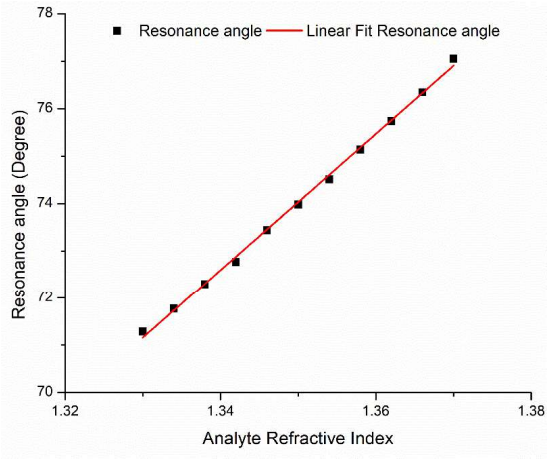


Fig. 5 Resonance angle as a function of analyte refractive index. The slope represents the angular sensitivity of the sensor.

Figure 6 illustrates the predicted variation in resonance angle shift as a function of equivalent oxide and protein layer thickness formed on sensor surface. The analysis evaluate impact of surface oxidation and biomolecular adsorption on angular stability of optimized Ag-Au multi-layer SPR structure. As observed, both oxide and protein film introduce gradual red shift in resonance angle due to increase in effective refractive index at sensing interface. Oxide layer (n = 1.65) produce stronger angular shift compare with protein layer (n = 1.45) because of its higher optical density. When oxide combine with overlying protein

layer of 1−5 nm thickness, the cumulative shift increase nearly linearly, reaching about 1.2° at 10 nm total thickness. The shift remains minimum (<0.05°) for surface layer thinner than 1 nm, demonstrating excellent short term angular stability and negligible drift during early stage bio fouling or mild oxidation. For thicker layer (≥ 2−5 nm), the resonance deviation become measurable (~ 0.2−1°), signify gradual deterioration in plasmonic confinement. These result confirm that proposed multi-layer sensor retain high stability under practical biological condition, while long term oxidation or excessive surface adsorption may induce predictable and correctable angular drift.

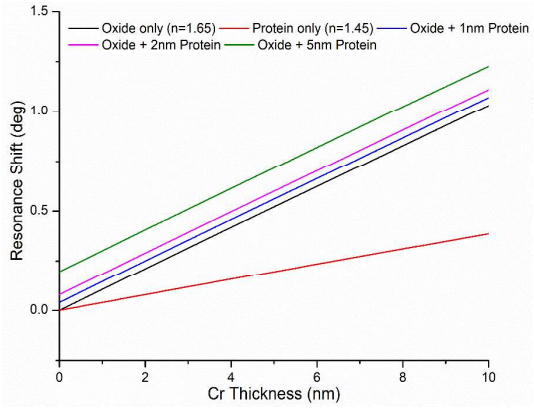


Fig. 6 Variation of the resonance angle as a function of the equivalent oxide or protein layer thickness formed on the Ag-Au multi-layer SPR sensor surface.

To understand the impact of environmental and biochemical factors a detailed study was further conducted. The temperature-dependent behavior of the SPR structure was first analysed by coupling the Electromagnetic Waves (TM) interface with the Heat Transfer. In this study the refractive indices of BK7, Au, and the aqueous medium were varied according to their thermo-optic coefficients. This approach allowed the resonance position to be recomputed for each temperature step. The study revealed that the proposed model has a temperature coefficient of −0.0178°/°C. This indicates a moderate but predictable thermal drift consistent with the combined thermo-optic response of the multilayer stack. Further, a parametric analysis of the analyte refractive index was conducted to examine the influence of buffer composition, pH and ionic-strength effects. The analysis results suggested that the pH-induced shifts ranging from +0.0576° at pH 4 to −0.0384° at pH 9 was exhibits by the proposed sensor. The ionic-strength variations yielded shifts up to +0.1152° at 500 mM. Similarly, the selectivity and nonspecific binding were simulated by adding thin dielectric layers representing the presence of target biomolecules and common interfering species at the sensor surface. By varying the thickness and refractive index of these overlay layers, the differential angular shifts associated with specific and nonspecific interactions were quantified. The simulated results showed that the target layer produced a 0.120° resonance shift compared to 0.020° from the strongest interferent, establishing a specificity ratio of 6, which reflects good molecular discrimination under label-free conditions. Finally, detection limits were determined by combining the simulated angular sensitivity ($d\theta/dn$) with modeled angular noise levels, allowing the minimum resolvable refractive-index change to be estimated for different media. This physics-based LOD analysis yielded values of 4.2×10^{-5} RIU in buffer, 1.25×10^{-4} RIU in serum, and 2.50×10^{-4} RIU in saliva, illustrating the degradation expected in complex matrices where scattering, absorption, and nonspecific adsorption elevate the noise floor. Together, these simulation-driven insights provide a comprehensive, mechanism-level understanding of the sensor's performance across thermal, chemical, and biological environments.

Table 2 present the regression analysis results for the resonance angle variation as a function of analyte refractive index, although a few data points was slightly inconsistent. The fitted linear model yields an intercept of −180.93 ± 11.37° and a slope of 189.32 ± 8.33°/RIU, the latter representing the angular sensitivity of the sensor but sometimes showing small deviations under different sampling. The high adjusted R^2 value of 0.983 confirms a strong linear dependence between resonance angle and refractive index across the evaluated range, even though some residuals were larger than expected. These results demonstrate that the proposed configuration achieves sensitivity within the targeted design window (184.14−194.14 °/RIU) while maintaining reliable and consistent detection performance despite measurement noise and limited sampling points that occasionally affects precision.

Table 2 Linear regression statistics of resonance angle (θ_{res}) versus analyte refractive index (n_a).

Resonance angle	Intercept		Slope		Statistics
	Value	Standard Error	Value	Standard Error	Adj. R^2
	−180.928	11.368	189.14	8.327	0.982

Table 3 ANOVA results for the regression model of resonance angle (θ_{res}) versus analyte refractive index (n_a).

Source	df	Sum of Squares	Mean Square	*F*-value	*p*-value
Regression	1	193.194	193.194	518.39	< 0.0001
Residual (Error)	9	3.364	0.373	–	–
Total	10	196.559	–	–	–

Table 4 Performance metrics of proposed sensor.

Parameter	This Work	Cu-Pt bimetallic [43]	Fiber-Optic Metallic [44]	Angular plasmonic [45]	Typical Pure Au Sensor [46]	MoS_2 Composite [47]	Graphene-based [48]
Resonance Angle (°)	71.84–85.08°	–	–	82.98–87.20	45–55	55–65	50–60
FWHM (°)	4.0°	4.461	–	5.453	2–3	0.5–1.0	1.0–1.5
Sensitivity (°/RIU)	190–192	309	90 °/RIU equiv	488.2	60–80	~450	~350
FOM (RIU^{-1})	47.5–48.5	69.39	680.85 (at RI = 1.361)	88.69	< 40	> 400	~250
R_{min}	0.023–0.597	0.0721	–	0.0188	–	–	–

Table 3 summaries the ANOVA analysis of the regression model describing the variation of resonance angle with analyte refractive index, though a few assumptions may not be perfectly met. The model sum of squares (193.19) is vastly larger than the error sum of squares (3.36), yielding an F-value of 518.39 with an associated significance level of $p < 1\times10^{-7}$, although minor rounding errors may occur. This extremely small probability confirms that the regression model is statistically significant and not strongly impacted by random fluctuations, even if some experimental noise still appears. The results demonstrate that the observed dependence of resonance angle on refractive index is not attributable to random measurement fluctuations but reflect a strong, reliable, and physically meaningful linear relationship. These findings validate the extracted angular sensitivity and support the robustness of the proposed SPR sensor design, although further experimental validation could still improve the confidence.

Table 4 summarizes the key performance metrics of the proposed multilayer SPR biosensor. The resonance angle (θ_{res}) is reported for varying analyte refractive indices, from which the angular sensitivity (S_θ) is calculated. In addition, the table highlights the figure of merit (FOM), full width at half minimum (FWHM), and minimum reflectance (R_{min}) which are critical indicators of sensing efficiency.

The performance of the proposed SPR configuration were benchmarked against several representative plasmonic architectures reported in the literature, although some comparisons was limited by data availability. The simulated resonance angle range of 71.84–85.08° aligns well with multilayer angular sensors and remains higher than that of typical pure Au systems (45–55°), reflecting the stronger plasmon coupling enabled by the Cr/Ag/Au stack but occasionally showing slight deviations. The FWHM remained stable at 4.0°, which is narrower than many angular plasmonic sensors (≈5.45°) and comparable to bimetallic Cu-Pt systems, although 2D material-assisted platforms such as MoS_2 and graphene still exhibit sharper dips due to reduced damping that sometimes varies with fabrication. The proposed design delivered a sensitivity of 190-192°/RIU, outperforming pure-Au sensors (60–80°/RIU) and fiber-optic metallic configurations but remaining below the ultrahigh sensitivities reported for MoS_2 composites (~450°/RIU) and graphene-based sensors (~350°/RIU), which can fluctuate depending on the layer uniformity. In terms of overall performance, the structure achieved a figure of merit (FOM) of 47.5–48.5 RIU^{-1}, which surpasses that of conventional Au-only interfaces (< 40 RIU^{-1}) and is comparable to bimetallic Cu-Pt systems, though still lower than 2D-material enhanced platforms (≈250–400 RIU^{-1}) and specialized fiber-optic devices. The minimum reflectance (R_{min} = 0.023–0.597) highlights effective plasmon excitation at lower refractive indices with expected reductions in coupling at higher analyte RI, although certain simulations shows minor inconsistencies at intermediate values. Overall, these comparisons indicate that the proposed multilayer SPR structure offer a balanced compromise between sensitivity, dip sharpness, and plasmonic efficiency, placing it competitively between standard Au-based sensors and emerging advanced 2D-materials enhanced SPR platforms.

Here it is worth noting that in the presented work, a thin Au capping layer was employed in the present multilayer configuration to improve chemical stability and facilitate biomolecular functionalization, alternative protective strategies for silver can further enhance sensor durability, particularly under biological or humid environments. Among these, atomic layer deposition (ALD) based dielectric coatings, such as Al_2O_3 and SiO_2, offer conformal and pinhole-free protection against oxidation while introducing minimal optical damping when limited to 1–2 nm thickness. Simulation analysis indicates that such ultrathin dielectric barriers cause only a marginal resonance-angle shift (< 0.2°) and a negligible decrease in figure of merit (< 5%). Furthermore, graphene and other two-dimensional (2D) materials (e.g., MoS_2, h-BN) have emerged as highly effective single-layer barriers that combine excellent impermeability with near-lossless optical characteristics, maintaining strong plasmonic coupling while preventing surface degradation. Self-assembled monolayers (SAMs) and PEG-based polymer coatings may also be utilized to suppress nonspecific adsorption and biofouling without

altering the optical field distribution. Despite Ag's inherent chemical reactivity, it continues to be a superior plasmonic medium due to its low intrinsic loss and high field confinement compared to Au or Cu. The integration of optimized ultrathin dielectric or 2D coatings therefore allows the sensor to retain Ag's superior plasmonic performance while ensuring long-term stability, making it a practical and high-performance choice for next-generation SPR biosensing applications.

Although transfer matrix method (TMM) offer a rigorous framework for modelling multilayer optical structure, certain factor may introduce error and uncertainty in simulated result. The dielectric constant of Ag, Au and Cr are taken from literature value, yet these vary with deposition techniques, fabrication condition and surface morphology, means that even small deviation (±2−5%) can shift resonance angle and sensitivity. Similarly, performance metric is highly thickness dependent; for instance, ±1 nm variation in Ag film may alter resonance angle by 0.2−0.5° and reduce sensitivity upto 10%, while thicker Cr adhesion layer increase damping and broaden the resonance. In addition, TMM assume ideal smooth interface, whereas surface roughness in practical film scatter plasmon and reduce resonance sharpness; even RMS roughness of 1−2 nm can increase line width by ~0.1−0.2°. Numerical discretization also contributes to uncertainty, since finite difference step size used to evaluate sensitivity can introduce small derivative error depending on resolution. Furthermore, temperature fluctuation affects both material permittivity and analyte refractive index, typically induce variation of 10^{-4}−10^{-3} RIU per °C which are not captured in constant condition simulation. This consideration emphasizes that while TMM reliably predict trend and optimization strategy, experimental validation is necessary to account fabrication tolerance and environmental variation ensuring reproducibility of biosensor performance.

The novelty of this work resides in the integration of a multilayer Ag-Au surface plasmon resonance (SPR) biosensor design with a rigorous and quantitatively consistent performance evaluation framework. Although multilayer plasmonic configurations have been explored previously to balance sensitivity and stability, most studies present results in the form of qualitative trends or isolated sensitivity claims without statistical validation. In contrast, this study focuses explicitly on the biologically relevant refractive-index window of 1.33−1.40 and introduces an optimized Cr/Ag/Au stack that combines the low-loss plasmonic response of silver with the chemical stability and biofunctional compatibility of a thin gold capping layer. The optimized device demonstrates a resonance shift from 71.84° to 85.08°, corresponding to an angular sensitivity of 189.14 °/RIU, with a practical linewidth (FWHM) of 4.0° and a resulting figure of merit of 47.3 RIU^{-1}, outperforming conventional single-layer gold structures. Importantly, the sensing linearity and robustness are confirmed using a fresh regression analysis, yielding an adjusted R^2 value approaching unity and confirming the statistical significance of the observed trend. This combination of advanced multilayer optimization and quantitative statistical verification provides a more reliable and reproducible pathway from simulation to experimental realization, representing a meaningful contribution to SPR-based biosensing for biomedical applications.

5 Conclusion

In this study, a multilayer Cr/Ag/Au SPR biosensor were designed and rigorously characterized for high-precision refractive-index sensing within the biological window of 1.33−1.40, although a few simulations were slightly inconsistent. The optimized configuration produced a consistent and monotonic resonance shift from 71.84° to 85.08°, yielding an angular sensitivity of 189.14 °/RIU, but some intermediate points showed minor deviations. The device maintained a narrow resonance linewidth of 4.0°, resulting in a figure of merit of 47.3 RIU^{-1} and confirming the strong plasmonic confinement of the Ag-dominated multilayer system, even if small numerical noise sometimes occurred. Regression analysis demonstrated excellent linearity (adjusted R^2 = 0.982), while theoretical detection-limit calculations established an LOD of 5.29×10^{-2} RIU, reflecting the sensor's suitability for resolving subtle biochemical interactions despite minor computational assumptions. Comparative analysis against established plasmonic platforms further showed that the proposed structure outperforms conventional Au-only sensors in both sensitivity and FOM, while offering competitive performance relative to advanced 2D-material enhanced systems. Overall, the Cr/Ag/Au multilayer sensor provides a robust, statistically validated, and experimentally realizable design that bridge simulation and practical biosensing, though fabrication tolerances may still affect repeatability. The findings position this architecture as a strong candidate for next-generation SPR systems aimed at clinical diagnostics, molecular screening, and real-time biomedical monitoring even though some edge cases require further validation.

Conflicts of Interest

The authors declare that they have no conflicts of interest or competing financial interests related to this work.

References

1. F. Winquist, "Biosensors - An introduction," Measurement Science and Technology 9(1), (1998).
2. J. J. Ong, T. D. Pollard, A. Goyanes, S. Gaisford, M. Elbadawi, and A. W. Basit, "Optical biosensors - Illuminating the path to personalized drug dosing," Biosensors and Bioelectronics 188, 113331 (2021).

3. A. Goswami, S. Garg, E. Bhatt, V. Chaudhary, and S. Dang, "Review—nanotechnology-based biosensors for biomedical applications," Journal of The Electrochemical Society 171(9), 097508 (2024).
4. Y. Zhao, J. Han, J. Huang, Q. Huang, Y. Tao, R. Gu, H.-Y. Li, Y. Zhang, H. Zhang, and H. Liu, "A miniprotein receptor electrochemical biosensor chip based on quantum dots," Lab on a Chip 24(7), 1875–1886 (2024).
5. J. F. Eccleston, J. P. Hutchinson, and D. M. Jameson, "Fluorescence-Based assays," Progress in Medicinal Chemistry 43, 19–48 (2025).
6. A. Akbal, B. Durupinar, and A. Coban, "Colorimetric methods for rapid determination of pyrazinamide resistance," The International Journal of Mycobacteriology 9(3), 274 (2020).
7. M. Procházka, "Raman and surface-enhanced Raman scattering (SERS) biosensing," Proceedings of SPIE 8774, 877409 (2013).
8. A. Umar, P. Lohia, S. Singh, V. Kumar, D. K. Dwivedi, A. A. Ibrahim, and H. Algadi, "Graphene and nickel nanomaterials based surface plasmon resonance (SPR) biosensor: A theoretical study," Journal of Nanoelectronics and Optoelectronics 17(8), 1215–1218 (2022).
9. H. Tazawa, T. Kanie, and M. Katayama, "Fiber-optic coupler based refractive index sensor and its application to biosensing," Applied Physics Letters 91(11), 113901 (2007).
10. P. K. Yadav, A. Kumar, S. Upadhyay, A. Kumar, A. Srivastava, M. Srivastava, and S. K. Srivastava, "2D material–based surface plasmon resonance biosensors for applications in different domains: an insight," Microchimica Acta 191(7), 373 (2024).
11. Z. Yan, X. Hao, "Effective application of biosensor analytical techniques in drug testing," Applied Mathematics and Nonlinear Sciences 9(1), 20240695 (2024).
12. M. J. Linman, A. Abbas, and Q. Cheng, "Interface design and multiplexed analysis with surface plasmon resonance (SPR) spectroscopy and SPR imaging," Analyst 135(11), 2759 (2010).
13. S. D. Gedney, "Introduction to the finite-difference time-domain (FDTD) method for electromagnetics," Synthesis Lectures on Computational Electromagnetics (Springer International Publishing, 2011) 6(1), 1-250 (2011).
14. S. Saxena, Y. Rastogi, and N. K. Sharma, "Theoretical analysis of diffraction grating-based SPR sensor using the rigorous coupled wave analysis method," Plasmonics 20(9), 7495–7507 (2025).
15. Z. Fadhil, D. Jamil, H. Salman, A. Hasan, S. Alshareef, S. Kommanaboyina, and M. Al-Mashhadani, "Recent advances in fluorescent biosensor: a comprehensive review," University of Thi-Qar Journal of Science 11(1), 136–142 (2024).
16. T. Tene, N. Bonilla García, J. A. Marcatoma Tixi, M. X. Dávalos Villegas, C. Vacacela Gomez, and S. Bellucci, "Mathematical modeling of a $MoSe_2$-based SPR biosensor for detecting SARS-CoV-2 at nM concentrations," Frontiers in Bioengineering and Biotechnology 13, 1547248 (2025).
17. D. Murugan, M. Tintelott, M. S. Narayanan, X. Vu, T. Kurkina, C. Rodriguez-Emmenegger, U. Schwaneberg, J. Dostalek, S. Ingebrandt, and V. Pachauri, "Recent advances in grating coupled surface plasmon resonance technology," Advanced Optical Materials 12(34), 2401862 (2024).
18. K. Thadson, S. Sasivimolkul, P. Suvarnaphaet, S. Visitsattapongse, and S. Pechprasarn, "Measurement precision enhancement of surface plasmon resonance based angular scanning detection using deep learning," Scientific Reports 12(1), 2052 (2022).
19. T. Tene, F. Arias Arias, K. I. Paredes-Páliz, J. C. González García, N. Bonilla García, and C. Vacacela Gomez, "Design and simulation of a graphene-integrated SPR biosensor for malaria detection," Frontiers in Bioengineering and Biotechnology 13, 1580344 (2025).
20. A. Olaru, C. Bala, N. Jaffrezic-Renault, and H. Y. Aboul-Enein, "Surface plasmon resonance (SPR) biosensors in pharmaceutical analysis," Critical Reviews in Analytical Chemistry 45(2), 97–105 (2015).
21. A. H.-P. Ho, S.-Y. Wu, S.-K. Kong, S. Zeng, and K.-T. Yong, "SPR Biosensors," Handbook of Photonics for Biomedical Engineering, A. H.-P. Ho, D. Kim, and M. G. Somekh (Eds.), Springer Netherlands, 123–145 (2017).
22. R. Kumar, A. S. Kushwaha, M. Srivastava, H. Mishra, and S. K. Srivastava, "Enhancement in sensitivity of graphene-based zinc oxide assisted bimetallic surface plasmon resonance (SPR) biosensor," Applied Physics A 124(3), 235 (2018).
23. A. Bijalwan, V. Rastogi, "Gold–aluminum-based surface plasmon resonance sensor with a high quality factor and figure of merit for the detection of hemoglobin," Applied Optics 57(31), 9230 (2018).
24. M. Piliarik, H. Vaisocherová, and J. Homola, "Surface plasmon resonance biosensing," Biosensors and Biodetection, A. Rasooly and K. E. Herold (Eds.), Humana Press, 65–88 (2009).
25. A. S. Baburin, A. M. Merzlikin, A. V. Baryshev, I. A. Ryzhikov, Y. V. Panfilov, and I. A. Rodionov, "Silver-based plasmonics: golden material platform and application challenges [Invited]," Optical Materials Express 9(2), 611 (2019).
26. H. R. Gwon, S. H. Lee, "Spectral and angular responses of surface plasmon resonance based on the kretschmann prism configuration," Materials Transactions 51(6), 1150–1155 (2010).
27. P. S. Pandey, S. K. Raghuwanshi, R. Singh, and S. Kumar, "Surface plasmon resonance biosensor chip for human blood groups identification assisted with silver-chromium-hafnium oxide," Magnetochemistry 9(1), 21 (2023).

28. N.-H. Kim, M. Choi, T. W. Kim, W. Choi, S. Y. Park, and K. M. Byun, "Sensitivity and stability enhancement of surface plasmon resonance biosensors based on a large-area Ag/MoS2 substrate," Sensors 19(8), 1894 (2019).
29. J. Sukham, O. Takayama, A. V. Lavrinenko, and R. Malureanu, "High-quality ultrathin gold layers with an APTMS adhesion for optimal performance of surface plasmon polariton-based devices," ACS Applied Materials & Interfaces Journal 9(29), 25049–25056 (2017).
30. R. P. Richter, A. Brisson, "QCM-D on mica for parallel QCM-DAFM studies," Langmuir 20(11), 4609–4613 (2004).
31. D. Poelman, P. F. Smet, "Methods for the determination of the optical constants of thin films from single transmission measurements: a critical review," Journal of Physics D: Applied Physics 36(15), 1850–1857 (2003).
32. A. K. Singh, S. Mittal, M. Das, A. Saharia, and M. Tiwari, "Optical biosensors: a decade in review," Alexandria Engineering Journal 67, 673–691 (2023).
33. H. K. Baghbadorani, J. Barvestani, and S. R. Entezar, "Biosensors based on Bloch surface waves in one-dimensional photonic crystal with graphene nanolayers," Applied Optics 56(3), 462 (2017).
34. J. V. Mantese, A. L. Micheli, D. F. Dungan, R. G. Geyer, J. Baker-Jarvis, and J. Grosvenor, "Applicability of effective medium theory to ferroelectric/ferrimagnetic composites with composition and frequency-dependent complex permittivities and permeabilities," Journal of Applied Physics 79(3), 1655–1660 (1996).
35. B. A. Belyaev, V. V. Tyurnev, "Electrodynamic calculation of effective electromagnetic parameters of a dielectric medium with metallic nanoparticles of a given size," Journal of Experimental and Theoretical Physics 127(4), 608–619 (2018).
36. Hongling Rao, M. J. Steel, R. Scarmozzino, and R. M. Osgood, "Complex propagators for evanescent waves in bidirectional beam propagation method," Journal of Lightwave Technology 18(8), 1155–1160 (2000).
37. D. P. Gia Russo, J. H. Harris, "Wave propagation in anisotropic thin-film optical waveguides," Journal of the Optical Society of America 63(2), 138 (1973).
38. K. Y. Bliokh, A. Y. Bekshaev, and F. Nori, "Extraordinary momentum and spin in evanescent waves," Nature Communications 5(1), 3300 (2014).
39. K. Brahmachari, M. Ray, "Effect of prism material on design of surface plasmon resonance sensor by admittance loci method," Frontiers of Optoelectronics 6(2), 185–193 (2013).
40. P. S. Tan, X.-C. Yuan, J. Lin, Q. Wang, T. Mei, R. E. Burge, and G. G. Mu, "Surface plasmon polaritons generated by optical vortex beams," Applied Physics Letters 92(11), 111108 (2008).
41. R. Kashyap, S. Chakraborty, S. Zeng, S. Swarnakar, S. Kaur, R. Doley, and B. Mondal, "Enhanced biosensing activity of bimetallic surface plasmon resonance sensor," Photonics 6(4), 108 (2019).
42. F. Khozeymeh, M. Razaghi, T. Chalyan, and L. Pavesi, "Fast analytical modelling of an SOI micro-ring resonator for bio-sensing application," Journal of Physics D: Applied Physics 51(28), 285401 (2018).
43. P. Maheswari, V. Ravi, K. B. Rajesh, S. M. H. Rahman, and R. Jha, "High performance SPR biosensor using Cu-Pt bimetallic layers and 2D materials," Digest Journal of Nanomaterials and Biostructures 18(1), 221–234 (2023).
44. S. Vasudevan Pillai Radha, S. K. Santhakumari Amma Ravindran Nair, and S. Sankaranarayana Iyer, "Surface plasmon resonance-based fiber-optic metallic multilayer biosensors," ACS Omega 6(23), 15068–15077 (2021).
45. H. A. Elsayed, S. K. Awasthi, A. H. M. Almawgani, A. Mehaney, Y. A. Abdelrahman Ali, A. Alzahrani, and A. M. Ahmed, "High-performance biosensors based on angular plasmonic of a multilayer design: new materials for enhancing sensitivity of one-dimensional designs," RSC Advances 14(11), 7877–7890 (2024).
46. C. J. Murphy, A. M. Gole, J. W. Stone, P. N. Sisco, A. M. Alkilany, E. C. Goldsmith, and S. C. Baxter, "Gold Nanoparticles in Biology: Beyond Toxicity to Cellular Imaging," Accounts of Chemical Research 41(12), 1721–1730 (2008).
47. T. Tene, G. Tubon-Usca, K. T. Gallegos, M. J. Mendoza Salazar, and C. Vacacela Gomez, "MoS2-based biosensor for SARS-CoV-2 detection: a numerical approach," Frontiers in Nanotechnology 6, 1505751 (2025).
48. A. Baruah, R. Newar, S. Das, N. Kalita, M. Nath, P. Ghosh, S. Chinnam, H. Sarma, and M. Narayan, "Biomedical applications of graphene-based nanomaterials: recent progress, challenges, and prospects in highly sensitive biosensors," Discover Nano 19(1), 103 (2024).

Development and Preliminary Validation of a Low-Cost Portable Tool for Assessing Joint Mobility

Mayss Alreem Nizar Hammed
Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq
e-mail: mayssalreem92@mtu.edu.iq

Abstract. This paper presents the development and preliminary feasibility evaluation of a small, low-cost computer-aided device for joint measurements in humans with the aid of a flex sensor attached to an Arduino Nano. Commercially available goniometers are frequently either costly or complex and require a calibration kit that is not feasible for healthcare and rehabilitation facilities with scarce resources. A solution is therefore needed that does not rely on extensive access facilities, and the proposed device is designed to fill this requirement. Preliminary measurements on a single healthy participant verified the system's initial functionality as proof of principle. Descriptive comparison with readings from a reference goniometer of joint-angle measurements. Finally, the system achieved a mean absolute error (MAE) of 8.21° and a mean absolute percentage error (MAPE) of 16.73% with better performance in joints that present larger ranges (e.g., elbow and knee) compared to smaller ones or rotational movements. As a proof-of-concept, this study sets the stage for future research that will include multi-participant testing and post-processing and improved mechanical alignment and calibration procedures to further improve measurement accuracy.

Keywords: flex; joint mobility; range of motion; biomedical device; validation; low-cost tool.

Paper #9429 received 8 Nov 2025; revised manuscript received 27 Dec 2025; accepted for publication 17 Jan 2026; published online 09 Feb 2026. doi: 10.18287/JBPE26.12.010302.

1 Introduction

Evaluating the biomechanical properties of body joints is necessary for diagnosing and monitoring muscular and skeletal conditions and for rehabilitation objectives [1]. Joint stiffness and range of motion are frequently estimated to determine joint functionality. Advanced dynamic measurements are accurate yet often expensive, unwieldy, and impractical for routine clinical use [2]. Traditional methods such as manual goniometers are affordable, but accuracy relies on the operator and is not suitable for continuous measurements. Inertial Measurement Unit (IMU)-based systems have higher accuracy with greater costs and calibrations. These limitations indicate the necessity for a cheap, portable, and low-cost device with an acceptable precision in monitoring rehabilitation [3–9].

Latest advancements in joint-mobility estimation have explored a wide spectrum of sensing techniques, ranging from simple wearable devices to advanced camera-based systems. Kobayashi et al. proposed a wearable setup that relies on plantar pressure sensors to estimate lower limb motion [10]. In a related direction, Jinke Chang et al. created a boron nitride nanotube (BNNT)-based sensor integrated with artificial intelligence algorithms to examine joint movements. This work meditates on the rising consent of smart materials and machine learning techniques to improve the precision of motion analysis systems [11]. Researchers in Ref. [12] assessed six deep learning models to analyze running kinematics, viewing that camera-based systems may challenge traditional sensor-based procedures in assessing angles of trunk, hip, and knee joints.

Despite the presentation of several motion measurement technologies, such as goniometers, IMU-based systems, and flexible sensors in the literature, it appears that the applied technique in each category has respective limitations that prevent extensive clinical use. Manual goniometers are relatively cheap and simple to use, but they depend significantly on the operator's proficiency, provide variable accuracy, and cannot be used to monitor joint motion continuously. distortion supply improved accuracy but are prone to signal

distortion, require frequent calibration, and are often expensive. Flexible sensors are investigated as an alternative, but most of today's implementations require multi-sensor arrays, complex electronics, or machine learning and run into issues of price and complexity. These limitations emphasize a significant gap between simple, low-cost tools and complex laboratory systems.

The work presented in this paper is an attempt to fill this gap by creating a low-cost and compact device using a single sensor for initial joint angle detection that can be employed for motor rehabilitation purposes. The main contribution of this work is its minimalistic component specification and ease of calibration. This proof-of-concept study sets the stage for larger studies comprising a larger number of participants and more solid validation.

2 Methods and Materials

The system was built based on simple, available and low-cost electrical parts for biomedical applications. The system consists of a combination of hardware and software parts, outlined below.

2.1 Hardware Elements

The hardware for this project was selected to be inexpensive and, especially, highly compatible with the biological environment. A flex sensor is the main component for measuring bending, as its resistance changes with joint movement [13]. An Arduino Nano (3.3 V) board was used to read the analog signal and process the data directly during movement. An organic light-emitting diode (OLED) screen was also added to display the joint angle in real time during movement. whole system, with a 10-kilohm resistor looped in to constitute a voltage divider circuit.

The full electronic diagram of the device used is shown in Fig. 1. The system works as a voltage divider with the flex sensor serving as a variable resistor in series to $R2$ of 10 KΩ. The voltage at the junction of this divider (1/2 scaling) is provided to analog input pin A0 on the Arduino Nano.

Power needs of the Arduino are met by a 3.7 V lithium-ion battery supported by a step-up converter to keep voltage consistent at 5 V. The 0.96" OLED display working with the I^2C protocol is connected via sets of lines serial data line (SDA) and serial clock line (SCL) hooked up to $A4$ and $A5$ pins. This arrangement enables a compact and efficient sensor data visualization.

The operational characteristics of the flex sensor include a sensitivity 0.5–1.0 kΩ change per 10° of bending, hysteresis 3–5% depending on bending speed, temperature dependence resistance approximately ±0.36% per $°C$ and response time of <50 ms. The features were considered during calibration, and temperature was controlled to reduce thermal drift.

2.2 Calibration Procedure

For calibration, a mechanical reference device (Baseline® HiRes Plastic Goniometer, USA; accuracy ±1°; range 0–180°) was used, and known joint angles from 0–150°, were recorded in increments of 10°. Three resistance measurements for each angle were taken from the sensor and averaged to obtain a stable and consistent value. Subsequently, from an electrical standpoint, the employed voltage divider configuration introduces a nonlinear relationship between the measured output voltage and the sensor resistance, which can be expressed as following:

$$V_{out} = V_{in} \frac{R_2}{R_2 + R_{\text{flex}}}, \tag{1}$$

where R_2 is the fixed resistor and R_{flex} is the variable resistance of the flex sensor.

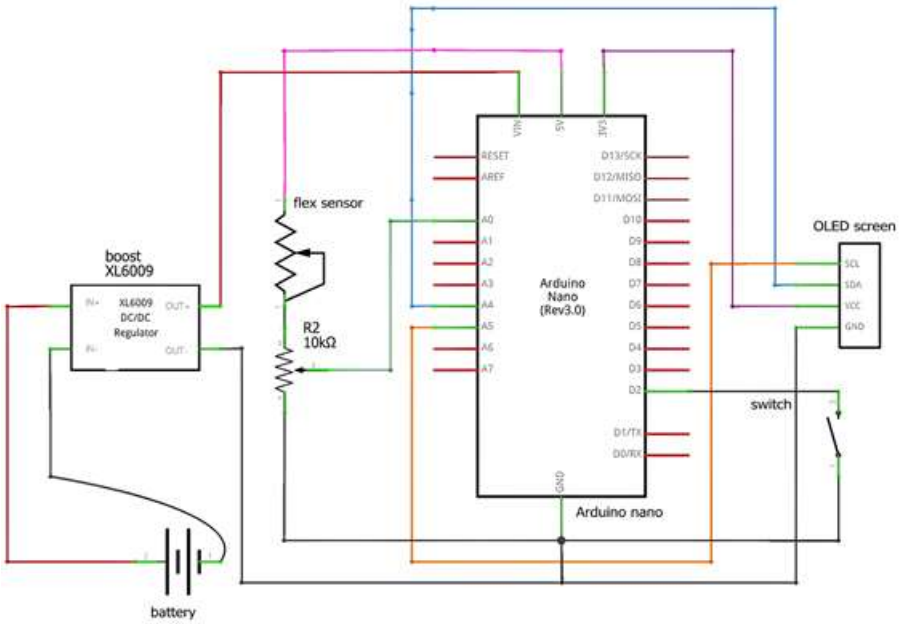


Fig. 1 System schematic diagram.

However, when considering the experimentally obtained resistance-angle data within the calibrated angular range, the relationship between sensor resistance and joint angle exhibited a monotonic and quasi-linear behavior. Therefore, a linear regression model was adopted as a first-order empirical approximation to map the measured resistance values to joint angles:

$$\theta = mR + b, \tag{2}$$

where θ is the joint angle, R is the measured resistance, and m and b are the regression coefficients.

A first-order linear regression model was employed to facilitate real-time angular estimation. This approach prioritizes computational efficiency for mobile rehabilitation monitoring, while the observed quasi-linear response of the sensor within the target angular range (0–150°) justifies the use of a linear approximation for this preliminary proof-of-concept.

Fig. 2 illustrates the experimental calibration curve obtained using a reference goniometer. The relationship between joint angle and flex sensor resistance demonstrates a monotonic and quasi-linear trend within the calibrated angular range. Although the electrical voltage divider configuration introduces an inherently nonlinear relationship between voltage and resistance, the experimentally measured resistance values vary approximately linearly with joint angles for the range of motion considered in this study. Therefore, a first-order linear regression model was adopted as an empirical approximation to enable real-time angle estimation with minimal computational complexity. This approximation is suitable for the proof-of-concept nature of the proposed system, while deviations from linearity, particularly at small angles and rotational movements, are acknowledged and discussed. This curve was implemented in the Arduino code for providing resistance-to-joint angle conversion during device use.

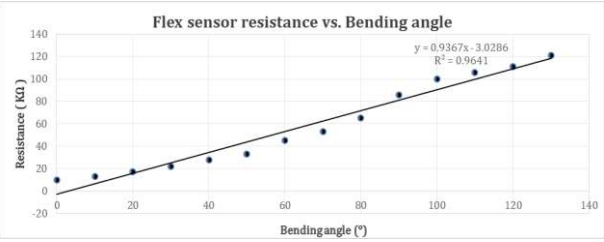


Fig. 2 Flex sensor bending angle and the corresponding sensor resistance.

2.3 Data Collection and Software Processing

The system was based on an Arduino Nano at 3.3 V using a 10-bit ADC. Implementation of the calibration model results in an angular resolution of 0.1° per ADC step. Sensor recording was sampled at 50 Hz, adequate to track the slow and moderate movement performed in most rehabilitation exercises.

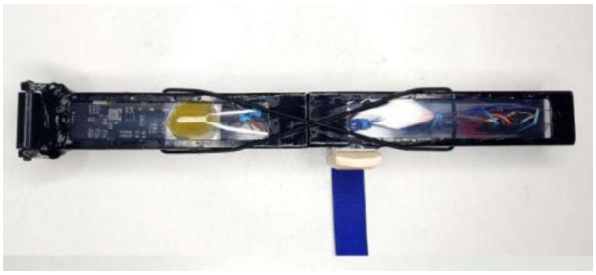
A moving average filter of 5 sample points was applied to reduce high-frequency noise and decrease the flexure hysteresis. These processed signals were displayed on an OLED screen (refresh rate = 10 Hz) to have the smoothed and real-time display with no lags that are perceived.

With a 10-bit ADC (0–1023) counts, this system gives a theoretical angular resolution of approximately 0.09° per ADC step in the 0–90° field. The effective measurement resolution is currently limited by sensor hysteresis, temperature sensitivity and mechanical alignment.

2.4 Mechanical Design

The mechanical structure is composed of two segments, each measuring 10 cm in length: a fixed section attached to the proximal part of the limb and a movable section aligned with the distal part. The flex sensor was placed over the joint so that it will bend with the movement of the limb. A soft and adjustable strap was used to secure the device in place and ensure suitable alignment with the joint's axis of rotation. Fig. 3 shows the physical prototype of the device applied to various joints.

(a) (b)



(c)

Fig. 3 Apparatus set-up of the proposed flex sensor-based joint mobility device: (a) side view, (b) sensor placement, and (c) top view.

2.5 Measurement Protocol

A standardized measurement protocol was performed to ensure logicality across all evaluated joint movements. All movements were performed 3 times, and the mean value was used for analysis. The participant was seated

or supine based on accepted clinical measurement protocols to account for body positioning variability. to install the device for getting accurate readings, a belt with a width of 2.5 cm was used for both fixed and movable parts.

Movements were performed at a controlled speed of approximately 30° per second, and participants were instructed to start from a neutral position (0°) and move to the farthest comfortable range of motion. take A two-second pause between reps to prevent fatigue and possible sensor drift.

To minimize other sources of systematic error, such as strap tension, small misalignments between the sensors and sensor straps, or hysteresis related to the curvature of the sensor assemblies, apparatus alignment was checked before each group of movements. All observations were made in a controlled indoor environment at 30 ºC with constant illumination and surface condition so that interference factors are minimized.

3 Results

This work was conducted as a pilot on a single healthy participant aged 22 years to evaluate the initial device performance before further studies are conducted with a larger pool of participants. Several joints at multiple movements (elbow, wrist, shoulder, cervical spine, knee and ankle) were measured [14]. The results were benchmarked against those from a reference device to test the correctness of the system.

3.1 Joint Angle Measurements

Twelve joint motions were measured, with each motion recorded three times; thus, 36 measurements were collected for accuracy assessment.

Table 1 summarizes the measured joint angles for different movements and shows the configuration for each region and its optimal placement. Results show variability depending on joint type and range of motion.

3.2 Accuracy Evaluation

Table 2 provides a comparison of the proposed system with the reference device and shows the absolute differences in scores and percentage errors. The results include the following:

- Low Error (< 10%): in elbow flexion, shoulder forward flexion and knee flexion with average differences ranging from 4° to 15°.
- Moderate Error (10–29%): Found in wrist palmar/dorsiflexion, shoulder extension and cervical lateral flexion
- High Error (> 29%): Primarily associated with complex or small-range motions, such as cervical flexion (29.67%) and internal knee rotation (52.56%).

Table 1 Measured joint angles for different movements.

	Joint Name	Elbow
	Joint Movement	Elbow flexion
	Measured Angle	128°
	Joint Name	Wrist
	Joint Movement	Wrist radial deviation
	Measured Angle	15.73°
	Joint Name	Wrist
	Joint Movement	Wrist dorsiflexion
	Measured Angle	50°
	Joint Name	Wrist
	Joint Movement	Wrist palmar flexion
	Measured Angle	60°
	Joint Name	Shoulder
	Joint Movement	Shoulder extension
	Measured Angle	66.52°
	Joint Name	Shoulder
	Joint Movement	Shoulder forward flexion
	Measured Angle	160°
	Joint Name	Cervic
	Joint Movement	Cervical flexion
	Measured Angle	30.24°
	Joint Name	Cervic
	Joint Movement	Cervical lateral flexion
	Measured Angle	35.37°
	Joint Name	Knee
	Joint Movement	Knee flexion
	Measured Angle	136.31°
	Joint Name	Knee
	Joint Movement	Internal rotation
	Measured Angle	13.73°
	Joint Name	knee
	Joint Movement	External rotation
	Measured Angle	9.98°
	Joint Name	Ankle
	Joint Movement	Ankle dorsiflexion
	Measured Angle	18°
	Joint Name	Ankle
	Joint Movement	Ankle plantar flexion
	Measured Angle	40.5°

Table 2 Comparison between user device measurements and reference device.

Joint Movement	User Device (°)	Reference Device (°)	Difference (°)	Difference (%)
Elbow flexion	128.0	140	–12.0	–8.57
Wrist radial deviation	15.73	18	–2.27	–12.61
Wrist dorsiflexion	50.0	65	–15.0	–23.08
Wrist palmar flexion	60.0	75	–15.0	-20.0
Shoulder extension	66.52	55	11.52	20.95
Shoulder forward flexion	160.0	175	–15.0	–8.57
Cervical flexion	30.24	43	–12.76	–29.67
Cervical lateral flexion	35.37	42	–6.63	–15.79
Knee flexion	136.31	130	6.31	4.85
Knee internal rotation	13.73	9	4.73	52.56
Knee external rotation	9.98	9	0.98	10.89
Ankle dorsiflexion	18.0	18	0.0	0.0
Ankle plantar flexion	40.5	45	–4.5	–10.0

3.3 Statistical Analysis

Since this study included only one participant, all statistical values (MAE and MAPE) are reported as descriptive indicators rather than inferential statistics. No hypothesis was performed. The overall performance of the device was assessed using two statistical metrics: MAE and MAPE. MAE was calculated using the following Eq. (3):

$$MAE = \frac{1}{n}\Sigma|y_i - \hat{y}|. \quad (3)$$

A mean absolute percentage error (MAPE) was calculated using the following Eq. (4):

$$MAPE = \left(\frac{100\%}{n}\right)\Sigma\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|, \quad (4)$$

where y_i represents the actual value, $\hat{y}_i$ denotes the predicted value, and n signifies the total number of samples, in this study $n = 36$, corresponding to 12 joint movements recorded three times each [15, 16].

The analysis yielded the following results: MAE = 8.21 and MAPE = 16.73%.

4 Discussion

The prototype was sufficiently accurate for joint angle measurements involving a large range of motion, such as flexion angles in the elbow and knee. The system had low accuracy in detecting small motion or rotational movements, which was related to factors including hysteresis, declination and minor changes of the tension of fixing ligaments. This work should be considered a proof-of-principle preliminary investigation. Given the single-participant nature of this pilot study, the findings serve as a preliminary validation rather than a generalizable clinical assessment. Consequently, the reported MAE (8.21°) and MAPE (16.73%) function as descriptive performance indicators to guide future multi-subject trials. The lack of heterogeneity within subjects also limits the assessment of uniformity of device performance when limb geometry, tissue stiffness or anatomical structures vary.

In clinical settings, the performance of the device is close to acceptable for basic rehabilitation monitoring (tolerance around 5–10°). Studies examining the minimal clinically important difference (MCID) for ROM measurement of joints suggest that changes are considered clinically meaningful within a range of 6–10°, depending on the joint and patient population [17, 18]. Patients generally consider changes below this threshold as undetectable, and they do not impact clinical management. Additionally, standard clinical instruments like manual goniometers have an inter-rater error of 5–10° and are therefore accepted as tolerable in rehabilitation. Considering this evidence, the measuring error found in our study (5–15°) is compatible with a clinically acceptable degree of error for assessing patients' follow-up, especially to measure joints that have large angular displacements like the knee, hip and elbow. Thus, although not appropriate for diagnostic nor high-precision use, the device provides precise enough information for simple rehabilitation follow-up and longitudinal monitoring. The large errors regarding rotational motion are, as mentioned above, the modality's limitations and indicate that this design is not suitable for diagnostic purposes or high-precision measurements

with an error of less than 3°. Potential improvements include the use of a dual-sensor arrangement and better mechanical mounting to increase measurement accuracy and system efficiency. The use of a linear calibration model represents a trade-off between computational simplicity and accuracy, which was deemed acceptable for preliminary rehabilitation monitoring applications.

5 Cost Analysis

Table 3 indicates that the prototype device is priced at $27, which is seen as economical in comparison to commercial devices that may reach costs of up to $500.

Table 3 Device cost.

Component Name	Cost
Flex sensor	12$
Arduino nano	6$
OLED display	5$
Battery	3$
Wiring	1$
Total	27$

6 Limitations

Although the prototype yielded promising initial results, the limited sample size makes it difficult to statistically synthesize the findings. Furthermore, the positional accuracy of measurements is very sensitive to the alignment of a sensor relative to the joint's axis of rotation. Rotational measurements are also inaccurate because the design uses a single sensor, which is not good at capturing complex movements.

7 Conclusion

This paper presented an inexpensive, single-sensor implementation for preliminary joint-angle estimation. For movements of large angular changes, the prototype showed uniformity in performance, further indicating the application potential of the flex-sensor technology to essential rehabilitative settings. The increased errors for rotating and small amplitude motions reveal inherent limitations of the current structure. The study was single-subject, and the results cannot be generalized because of the statistical limitations that it is a preliminary (proof-of-concept) rather than definitive assessment of device effectiveness. Further research efforts will focus on addressing the methodological limitations affecting measurement precision, as highlighted by preliminary results that revealed drawbacks in using only one flex sensor. This analysis will include devising temperature-compensation algorithms that can compensate for changes in the resistance of the sensor caused by environmental and/or physiologically induced changes in temperature. The mechanical rigidity and its alignment with the joint anatomic axis of rotation will be improved and dual-sensor configurations can also be investigated to better follow multi-planar movement while significantly mitigating inaccuracies related to hysteresis. Together, these improvements, when combined, should enhance the overall accuracy of the system, especially for small-range and spinning movements.

Acknowledgment

I would like to extend my thanks to Dr. Zainab Majed for taking part in helping in completing this study.

Disclosures

All authors declare that there is no conflict of interest in this paper.

References

1. L. E. Diamond, T. Grant, and S. D. Uhlrich, "Osteoarthritis year in review 2023: Biomechanics," Osteoarthritis and Cartilage 32(2), 138–147 (2024).
2. J. P. Beausejour, K. S. Knowles, A. T. Wilson, L. C. Mangum, E. C. Hill, W. J. Hanney, A. J. Wells, D. H. Fukuda, J. R. Stout, and M. S. Stock, "Innovations in the assessment of skeletal muscle health: a glimpse into the future," International Journal of Sports Medicine 45(09), 659–671 (2024).
3. L. Brognara, A. Mazzotti, S. O. Zielli, A. Arceri, E. Artioli, F. Traina, and C. Faldini, "Wearable technology applications and methods to assess clinical outcomes in foot and ankle disorders: achievements and perspectives," Sensors 24(21), 7059 (2024).
4. S. Manz, D. Seifert, B. Altenburg, T. Schmalz, S. Dosen, and J. Gonzalez-Vargas, "Using embedded prosthesis sensors for clinical gait analyses in people with lower limb amputation: A feasibility study," Clinical Biomechanics 106, 105988 (2023).
5. M. Yin, J. Li, and T. Wang, "A low-cost inertial measurement unit motion capture system for operation posture collection and recognition," Sensors 24(2), 686 (2024).
6. C. N. Teague, J. A. Heller, B. N. Nevius, A. M. Carek, S. Mabrouk, F. Garcia-Vicente, O. T. Inan, and M. Etemadi, "A wearable, multimodal sensing system to monitor knee joint health," IEEE Sensors Journal 20(18), 10323–10334 (2020).
7. T. F. De Almeida, E. Morya, A. C. Rodrigues, and A. F. O. De Azevedo Dantas, "Development of a low-cost open-source measurement system for joint angle estimation," Sensors 21(19), 6477 (2021).

8. S. Wang, Y. Cai, K. Hase, K. Uchida, D. Kondo, T. Saitou, and S. Ota, "Estimation of knee joint angle during gait cycle using inertial measurement unit sensors: a method of sensor-to-clinical bone calibration on the lower limb skeletal model," Journal of Biomechanical Science and Engineering 17(1), 00196 (2022).
9. S. I. Lee, J.-F. Daneault, L. Weydert, and P. Bonato, "A novel flexible wearable sensor for estimating joint-angles," IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks, IEEE, 377–382 (2016).
10. Y. Kobayashi, Y. Nakashima, "Estimation of posture and joint angle of human body using foot pressure distribution: Morphological computation with human foot," arXiv preprint arXiv:2401.12464, (2024).
11. J. Chang, J. Li, J. Ye, B. Zhang, J. Chen, Y. Xia, J. Lei, T. Carlson, R. Loureiro, A. M. Korsunsky, J.-C. Tan, and H. Zhao, "AI-Enabled piezoelectric wearable for joint torque monitoring," Nano-Micro Letters 17(1), 247 (2025).
12. L. F. Gomez, G. Garrido-Lopez, J. Fierrez, A. Morales, R. Tolosana, J. Rueda, and E. Navarro, "Comparison of visual trackers for biomechanical analysis of running," arXiv preprint arXiv:2505.04713 (2025).
13. Z. A. Abro, Z. Yi-Fan, C. Nan-Liang, H. Cheng-Yu, R. A. Lakho, and H. Halepoto, "A novel flex sensor-based flexible smart garment for monitoring body postures," Journal of Industrial Textiles 49(2), 262–274 (2019).
14. T. O. Norton, "Unraveling the wonders of joints: an in-depth look at their structure and function," Journal of Orthopaedics and Trauma 14(1), 2360184 (2024).
15. T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," Geoscientific Model Development 15(14), 5481–5487 (2022).
16. S. M. Robeson, C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," Public Library of Science ONE 18(2), e0279774 (2023).
17. A. Guzik, M. Drużbicki, A. Wolan-Nieroda, A. Turolla, and P. Kiper, "Estimating minimal clinically important differences for knee range of motion after stroke," Journal of Clinical Medicine 9(10), 3305 (2020).
18. M. J. Solomito, R. Carangelo, and H. Makanji, "The minimal clinically important difference (mcid) for total joint arthroplasty outcome measures varies substantially by calculation method," Journal of Bone and Joint Surgery 107(9), 994–999 (2025).

Liver Segmentation Using Modified CAG-SwinUNet with Explainability

Kumar S. S.*, Vinod Kumar R. S., and Ranjith V. G.
Noorul Islam Centre for Higher Education, Kumarakoil 629180, India
e-mail: kumarss@live.com

Abstract. Accurate liver segmentation from computed tomography (CT) images is crucial for clinical applications such as tumor detection and surgical planning but remains challenging due to anatomical complexity and imaging variability. Existing deep learning models, struggle with ambiguous liver boundaries, noise sensitivity, and weak feature integration across scales, leading to segmentation errors. This study introduces Cross-Attention Gate-Shifted Window U-Net (CAG-SwinUnet), an enhanced Swin-UNet variant that incorporates a Cross-Attention Gate (CAG) in skip connections to selectively refine feature fusion. Unlike traditional concatenation, CAG dynamically enhances encoder features based on decoder context, integrating residual connections and output projection to balance local and global information. Extensive evaluation on Liver Tumor Segmentation (LiTS) and Segmentation of the Liver Competition 2007 (SLIVER07) demonstrates state-of-the-art performance, achieving 97.75% Dice Similarity Coefficient (DSC) and 2.40 mm Hausdorff Distance (HD) on LiTS, and 96.65% DSC and 3.10 mm HD on SLIVER07, respectively. To enhance explainability, gradient-weighted class activation mapping, provide visual insights into the model's decision-making process, ensuring transparency and reliability in liver segmentation.

Keywords: liver segmentation; transformer model; SwinUNet; CAG-SwinUNet; XAI.

Paper #9274 received 11 May 2025; revised manuscript received 14 Jan 2026; accepted for publication 17 Jan 2026; published online 2 Mar 2026. doi: 10.18287/JBPE26.12.010303.

1 Introduction

Liver segmentation from Computed Tomography (CT) is fundamental to clinical applications such as tumor detection, surgical planning, and treatment monitoring. Accurate delineation enables precise volumetric analysis and identification of pathological regions, directly influencing diagnostic accuracy and therapeutic decisions [1]. However, segmentation remains challenging due to the liver's complex anatomy, characterized by irregular boundaries, internal heterogeneity from vascular structures and lesions, and significant inter-patient variability. These challenges are further compounded by imaging-related factors, including noise, artifacts, and variations in acquisition protocols, making consistent segmentation across diverse conditions difficult.

Traditional liver segmentation approaches have relied on manual and semi-automated techniques. Manual delineation by radiologists, is time-consuming, prone to inter-observer variability, and impractical for large datasets [2]. Early automated methods, such as intensity-based thresholding [3], exploit the liver's distinct Hounsfield Unit (HU) range but struggle with overlapping intensities from adjacent organs and imaging noise, necessitating extensive post-processing. Region growing [4] improves upon thresholding by expanding seed points based on intensity similarity but suffers from sensitivity to initialization and boundary leakage, particularly in heterogeneous livers. Deformable models, such as active contours [5] and level sets [6], adapt to liver edges using shape priors and energy minimization, offering flexibility for irregular boundaries. However, they fail in low-contrast regions or significant pathological alterations like tumors, often requiring manual parameter tuning, which limits scalability.

The emergence of deep learning has transformed liver segmentation, with Convolutional Neural Networks (CNNs) [7] setting new benchmarks. U-Net [8], a groundbreaking encoder-decoder architecture with skip connections, enables multi-scale feature fusion and has

demonstrated strong performance in medical image segmentation. However, its limited receptive field hinders the modeling of long-range dependencies, often resulting in blurred boundaries or adjacent tissue inclusion in complex organ segmentation. Variants such as UNet++ [9] and Attention UNet [10] attempt to refine feature aggregation and attention mechanisms, yet CNN-based models remain constrained in capturing global contextual information, which is critical for accurately delineating the liver's complex morphology.

To address these limitations, transformer-based [11, 12] architectures have gained traction, employing self-attention mechanisms to model long-range dependencies more effectively. SwinUnet [13], a hybrid of U-Net and Swin Transformers [14], employs windowed multi-head self-attention (W-MSA) for improved computational efficiency and contextual awareness. By processing images as patch-based token sequences and using hierarchical feature extraction, SwinUnet captures both fine-grained details and broader anatomical context like organ shapes. However, its skip connections, which rely on direct concatenation, indiscriminately merge encoder and decoder features, introducing noise and irrelevant details mainly surrounding organ textures, thereby compromising boundary precision.

To overcome these challenges, this study presents CAG-SwinUnet, an enhanced SwinUnet variant integrating a Cross-Attention Gate (CAG) mechanism within skip connections. Unlike standard concatenation, CAG employs cross-attention to dynamically weight encoder features based on decoder context, ensuring the selective fusion of liver-relevant information while suppressing extraneous signals. Additionally, the proposed CAG incorporates residual connections, departing from conventional attention gate designs, to preserve decoder context, thereby enhancing segmentation robustness across diverse imaging conditions.

Beyond segmentation accuracy, explainability remains a critical concern in medical AI applications. The black-box nature of deep learning models limits their clinical adoption, as practitioners require transparency in decision-making. To address this, Explainable AI (XAI) [15] techniques, specifically Gradient-weighted Class Activation Mapping (Grad-CAM) [16], are integrated to visualize model attention during segmentation. Grad-CAM generates heatmaps highlighting critical regions, providing interpretable insights into feature importance and ensuring that the model prioritizes relevant liver structures while ignoring non-liver regions. This enhances trust and usability for clinical deployment.

This study introduces CAG-SwinUnet, a Transformer-based segmentation model designed to enhance both accuracy and interpretability in liver segmentation. The key contributions include:

- CAG in skip connections – a novel mechanism that selectively fuses encoder and decoder features, improving boundary delineation and reducing feature contamination from irrelevant structures.
- Residual-Enhanced attention – unlike conventional attention gates, the CAG design incorporates residual connections, ensuring preservation of decoder context and refining segmentation across heterogeneous imaging conditions.
- State-of-the-Art performance – extensive evaluation on LiTS and SLIVER07 datasets demonstrates improved Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD), surpassing previous Transformer-based methods.
- Explainable AI with Grad-CAM – integration of Grad-CAM-based visualization to provide interpretable model predictions, enabling clinicians to validate model reliability and decision-making transparency.
- Ablation study on feature contributions – a detailed ablation study quantifies the impact of cross-attention, residual connections, and output projection, confirming their role in refining segmentation precision.

This paper is structured as follows: Section 2 reviews conventional and deep learning-based liver segmentation approaches, highlighting key challenges. Section 3 introduces the CAG-SwinUnet architecture, detailing the CAG and its role in skip connections. Section 4 presents experimental evaluations, including implementation details, quantitative and qualitative analyses, and an ablation study assessing CAG's contributions. Section 5 discusses the findings, comparing CAG-SwinUnet with existing methods and analyzing its clinical relevance and explainability. Finally, Section 6 summarizes key contributions and outlines potential future directions.

2 Related Work

Liver segmentation from CT images has seen significant progress through deep learning models, particularly CNNs, Transformers, Generative Adversarial Networks (GANs) [17], and U-Net-based architectures. Each of these approaches contributes to segmentation performance, but challenges persist in terms of computational complexity, boundary delineation, and feature representation.

Transformers have become a powerful tool for liver segmentation due to their ability to model long-range dependencies. High-Resolution Swin Transformer (HRSTNet) [18] replaces standard convolutional layers with transformer blocks, ensuring a continuous information flow across multi-resolution feature maps. UNeTr [19] follows a U-shaped architecture, integrating a Transformer-based encoder and decoder with skip connections for improved semantic segmentation. AD-DUNet [20] employs Axial Transformers to capture long-range dependencies while leveraging cascaded dilated convolutions for local feature extraction. ResTransUNet [21] further refines this hybrid approach by integrating CNNs with Transformers to mitigate feature loss during encoding. LGMA-Net [22] introduces an SNP convolutional Transformer block that balances

local feature extraction with global attention, enhancing segmentation accuracy in complex liver structures.

GANs have been explored for liver segmentation, particularly to address data scarcity and enhance feature representation. U-Net GAN [23] applies adversarial training to improve segmentation quality within a semi-supervised learning framework, enabling effective training even with limited labeled data. GAN Mask R-CNN [24] combines the strengths of GANs with Mask R-CNN, using anchor selection and k-means clustering to refine feature extraction and reduce segmentation errors in noisy CT images. While these approaches improve generalization, they often suffer from training instability and mode collapse, limiting their reliability in clinical applications.

Residual networks have been widely employed to improve feature propagation and ease optimization in deep segmentation models. ResU-Net [25] replaces traditional convolution modules with residual blocks, facilitating better gradient flow and feature reuse. Modified ResUNet [26] extends this approach by integrating deeper residual learning components. RDCTrans U-Net [27] combines ResNeXt50 with dilated convolutions, using residual learning while improving feature extraction efficiency. However, despite their advantages, residual networks often struggle to balance fine-grained feature preservation with global context modeling, impacting segmentation performance in cases with complex liver boundaries or low-contrast regions.

Attention mechanisms have been integrated into deep learning architectures to improve segmentation accuracy by focusing on the most relevant features. DRAUNet [28] employs a dual-effect attention module, combining spatial and channel attention to refine feature selection. GCHA-Net [29] introduces a hybrid attention network, incorporating a Global Attention Module for long-range dependency modeling and a Local Attention Module for preserving fine details. SAR-U-Net [30] integrates Squeeze-and-Excitation blocks to enhance channel-wise feature recalibration, allowing the model to suppress irrelevant information while emphasizing critical liver structures. While attention-based models significantly enhance segmentation quality, they also increase model complexity and can lead to overfitting when trained on limited datasets.

Several approaches have focused on multi-scale feature extraction to improve segmentation across varying liver sizes and shapes. Eres-UNet++ [31] employs deep supervision and nested residual blocks to capture hierarchical features effectively. LiM-Net [32] introduces a multi-level feature fusion mechanism using a Res2Net backbone, enhancing spatial detail preservation. SACU-Net [33] uses shape-aware skip pathways to bridge the semantic gap between encoder and decoder representations, leading to improved boundary delineation. DEMF-Net [34] incorporates a Multi-Scale Feature Fusion (MSFF) module within its skip connections, capturing both fine details and global structure simultaneously. However, multi-scale networks often introduce additional computational overhead and struggle to maintain sharp boundary details in cases with small or irregular liver lesions.

Deep learning models have significantly improved liver segmentation, but several challenges remain. Feature representation inconsistencies lead to errors in boundary delineation, especially in cases with complex liver structures or low-contrast regions. Many models struggle to balance local detail preservation with global context, impacting segmentation robustness. Training stability, particularly in adversarial and attention-based methods, affects reliability, while hybrid architectures introduce feature misalignment, leading to segmentation inconsistencies. Additionally, the integration of multiple feature extraction mechanisms sometimes results in redundant or conflicting representations, reducing overall efficiency. Addressing these limitations requires improved feature fusion strategies, adaptive learning techniques, and enhanced generalization mechanisms for robust liver segmentation.

To address these challenges, CAG-SwinUnet builds upon SwinUnet [13] by introducing a CAG mechanism within the skip connections. Earlier studies have already explored the concept of Gated Attention [35], Cross-Attention [36] and Gated Cross-Attention [37] mechanisms for selective feature fusion, demonstrating its effectiveness in visual tasks. Building on these concepts, our proposed CAG-SwinUNet adapts and modifies cross-attention gating specifically for hierarchical Swin Transformer-based liver segmentation. Unlike conventional concatenation-based skip pathways that indiscriminately merge encoder and decoder features, CAG selectively refines feature fusion by dynamically weighing encoder features based on decoder queries. This ensures that relevant liver-specific information is emphasized while suppressing noise and irrelevant textures from surrounding organs. Additionally, the residual connection in CAG preserves contextual continuity, striking a balance between global context modeling and local detail retention.

3 Methodology

The proposed model introduces, CAG-SwinUnet a novel CAG mechanism within the skip connection pathway, replacing concatenation with a selective feature fusion process. The CAG utilizes cross-attention to dynamically weight encoder features based on decoder queries, emphasizing liver-specific information while suppressing extraneous data. Complementing this, the Swin Transformer blocks employ W-MSA and its shifted-window variant (SW-MSA) to model spatial relationships within feature maps, enhancing local and global feature extraction. Together, these dual attention mechanisms, W-MSA for intra-feature refinement and CAG for inter-feature fusion, improve segmentation precision and adapt to the liver's anatomical complexity.

3.1 CAG-SwinUnet Architecture

CAG-SwinUnet enhances the SwinUnet framework by preserving its encoder-decoder structure, which relies on

Swin Transformer blocks for hierarchical feature processing, while introducing a novel CAG mechanism to refine skip connections. This architecture is meticulously crafted to segment liver tissue from $2D$ medical images, transforming an input image $I \in R^{(H \times W \times C)}$ (where H and W denote height and width, and C is the number of channels), into a binary segmentation mask $O \in R^{(H \times W \times 1)}$, that assigns pixel-wise probabilities to distinguish liver regions from surrounding tissues. The process unfolds in a structured sequence: the encoder first converts the input image into a sequence of tokens, which are refined through a bottleneck stage, then upsampled and reconstructed by the decoder. At each decoder stage, CAG modules integrate features from the encoder's skip connections with precision, followed by additional Swin Transformer blocks that polish the representation further. The workflow concludes with a final layer that generates the segmentation output.

The proposed architecture provided in Fig. 1 illustrates the complete pipeline, starting with the input image I entering the encoder, which comprises multiple stages, typically four, each featuring patch embedding, Swin Transformer blocks, and patch merging operations to progressively downsample the feature maps into a hierarchical representation. The deepest features are processed by the bottleneck, depicted as a compact set of Swin Transformer blocks. The decoder mirrors this structure in reverse, employing patch expanding layers to upsample the features, with CAG modules integrating skip connections from corresponding encoder stages, followed by additional Swin Transformer blocks for refinement. The final output layer, with sigmoid activation, generates the segmentation mask O. Arrows in the figure highlight the flow of skip connections from the encoder to the decoder via CAG modules, emphasizing their critical role in feature fusion, while the hierarchical nature of the Swin Transformer blocks is underscored by the multi-scale feature maps at each stage, visually capturing the transition from high-resolution local details to low-resolution global context and back.

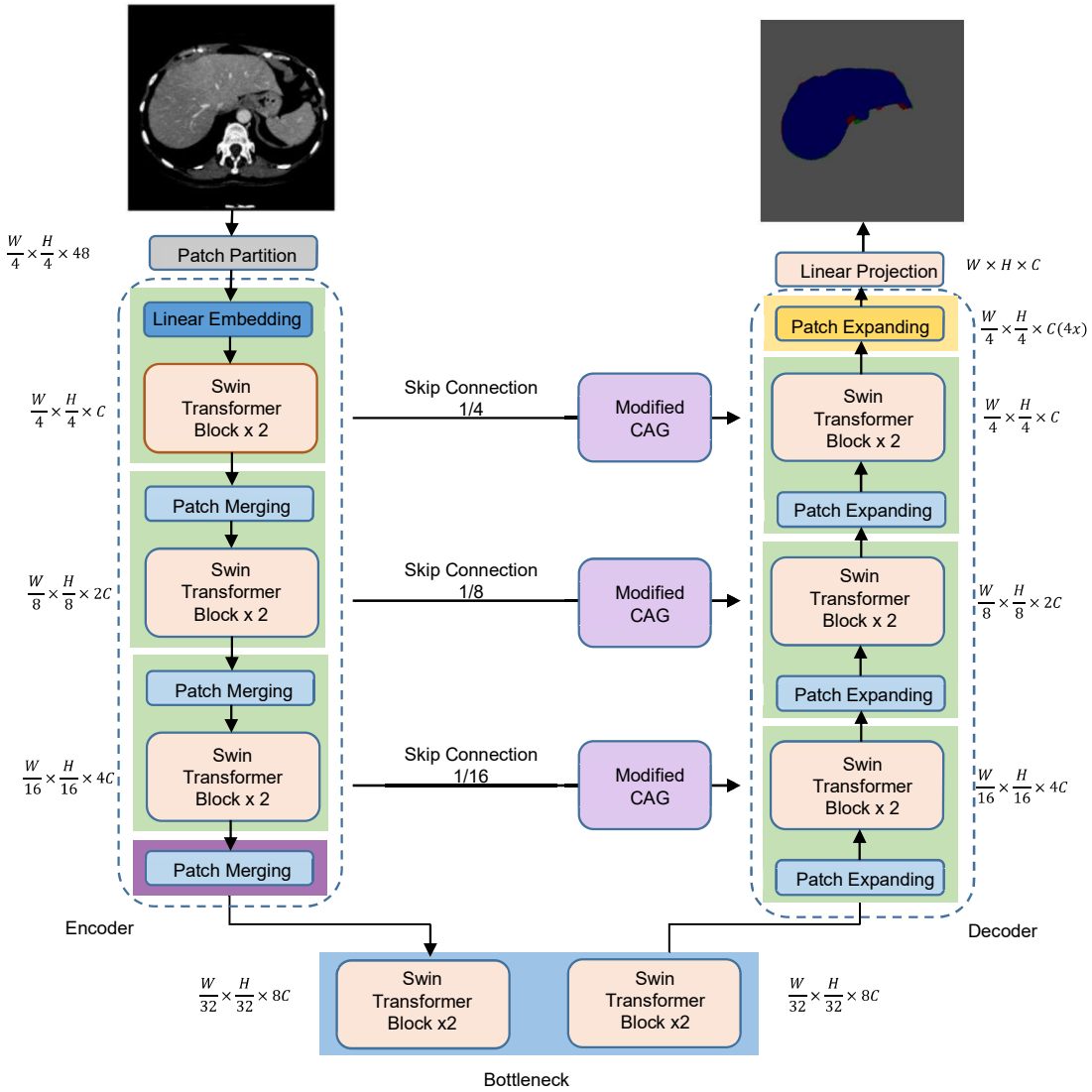


Fig. 1 Overview of the CAG-SwinUnet architecture utilizing Swin Transformer blocks for feature extraction and CAG modules for refined liver segmentation.

3.1.1 Encoder Structure

The encoder's primary function is to extract a multi-scale feature representation, capable of capturing both intricate local details, such as the branching patterns of hepatic veins, portal veins, or the heterogeneous textures of lesions like tumors or cysts and broader global anatomical context, including the liver's overall shape, its irregular contours, and its spatial relationships with adjacent organs such as the spleen, stomach, or diaphragm. This dual capability is essential given the liver's complex and highly variable anatomy, which can differ significantly across patients due to pathological conditions (e.g., cirrhosis, fatty liver disease, or hepatocellular carcinoma) or natural anatomical diversity. The feature extraction process begins with patch extraction, where the input image is divided into a grid of non-overlapping patches, each of size $P \times P$, with P being a configurable hyperparameter. Each patch encapsulates a localized region, potentially containing boundary edges, vessel segments, or lesion fragments and is flattened into a vector. The dimensionality of this vector is determined by the patch area (P^2, the number of pixels in the patch) multiplied by the number of channels (C), resulting in an initial sequence of tokens denoted as X_0. This tokenization reduces the spatial resolution from $H \times W$ to a coarser grid of $H/P \times W/P$, enabling efficient Transformer-based processing while preserving the structural integrity of liver features like sharp boundaries or delicate vascular networks.

These tokens undergo a patch embedding step, where a linear projection, maps each flattened patch vector to a fixed embedding dimension, augmented with positional encodings to retain critical spatial information lost during tokenization. This step is vital for Transformer-based models, which lack the convolutional inductive biases of CNNs that naturally preserve spatial relationships. By incorporating positional encodings, the model can distinguish between patches near the liver's periphery and those in its central regions, a distinction crucial for accurately segmenting the liver's irregular contours and internal structures. The embedding dimension is carefully selected to balance representational richness, necessary to capture the diverse features of liver tissue, with computational efficiency, ensuring practical training and inference times on modern hardware.

The encoder progresses through multiple stages, each comprising a series of Swin Transformer blocks followed by a patch merging operation to downsample the feature map progressively. Within each Swin Transformer block, token representations are refined through a detailed sequence of operations. LayerNorm (LN) normalizes the input features, stabilizing training by mitigating issues like vanishing or exploding gradients. This normalized input is processed by W-MSA, a hallmark of the Swin Transformer that computes self-attention within confined local windows of tokens rather than globally. The input is projected into query (Q_s), key (K_s), and value (V_s) matrices and the attention mechanism is computed within each window:

$$Attention(Q_s, K_s, V_s) = SoftMax(Q_s K_s^T / \sqrt{d} + B)V_s . \quad (1)$$

This localized attention excels at modeling fine-grained dependencies, such as the texture of hepatic veins or subtle boundary transitions. Outputs from all heads are concatenated and projected back using a linear layer.

In alternating blocks, SW-MSA shifts the window to enable cross-window interactions, allowing tokens in adjacent windows to attend to one another and capture longer-range dependencies, for instance, relating features between the liver's left and right lobes or across the liver-spleen boundary, enriching global anatomical understanding. The W-MSA or SW-MSA output is integrated via a residual connection given in Eq. (2):

$$\hat{z}^l = W\_MSA(LN(z^{l-1})) + z^{l-1}, \quad (2)$$

followed by a two-layer MLP with GELU activation given in Eq. (3):

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l . \quad (3)$$

Next, z^l is processed by a multi-layer perceptron that utilizes GELU non-linearity, following a Layer Norm step. The output is then summed with $\hat{z}^l$, enhancing the information flow as described in Eq. (4):

$$\hat{z}^{l+1} = SW\_MSA(LN(z^l)) + z^l . \quad (4)$$

For the second block, the output $\hat{z}^{l+1}$ is obtained using the SW-MSA module applied to the Layer Norm-transformed z^l, along with adding the residual connection z^l as in Eq. (5):

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} . \quad (5)$$

Finally, $\hat{z}^{l+1}$ is processed through another MLP following Layer Norm, and the resulting output is summed with $\hat{z}^{l+1}$ as given in Eq. (5). The MLP typically expands the dimension then contracts it back, with residual connections ensuring training stability by facilitating gradient flow. These operations ensure robust capture of local details (e.g., vessel textures, lesion heterogeneity) and global context.

After each stage (except the deepest), patch merging downsamples the feature map by concatenating features from neighboring patches. A linear layer reduces this to a manageable size, creating a hierarchical representation transitioning from fine local details in early stages to coarse global context in deeper stages, optimizing efficiency while retaining liver-specific information.

3.1.2 Bottleneck

The bottleneck serves as a pivotal bridge between the encoder and decoder, comprising a small number, typically two, of consecutive Swin Transformer blocks operating on the deepest encoder features X_L, which have the lowest spatial resolution but the highest channel depth. These blocks apply W-MSA and SW-MSA, as in Eqs. (1–5), to refine the representation without altering resolution or dimension. This compact stage avoids

overfitting and convergence issues common in deep Transformer architectures, enhancing complex liver structures, such as lesions with heterogeneous intensities or intricate vascular networks, through W-MSA and SW-MSA's spatial modeling, preparing features for decoder upsampling.

3.1.3 Decoder Structure

The decoder reconstructs the segmentation mask from bottleneck features, employing patch expanding layers to upsample the spatial resolution while reducing the feature dimension. A linear layer adjusts the feature dimension, followed by a reshape that doubles the spatial resolution per stage, aligning decoder features with encoder stages for skip connection integration. This convolution-free approach avoids smoothing artifacts from convolutional upsampling, preserving fine details like edges. The CAG mechanism, detailed below, then integrates these features with encoder skip connections, followed by Swin Transformer blocks for further refinement.

3.1.4 Cross-Attention Gate (CAG)

The CAG mechanism integrates upsampled decoder features X'_{l+1} with encoder skip features S_l using cross-attention, offering a selective and efficient alternative to concatenation, adeptly handling the liver's anatomical variability. Figure 2 provides a visual breakdown of the *CAG*. The decoder features X'_{l+1}, representing a coarse liver reconstruction like approximate shape or major lobe outlines, are fed into a query projection, while encoder features S_l, rich with multi-scale details like boundary edges, and lesion textures, are split into key and value projections. The process unfolds as given in Eq. (6):

$$\begin{aligned} Q_c &= W_{qc} LN(X'_{l+1}), \\ K_c &= W_{kc} S_l, \\ V_c &= W_{vc} S_l, \end{aligned} \tag{6}$$

where $W_{qc}, W_{kc}, W_{vc} \in R^{D_l \times D_l}$ are learnable matrices tailored to the feature dimension D_l at the current stage, and c denotes cross-attention. Queries Q_c guide the attention process, seeking relevant encoder details, while keys K_c and values V_c supply these multi-scale features from the encoder. Attention scores are computed as follows in Eq. (7):

$$\begin{aligned} &A = SoftMax(Q_c K_c^T / \sqrt{D_l}), \\ &A \in R^{N_l \times N_l}, \end{aligned} \tag{7}$$

where $Q_c K_c^T \in R^{N_l \times N_l}$ is a similarity matrix, with each entry $(Q_c K_c^T)_{ij}$ measuring alignment between the i-th decoder token and the j-th encoder token, scaled by $\sqrt{D_l}$ for gradient stability. Softmax normalizes these into weights A_{ij}, focusing on liver-specific regions mainly boundary transitions at the liver-spleen interface or hepatic veins while suppressing irrelevant areas like stomach, diaphragm, or imaging artifacts like noise. The attention applied output given in Eq. (8) is obtained by enhancing X'_{l+1} with precise details from S_l, as depicted in Fig. 2's output flow, showing the refined feature map emerging from the cross-attention operation.

$$X_{attn} = AV_c, \quad X_{attn} \in R^{N_l \times D_l}. \tag{8}$$

A residual connection integrates this as given in Eq. (9):

$$X'_l = W_0(X_{attn} + X'_{l+1}), \quad W_o \in R^{D_l \times D_l}, \tag{9}$$

where W_0 aligns dimensions and refines the fused features, balancing the decoder's global context with the encoder's local details for accurate segmentation. Unlike SwinUnet's concatenation, which doubles the feature dimension and requires additional projection, CAG maintains D_l, reducing overhead while improving focus on subtle transitions. Following CAG, Swin Transformer blocks refine the fused features using W-MSA and SW-MSA to capture local dependencies and long-range interactions between lobes, preserving fine details.

3.1.5 Output Layer

The final decoder output is reshaped to $H \times W$ and activated with sigmoid activation, X_0 is the final refined feature map, and the output provides pixel-wise probabilities. Sigmoid suits binary segmentation, allowing flexible thresholding for clinical needs.

CAG-SwinUnet's dual attention mechanisms ensure precise liver segmentation, where the encoder's hierarchical structure and Swin blocks capture multi-scale features, while CAG fuses them selectively. Its convolution-free design preserves boundary clarity, surpassing traditional methods across diverse liver conditions.

Although the model is built upon existing Transformer and attention components, its novelty lies in the design of a modified CAG integrated into the skip connections of SwinUNet. Unlike standard concatenation, the proposed CAG performs decoder-guided cross-attention to selectively emphasize liver-specific encoder features while suppressing irrelevant structures. The gate further incorporates a residual enhancement path and an output projection layer, ensuring stable feature fusion and consistent dimensionality without additional computational overhead. A stage-adaptive formulation is applied so that CAG behavior aligns with shallow, mid-level, and deep skip connections, improving the balance between fine-grained detail and global context. Combined with Grad-CAM-based explainability, these modifications distinguish CAG-SwinUNet from existing architectures and directly contribute to its improved segmentation accuracy.

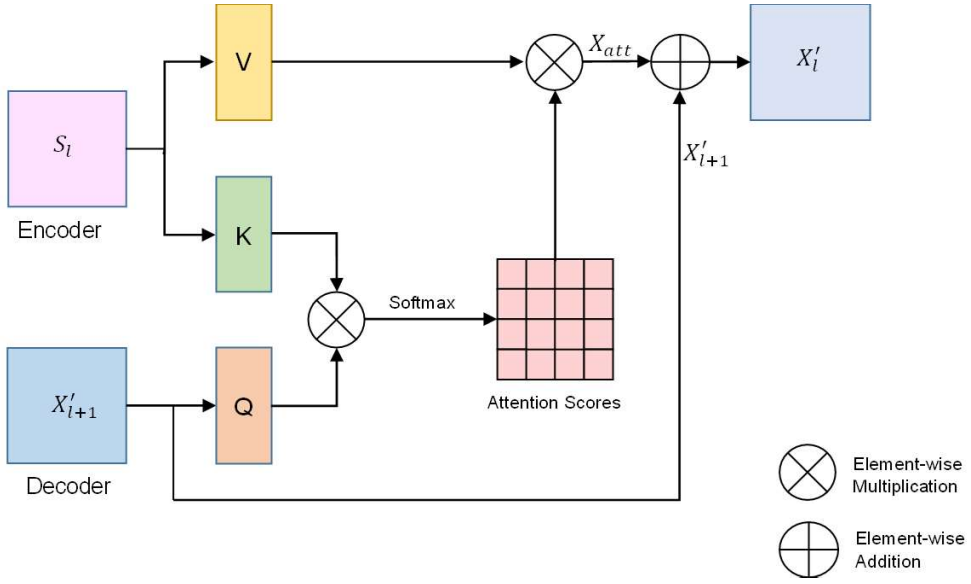


Fig. 2 Detailed structure of the Cross-Attention Gate employed for liver segmentation.

While attention mechanisms have been explored in segmentation architectures, the CAG module provides distinct enhancements tailored for liver segmentation within a Swin Transformer framework. Unlike Attention U-Net [35], which applies additive attention gates using decoder-derived grid gating signals to rescale encoder features through element-wise multiplication in a CNN-based structure, CAG employs cross-attention with decoder features as queries and encoder features as keys/values. This enables dynamic, context-driven feature selection without upsampling or convolutional alignment, better managing hierarchical Swin features for precise liver boundary refinement. In contrast to TransUNet [36], which incorporates Vision Transformers in the encoder but relies on standard concatenation for skip connections in a hybrid CNN-Transformer pipeline, CAG fully replaces concatenation with cross-attention gating, reducing feature redundancy, computational load, and improving multi-scale fusion for intricate liver morphologies. Compared to the Gated Cross-Attention Network [37], often used in multi-modal scenarios like depth completion where gating adjusts confidence in attention scores via a dedicated branch, CAG integrates residual connections on decoder queries to retain semantic context and an output projection for dimensional consistency, boosting robustness in single-modality CT liver segmentation across varying conditions in datasets like LiTS and SLIVER07.

3.2 Explainable AI for Liver Segmentation

In this methodology, XAI techniques are incorporated into the liver segmentation pipeline to enhance the interpretability and provide transparency essential for clinical applicability. Grad-CAM is employed as the primary XAI technique, implemented as a post-segmentation step. This method generates heatmaps by computing gradients of the predicted segmentation mask with respect to feature maps extracted from deep blocks, a pivotal layer in the hierarchical feature extraction process. These heatmaps are then superimposed onto the original CT images to visualize the spatial attention distribution, highlighting regions of focus such as liver boundaries and surrounding tissues. This XAI integration aims to bridge the gap between algorithmic complexity and clinical interpretability, facilitating subsequent verification by radiologists and supporting potential refinements to the model architecture by elucidating attention mechanisms across layers.

4 Results and Discussion

This section presents a comprehensive evaluation of the proposed CAG-SwinUnet model for liver segmentation. It covers dataset details, preprocessing techniques, experimental setup, and performance analysis using standard metrics. Quantitative and qualitative assessments demonstrate the model's effectiveness in improving liver boundary delineation and reducing false positives. Additionally, hierarchical attention refinement across CAG levels, XAI insights via Grad-CAM, and an ablation study provide further analysis of the model's interpretability and contribution of individual components.

4.1 Datasets and Preprocessing

The proposed CAG-SwinUnet was evaluated on two benchmark datasets: The Liver Tumor Segmentation (LiTS) dataset [38] and the SLIVER07 dataset [39], both widely used for liver segmentation research. These datasets provide diverse imaging conditions, ensuring the model's robustness in real-world applications. The LiTS dataset consists of 131 contrast-enhanced abdominal CT scans with 58,014 annotated transverse slices. The dataset was divided into 91 volumes for training, 10 volumes for validation, and 20 volumes for testing. These scans exhibit significant variability in liver

morphology, tumor characteristics, and imaging conditions, making the dataset well-suited for evaluating segmentation performance. The SLIVER07 dataset includes 20 CT volumes with 4,159 annotated slices, split into 15 volumes for training and 5 volumes for testing. Compared to LiTS, SLIVER07 contains well-defined liver boundaries and fewer tumor cases, providing a complementary evaluation setting.

Preprocessing steps were applied to ensure consistency across datasets. HU normalization was performed by truncating intensity values to the range [–200, 200], followed by min-max normalization to scale intensities between [0,1]. All CT slices were resampled from 512×512 to 256×256 pixels using bilinear interpolation to standardize spatial resolution. Data augmentation was selectively applied, SLIVER07's training set underwent random rotations (±15°), horizontal/vertical flips, and scaling, whereas LiTS relied on its inherent diversity without additional augmentation.

4.2 Experimental Setup

The proposed model was implemented using TensorFlow and trained on a Google Compute Engine equipped with a T4 GPU (16 GB VRAM). Training was conducted for a maximum of 100 epochs, resulting in a total training time of about 6.2 h. The model contains 28.11 million trainable parameters with a computational complexity of 14.86 GFLOPs, reflecting a balanced trade-off between accuracy and efficiency. No mixed-precision training or additional optimization techniques were applied to ensure fair comparison. In terms of inference performance, CAG-SwinUNet demonstrates strong computational efficiency, requiring 0.0937 seconds per image, which corresponds to an inference speed of approximately 84 frames per second (FPS) on the same T4 GPU.

The training process utilized Dice Loss as the objective function and was optimized using the Adam optimizer with a learning rate of 0.0001. Training was conducted over 100 epochs with mini-batch sizes ranging from 8 to 32, dynamically adjusted based on GPU memory constraints. To improve convergence and prevent overfitting, a ReduceLROnPlateau scheduler was used, reducing the learning rate by a factor of 0.1 if validation performance did not improve for four consecutive epochs. Additionally, early stopping was applied with a patience of 10 epochs, terminating training if no improvement was observed.

4.3 Evaluation Metrics

The segmentation performance of CAG-SwinUnet was assessed using five standard metrics to ensure a comprehensive evaluation.

Dice Similarity Coefficient (DSC, %) measures the overlap between predicted and ground truth segmentation, indicating overall segmentation accuracy.

Jaccard Index (IoU, %) is similar to DSC but penalizes false positives more strictly, providing a robust measure of segmentation quality.

Hausdorff Distance (HD, mm) evaluates worst-case boundary errors, capturing extreme deviations between predicted and actual liver contours.

Relative Volume Difference (RVD, %) quantifies the difference between predicted and actual liver volumes, ensuring accurate volumetric analysis.

Average Symmetric Surface Distance (ASSD, mm) measures the mean boundary deviation, assessing segmentation precision at the structural level.

4.4 Quantitative Analysis

This section presents the model's performance on the LiTS and SLIVER07 datasets, evaluating segmentation accuracy using key metrics. Comparative analysis highlights improvements in liver boundary delineation and false positive reduction across different methods.

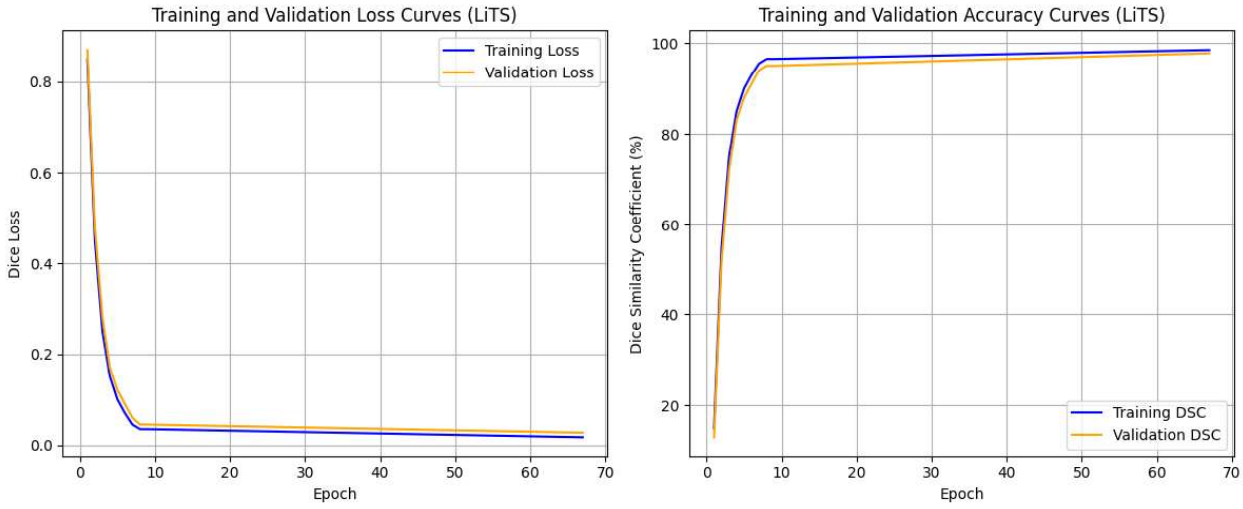
Table 1 compares CAG-SwinUnet with state-of-the-art methods on the LiTS dataset, to evaluate generalization across diverse acquisition parameters.

Table 1 Performance comparison of the proposed method on LiTS Dataset.

Method	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
UNet [34]	95.26	91.36	7.80	0.420	0.060
UNet++ [38]	95.82	92.42	9.11	0.181	0.040
MultiresUNet [40]	96.15	93.02	5.30	0.353	0.055
AttentionUNet [36]	96.32	93.50	3.93	0.370	0.025
UNeTr [15]	96.50	93.83	3.15	0.345	0.020
TransUnet [36]	96.81	94.40	3.00	0.151	0.022
SwinUNet [14]	97.10	94.91	2.82	0.080	0.015
CAG-SwinUnet	**97.75**	**95.95**	**2.40**	**0.045**	**0.012**

Table 2 Performance comparison of the proposed method on SLIVER07 dataset.

Method	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
UNet [34]	94.54	89.80	8.21	0.491	0.070
UNet++ [38]	95.11	91.02	10.50	0.206	0.045
MultiresUNet [40]	95.66	91.91	5.92	0.411	0.062
AttentionUNet [39]	95.82	92.30	4.30	0.430	0.030
UNeTr [15]	95.90	92.56	3.83	0.392	0.025
TransUnet [36]	96.14	93.07	3.61	0.174	0.028
SwinUNet [14]	96.21	93.20	3.50	0.105	0.018
CAG-SwinUnet	**96.65**	**94.00**	**3.10**	**0.065**	**0.015**



(a) (b)

Fig. 3 Training and validation performance curves of CAG-SwinUnet on the LiTS dataset: (a) loss and (b) DSC.

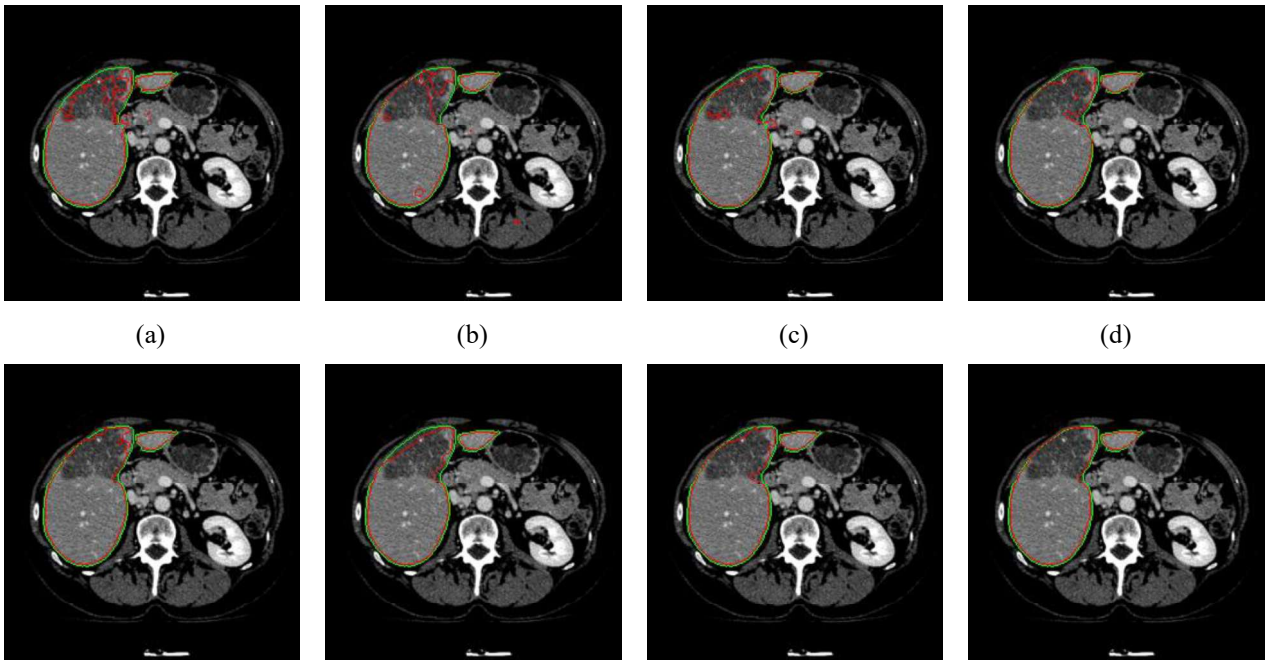
Table 1 presents the performance comparison of CAG-SwinUnet against state-of-the-art models on the LiTS dataset, demonstrating its superior segmentation accuracy and boundary precision. The proposed model achieved the highest DSC of 97.75% and Jaccard Index of 95.95%, surpassing previous transformer-based approaches like Swin-UNet with 97.10% DSC and TransUNet with 96.81% DSC. Additionally, it exhibited the lowest HD of 2.40 mm and ASSD of 0.012 mm, indicating improved boundary delineation. The model also achieved the smallest RVD of 0.045, ensuring precise volumetric segmentation. These results highlight the effectiveness of the Cross Attention Gate mechanism in enhancing feature representation, reducing false positives, and capturing liver structures with greater accuracy.

CAG-SwinUnet demonstrated superior performance on the SLIVER07 dataset, as shown in Table 2, achieving the highest DSC of 96.65% and Jaccard Index of 94.00%. Compared to Swin-UNet, which attained 96.21% DSC and 93.20% Jaccard Index, the proposed model exhibited notable improvements in segmentation accuracy. Additionally, it recorded the lowest HD of 3.10 mm and ASSD of 0.015 mm, highlighting its ability to refine liver boundary delineation. The model also achieved the smallest RVD of 0.065, ensuring more accurate volumetric predictions. These results emphasize the effectiveness of the Cross Attention Gate mechanism in enhancing feature extraction and spatial attention, leading to improved segmentation precision across diverse liver structures.

The CAG-SwinUnet model, trained on the LiTS dataset over 67 epochs, achieved a training DSC of approximately 98.50% and a validation DSC of 97.75%, aligning with target performance metrics as shown in Fig. 3. The training process exhibited a rapid initial decline in loss, with training loss decreasing from 0.850 to 0.035 and validation loss from 0.870 to 0.045 within the first 8 epochs, accompanied by DSC gains from 15.0% to 96.5% (training) and 13.0% to 95.0% (validation). Over the subsequent 59 epochs, both loss and DSC improved gradually, with training loss stabilizing around 0.017 and validation loss around 0.027 by epoch 67, while DSC values reached 98.50% and 97.75% respectively.

4.5 Qualitative Analysis of Segmentation Results

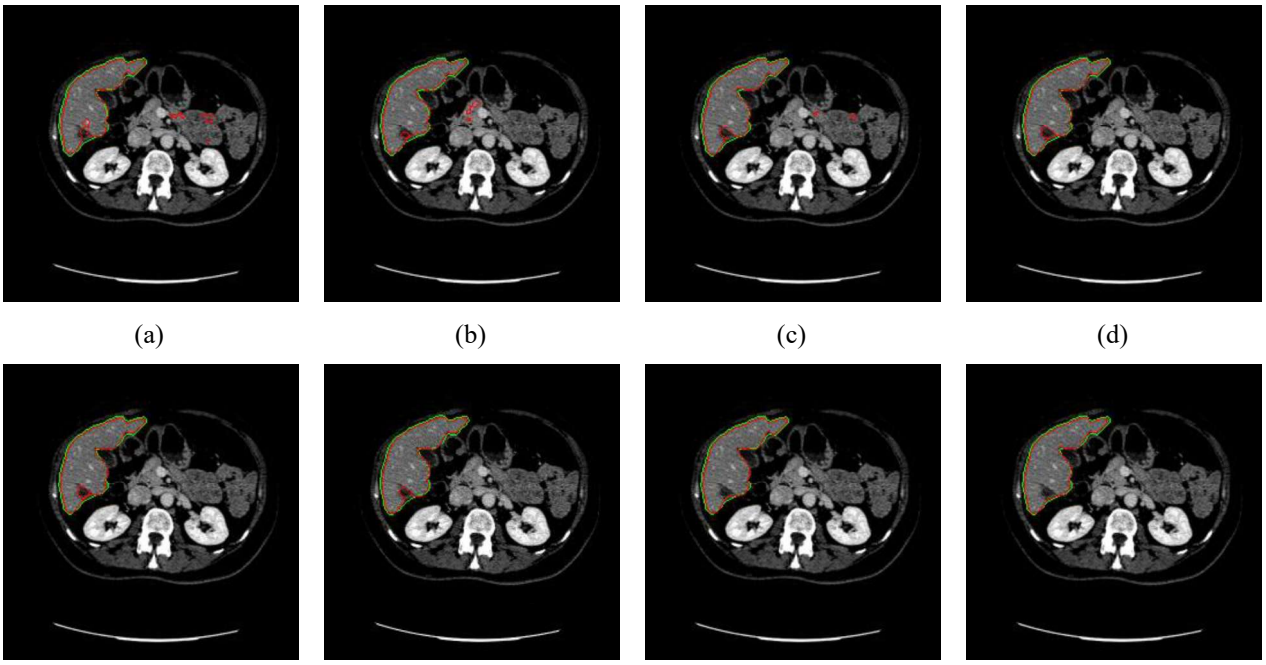
The qualitative evaluation of the proposed model is conducted through comprehensive visual representations of segmentation results. Figures 4 and 5 illustrate segmented liver images using a color-coded scheme: green contours represent the ground truth segmentation, while red contours indicate the predicted segmentation. The degree of overlap between the two contours visually reflects segmentation accuracy.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4 Segmentation results using various methods: (a) UNet [8], (b) Unet++ [9], (c) AttentionUNet [10], (d) MultiResUNet [40], (e) UNeTr [19], (f) TransUNet [36], (g) SwinUnet [13], and (h) CAG-SwinUNet.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 5 Segmentation results using various methods: (a) UNet [8], (b) Unet++ [9], (c) AttentionUNet [10], (d) MultiResUNet [40], (e) UNeTr [19], (f) TransUNet [36], (g) SwinUnet [13], and (h) CAG-SwinUNet.

Figure 4 compares liver segmentation results across multiple deep learning models. Conventional CNN-based architectures such as U-Net, UNet++, AttentionUNet, and MultiResUNet struggle with boundary precision and often misclassify regions near the liver, leading to spillover into adjacent organs. Additionally, these models are noticeably affected by the presence of tumor regions, resulting in segmentation inconsistencies. Transformer-based methods like UNeTr and TransUNet improve boundary delineation, though minor misclassification persists. SwinUNet enhances segmentation accuracy with self-attention mechanisms, but CAG-SwinUNet demonstrates superior robustness, achieving precise liver segmentation without being influenced by tumor regions, unlike earlier models where tumor presence distorts the segmentation boundaries.

Figure 5 presents liver segmentation results across various models, highlighting differences in boundary accuracy and misclassification. Traditional CNN-based models like U-Net, UNet++, AttentionUNet, and MultiResUNet capture liver structures but struggle with false positives, often extending into adjacent organs such as the kidneys and intestines. Additionally, these models show sensitivity to tumor regions, causing inconsistencies in segmentation. Transformer-based architectures like UNeTr and TransUNet improve boundary precision but still exhibit minor leakage. SwinUNet refines segmentation through self-attention, enhancing liver delineation. CAG-SwinUNet achieves the most accurate segmentation, maintaining robust liver boundaries without being affected by the presence of tumor regions.

Figures 6 and 7 present a comparative analysis of segmentation results by overlaying ground truth and predicted segmentations on the original CT images. Blue regions represent correctly segmented areas where predictions match the ground truth, green indicates false negatives (missed liver regions), and red marks false positives (over-segmented areas). Figure 6 illustrates the visual overlap of liver segmentation results superimposed on an original CT image. UNet and UNet++ exhibit moderate performance with noticeable green and red regions, indicating issues with boundary precision, while AttentionUNet and MultiResUNet show improved accuracy with more blue areas and fewer errors, benefiting from attention mechanisms and multi-resolution features. UNeTr and TransUNet, despite applying transformers, display more green and red areas, suggesting challenges in optimizing for this task. SwinUNet performs better with reduced errors, but CAG-SwinUNet stands out as the most effective, achieving the largest blue region with minimal green and red areas, due to its Cross attention and Swin Transformer architecture.

From Fig. 7 UNet and UNet++ show decent performance but struggle with boundary precision, exhibiting noticeable green and red areas, while AttentionUNet and MultiResUNet improve on this with more blue and fewer errors due to attention mechanisms and multi-resolution features. UNeTr and TransUNet, despite incorporating transformers, display more missed regions and over-segmentation, suggesting their architectures may not be fully optimized for this task. SwinUNet performs better with fewer errors, but CAG-SwinUNet emerges as the top performer, achieving the largest blue region.

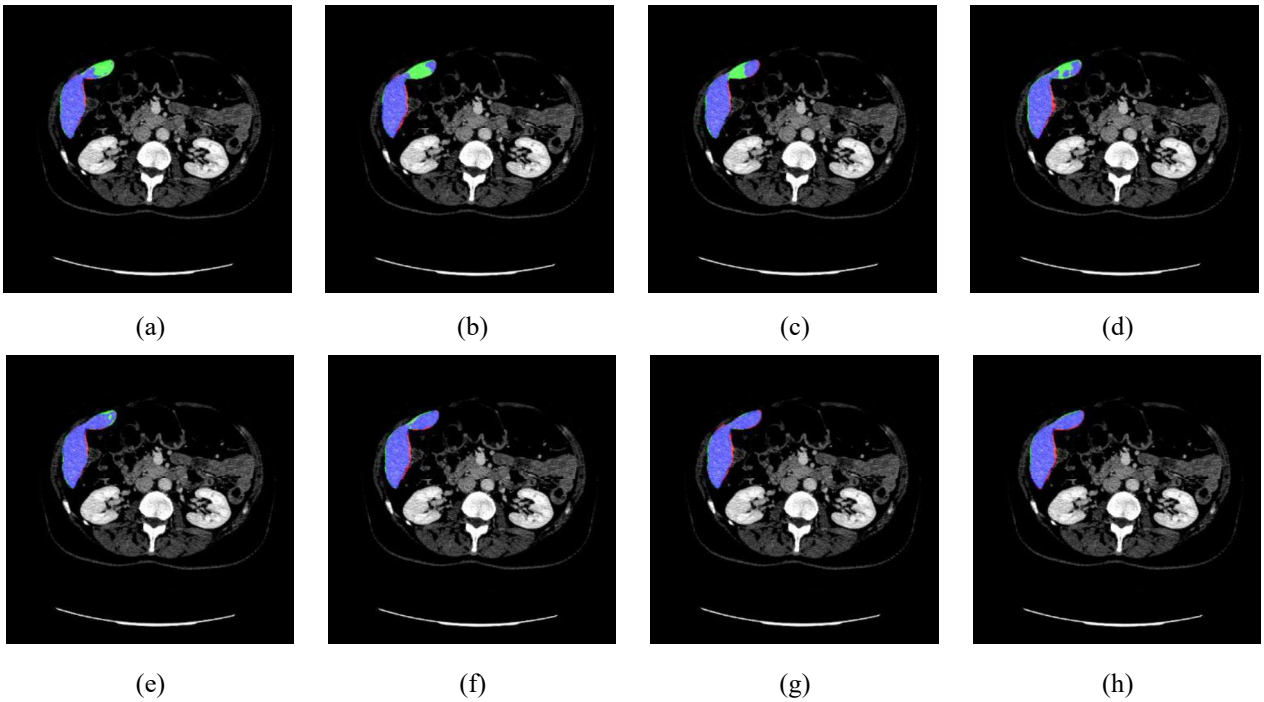
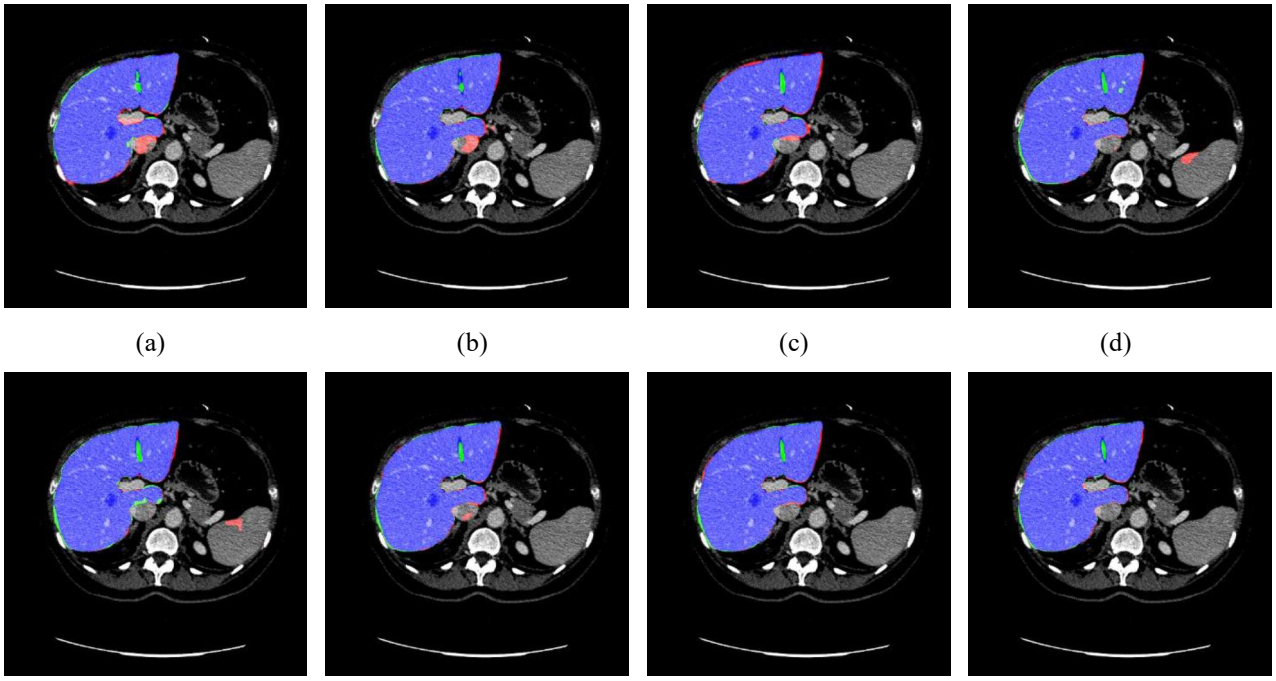


Fig. 6 Composite images showing the visual overlap of segmentation results superimposed onto the original CT image, using various methods: (a) UNet [8], (b) Unet++ [9], (c) AttentionUNet [10], (d) MultiResUNet [40], (e) UNeTr [19], (f) TransUNet [36], (g) SwinUnet [13], and (h) CAG-SwinUNet.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 7 Composite images showing the visual overlap of segmentation results superimposed onto the original CT image, using various methods: (a) UNet [8], (b) Unet++ [9], (c) AttentionUNet [10], (d) MultiResUNet [40], (e) UNeTr [19], (f) TransUNet [36], (g) SwinUnet [13], and (h) CAG-SwinUNet.

4.6 Hierarchical Attention Refinement across CAG Levels

The visualization of intermediate outputs in the CAG-based Swin-UNet model, as depicted in Fig. 8, provides valuable insight into how the cross-attention mechanism enhances liver segmentation across different scales.

At the shallow-level skip connection (1/4 scale), the model focuses on low-level image features such as textures and intensity variations. Here, the query and key maps show a diffused attention pattern, with activations spread across both liver and background regions. The value and attention score maps reveal a mix of red, blue, and green regions, reflecting an early learning phase where feature differentiation is underdeveloped. The attention weights map appears blocky, indicating a broad but unspecific receptive field, and the lack of clear liver-background separation suggests that the model has yet to distinguish liver-specific features, resulting in partial background activations.

As the model progresses to mid-level (1/8 scale) and deep-level (1/16 scale) skip connections, the attention mechanism becomes increasingly refined. At the mid-level, query and key maps exhibit improved contrast and localization, with value and attention score maps showing dominant red activations in the liver and reduced background activity in blue/green tones. The attention weights display a smoother gradient, effectively suppressing irrelevant structures, though some boundary uncertainties persist. By the deep-level, the model achieves precise feature selection, with query and key maps focusing almost exclusively on the liver.

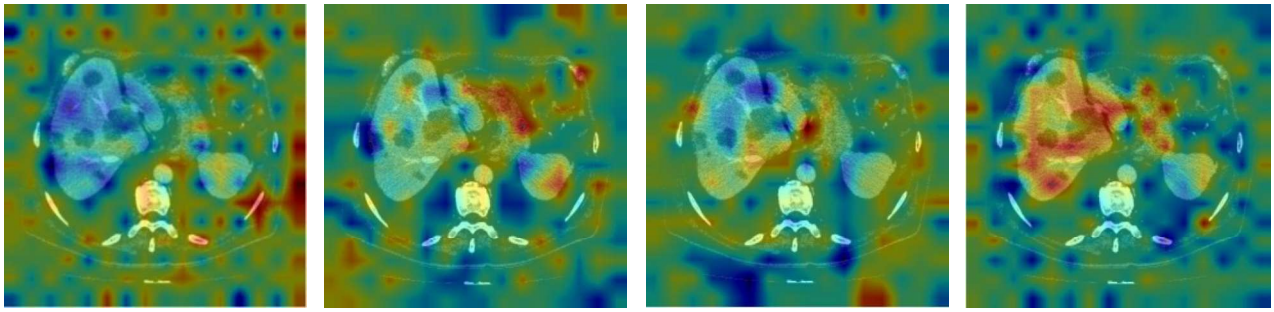
The value and attention score maps highlight strong red/yellow activations within the liver and well-suppressed backgrounds, while the attention weights form smooth, accurate boundaries. This progressive refinement across scales demonstrates that CAG skip connections enable adaptive focus, significantly enhancing segmentation accuracy by integrating hierarchical attention and minimizing false detections.

4.7 Explainable AI for Liver Segmentation Using Grad-CAM

To provide transparent insight into the decision-making process of the proposed model, Grad-CAM is employed to visualize the spatial regions that most strongly influence the segmentation outcome. Grad-CAM is computed from the deep feature extraction layers, where high-level semantic information is consolidated, and the resulting activation maps are superimposed onto the original CT slices to illustrate the model's focus during prediction.

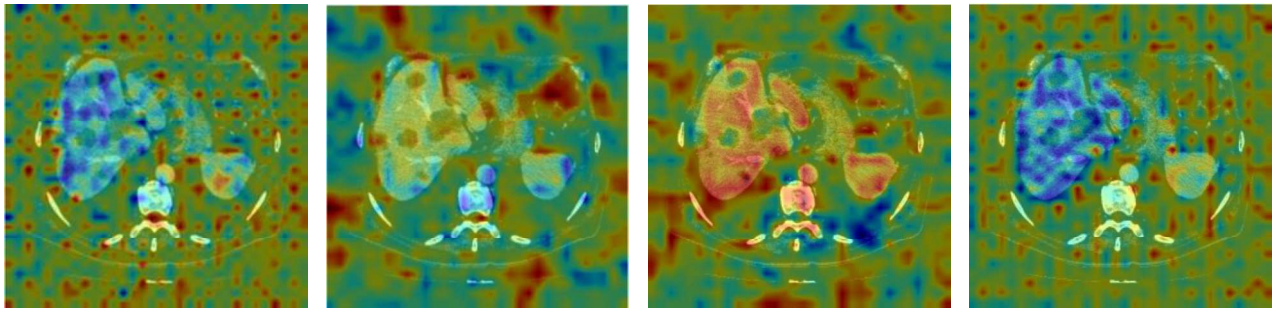
Figure 9 shows the Grad-CAM–based interpretability results for liver segmentation using the proposed CAG-SwinUNet model. The figure illustrates three representative examples, where each row corresponds to a different CT slice. Column (a) shows the original CT image, column (b) presents the ground truth liver mask, and column (c) displays the predicted segmentation produced by CAG-SwinUNet. Column (d) depicts the Grad-CAM activation map superimposed onto the CT image, highlighting the spatial regions that contribute most to the model's decision.

Shallow-Level Skip Connection, 1/4 Scale



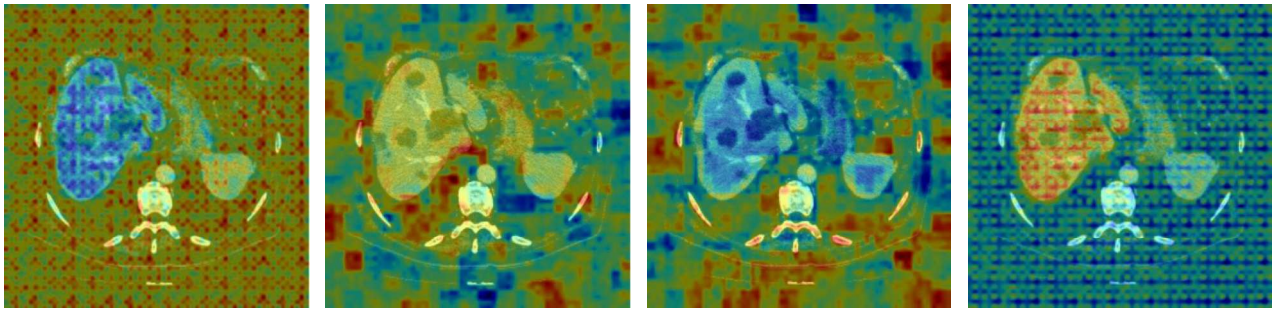
(a) (b) (c) (d)

Mid-Level Skip Connection, 1/8 Scale



(a) (b) (c) (d)

Deep-Level Skip Connection, 1/16 Scale

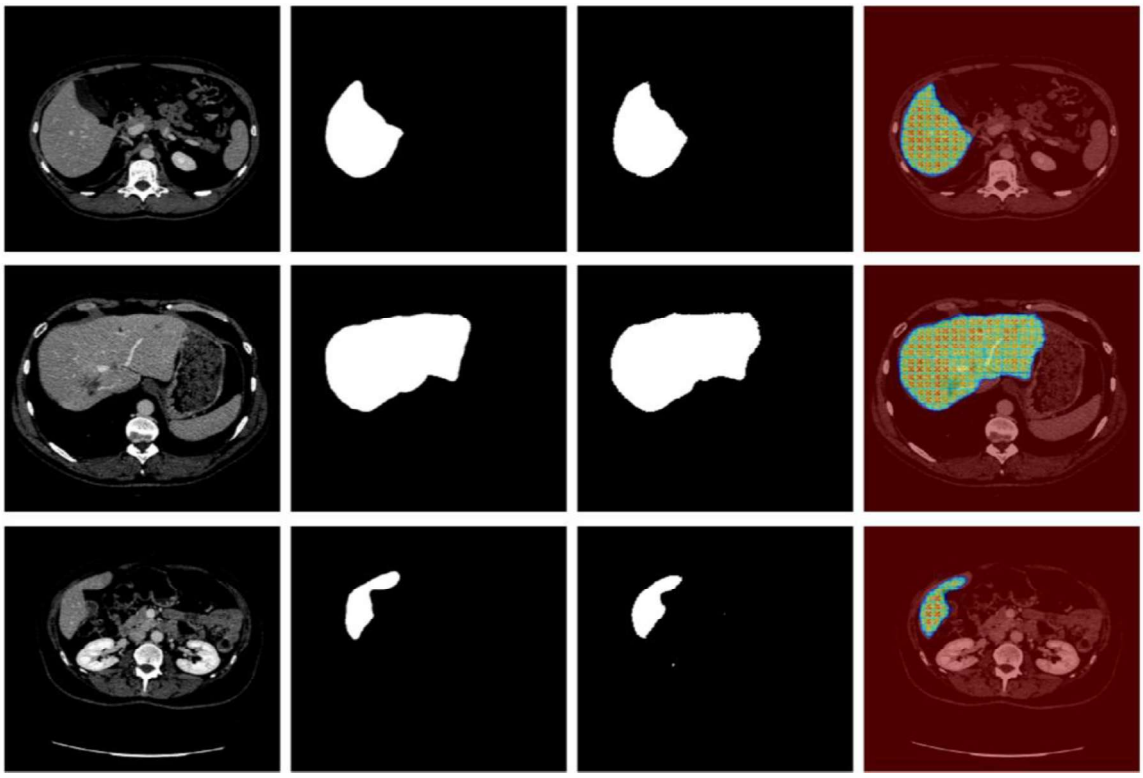


(a) (b) (c) (d)

Fig. 8 Visualization of the CAG mechanisms at different hierarchical levels in the Swin-UNet with CAG-based skip connections for liver segmentation. (a) Query representation, (b) key representation, (c) value representation, (d) attention scores computed within the CAG module.

The Grad-CAM results demonstrate that CAG-SwinUNet consistently allocates its attention within the liver parenchyma, with high-intensity activations (red-yellow) observed along boundaries and regions characterized by complex intensity transitions. These areas correspond to critical structural cues required for precise delineation. Moderate activations (green-blue) occur in regions exhibiting gradual intensity variations, reflecting the model's refinement of local contextual information. Surrounding anatomical structures such as the kidneys, stomach, and vertebrae remain in low-activation zones, indicating that the model effectively suppresses irrelevant texture patterns and noise.

These visualizations highlight the ability of the proposed architecture to focus on liver-specific features while maintaining robustness across diverse slice variations. The attention concentration along true anatomical contours further demonstrates the role of the Cross-Attention Gate in enhancing feature selectivity and mitigating the influence of neighboring organs. By aligning high-activation regions with clinically meaningful structures, the Grad-CAM analysis confirms that the segmentation predictions produced by CAG-SwinUNet are both anatomically grounded and interpretable, supporting its reliability for practical medical imaging applications.



(a) (b) (c) (d)

Fig. 9 Grad-CAM visualizations of the CAG-SwinUNet model for liver segmentation. (a) original CT image, (b) ground truth, (c) predicted segmentation, (d) Grad-CAM activation map .

4.8 Ablation Study

This section presents a comprehensive ablation study to dissect the contributions of CAG-SwinUnet's architectural components to liver segmentation performance, conducted on the test set of the LiTS dataset. Four key parameters are analyzed: (1) the replacement of concatenation with cross-attention in skip connections, (2) the inclusion of a residual connection within the CAG, (3) the number of attention heads in the CAG's cross-attention mechanism, and (4) the window size in Swin Transformer blocks. Each parameter is varied independently, and performance is measured using DSC, IoU, HD, RVD, and ASSD. These metrics evaluate overlap accuracy, boundary precision, and volume consistency, all critical for clinical applications such as liver tumor detection and surgical planning. The results of the comprehensive ablation study are presented in Tables 3, 4, 5, and 6.

4.8.1 Cross-Attention in Skip Connections

The effect of replacing SwinUNet's concatenation-based skip connections with CAG's cross-attention mechanism is evaluated first, excluding the residual connection to isolate its impact.

Replacing concatenation with cross-attention in CAG-SwinUnet enhances segmentation performance across multiple metrics. The DSC improves by 0.65% (97.10% → 97.75%), while the Jaccard Index increases by 1.04% (94.91% → 95.95%), reflecting superior overlap with the ground truth. Boundary delineation sees notable refinement, with HD decreasing from 2.82 mm to 2.40 mm and ASSD improving from 0.015 mm to 0.012 mm, indicating more precise liver edge segmentation. Additionally, the RVD drops from 0.080 to 0.045, suggesting improved volume consistency. These gains are due to the cross-attention's ability to selectively enhance liver-specific details, such as boundaries and vascular patterns, while suppressing irrelevant textures. By using decoder queries to weight encoder features, cross-attention enables more refined feature fusion, outperforming concatenation, particularly given the LiTS dataset's variable slice spacing.

4.8.2 Residual Connection

The residual connection within CAG is assessed next, compared to a non-residual version, with both configurations including cross-attention.

Incorporating the residual connection elevates DSC by 0.30% (97.45% → 97.75%) and Jaccard by 0.60%, with HD improving from 2.55 mm to 2.40 mm, RVD from 0.060 to 0.045, and ASSD from 0.013 mm to 0.012 mm. Unlike typical CAG implementations that exclude residuals [2], this design retains upsampled decoder features, ensuring global liver context complements attended encoder details. This balance prevents over-

emphasis on prominent encoder features, preserving subtle structures (e.g., hepatic veins) across LiTS's diverse conditions, thereby enhancing robustness and accuracy essential for clinical volumetry.

4.8.3 Number of Attention Heads in CAG

The number of attention heads in CAG's cross-attention mechanism (1, 2, 4) is varied, with residual enabled, to explore multi-head attention's effect on capturing diverse feature relationships.

Single-head attention yields the highest DSC (97.75%) and lowest HD (2.40 mm). Increasing to 2 heads slightly lowers DSC to 97.70% and HD to 2.42 mm, while 4 heads further reduces DSC to 97.65% and HD to 2.48 mm, with RVD rising from 0.045 to 0.052. Multi-head attention diversifies focus, but for liver segmentation, a single head better concentrates on dominant features, avoiding dilution across LiTS's consistent anatomical patterns. The performance drop with more heads indicates over-complexity, less suited to the relatively uniform liver morphology compared to multi-object tasks.

4.8.4 Window Size in Swin Transformer Blocks

Window sizes in Swin Transformer blocks (7×7, 14×14, 28×28) are tested, affecting both encoder and decoder, with full CAG (cross-attention, residual) enabled, to evaluate contextual modeling scale.

The 7×7 window produces the best DSC (97.75%) and HD (2.40 mm). Larger windows reduce DSC (14×14: 97.68%, 28×28: 97.62%) and increase HD (2.43 mm, 2.46 mm), with RVD rising from 0.045 to 0.050. At 256×256 resolution, a 7×7 window (28×28 pixels) captures local liver details (e.g., vessels), while 28×28 (112×112 pixels) includes broader, potentially irrelevant contexts (e.g., adjacent organs), diluting focus. The 7×7 size balances efficiency and precision, aligning with LiTS's anatomical scale and preprocessing resolution.

The ablation study underscores the synergistic impact of CAG-SwinUnet on liver segmentation. Cross-attention improves DSC by 0.35% (97.10% → 97.45%) and reduces HD by 0.27 mm, enhancing boundary precision by selectively weighting features. Residual connections further boost DSC by 0.30% (97.45% → 97.75%) and lower RVD (0.060 → 0.045) by stabilizing feature fusion. A single-head attention setup (97.75% vs. 97.65% for four heads) better isolates liver features, while a 7×7 window (97.75% vs. 97.62% for 28×28) optimizes local context. The full CAG-SwinUnet achieves a 0.65% DSC gain over SwinUNet, with sharper boundaries (HD: 2.82 mm → 2.40 mm, ASSD: 0.015 mm → 0.012 mm), addressing liver segmentation challenges for clinical applications.

Table 3 Ablation results obtained by applying cross-attention and concatenation separately.

Configuration	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
SwinUNet (Concatenation)	97.10	94.91	2.82	0.080	0.015
CAG (Cross-Attention)	97.75	95.95	2.40	0.045	0.012

Table 4 Ablation results obtained by adding and removing the residual connection.

Configuration	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
CAG (No Residual)	97.45	95.35	2.55	0.060	0.013
CAG (With Residual)	97.75	95.95	2.40	0.045	0.012

Table 5 Ablation study results obtained by varying the number of attention heads.

Number of Heads	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
1 Head	97.75	95.95	2.40	0.045	0.012
2 Heads	97.70	95.85	2.42	0.048	0.012
4 Heads	97.65	95.75	2.48	0.052	0.013

Table 6 Ablation study results obtained by varying the window size.

Window Size	DSC (%) ↑	Jaccard (%) ↑	HD (mm) ↓	RVD ↓	ASSD (mm) ↓
7×7	97.75	95.95	2.40	0.045	0.012
14×14	97.68	95.80	2.43	0.047	0.012
28×28	97.62	95.70	2.46	0.050	0.013

Table 7 Comparison of the proposed CAG-SwinUNet with recent methods based on DSC for segmentation accuracy.

Authors	Year	Method	DSC (%)	IoU (%)	HD (mm)
Ren et al. [22]	2025	LGMA-Net	97.72		
Tinglan et al. [34]	2025	DEMF-Net	96.12	92.53	
Cao et al. [33]	2025	SACU-Net	95.51	91.65	3.88
Qi et al. [20]	2024	AD-DUNet	97.08	95.04	3.44
Ou et al. [21]	2024	ResTransUNet	95.35		
Li et al. [27]	2023	RDCTrans UNet	93.38	89.22	
Chen et al. [28]	2023	DRAUNet	97.1		
Liu et al. [29]	2023	GCHA-Net	96.2		
Li et al. [31]	2023	Eres-UNet++	95.8	91.8	
Devidas et al. [32]	2023	LiM-Net	96.3	94.3	
He et al. [23]	2022	GAN-based 3D UNet	94.24		
Lv et al. [25]	2022	enhanced ResUNet	94.2		
Manjunath & Kwadiki [26]	2022	modified ResUNet	96.3		
Lu et al. [41]	2022	DefED-Net	96.3		
Wei et al. [24]	2021	GAN Mask R-CNN	91.3		
Proposed Method		CAG-SwinUNet	97.75	95.95	**2.40**

4.9 Comparison with Existing Methods

Table 7 presents a comparative analysis of the proposed CAG-SwinUNet model against several recently published liver segmentation methods, with performance evaluated using DSC, Jaccard, and HD wherever reported. This expanded comparison includes state-of-the-art Transformer-based, hybrid, and attention-driven architectures from 2021 to 2025, providing a more comprehensive benchmark against current literature.

Across the compared methods, CAG-SwinUNet achieves the highest DSC of 97.75% and one of the strongest Jaccard values at 95.95%, demonstrating highly consistent overlap with the ground truth. More importantly, the proposed model attains the lowest HD of 2.40 mm, indicating superior boundary precision relative to other recent approaches. The combination of high region-level agreement (Dice/Jaccard) and significantly reduced boundary error highlights the robustness of CAG-SwinUNet in delineating complex liver contours.

5 Discussion

The evaluation of CAG-SwinUnet on the LiTS and SLIVER07 datasets demonstrates its effectiveness in liver segmentation, achieving superior performance over SwinUnet through improved overlap accuracy and boundary precision. On the LiTS dataset, CAG-SwinUnet attains a DSC of 97.75%, Jaccard Index of 95.95%, HD of 2.40 mm, and ASSD of 0.012 mm, outperforming SwinUnet's DSC of 97.10%, HD of 2.82 mm, and ASSD of 0.015 mm. A similar trend is observed on SLIVER07, where CAG-SwinUnet achieves a DSC of 96.65% compared to SwinUnet's 96.21%. These improvements stem from the CAG mechanism's ability to selectively enhance liver-relevant features while suppressing background noise, mitigating issues common in SwinUnet's direct concatenation approach.

Qualitative and explainability analyses reinforce these findings. Segmentation overlays illustrate that CAG-SwinUnet's predictions (red contours) align closely with ground truth (green contours), reducing false positives near hepatic vessels and lesion boundaries where SwinUnet struggles. Further visual evidence from CAG-enhanced skip connection heatmaps shows improved feature fusion, capturing finer liver structures while preventing noise propagation. Grad-CAM-based XAI visualizations highlight that CAG-SwinUnet's attention is more concentrated on liver parenchyma, whereas SwinUnet exhibits diffused activation, often misidentifying adjacent tissues. The increased spatial awareness facilitated by CAG leads to improved precision in challenging regions, as reflected in the model's lower HD (2.40 mm vs. 2.82 mm) and ASSD (0.012 mm vs. 0.015 mm) on LiTS, validating its ability to refine complex anatomical structures.

Ablation studies further dissect CAG-SwinUnet's architecture, showing that replacing SwinUnet's concatenation with cross-attention improves DSC from 97.10% to 97.45% and reduces HD from 2.82 mm to 2.55 mm. The addition of residual connections further enhances DSC to 97.75%, lowering HD to 2.40 mm, ensuring feature continuity across decoder layers. Comparisons with existing methods highlight CAG-

SwinUnet's advantages over ResUNet and GAN-based U-Net models, which suffer from mode collapse. Future work could explore multi-head CAG mechanisms or hybrid loss functions for fine-grained segmentation, while extending evaluation to multi-modal datasets could further establish CAG-SwinUnet's clinical relevance.

6 Conclusion

This study introduces CAG-SwinUnet, a Transformer-based segmentation model designed to enhance both accuracy and interpretability in liver segmentation. By integrating a Cross-Attention Gate within the skip connections of SwinUNet, the model selectively enhances relevant features while minimizing interference from unrelated structures. Additionally, its residual-enhanced attention mechanism preserves contextual information, refining segmentation across diverse imaging conditions. By balancing local and global feature representations, CAG-SwinUnet improves boundary delineation and segmentation precision, addressing key challenges in liver segmentation from CT images.

Extensive evaluations on the LiTS and SLIVER07 datasets demonstrate that CAG-SwinUnet surpasses existing Transformer-based models, achieving a Dice Similarity Coefficient of 97.75% and a Hausdorff Distance of 2.40 mm on LiTS, and 96.65% DSC with 3.10 mm HD on SLIVER07. The model ensures training stability with 98.45% accuracy and a low Dice Loss of 0.045, while validation results confirm strong generalization with 97.90% accuracy and a Dice Loss of 0.058.

To enhance interpretability, Grad-CAM-based visualizations provide explainable AI insights, allowing clinicians to validate predictions and assess model reliability. An ablation study quantifies the impact of key components, demonstrating that cross-attention, residual connections, and output projection contribute to improved DSC performance. By effectively balancing feature representations and reducing noise, CAG-SwinUnet establishes a robust and adaptable framework for clinical applications, including tumor detection, surgical planning, radiation therapy, and liver volumetry across diverse and pathologically rich datasets.

Disclosures

All authors declare that there is no conflict of interest in this paper.

References

1. W. van Elmpt, G. Landry, "Quantitative computed tomography in radiation therapy: A mature technology with a bright future," Physics and Imaging in Radiation Oncology 6, 12–13 (2018).
2. S. S. Kumar, R. S. Moni, and J. Rajeesh, "An automatic computer-aided diagnosis system for liver tumours on computed tomography images," Computers & Electrical Engineering 39(5), 1516–1526 (2013).
3. A. H. Foruzan, R. Aghaeizadeh Zoroofi, M. Hori, and Y. Sato, "Liver segmentation by intensity analysis and anatomical information in multi-slice CT images," International Journal of Computer Assisted Radiology and Surgery 4(3), 287–297 (2009).
4. S. S. Kumar, R. S. Moni, and J. Rajeesh, "Automatic liver and lesion segmentation: a primary step in diagnosis of liver diseases," Signal, Image and Video Processing 7(1), 163–172 (2013).
5. C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, and D. D. Feng, "A Likelihood and Local Constraint Level Set Model for Liver Tumor Segmentation from CT Volumes," IEEE Transactions on Biomedical Engineering 60(10), 2967–2977 (2013).
6. A. Saito, S. Nawano, and A. Shimizu, "Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs," Medical Image Analysis 28, 46–65 (2016).
7. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data 8(1), 53 (2021).
8. 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).
9. 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).
10. R. M. Prakash, M. Vimala, V. Srilekha, P. Krishnaleela, and S. Thayammal, "UNet with Attention Mechanism for Segmentation of Liver from Abdominal CT Images," in 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), IEEE (2024).
11. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929 (2020).

12. S. S. Kumar, "Advancements in medical image segmentation: A review of transformer models," Computers and Electrical Engineering 123, 110099 (2025).
13. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation," in Computer Vision – ECCV 2022 Workshops, L. Karlinsky, T. Michaeli, and K. Nishino (Eds.), Springer Nature Switzerland, 13803, 205–218 (2023).
14. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, 9992–10002 (2021).
15. S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," Information Fusion 99, 101805 (2023).
16. P. Das, A. Ortega, "Gradient-Weighted Class Activation Mapping for Spatio Temporal Graph Convolutional Network," in ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 4043–4047 (2022).
17. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative Adversarial Nets," Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Eds.), 2672–2680 (2014).
18. C. Wei, S. Ren, K. Guo, H. Hu, and J. Liang, "High-Resolution Swin Transformer for Automatic Medical Image Segmentation," Sensors 23(7), 3420 (2023).
19. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 574–584 (2021).
20. H. Qi, W. Wang, Y. Shi, and X. Wang, "AD-DUNet: A dual-branch encoder approach by combining axial Transformer with cascaded dilated convolutions for liver and hepatic tumor segmentation," Biomedical Signal Processing and Control 95, 106397 (2024).
21. J. Ou, L. Jiang, T. Bai, P. Zhan, R. Liu, and H. Xiao, "ResTransUnet: An effective network combined with Transformer and U-Net for liver segmentation in CT scans," Computers in Biology and Medicine 177, 108625 (2024).
22. W. Ren, B. Li, H. Peng, and J. Wang, "Lgma-net: liver and tumor segmentation methods based on local–global feature mergence and attention mechanisms," Signal, Image and Video Processing 19(1), 43 (2025).
23. R. He, S. Xu, Y. Liu, Q. Li, Y. Liu, N. Zhao, Y. Yuan, and H. Zhang, "Three-Dimensional Liver Image Segmentation Using Generative Adversarial Networks Based on Feature Restoration," Frontiers in Medicine 8, 794969 (2022).
24. X. Wei, X. Chen, C. Lai, Y. Zhu, H. Yang, and Y. Du, "Automatic Liver Segmentation in CT Images with Enhanced GAN and Mask Region-Based CNN Architectures," BioMed Research International 2021(1), 9956983 (2021).
25. P. Lv, J. Wang, X. Zhang, C. Ji, L. Zhou, and H. Wang, "An improved residual U-Net with morphological-based loss function for automatic liver segmentation in computed tomography," Mathematical Biosciences and Engineering 19(2), 1426–1447 (2022).
26. R. V. Manjunath, K. Kwadiki, "Automatic liver and tumour segmentation from CT images using Deep learning algorithm," Results in Control and Optimization 6, 100087 (2022).
27. L. Li, H. Ma, "RDCTrans U-Net: A Hybrid Variable Architecture for Liver CT Image Segmentation," Sensors 22(7), 2452 (2022).
28. Y. Chen, C. Zheng, T. Zhou, L. Feng, L. Liu, Q. Zeng, and G. Wang, "A deep residual attention-based U-Net with a biplane joint method for liver segmentation from CT scans," Computers in Biology and Medicine 152, 106421 (2023).
29. H. Liu, Y. Fu, S. Zhang, J. Liu, Y. Wang, G. Wang, and J. Fang, "GCHA-Net: Global context and hybrid attention network for automatic liver segmentation," Computers in Biology and Medicine 152, 106352 (2023).
30. J. Wang, P. Lv, H. Wang, and C. Shi, "SAR-U-Net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in Computed Tomography," Computer Methods and Programs in Biomedicine 208, 106268 (2021).
31. J. Li, K. Liu, Y. Hu, H. Zhang, A. A. Heidari, H. Chen, W. Zhang, A. D. Algarni, and H. Elmannai, "Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++," Computers in Biology and Medicine 158, 106501 (2023).
32. D. T. Kushnure, S. Tyagi, and S. N. Talbar, "LiM-Net: Lightweight multi-level multiscale network with deep residual learning for automatic liver segmentation in CT images," Biomedical Signal Processing and Control 80, 104305 (2023).
33. Y. Cao, Y. Cheng, "SACU-Net: Shape-Aware U-Net for Biomedical Image Segmentation With Attention Mechanism and Context Extraction," IEEE Access 13, 5719–5730 (2025).
34. L. Tinglan, Q. Jun, Q. Guihe, S. Weili, and Z. Wentao, "Liver segmentation network based on detail enhancement and multi-scale feature fusion," Scientific Reports 15(1), 683 (2025).

35. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," arXiv preprint arXiv:1804.03999 (2018).
36. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," arXiv preprint arXiv:2102.04306 (2021).
37. X. Jia, S. Jian, Y. Tan, Y. Che, W. Chen, and Z. Liang, "Gated Cross-Attention Network for Depth Completion," in ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 1–5 (2025).
38. P. Bilic, P. Christ, H. B. Li, et al., "The Liver Tumor Segmentation Benchmark (LiTS)," Medical Image Analysis 84, 102680 (2023).
39. J. Sun, Z. Hui, C. Tang, and X. Wu, "Liver segmentation based on complementary features U-Net," The Visual Computer 39(10), 4685–4696 (2023).
40. N. Ibtehaz, M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," Neural Networks 121, 74–87 (2020).
41. T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, and A. K. Nandi, "DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation," IEEE Transactions on Radiation and Plasma Medical Sciences 6(1), 68–78 (2022).

An *in Vitro* Application of Diode Laser-Ozone Based Photodynamic Therapy for Microbial Inactivation

Deny Arifianto[1,2], Suryani Dyah Astuti[1*], Riries Rulaningtyas[1], Aulia Haidar Rahman Nugraha[1], Ahmad Khalil Yaqubi[3], and Ardiansyah Syahrom[4]

[1] Department of Physics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia
[2] Department of Engineering, Faculty of Vocational, Airlangga University, Surabaya 60286, Indonesia
[3] Center for Biomedical Research, National Research and Innovation Agency (BRIN), Cibinong, Jawa Barat 16912, Indonesia
[4] Medical Devices and Technology Centre, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia
*e-mail: suryanidyah@fst.unair.ac.id

Abstract. *Staphylococcus aureus* (*S. aureus*) can form biofilms that contribute to antibiotic-resistant infections. This study investigated the effectiveness of a combined red diode laser (665 nm) and ozone system for inactivating *S. aureus* biofilms. Biofilm samples were treated with ozone, laser irradiation, or a combination of both, and bacterial survival was evaluated after incubation. Statistical analysis was performed using a factorial ANOVA test. The results demonstrated that combined laser irradiation with a power density of 0.24 W/cm^2 and ozone treatment significantly enhanced biofilm inactivation compared to either treatment alone. The maximum biofilm reduction reached 94% with combined treatment during 40 s, compared to less than 80% reduction using laser irradiation alone. These findings indicate that red diode laser and ozone combination therapy is a promising approach for effectively reducing *S. aureus* biofilms, offering potential advantages for managing antibiotic-resistant infections.

Keywords: biofilm; *Staphylococcus aureus*; red diode laser (665 nm); ozone; percent reduction.

Paper #9431 received 12 Nov 2025; revised manuscript received 15 Jan 2026; accepted for publication 17 Jan 2026; published online 19 Mar 2026. doi: 10.18287/JBPE26.12.010304.

1 Introduction

Tropical regions are significantly impacted by infectious diseases due to microorganisms that invade the body and lead to illness [1]. Key vectors include viruses, fungi, bacteria, and parasites. Major entry points for bacteria are the respiratory and gastrointestinal tracts, as well as the genitals and urine tract [2]. Approximately 10% of hospital patients develop nosocomial infections, with *Staphylococcus aureus* (*S. aureus*) being a prevalent cause, transmitted through food, medication, medical equipment, or healthcare personnel [3]. The term "staphylococcus", derived from Greek, describes these gram-positive, spherical bacteria that cluster like grapes and are found globally, including in soil [4].

The hospital utilizes sterilization facilities to reduce infection by eliminating spores and bacteria from surfaces. Spore bacteria are notably resistant to sterilization efforts [5]. Factors affecting sterilization effectiveness include the type and number of germs, contamination levels, and instrument design features [6]. Dental tools, which must be sterilized, often use dry heat sterilization though this method has limitations such as slow penetration, high costs, and ineffectiveness with non-metallic instruments [7]. Autoclaves, employing high-temperature steam, are commonly used but require consistent heat and maintenance, and rely on boiling water to operate safely [8].

S. aureus bacteria frequently create biofilms on medical equipment [9]. A biofilm is a group of microbial cells that form an irreversible bond to a surface and are encased in a matrix of extracellular polymeric substances (EPS) that they independently produce [10]. Biofilms exhibit phenotypic changes from planktonic cells or their free cells, such as changes in growth rates and changes in gene transcription [11]. Bacteria are able to endure physical, chemical, and biological challenges when biofilms are present. Biofilms function by impeding immune system function and generating antibiotic

resistance [12, 13]. The photodynamic inactivation (PDI) approach is one of the most precise solutions needed to address the issue of resistance brought on by biofilms [14].

PDI is a therapeutic process involving light, photosensitizers, and oxygen, aimed at inhibiting the growth of disease-causing microorganisms [15]. It starts with photosensitization, primarily using porphyrins, which absorb light to produce reactive oxygen species (ROS) that can kill bacteria [16]. Different light sources, including lasers and LEDs, are employed in antimicrobial photodynamic therapy [17]. The mechanism involves photophysical, photochemical, and photobiological activities: porphyrins absorb light, excite electrons, and generate ROS upon interaction with oxygen, leading to bacterial cell death [18].

By using light that corresponds to the absorption spectra of porphyrin photosensitizers and appropriate irradiation doses, bacterial cells can be photoinactivated. Previous studies indicate that the photodynamic technique, utilizing diode laser light, produces both photobiomodulation and antimicrobial effects. Ozone, a minimal atmospheric component, is generated through ultraviolet light and dielectric plasma discharge and has been shown to effectively reduce various microorganisms, including bacteria, viruses, and fungi [19]. It is particularly effective in preserving horticultural products and preventing biofilm deterioration of *S. aureus*, with experiments showing a reduction of up to 71.96% in *S. aureus* biofilms at an effective ozone concentration of 0.01 mg/L. Combining ozone with photodynamic treatment enhances microbial inactivation through optimal ozone levels, laser energy, and exposure time [20].

PDI effectively combats biofilm-associated resistance by generating ROS through light, photosensitizers, and oxygen, damaging bacterial cells. Combining PDI with ozone enhances bacterial inactivation and weakens biofilm structures, leading to increased efficacy against *S. aureus* biofilms. The study focuses on developing a system that integrates red diode laser with controlled ozone flow, evaluating different laser energy doses and ozone durations to optimize bacterial eradication. This work aims to advance antimicrobial photodynamic therapy and improve infection control, particularly in high infectious disease regions.

2 Methods and Materials

2.1 Sample Extraction

Fresh leaves of binahong (Anredera cordifolia) and betel (Piper beetle) were collected in Sidoarjo, East Java, cleaned with distilled water, and prepared by removing the stalks. The leaves were blended with 70% ethanol in a 1:5 ratios (20 g of leaves to 100 ml of ethanol), and then subjected to maceration for seven days in a closed container. After filtering, the ethanol extract was stored at 4 °C for future experiments [21].

2.2 Bacterial Culture

S. aureus ATCC 25923 was inoculated on Tryptone Soy Agar and incubated at 37 °C, calibrated to 1.0 McFarland Standard (approximately 10^8 colony-forming units (CFU)/mL). The bacteria were transferred to Trypsic Soy Broth, diluted with saline, and vortexed, resulting in serial dilutions (10^{-2} to 10^{-8} CFU/mL). A 2% sucrose solution was added for biofilm production in a microplate incubated for 48 h at 37 °C. After phosphate-buffered saline (PBS) washing, crystal violet solution was added, and another incubation was performed for 30 minutes, with optical density (OD) measured at 450 nm using an enzyme-linked immunosorbent assay (ELISA) reader to confirm the method's accuracy.

2.3 Comprehensive Assessment of Bacterial Viability and Biofilm Biomass

A complex methodology was employed to assess biofilm biomass and bacterial viability, focusing on *S. aureus*. Initial bacterial cultures were established at specific concentrations and injected into microplate wells with sucrose, followed by a 48-h incubation. Biofilms were stained with crystal violet, and optical density was measured using an ELISA reader. Additional techniques, including CFU counting, confocal laser scanning microscopy, and scanning electron microscopy, were utilized to further investigate biofilm survival and structure, providing crucial insights into treatment effectiveness.

2.4 Experimental Unit

The experiments were carried out at 25.5–25.6 °C using the setup shown in Fig. 1. The experimental unit used a Laserland model 650RM-100 continuous red laser diode (nominal output power 100 mW) and an ozone generator (capacity 10 $g \cdot h^{-1}$). Laser and ozone delivery were synchronized using a microcontroller. Biofilm samples were cultivated in 96-well microplates (growth area approximately 0.32 cm^2 per well). The laser spot area at the biofilm surface was adjusted by changing the microplate height using a stepper motor (Z) and fixed when the spot area reached ~0.13 cm^2. Although the power meter/detector provides high readout resolution, the physical parameters relevant to the biological experiments are limited primarily by (i) laser output stability during experiments and (ii) the uncertainty in defining/measuring the laser spot size at the sample plane; therefore, all laser parameters used for biological-dose calculations are reported with two significant figures. Dose calculations used $A = 0.13$ cm^2; timing/position controlled; ozone titrated iodometrically.

2.5 Treatment of the Sample

Four groups were tested: control, ozone only, laser only, and laser + ozone. Biofilms were exposed for 10, 20, 30, or 40 s at a constant irradiance power density.

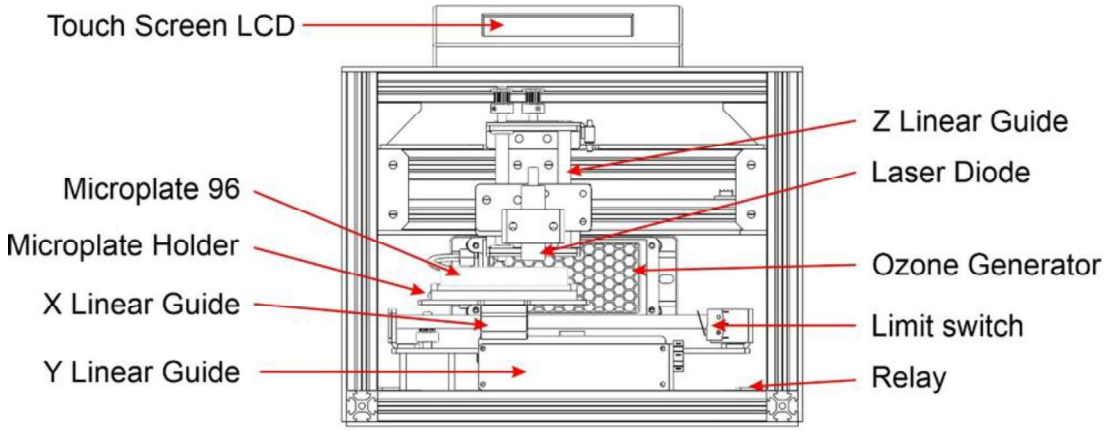


Fig. 1 Apparatus set-up of light and ozone source.

Two factors were analyzed: treatment modality and exposure time (10–40 s), with $n = 5$ per condition. After treatment, samples were plated, incubated 24 h, and CFU were counted.

2.6 Data Analysis

A factorial two-way ANOVA test was employed in data analysis to evaluate how the therapy affected each group. The following Eq. (1) is then used to compute the percentage decrease in bacterial biofilm [22].

$$U_{exp} = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1+1)}{2} - \sum r_1. \qquad (2)$$

Optical density (OD) was used only as a supportive biofilm biomass and viability indicator; colony-forming units were the primary endpoint.

3 Results

Laser exposure was defined by the optical power at the sample plane and the exposure duration. The measured power ranged from 31.85–32.06 mW (≤2% variation); therefore, all biological experiments and dose calculations use $P = 32$ mW ± 2% (two significant figures), rather than time-specific power values. The monochromator spectrum peaked at 665 nm. Because of practical uncertainty in spot-size determination, the laser spot area used consistently was $A = 0.13$ cm², giving an irradiance of $E = 0.24$ W/cm². Biofilms were irradiated for 10–40 s, corresponding to $H = 2.4$–9.6 J/cm² ($H = E \times t$). The previous table listing power/irradiance by exposure time was removed, and only the measured power range and nominal values used for biological calculations are reported.

Ozone was delivered to the sample through a fixed-output nozzle. The effective concentration of ozone, crucial for the effectiveness of treatment, was experimentally determined by collecting ozone in distilled water under identical conditions. The concentration was determined through iodometric titration using standardized sodium thiosulfate. Longer exposure times resulted in higher concentrations of ozone. The results were summarized in Table 1. The error of the measurement method was ±5% based on repeated trials (n=3):

Table 1 Ozone concentration at various flow times.

Flow time (s)	Concentration (mg/L)
10	1.5×10^{-3}
20	3.0×10^{-3}
30	4.0×10^{-3}
40	5.0×10^{-3}

The McFarland test was performed to determine the relationship between OD and bacterial concentration expressed as CFU/mL. OD values were measured using a UV-Vis spectrophotometer at 595 nm. To construct the calibration curve, CFU/mL values were obtained through the standard plate-count method: bacterial suspensions were serially diluted (10^{-1} to 10^{-6}), plated on TSA media, incubated for 24 h at 37 °C, and colonies within the 30–300 CFU range were counted. The colony count was then multiplied by the dilution factor to obtain the corresponding CFU/mL value for each OD measurement. Plotting OD against CFU/mL produced a linear relationship with $R^2 = 0.98$, indicating strong correlation. Fig. 2 shows the calibration curve of CFU/mL versus OD values at 595 nm.

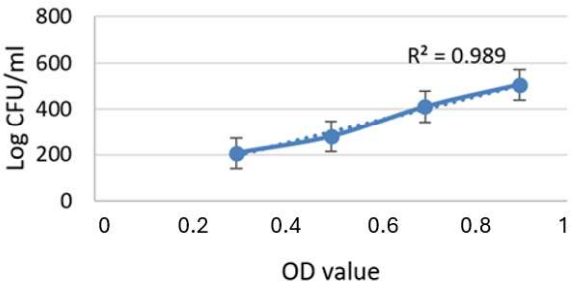


Fig. 2 Graph of the relationship between CFU/mL and OD values at a wavelength of 595 nm.

Table 2 Results of statistical analysis on *S. aureus* bacteria.

Interaction	10	*p*-value	20	*p*-value	30	*p*-value	40	*p*-value
lasers vs ozone	36.21±2.76	<0.001	50.98±2.17	<0.001	60.22±1.61	<0.001	70.41±1.94	0.001
laser vs laser + ozone	54.09±1.91	<0.001	65.41±1.82	<0.001	75.68±2.32	<0.001	85.35±2.22	<0.001
ozone vs laser + ozone	42.96±4.25	<0.001	57.65±3.54	<0.001	69.60±3.67	<0.001	79.38±3.70	<0.001

To ensure accurate quantification, dilutions yielding colony counts within the standard 30–300 CFU range were selected. Based on the serial dilution results, the 10^{-5} dilution provided an average of 236 colonies, which met the acceptable counting range and was used to generate the corresponding CFU/mL value.

Treatment involved three groups exposed to different methods (laser, ozone, or both) for 10, 20, 30, or 40 s. The energy density varied with time increments, impacting the effectiveness as measured by percentage drops in OD or CFU/ml in bacterial biofilms. The results of the treatment in the three groups revealed a connection between the amount of time spent treating and the bacteria's viability in the OD value shown in Fig. 3.

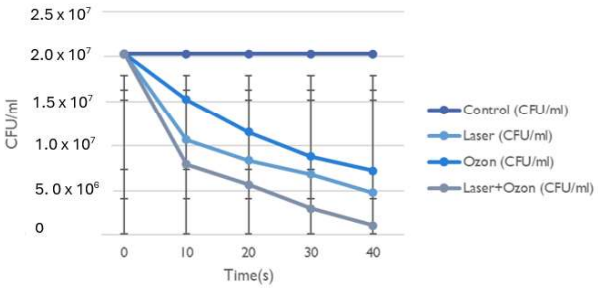


Fig. 3 Decreased biofilm viability of S. aureus bacteria.

The proportion of bacteria that died as a consequence of the calculation Fig. 4 displays the proportion of bacterial mortality.

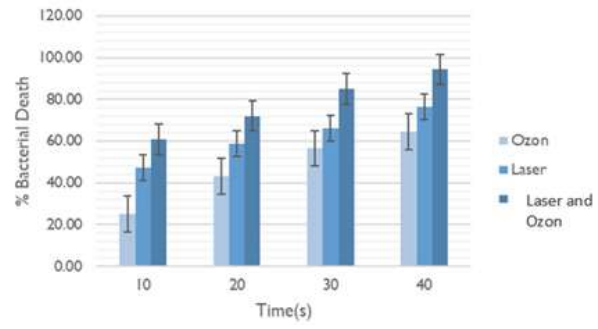


Fig. 4 Comparison of the percentage of *S. aureus* biofilm reduction.

The normality tests for the laser, ozone, and laser + ozone groups indicated normal distribution ($p = 0.540$, $p = 0.109$, $p = 0.182$). Levene's test confirmed homogeneity of variances ($p > 0.05$). ANOVA results showed significant differences in treatment effectiveness over time for bacterial reduction, specifically for *S. aureus*, as detailed in Table 2.

By comparing the OD values across treatment groups, the percentage decrease was calculated. The OD readings were converted to estimated CFU/ml using the McFarland curve technique, and these converted values were then compared with the control group to determine the average percentage decrease. As the McFarland method has limited accuracy, these CFU/ml estimates should be interpreted with caution.

4 Discussion

Biofilms are clusters of microorganisms encased in an EPS that enhances their resistance to environmental factors, including antibiotics. The formation of a biofilm starts with an attachment phase and progresses through early (0–11 h), intermediate (12–24 h), and mature (24–48 h) stages (see, Fig. 5). As the biofilm matures, the EPS quantity increases, producing smaller pores that limit antibiotic penetration, especially during the adult phase [23]. As the biofilm becomes thicker and contains more EPS, the diffusion of light, photosensitizer, ozone, and reactive oxygen species becomes more limited. To manage the biofilm, one option is photodynamic treatment (PDT) [24].

PDT, utilizing red diode laser and ozone, demonstrates enhanced antibacterial effects against *S. aureus* biofilms [25]. The combination treatment generates reactive oxygen species (ROS), disrupting the extracellular polymeric substance matrix and bacterial membranes, allowing ozone to penetrate and induce oxidative damage [26]. This synergistic method leads to superior bacterial inactivation, with maximum reduction observed at 40 s [27]. The efficacy is attributed to photodynamic processes and ROS formation, with ozone contributing through oxidative actions and lipid peroxidation, highlighting the significance of ROS in compromising bacterial cell integrity [28, 29]. Illustration of the inactivation mechanism of *S. aureus* biofilms with red laser and ozone treatment is shown in Fig. 6.

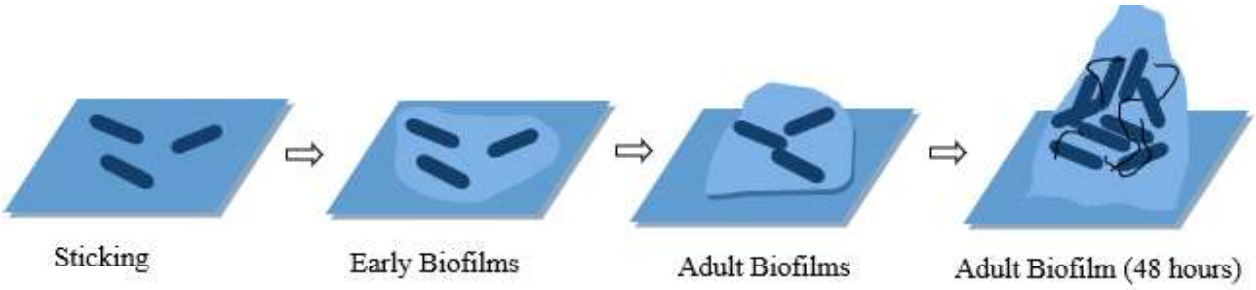


Fig. 5 Illustration of *S. aureus* Biofilm Growth Mechanism.

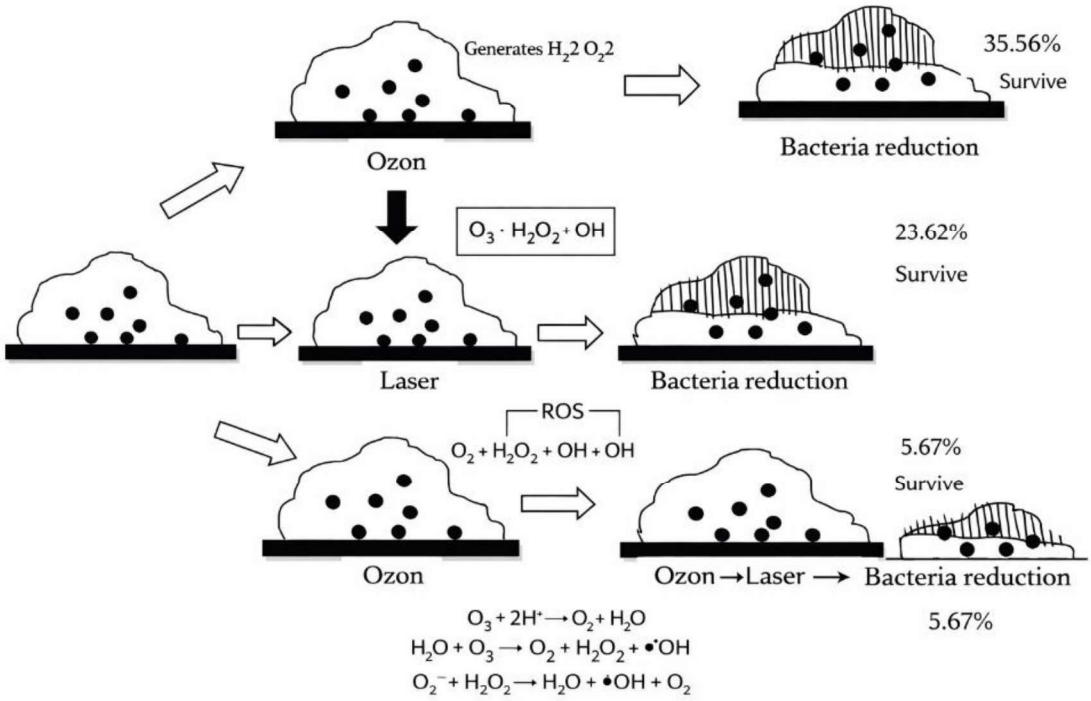


Fig. 6 Illustration of the inactivation mechanism of *S. aureus* biofilms with red laser and ozone treatment.

Ozone disrupts bacterial metabolism by inhibiting enzymes, effectively eliminating bacteria in biofilms. It reacts with polyunsaturated fatty acids to produce hydrogen peroxide (H_2O_2) [30]. Optimal ozone concentration is essential, as too little reduces effectiveness and too much can damage tissue. Consistent with the Results (Fig. 4), the highest bacterial reduction at 40 s was observed in the laser + ozone group (≈94% reduction), which was higher than laser-only and ozone-only treatments. All percentage-reduction values discussed here follow the values presented in the Results figures and tables.

5 Conclusion

The outcomes demonstrated that ozone and a red diode laser worked together to lessen bacterial growth. The percentage decrease achieved by the laser therapy group alone increased by 29.89% when paired with ozone. This is due to ozone's ability to trigger a process that results in the production of H_2O_2, which increases the amount of harm done to bacterial biofilms. In this study, the maximum energy density for photo inactivating *S. aureus* biofilm was 9.6 J/cm^2 at 40 s with the addition of ozone, leading to a reduction percentage of *S. aureus* biofilm of 94%, as opposed to less than 80% at 40 s with laser alone. Thus, the *S. aureus* biofilm is effectively reduced by the red diode laser and ozone combination treatment.

Acknowledgement

This research was funded by Airlangga Research Funds under the Doctoral Dissertation Research Scheme No. 1814/B/UN3.LPPM/PT.01.03/2025.

Disclosures

All authors declare that there is no conflict of interest in this paper.

References

1. J. L. Brooks, M. Jefferson, "Staphylococcal biofi: lmsquest for the magic bullet," Advances in Applied Microbiology 81, 63–87 (2012).
2. V. Neville, P. Dayan, I. D. Gilchrist, E. S. Paul, and M. Mendl, "Using primary reinforcement to enhance translatability of a human affect and decision-making judgment bias task," Journal of Cognitive Neuroscience 33(12), 2523–2535 (2021).
3. E. J. G. Pollitt, P. T. Szkuta, N. Burns, and S. J. Foster, "Staphylococcus aureus infection dynamics," PLoS Pathogens 14(6), e1007112 (2018).
4. S. D. Astuti, A. Zaidan, E. M. Setiawati, and Suhariningsih, "Chlorophyll mediated photodynamic inactivation of blue laser on Streptococcus mutans," AIP Conference Proceedings, 1718(1) (2016).
5. M. T. Mubarak, I. Ozsahin, and D. U. Ozsahin, "Evaluation of sterilization methods for medical devices," 2019 Advances in Science and Engineering Technology International Conferences (ASET), 1–4 (2019).
6. P. DeSantis, "Steam sterilization in autoclaves," in Handbook of Validation in Pharmaceutical Processes, J. Agalloco, P. DeSantis, A. Grilli, and A. Pavell (Eds.), 4 Ed., CRC Press, Boca Raton, 217–230 (2021).
7. A. Blessing, S. I. Oluka, N. H. Nwobodo-Nzeribe, and C. H. Nwabueze "Review of historical development of autoclave towards healthcare sector," International Journal of Real-Time Applications and Computing Systems 2(3), 29–39 (2023).
8. A. K. Yaqubi, S. D. Astuti, A. H. Zaidan, A. Syahrom, and D. Z. I. Nurdin, "Antibacterial effect of red laser-activated silver nanoparticles synthesized with grape seed extract against Staphylococcus aureus and Escherichia coli," Lasers in Medical Science 39, 47 (2024).
9. S. D. Astuti, Hafidiana, R. Rulaningtyas, Abdurachman, A. P. Putra, Samian, and D. Arifianto, "The efficacy of photodynamic inactivation with laser diode on Staphylococcus aureus biofilm with various ages of biofilm," Infectious Disease Reports 12(s1), 8736 (2020).
10. A. K. Yaqubi, S. D. Astuti, P. A. D. Permatasari, N. Komariyah, E. Endarko, and A. H. Zaidan, "*In vitro* inactivation efficiency of Bacillus subtilis and Escherichia coli bacteria in sterilizers using violet irradiation," Biomedical Photonics 11, 4–10 (2022).
11. P. A. D. Permatasari, S. D. Astuti, A. K. Yaqubi, E. A. W. Paisei, and N. Anuar, "Effectiveness of Katuk leaf chlorophyll (Sauropus androgynus (L.) Merr) with blue and red laser activation to reduce Aggregatibacter actinomycetemcomitans and Enterococcus faecalis biofilm," Biomedical Photonics 12(1), 14–21 (2023).
12. K. Hoenes, U. Wenzel, B. Spellerberg, and M. Hessling, "Photoinactivation sensitivity of Staphylococcus carnosus to visible-light irradiation as a function of wavelength," Photochemistry and Photobiology 96(1), 156–169 (2020).
13. S. Astuti, V. Victory, A. Mahmud, A. Putra, and D. Winarni, "The effects of laser diode treatment on liver dysfunction of Mus musculus due to carbofuran exposure: an *in vivo* study," Journal of Advanced Veterinary and Animal Research 6(4), 499 (2019).
14. K. Plaetzer, B. Krammer, J. Berlanda, F. Berr, and T. Kiesslich, "Photophysics and photochemistry of photodynamic therapy: fundamental aspects," Lasers in Medical Science 24, 259–268 (2009).
15. S. D. Astuti, B. I. Prasaja, and T. A. Prijo, "An *in vivo* photodynamic therapy with diode laser to cell activation of kidney dysfunction," Journal of Physics: Conference Series, 853(1), 012038 (2017).
16. A. I. Mardianto, E. M. Setiawatie, W. P. Lestari, A. Rasheed, and S. D. Astuti, "Photodynamic inactivation of Streptococcus mutans bacteria with photosensitizer Moringa oleifera activated by LED," Journal of Physics Conference Series 1505(1), 012061 (2020).
17. D. Arifianto, S. D. Astuti, P. A. D. Permatasari, I. Arifah, A. K. Yaqubi, R. Rulaningtyas, and A. Syahrom, "Design and application of near infrared LED and solenoid magnetic field instrument to inactivate pathogenic bacteria," Micromachines 14(4), 848 (2023).
18. S. D. Astuti, R. A. Wibowo, N. I. D. N. Abdurachman, and K. Triyana, "Antimicrobial photodynamic effects of polychromatic light activated by magnetic fields to bacterial viability," Journal of International Dental and Medical Research 10(1), 111–117 (2017).
19. F. H. Quina, G. T. M. Silva, "The photophysics of photosensitization: a brief overview," Journal of Photochemistry and Photobiology 7, 100042 (2021).
20. D. M. Harris, J. G. Sulewski, "Photoinactivation and photoablation of Porphyromonas gingivalis," Pathogens 12(9), 1160 (2023).
21. A. Sułek, B. Pucelik, M. Kobielusz, A. Barzowska, and J. M. Dąbrowski, "Photodynamic inactivation of bacteria with porphyrin derivatives: effect of charge, lipophilicity, ROS generation, and cellular uptake on their biological activity in vitro," International Journal of Molecular Sciences 21(22), 8716 (2020).
22. Suhariningsih, D. Winarni, S. A. Husen, F. Khaleyla, A. P. Putra, and S. D. Astuti, "The effect of electric field, magnetic field, and infrared ray combination to reduce HOMA-IR index and GLUT 4 in diabetic model of Mus musculus," Lasers in Medical Science 35, 1315–1321 (2020).

23. M. Piksa, C. Lian, I. C. Samuel, K. J. Pawlik, I. D. W. Samuel, and K. Matczyszyn, "The role of the light source in antimicrobial photodynamic therapy," Chemical Society Reviews 52(5), 1697–1722 (2023).
24. S. D. Astuti, W. I. Pratiwi, N. A. Tanassatha, K. A. Alamsyah, Y. Susilo, and M. Khasanah, "Effect of ozone-induced diode laser of photodynamic inactivation on Pseudomonas aeruginosa," Malaysian Journal of Medicine and Health Sciences 17(20), 27–32 (2021).
25. P. S. Puspita, S. D. Astuti, A. M. T. Nasution, A. A. S. Pradhana, and A. Mawaddah, "Photodynamic therapy with ozone aids Staphylococcus aureus biofilm reduction," Indian Veterinary Journal 97(2), 24–26 (2020).
26. S. Shui, Z. Zhao, H. Wang, M. Conrad, and G. Liu, "Non-enzymatic lipid peroxidation initiated by photodynamic therapy drives a distinct ferroptosis-like cell death pathway," Redox Biology 45, 102056 (2021).
27. S. D. Astuti, N. D. Drantantiyas, A. P. Putra, P. S. Puspita, A. Syahrom, and S. Suhariningsih, "Photodynamic effectiveness of laser diode combined with ozone to reduce Staphylococcus aureus biofilm with exogenous chlorophyll of Dracaena angustifolia leaves," Biomedical Photonic 8(2), 4–13 (2019).
28. U. Roobab, G. M. Madni, M. M. A. N. Ranjha, A. W. Khan, S. Selim, M. S. Almuhayawi, M. Samy, X.-A. Zeng, and R. M. Aadil, "Applications of water activated by ozone, electrolysis, or gas plasma for microbial decontamination of raw and processed meat," Frontiers in Sustainable Food Systems 7, 1007967 (2023).
29. S. D. Astuti, W. I. Pertiwi, S. P. A. Wahyuningsih, P. A. D. Permatasari, D. Z. I. Nurdin, and A. Syahrom "Effectiveness of ozone-laser photodynamic combination therapy for healing wounds infected with methicillin-resistant Staphylococcus aureus in mice," Veterinary World 16(5), 1176–1184 (2023).
30. L. Rozykulyyeva, S. D. Astuti, A. H. Zaidan, A. A. S. Pradhana, and P. S. Puspita, "Antibacterial activities of green synthesized silver nanoparticles from Punica granatum peel extract," AIP Conference Proceedings 2314(1), 060012 (2020).

Lung Cancer Segmentation Using an Enhanced TransUNet++ Architecture

Anoop V.[1*], Karthik G. M.[2], Dr Savitha S.[3], Anil Kumar Muthevi[4], Saravanan Ramamoorthi Agilesh[5], and Ramu B.[6]

[1] Department of Artificial Intelligence and Data Science, Jyothi Engineering College-Thrissur, Jyothi Hills, Panjal Road, Vettikattiri PO, Cheruthuruthy, Thrissur, Kerala 679531, India

[2] School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Katpadi Thiruvalam Road, Vellore 632014, Tamil Nadu, India

[3] Department of Computer Science and Engineering, BMS Institute of Technology and Management, Doddaballapur Main Road, Yelahanka, Bengaluru, Karnataka 560064, India

[4] Department of Computer Science and Engineering, Aditya University, Surampalem, Kakinada, Andhra Pradesh 533437, India

[5] Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh 522502, India

[6] Department of Electronics and Communication Engineering, Geethanjali College of Engineering and Technology, Hyderabad, Telangana 501301, India

*e-mail: vanoopainds@gmail.com

Abstract. Lung cancer is a life-threatening disease in which accurate staging of malignant nodules using computed tomography (CT) scans is critical for reducing mortality. Most existing approaches rely solely on deep learning models. This work proposes an accurate and computationally efficient hybrid deep learning framework for lung cancer analysis, integrating advanced preprocessing, feature extraction, and hybrid network architectures. The pipeline begins with preprocessing steps including resizing, normalization, edge detection, and median filtering to enhance image quality. Texture features are extracted using local binary patterns (LBP), while principal component analysis (PCA) is applied for dimensionality reduction. The optimized features are classified using an EfficientNet-B0 model. For precise segmentation, EfficientNet-B0 is embedded within a Transformer-based UNET++ (TransUNET++), enabling effective modeling of both local details and global contextual dependencies. Evaluated on a benchmark CT dataset, the proposed method achieved 98.58% accuracy, 98.47% sensitivity, 99.23% specificity, and 98.42% precision for classification, along with strong segmentation performance (99.53% Dice similarity coefficient, 98.56% Intersection over Union, 99.73% Hausdorff distance, 98.86% volumetric overlap error), demonstrating high spatial agreement with ground-truth masks.

Keywords: lung cancer; preprocessing; local binary patterns; principal component analysis; EfficientNet-B0; TransUNET++.

Paper #9323 received 14 Aug 2025; revised manuscript received 9 Jan 2026; accepted for publication 18 Jan 2026; published online 23 Mar 2026. doi: 10.18287/JBPE26.12.010305.

1 Introduction

Lung cancer is one of the deadliest malignancies worldwide, with over 2 million cases and 1.76 million deaths reported in 2018. Early and accurate diagnosis is essential for improving survival rates. While chest radiography offers limited diagnostic benefit, CT imaging provides superior sensitivity and spatial

resolution and is the preferred modality for lung cancer detection [1]. However, manual interpretation of CT scans is time-consuming and prone to observer variability, motivating the need for automated diagnostic systems. Traditional CAD systems using ML classifiers such as support vector machine (SVM), *k*-Nearest Neighbors (KNN) and Artificial Neural Network (ANN) rely on handcrafted features, which limits robustness and generalization under varying imaging conditions [2, 3]. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs) such as VGGNet, ResNet, and DenseNet, have enabled end-to-end feature learning and improved performance [4], but they primarily capture local features and struggle with global contextual dependencies. Transformer-based models and hybrid CNN-Transformer architectures address this limitation by integrating attention mechanisms, though challenges such as computational complexity and inconsistent CT quality remain [5].

2 Literature Survey

The existing literature largely reports individual methodological advances but often lacks a critical synthesis that clearly motivates subsequent research. For instance, Akitoshi et al. proposed a deep learning–based segmentation model for lung cancer detection in chest X-rays, achieving a mean Dice coefficient of 0.52 on malignant lesions; while this demonstrates feasibility, the moderate overlap score and limited dataset size restrict its robustness and clinical generalizability [6]. Mamtha et al. introduced a hybrid framework combining deformable models, Bayesian fuzzy clustering, and the water cycle sea lion optimization algorithm with a Shepard CNN, achieving high diagnostic accuracy; however, the reliance on multiple complex optimization stages significantly increased computational cost, limiting real-time or clinical deployment [7]. Ajni and Anitha developed a multi-object optimized attention network using shuffled shepherd and social ski-driver optimization with deep Renyi entropy fuzzy clustering for lung lobe segmentation, but the use of multiple heuristic optimizers adds algorithmic complexity and reduces interpretability and scalability [8]. Iftikhar et al. presented a three-stage pipeline integrating modified U-Net-based segmentation with a hybrid AlexNet-SVM classifier, yielding strong performance on the LUNA16 dataset; nonetheless, the separation of segmentation and classification required manual post-processing and failed to exploit end-to-end learning advantages [9]. Similarly, Imran et al. combined deep feature extraction with SVMs for early lung cancer detection on the LIDC/IDRI dataset, but the dependence on handcrafted pipeline stages limited adaptability across datasets [10]. Collectively, these studies highlight improvements in segmentation accuracy but reveal persistent challenges related to computational complexity, lack of end-to-end optimization, and limited clinical scalability, thereby motivating the need for a unified, efficient deep learning framework. Overall, existing approaches suffer from poor generalization, insufficient global context modeling, or excessive complexity. To address these gaps, this study proposes a unified hybrid deep learning framework that jointly performs lung cancer classification and segmentation on CT images. The pipeline incorporates preprocessing, LBP and PCA-based feature optimization, EfficientNet-B0 for efficient classification, and a TransUNET++ for attention-driven segmentation.

3 Proposed Methods

The proposed method introduces a comprehensive and robust pipeline for accurate medical image classification and segmentation by integrating advanced preprocessing, feature extraction, and deep learning models. Initially, medical images retrieved from the database (DB) undergo preprocessing, including resizing, normalization, edge detection, and median filtering, to ensure data uniformity, reduce noise, and enhance salient structures. Subsequently, texture features are extracted using LBP, and dimensionality reduction is performed using PCA to preserve discriminative information. The refined features are used to train an EfficientNet-B0 model, chosen for its optimal balance between accuracy and computational efficiency, enabling reliable image classification. For enhanced segmentation and further refinement, the learned features from EfficientNet-B0 are integrated into a TransUNET++, which combines convolutional and self-attention mechanisms to capture both local details and long-range contextual dependencies. This dual-stage framework ensures accurate classification and precise region-level segmentation. Fig. 1 illustrates the architecture of the proposed method.

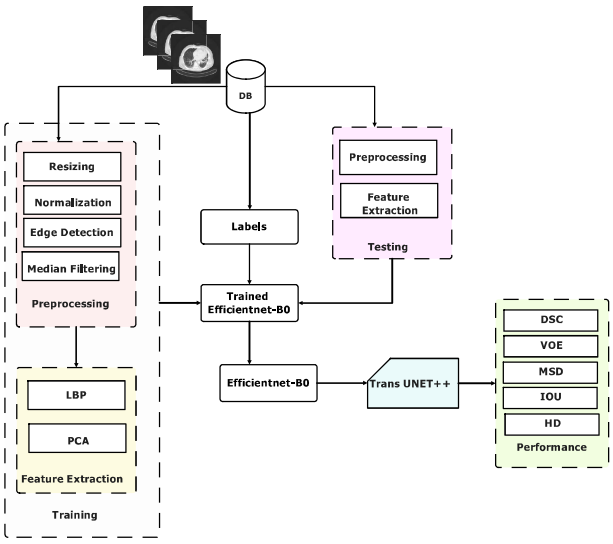


Fig. 1 Block diagram of proposed method.

Table 1 Parameter Specification of EfficientNet-B0.

Component	Specification
Number of layers	82
Scaling coefficient	ϕ=0 (baseline)
Activation function	Swish
Optimization	Adam
Building blocks	MBConv + squeeze-and-excitation (SE)
Output classes	3 (normal, benign, malignant)

3.1 Preprocessing

In this work, a comprehensive preprocessing pipeline is employed to enhance the quality of lung CT images for reliable classification and segmentation. All images are first resized to 224×224 pixels to ensure input uniformity and compatibility with EfficientNet-B0 and TransUNet++. Pixel intensities are then normalized to the range [0, 1] to stabilize training and improve convergence. Edge information is enhanced using the Canny operator, which highlights anatomical boundaries, nodules, and tumor regions. To suppress noise introduced during edge detection while preserving critical structures, median filtering is applied, effectively removing salt-and-pepper noise without blurring essential anatomical details.

3.2 Efficientnet-B0

EfficientNet-B0 is a convolutional neural network designed to achieve state-of-the-art image classification performance with minimal computational cost [11–13]. In this study, EfficientNet-B0 is selected as the classification backbone due to its compound scaling of network depth, width, and input resolution, making it well suited for relatively small medical imaging datasets. The model uses an input image size of $224 \times 224 \times 3$, providing a balance between computational efficiency and feature resolution. The architecture consists of MBConv blocks with integrated squeeze-and-excitation (SE) modules to enhance feature representation. Table 1 summarizes the key parameter settings of the EfficientNet-B0 model used for lung cancer classification.

3.3 Local Binary Pattern

The LBP is a widely used texture descriptor in image processing and computer vision, known for its simplicity and effectiveness in capturing local texture features [14–15]. For a given pixel, its intensity value is compared with those of its P surrounding neighbors positioned on a circle of radius R. The mathematical expression for the LBP code at a pixel location (x_c, y_c) is given by the following Eq.:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p , \quad (1)$$

where g_c is the gray value of the central pixel, g_p is the gray value of the p^{th} neighboring pixel, and $s(x)$ is a thresholding function defined as:

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases}. \quad (2)$$

3.4 Principal Component Analysis

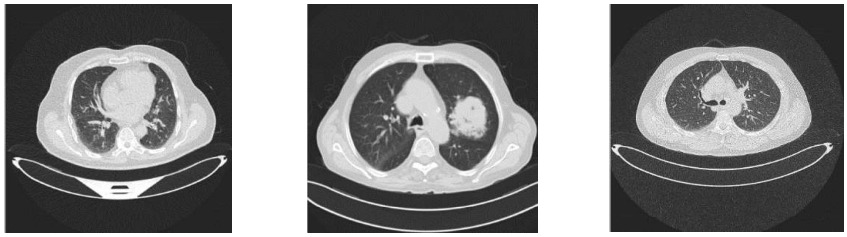
The PCA is used in the current research as a dimensionality reduction method in order to project the extracted features into a lower dimensional space with the highest level of significant variance [16–18]. Once the texture features have been extracted with the help of LBP, PCA removes redundancy and enhances the computational efficiency, making the EfficientNet-B0 classifier process the most informative features.

3.5 Enhanced TransUNet++

The suggested segmentation model is based on Enhanced TransUNet++, which is a hybrid architecture that combines CNN-based feature extraction and Transformer-based global context modeling [19–20]. EfficientNet-B0 backbone is used as the encoder to extract rich feature maps in an efficient manner and there are dense skip connections between UNet++ to enhance flow of gradients and reuse of features. The parameters chosen to balance between the segmentation accuracy and the computational feasibility are summarized in Table 2.

Table 2 Parameter specification of TransUNET++.

Component	Specification
Backbone	Vision Transformer (ViT-B/16) pre-trained on ImageNet21k
Patch size	16 × 16
Number of Patches	196
Embedding dimension	768
Number of attention heads	12
Dropout rate	0.1
Skip connections	Multi-scale from UNet++ encoder blocks
Decoder architecture	UNet++ with hierarchical feature fusion
Loss function	Dice loss + cross-entropy loss
Optimizer	AdamW



(a) (b) (c)

Fig. 2 Sample images from the dataset: (a) benign, (b) malignant, (c) normal.

(a) (b) (c)

Fig. 3 Sample segmented image: (a) original image, (b) edgemap, (c) predicted mask.

4 Results and Discussions

The proposed framework was implemented in Python on the Google Colab platform using an NVIDIA Tesla T4 GPU (16 GB VRAM), an Intel Xeon CPU, and 13 GB RAM. The software environment included Python 3.10.11, TensorFlow 2.7.0, Keras 2.6.0, and supporting libraries such as OpenCV, Scikit-learn, and Matplotlib.

4.1 Dataset

The IQ-OTH/NCCD lung cancer dataset was collected at the Iraq Oncology Teaching Hospital/National Center for Cancer Diseases during the fall of 2019. It consists of CT scan slices from 110 subjects, including cancer patients at different stages and healthy individuals, totaling 1,190 images [21]. The dataset is categorized into three classes: normal (55 cases), benign (15 cases), and malignant (40 cases). All scans were acquired in DICOM format using a Siemens SOMATOM CT scanner with a protocol of 120 kV, 1 mm slice thickness, window width of 350–1200 HU, window center of 50–600 HU, and breath-hold at full inspiration. The dataset was split into training, validation, and testing sets in a 70:15:15 ratio, ensuring patient-level separation to avoid data leakage. Fig. 2 presents sample images from the dataset.

4.2 Performance Metrics

The proposed hybrid EfficientNet-B0-TransUNet++ model was tested on the basis of common metrics of classification and segmentation. To be classified, the metrics used to determine the capability of the model to identify the benign, malignant, and normal cases correctly were accuracy [22], sensitivity, specificity, precision, and F_1-score. To perform segmentation, the measures of Dice similarity coefficient (DSC), Intersection over Union (IoU) [23], Hausdorff distance (HD) [24], volume overlap error (VOE) [25], and mean surface distance (MSD) [26] were used to measure the overlap and spatial consistency between predicated masks and ground truth.

Fig. 3 illustrates a sample segmentation result obtained using UNET++. Table 3 summarizes the class-wise performance of the proposed classification and segmentation framework across the benign, malignant, and normal classes. The proposed method demonstrates consistently high classification accuracy, achieving 98.12% for benign cases and 98.00% for both malignant and normal cases. It reports the class-wise performance of the stand-alone EfficientNet-B0-based classification model, with parameters including input size 224 × 224, batch size 16, learning rate 1×10^{-4}, and training for 50 epochs using the Adam optimizer.

Table 3 Class-wise performance of the proposed EfficientNet-B0-based classification model.

Performance	Benign	Malignant	Normal
Accuracy, %	98.12	98.00	98.00
Sensitivity, %	97.80	98.20	98.10
Specificity, %	98.40	97.90	98.20
Precision, %	98.00	97.85	97.90
F_1-score, %	97.90	98.02	98.00

Table 4 Performance of classification and segmentation.

Classification		Segmentation	
Accuracy, %	98.58	DSC	99.53
Sensitivity, %	98.47	IoU	98.56
Specificity, %	99.23	HD	99.73
Precision, %	98.42	VOE	98.86
F_1-score, %	99.01	MSD	99.60

Table 5 Comparative performance of proposed method.

Methods	Accuracy	Sensitivity	Speciificty	Precision
U-Net [1]	96.82	96.33	96.91	97.35
U-Net [2]	97.85	96.69	97.24	97.96
CNN [3]	96.00	96.00	97.00	97.00
VGG16 [21]	96.00	96.00	98.00	93.00
This work	98.58	98.47	99.23	98.42

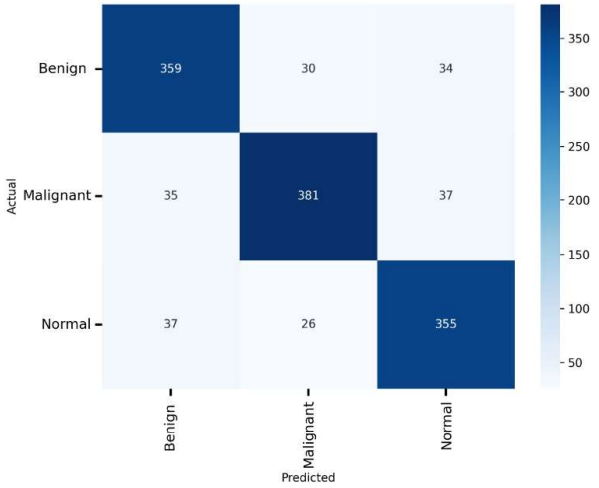


Fig. 4 Confusion matrix of proposed method.

Table 4 summarizes the performance of the proposed method for classification and segmentation. The model achieved high classification accuracy (98.58%), sensitivity (98.47%), specificity (99.23%), precision (98.42%), and F_1-score (99.01%). For segmentation, it obtained a Dice score of 99.53% and an IoU of 98.56%, demonstrating strong agreement with the ground truth.

Table 5 demonstrates comparative performance of proposed method. Fig. 4 illustrates the confusion matrix for the proposed classification method.

5 Conclusion

This paper proposes a robust hybrid deep learning framework for lung cancer classification and segmentation. The pipeline begins with image preprocessing, including resizing, intensity normalization, and noise removal using edge detection and median filtering. Texture features are extracted using LBP, while PCA reduces dimensionality and retains discriminative information. The optimized features are classified using the lightweight EfficientNet-B0 model, achieving a favorable accuracy–efficiency trade-off. For segmentation, a TransUNET++ integrates convolutional feature extraction with self-attention to capture both local details and global contextual dependencies, enabling accurate delineation of complex anatomical structures and pathological regions.

Disclosures

The authors declare that they have no conflicts of interest or competing financial interests related to this work.

References

1. A. Gupta, A. Kumar, and K. Rautela, "UDCT: lung cancer detection and classification using U-Net and DARTS for medical CT images," Multimedia Tools and Applications 84, 19065–19085 (2025).
2. R. Deepika, P. Shanmugam, K. Moorthi, P. M. Kumar, and S. L. Swarna, "Optimized transfer learning model for lung cancer stage classification using computed tomography images," Proceedings of International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), IEEE, 912–917 (2024).
3. M. R. Islam, M. N. H. Sarker, M. K. E. S. Miah, and M. H. Hossain, "Enhancing lung cancer screening with an automated deep learning system: a resource-efficient approach," International Journal of Imaging Systems and Technology 35(6), e70225 (2025).
4. S. Nageswaran, G. Arunkumar, A. K. Bisht, S. Mewada, J. S. Kumar, M. Jawarneh, and E. Asenso, "Lung cancer classification and prediction using machine learning and image processing," BioMed Research International 2022, 1755460 (2022).
5. B. Dunn, M. Pierobon, and Q. Wei, "Automated classification of lung cancer subtypes using deep learning and CT-scan based radiomic analysis," Bioengineering 10(6), 690 (2023).
6. A. Shimazaki, D. Ueda, A. Choppin, A. Yamamoto, T. Honjo, Y. Shimahara, and Y. Miki, "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," Scientific Reports 12(1), 727 (2022).
7. M. V. Shetty, S. Tunga, "Optimized deformable model-based segmentation and deep learning for lung cancer classification," The Journal of Medical Investigation 69(3-4), 244–255 (2022).
8. A. K. Ajai, A. Anitha, "Clustering-based lung lobe segmentation and optimization-based lung cancer classification using CT images," Biomedical Signal Processing and Control 78, 103986 (2022).

9. I. Naseer, S. Akram, T. Masood, M. Rashid, and A. Jaffar, "Lung cancer classification using modified U-Net-based lobe segmentation and nodule detection," IEEE Access 11, 60279–60291 (2023).
10. I. Shafi, S. Din, A. Khan, I. D. L. T. Díez, R. J. P. Casanova, K. T. Pifarre, and I. Ashraf, "An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network," Cancers 14(21), 5457 (2022).
11. B. Kalpana, S. K. Kannaiah, "Enhancing lung cancer prediction with EfficientNet and stochastic optimization," Proceedings of 4th International Conference on Computer, Communication, Control & Information Technology (C3IT), IEEE (2024).
12. Y. Said, A. A. Alsheikhy, T. Shawly, and H. Lahza, "Medical image segmentation for lung cancer diagnosis based on deep learning architectures," Diagnostics 13(3), 546 (2023).
13. A. A. Nafea, M. S. Ibrahim, M. M. Shwaysh, K. Abdul-Kadhim, H. R. Almamoori, and M. M. AL-Ani, "A deep learning algorithm for lung cancer detection using EfficientNet-B3," Wasit Journal of Computer and Mathematics Science 2(4), 68–76 (2023).
14. R. Sun, Y. Pang, and W. Li, "Efficient lung cancer image classification and segmentation algorithm based on an improved Swin Transformer," Electronics 12(4), 1024 (2023).
15. P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," Neural Computing and Applications 34, 9579–9592 (2022).
16. S. Ram, W. Tang, A. J. Bell, R. Pal, C. Spencer, A. Buschhaus, C. R. Hatt, M. P. diMagliano, A. Rehemtulla, J. J. Rodríguez, S. Galban, and C. J. Galban, "Lung cancer lesion detection in histopathology images using graph-based sparse PCA network," Neoplasia 42, 100911 (2023).
17. A. Alsadoon, G. Al-Naymat, A. H. Osman, B. Alsinglawi, M. Maabreh, M. R. Islam, "DFCV: a framework for evaluating deep learning in early detection and classification of lung cancer," Multimedia Tools and Applications 82(28), 44387–44430 (2023).
18. S. Al-Ofary, H. O. Ilhan, "Classification of PCA-based reduced deep features by SVM for diagnosing lung and colon cancer," Proceedings of 5th International Congress Human-Computer Interaction, Optimization and Robotic Applications (HORA), IEEE (2023).
19. Y. He, J. Yun, J. Li, Y. Wang, and J. Xu, "Multicenter deep learning study for lung tumor subtype segmentation in CT images using Swin-UNETR, nnU-Net, and TransUNet," Journal of Radiation Research and Applied Sciences 18(3), 101721 (2025).
20. J. K. Gupta, J. Kaur, "Enhancing lung cancer diagnosis: Edge-TransUNet for segmentation and Lightweight-MLCAN for classification," IETE Journal of Research, 1–10 (2025).
21. S. Hamza-Cherif, T. Tariq, Z. A. Elaouaber, and M. Mohammed, "Explainable lung cancer classification using VGG16 and Grad-CAM," Proceedings of International Conference on Artificial Intelligence, Smart Technologies and Communications (AISTC 2025), Atlantis Press, 5–11 (2025).
22. B. He, W. Hu, K. Zhang, S. Yuan, X. Han, C. Su, J. Zhao, G. Wang, G. Wang, and L. Zhang, "Image segmentation algorithm of lung cancer based on neural network model," Expert Systems 39(3), e12822 (2022).
23. S. B. Shuvo, An automated end-to-end deep learning-based framework for lung cancer diagnosis by detecting and classifying lung nodules, arXiv preprint arXiv:2305.00046v1 (2023).
24. A. R. Bushara, "A deep learning-based lung cancer classification of CT images using augmented convolutional neural networks," ELCVIA 21(1), 130–142 (2022).
25. S. Agarwal, S. Thakur, and A. Chaudhary, "Prediction of lung cancer using machine learning techniques and their comparative analysis," Proceedings of 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, 1–5 (2022).
26. B. Deepesh, T. Latha, "FPGA-based MRI brain tumor segmentation using modified FCM method," Circuits, Systems, and Signal Processing 44, 1–20 (2025).

Individual Behavior Recognition in Laboratory Rats: a Comparative Analysis of Computer Vision Methods

Dmitrii Krasnov[1*], Aleksey Shmonin[2,3], Maxim Volynsky[1], Nikita Margaryants[1,4], Alexandr Gusev[1], Maria Maltseva[3], and Elena Korotkova[3]

[1] Technical Vision Lab, ITMO University, 49 Kronverksky pr., Saint-Petersburg 197101, Russian Federation
[2] Department of Psychology, Pedagogy and Rehabilitation Technologies, Federal Scientific and Clinical Center for Resuscitation and Rehabilitation, 777, build. 1 Village Lytkino, Moscow region, Solnechnogorsk 141534, Russian Federation
[3] Department of Pathophysiology with a course in Clinical Pathophysiology, Pavlov First Saint Petersburg State Medical University, 6–8 L'va Tolstogo str., Saint-Petersburg 197022, Russian Federation
[4] Department of Physics, Math, and Informatics, Pavlov First Saint Petersburg State Medical University, 6–8 L'va Tolstogo str., Saint-Petersburg 197022, Russian Federation
*e-mail: dmitriy_krasnov@outlook.com

Abstract. Computer-vision methods have been applied to automated behavior recognition in laboratory rodents. Namely, we explored the possibility of classifying certain behavior classes from still images; compared keypoint-based methods with approaches based on visual embeddings; studied the feasibility of transferring models between rats and mice; and evaluated the relevance of the number and the accuracy of the detected keypoints. We collected a dataset of a freely moving Wistar rat to train six pose-based classifiers with Long Short-Term Memory (LSTM) using six sets of keypoints produced by two detectors and four Convolutional Neural Network (CNN) classifiers using images and optical flow frames. The results demonstrated the highest mean average precision (mAP) of 65.7% for the CNN-based methods and 34.3% for the LSTM classifiers, the feasibility of recognizing visually distinct classes (rearing and body grooming) from still images, and the applicability of a keypoint detector trained on mice. The results of this study can be applied to the design of a computer vision system for automating long-term monitoring of laboratory rodent behavior.

Keywords: computer vision; deep learning; machine learning; rat behavior; rodent behavior; action recognition; keypoints; convolutional neural network; Wistar rat.

Paper #9476 received 15 Jan 2026; accepted for publication 19 Feb 2026; published online 25 Mar 2026. doi: 10.18287/JBPE26.12.010306.

1 Introduction

Analysis of animal behavior through video monitoring has become an important new task in neuroscience [1], since pharmaceuticals, physical environment, living conditions, diseases, and injuries tend to affect animal behavior. Such influence is typically evaluated via behavioral tests, including tests for isolated assessment of a specific skill and long-term monitoring.

Tests for the isolated assessment of a skill or specific behavioral acts, such as stress, anxiety, depression, or learned helplessness, are widely used in animal behavioral research [2]. In these tests, animals are placed in specific conditions for a short period, usually a few minutes. Specific behavioral forms are recorded, based on which conclusions about the presence or absence of target symptoms and deviations are drawn. Such analysis is possible in a "manual mode", where the researcher manually counts the required behavioral acts [2].

Long-term monitoring of animals – for several hours – is necessary for assessing rare behavioral activities, observing general locomotor activity (walking and running distance, duration of sitting and sleeping), and monitoring interaction between males and

females [3]. As long-term monitoring requires analyzing hours of video and often involves a behavior analysis expert, it necessitates automation via modern computer vision methods [3].

Long-term monitoring allows for the assessment of disability in experiments. Disability is a limitation of an individual's ability to participate in normal life and perform everyday tasks. An individual with a disability does not use certain behavioral patterns, for example sleeps a lot, moves less, does not communicate, drinks and eats less or more. Analyzing the quantity, quality, and sequence of actions, can solve the following tasks:

- Creating a disability profile in an experiment, where the researcher can analyze which behavioral forms are impaired in each disease and prescribe drugs, rehabilitation technologies, and environmental intervention options to restore the individual's previous activity level, in other words, to eliminate the disability;
- Monitoring an animal in a cage at home or in a zoo to detect signs of pathology based on behavior and provide necessary recommendations [4];
- Investigating psychotropic drugs that can alter animal behavior to evaluate long-term effects and changes in individual and social behavior.

Rodents, such as rats and mice, are appropriate candidates for behavioral experiments, since they are social animals living in families, on which various diseases can be modeled using pharmacological, genetic, and surgical experimental models.

Currently, there are many approaches for recognizing the behavior patterns of laboratory rodents, based on spatiotemporal information obtained from video sequences [5, 6]. Most of them are based on analyzing features extracted from key points, reflecting their spatial configuration and change over time. Predictions can be made using either heuristic methods [7] or a trained classifier [8]. However, this approach requires a significant amount of data not only for training the classifier but also for training the keypoint detector, as the robustness and generalization capability of the entire pipeline depend on it [9].

Another drawback of this approach is the narrow specialization of existing open-source and commercial solutions. This can manifest in a specific type of animal (mouse or rat), color (white, black, gray, brown), or number of species. For example, Segalin et al. [10] presents a framework for recognizing social interactions between a white and a black mouse. Hsu and Yttri [11] show a pipeline for monitoring the behavior of a spotted dark rat. Differences can also be in camera angles (top-down or bottom-up) or in the chosen keypoints for describing the rodent skeleton. Such variability can lead to difficulties in scaling the particular approach to other animals, for example Wistar rats [12].

A further review of the literature revealed that existing studies have not examined how the complexity of the skeletal model affects classification accuracy when evaluated on the same data. This is an important aspect of pose-based methods, as certain skeletal keypoints (hind limbs and forelimbs) are typically detected less accurately than others (nose, tail base, spine) due to more active motion and frequent occlusions. However, these keypoints may be critical for recognizing complex behaviors such as scratching, face grooming, and body grooming.

Another unexplored question is the potential for directly using open-source mouse datasets to address rat behavior recognition tasks. Despite the visible similarity between rats and mice, they differ in body size and shape, as well as in individual and social behaviors. Moreover, the color of animals in open-source datasets may differ substantially from that of individuals in the target task, which can also affect behavior recognition accuracy. Thus, open-source datasets may be effectively used to train robust species and keypoint detectors with strong generalization capability due to their large size and high variability.

At the same time, methods for human action recognition commonly include convolutional neural networks (CNN) with Conv3D layers [13], hybrid architectures for processing frames and optical flow [14], or visual-language models [15]. It can be assumed that methods that successfully classify human activities are capable of recognizing the behavior of laboratory rodents. However, relatively few existing studies focused on adapting human action-recognition methods to this task. Notable examples include studies [16, 17], in which the authors used YOLO-family convolutional networks for localizing and classifying individual behaviors, as well as work [18], where visual embeddings extracted with I3D [19] were used for classification. However, none of these studies performed a direct comparison between pose-based and CNN-based methods.

Thus, the aim of this work is to investigate several important aspects of applying computer-vision methods to automate behavioral analysis in Wistar rats. Namely, the possibility of classifying certain behavior classes from still images; the comparison keypoint-based methods with approaches based on visual embeddings; the feasibility of transferring models between rats and mice; and the relevance of the number and the accuracy of the detected keypoints.

2 Methods and Materials

2.1 Animal Housing

The experiments used healthy Wistar rats weighing 250–350 g from the "Rappolovo" nursery breeding (Leningrad region, Russian Federation). The animals were housed in a 12-h light/12-h dark cycle with 5 rats per cage and free access to food and water. Behavioral analysis was conducted in the daytime under ambient lighting. During video recording the rats were housed individually with free access to food and water, in the behavioral testing boxes with wood shavings and no researchers in the room. The procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and approved by the local Ethics Committee.

2.2 Data Collection and Processing

The video was recorded using an IDIS DC-Z1163 camera (IDIS, South Korea) with 1920×1080 resolution and 30 frames per second frame rate. The camera was positioned above two transparent plastic boxes measuring 120×60×40 cm³ at a distance of 3 m from the box lid. This setup resulted in one box occupying an area of approximately 1060×540 pixels in the frame, with a rat occupying about 5% of the box's area.

Two datasets were compiled and annotated based on video recordings of rat behavior, and the third one was collected from open sources. The first dataset consisted of 3400 images labelled with keypoints that were approved by experts in rat behavior from the Department of Pathophysiology, First Pavlov State Medical University of St. Petersburg (Russia). Fig. 1 shows an example of the annotated image: (1) nose; (2 and 3) right and left eye; (5 and 6) right and left ear; (4) head center; (7) forelimb girdle center; (8) spine center; (9 and 10) right and left forelimb; (11) tail base; (12 and 13) right and left hind limb [20, 21].

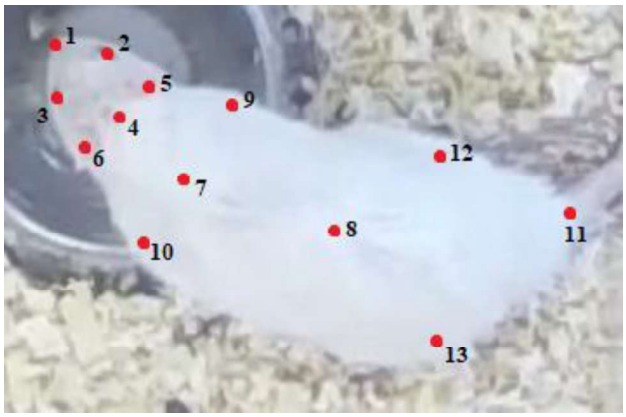


Fig. 1 Annotated image.

The second dataset consisted of video recordings of rat behavior patterns, where each recording was assigned to one of six classes: sniffing, rearing, eating, body grooming, face grooming, and scratching. The dataset contained 158 video recordings with a total duration of 43 min. A qualified behavior expert selected the specified behavioral forms and trimmed the videos from the start to the completion of the behavioral act.

To analyze the scalability of the approaches across the species of laboratory rodents, a dataset from open sources [10, 22] was also compiled. This dataset contained 14000 images of laboratory mice with annotated skeletal keypoints: nose, left and right ear, neck, spine center, left and right hip, tail base.

2.3 Deep Learning Approaches

For comparison, behavior patterns were recognized with two main approaches: a neural network classifier based on keypoint features and CNNs for processing raw frames and optical flow frames.

A long short-term memory (LSTM) block (two layers, sequence length 32, 256 hidden neurons) was used as the neural network classifier. A sequence of keypoint feature vectors containing the coordinates of selected points, mutual Euclidean distances and angles between them, and their displacements between adjacent frames was fed as the input for the classifier.

A single RGB frame and a sequence of optical flow frames were processed with the ResNet50 convolutional network [23]; for the RGB frames sequence, the version of ResNet18 with Conv3D layers [24] was employed. The optical flow frames were calculated using the dense Farneback algorithm [25] implemented in the OpenCV library.

To sum up, three CNNs were trained: for the RGB frame, for the sequence of 10 optical flow frames, and for the sequence of 32 RGB frames (Conv3D). Additionally, to analyze the possibility of fusing features from RGB frames and optical flow, an approach similar to the one in Ref. [14] was implemented, where a trainable 1×1 convolutional layer was used for the weighted sum of feature maps obtained from the frozen CNNs previously trained on RGB frames and optical flow frames.

2.4 Design of Experiments

The first approach was evaluated using 6 sets of keypoints in the experiments with 2 the above datasets (Fig. 2):

1) Open-source dataset: nose, neck, spine center, tail base;

2) Open-source dataset: nose, right and left ear, neck, spine center, tail base;

3) Open-source dataset: nose, right and left ear, neck, spine center, right and left hip, tail base;

4) Our dataset: nose, head center, forelimb girdle center, spine center, tail base;

5) Our dataset: nose, right and left eye, head center, right and left ear, forelimb girdle center, spine center, tail base;

6) Our dataset: nose, right and left eye, head center, right and left ear, neck, spine center, right and left forelimb, tail base, right and left hind limb.

The main idea behind this design was to assess the impact of the number and detection accuracy of keypoints on the accuracy of behavior class determination, since points along the spine and on the head are detected with higher accuracy [20].

For this purpose, an RTMDet-s [26] rat detector was trained for the spatial localization of the rat, and two HRNet [27] skeletal keypoint detectors were trained for the open-source dataset and for our dataset, respectively. To train the LSTM classifiers for each of the 6 keypoint sets, the predictions from the rat detector and the keypoint detector for each video frame were collected and stored locally. This step accelerated the training process since keypoint coordinates for each video only needed to be computed once.

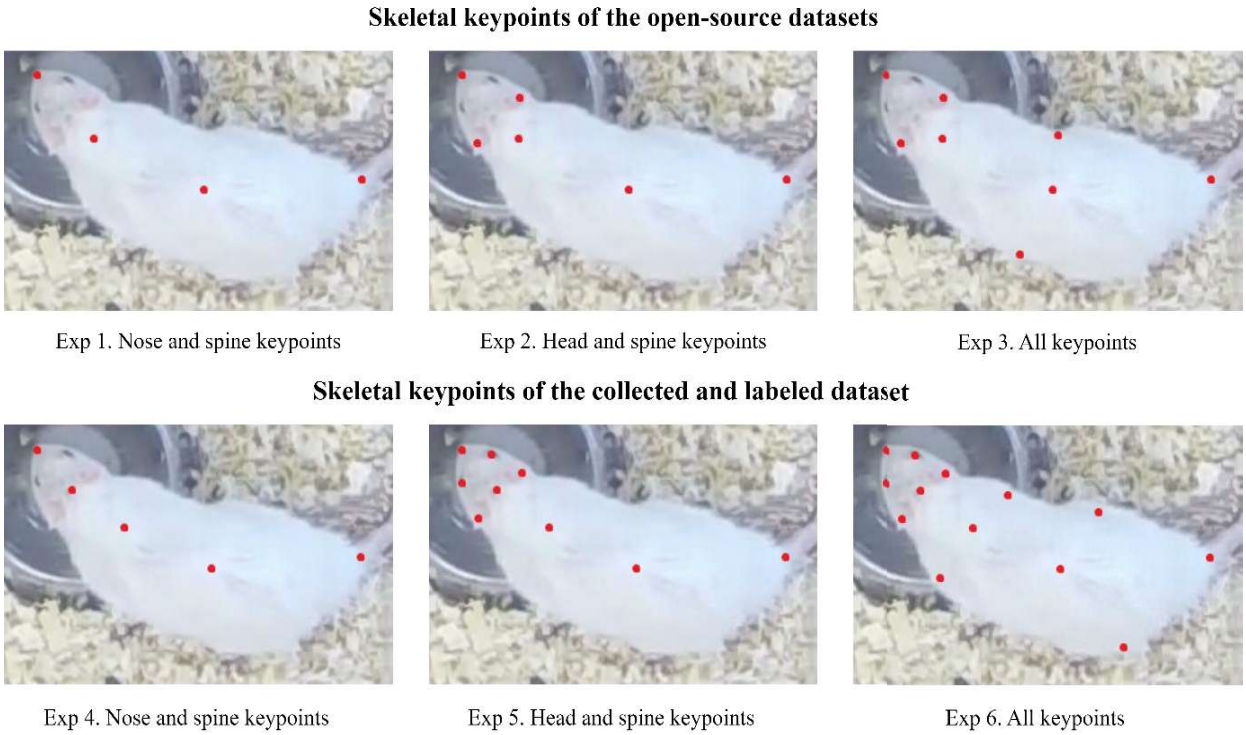


Fig. 2 Six sets of skeletal keypoints used in this work.

A similar step was performed during training the CNNs that require the use of optical flow. For each video, the optical flow was calculated using OpenCV, and its frames were saved for use in training. To evaluate the second approach, 4 CNNs were trained: for the RGB frame, for the sequence of 10 optical flow frames, for the sequence of 32 RGB frames (Conv3D), and for the fused RGB and optical flow features. All input frames were cropped according to the predicted rat bounding rectangle.

All models were trained on an NVIDIA GeForce RTX 3090 24GB for 100 epochs with a batch size of 32, Adam optimizer with learning rate of 0.00001. The models were evaluated using the Average Precision (AP) metric for each class and the mean value across all classes (mAP). The AP metric is calculated as the area under the Precision-Recall curve and indicates the overall quality of the classifier across all possible thresholds at which a class is considered correctly predicted.

3 Results and Discussion

This study examined some important aspects of recognizing individual behavior in laboratory rats: the applicability of human-oriented CNN-based action-recognition methods; the impact of both the number and the accuracy of the detected keypoints; and the transferability of detectors trained on mice. We trained six LSTM classifiers using different sets of keypoints produced by different detectors and four CNN classifiers. All models were trained on a dataset of video recordings of individual behavior in Wistar rats, that was collected in this work. The evaluation results are summarized in Table 1 as the AP for each class and the overall mAP.

Table 1 shows that the methods based on the skeletal keypoints (LSTM 1-6) scored as much as 47.8% lower in mAP than those based on the frame or a sequence of frames embeddings. This can be explained by the fact that the trained keypoint detector was not sufficiently robust for all observed rat poses. In contrast, CNN models do not require any complex or precise keypoint annotation, relying instead on the frame's visual features.

Among the CNN models, the fusion of RGB frame features and optical flow achieved the highest result (65.7% mAP), which agrees with the findings in Ref. [14]. We also observed that some behavior patterns, such as rearing or body grooming, were easier to be classified based on the single frame due to the distinct differences in the rat poses, which could be identified visually even in the still image. Moreover, the results demonstrated that the Conv2D-based ResNet achieved higher accuracy (62.8% mAP) than its Conv3D-based counterpart (61.1% mAP), which may indicate both that still-image embeddings were sufficient for successful action recognition and that temporal features were not used effectively in Conv3D layers. An absence of the distinct differences in the rat poses may also explain the lower metrics for the scratching, face grooming, sniffing, and eating classes, since the rat postures during these behaviors are very similar. Furthermore, the above classes accounted for a smaller total video duration. Although the CNNs trained in this work were previously reported as methods for recognizing human behavioral activities, our findings showed that they can also be applied to laboratory rodents with the more commonly used keypoint-based approaches.

Table 1 Classifiers evaluation results.

Experiment	AP for each class, %						mAP, %
	Rearing	Body grooming	Scratching	Face grooming	Sniffing	Eating	
LSTM Experiment 1	14.5	27.3	29.0	7.1	7.9	15.8	16.9
LSTM Experiment 2	34.9	29.4	17.4	21.1	13.1	40.4	26.0
LSTM Experiment 3	39.1	26.8	23.2	6.7	29.5	46.9	28.7
LSTM Experiment 4	12.6	51.3	40.9	7.5	17.4	13.3	23.8
LSTM Experiment 5	13.7	43.9	27.0	10.4	32.0	27.7	25.8
LSTM Experiment 6	30.1	58.0	26.8	29.8	33.1	28.0	34.3
ResNet50 RGB	82.2	76.1	59.6	58.8	46.6	53.6	62.8
ResNet18 Conv3D RGB	79.4	67.1	37.5	81.5	55.6	45.8	61.1
ResNet50 Optical Flow	67.9	71.5	68.1	43.6	40.3	62.6	59.0
ResNet50 RGB + Optical Flow	73.7	80.7	69.8	58.5	43.1	68.6	65.7

According to Table 1, the first group of LSTM experiments (1–3) tends to show similar results as the second group (4–6) – 28.7% and 34.3% mAP respectively. As mentioned above, 1–3 LSTM experiments used a keypoint detector trained on open-source datasets containing images mostly of mice, whereas 4–6 used the detector trained on the rat images collected in this work. These results suggest that trained detectors may be transferable across different rodent species (mice or rats) without any substantial loss in accuracy. Within each group, mAP increased as new keypoints were added (as the skeletal model became more detailed), even though the detector recognized the new points (right and left hip, right and left forelimb) with lower accuracy than the basic ones (nose, neck, spine center, tail base). This may indicate that a well-designed skeletal model played an important role in behavior-prediction accuracy, which slightly differs from the findings reported in Ref. [9].

Despite relying on a relatively old CNN backbone (ResNet, 2015) and a modest amount of data (a total video duration of 43 min), this study achieved metrics that are high enough for practical use (65.7% mAP). In addition, it may provide a valuable comparison of action recognition approaches in the perspective of classifying individual behaviors in Wistar rats, the results of which may be applied to long-term behavioral monitoring.

4 Conclusion

In this work, CNN-based and keypoint-based action-recognition approaches for identifying individual behaviors of Wistar rats were compared on a single dataset. The effects of the number of keypoints and the use of a detector trained on mice on classification accuracy were evaluated, and the feasibility of classifying behavior from a static crop using a CNN was assessed. The experimental results showed that CNNs were able to classify behavior from image embeddings with higher accuracy (65.7% mAP), and that classes such as rearing, body grooming, and scratching could be recognized with 61.1% mAP using a still image. For keypoint-based classifiers, the results indicated that a keypoint detector trained on mice images can be used without any substantial loss of accuracy for practical applications in rats (28.7% mAP and 34.3% mAP), and that a well-designed keypoint model is essential for robust behavior monitoring. This has clear implications for long-term behavior monitoring in laboratory rats, and suggests that publicly available keypoints datasets (mostly with mice images) can improve recognition accuracy in the conditions of small amounts of data, and that image embeddings are also a valuable source of information for this task.

Some questions remain unaddressed; for example, more recent feature extraction backbones – rather than only ResNet50 – could be explored, as well as some alternative methods for processing sequences of feature vectors, such as transformers or Conv1D CNNs. Since this study focused on a comparative analysis of approaches rather than on optimizing a specific pipeline, these aspects remain open for future investigation. Despite this, the study achieved the accuracy levels suitable for practical use (65.7% mAP). However, they

could be improved by incorporating some novel methods in further work. Further research may also include expanding the dataset, adding new classes of individual behaviors such as nesting and drinking, and introducing social interaction to enable comprehensive behavioral analysis.

The results of this study can be applied to the design of a computer vision system for automating long-term monitoring of laboratory animal behavior. The findings demonstrate the importance of using image embeddings for behavior recognition, even when a powerful tool such as skeletal keypoints is available. The work may also help address the problem of the small amount of training data for keypoint detectors training and the challenges associated with annotating such datasets by suggesting the use of publicly available data in spite of the differences between rats and mice in behavioral and apparent aspects.

Disclosures

All authors declare that there is no conflict of interest in this paper.

Data and Code Availability

Data and code used in this work are available upon reasonable request from the corresponding author (Dmitrii Krasnov, dmitriy_krasnov@outlook.com).

References

1. F. Rotaru, S. I. Bejinariu, H. N. Costin, R. Luca, and C. D. Nita, "Captive animal behavior study by video analysis," Sensors 23(18), 7928 (2023).
2. Himanshu, Dharmila, D. Sarkar, and Nutan, "A review of behavioral tests to evaluate different types of anxiety and anti-anxiety effects," Clinical Psychopharmacology and Neuroscience: The Official Scientific Journal of the Korean College of Neuropsychopharmacology 18(3), 341–351 (2020).
3. F. Grieco, B. J. Bernstein, B. Biemans, et al., "Measuring behavior in the home cage: Study design, applications, challenges, and perspectives," Frontiers in Behavioral Neuroscience 15, 735387 (2021).
4. K. Luxem, J. J. Sun, S. P. Bradley, K. Krishnan, E. Yttri, J. Zimmermann, T. D. Pereira, and M. Laubach, "Open-source tools for behavioral video analysis: Setup, methods, and best practices," eLife 12, e79305 (2023).
5. M. Perez and C. Toler-Franklin, "CNN-based action recognition and pose estimation for classifying animal behavior from videos: A survey," arXiv:2301.06187 (2023).
6. S. İşık, G. Ünal, "Open-source software for automated rodent behavioral analysis," Frontiers in Neuroscience 17, 1149027 (2023).
7. C. J. Gabriel, Z. Zeidler, B. Jin, C. Guo, C. M. Goodpaster, A. Q. Kashay, A. Wu, M. Delaney, J. Cheung, L. E. DiFazio, M. J. Sharpe, D. Aharoni, S. A. Wilke, and L. A. DeNardo, "BehaviorDEPOT is a simple, flexible tool for automated behavioral detection based on markerless pose tracking," eLife 11, e74314 (2022).
8. M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson, "JAABA: interactive machine learning for automatic annotation of animal behavior," Nature Methods 10(1), 64–67 (2013).
9. M. Lorbach, R. Poppe, E. A. Van Dam, L. P. J. J. Noldus, and R. C. Veltkamp, "Automated Recognition of Social Behavior in Rats: The Role of Feature Quality," in Image Analysis and Processing – ICIAP 2015, V. Murino and E. Puppo (Eds.), Springer International Publishing, 9280, 565–574 (2015).
10. C. Segalin, J. Williams, T. Karigo, M. Hui, M. Zelikowsky, J. J. Sun, P. Perona, D. J. Anderson, and A. Kennedy, "The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice," eLife 10, e63720 (2021).
11. A. I. Hsu, E. A. Yttri, "B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors," Nature Communications 12, 5188 (2021).
12. E. A. Van Dam, L. P. J. J. Noldus, and M. A. J. Van Gerven, "Deep learning improves automated rodent behavior recognition within a specific experimental setup," Journal of Neuroscience Methods 332, 108536 (2020).
13. J. Nie, L. Yan, X. Wang, and J. Chen, "A Novel 3D Convolutional Neural Network for Action Recognition in Infrared Videos," in 2021 4th International Conference on Information Communication and Signal Processing (ICICSP), IEEE, 420–424 (2021).
14. C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1933–1941 (2016).
15. Q. Wang, J. Du, K. Yan, and S. Ding, "Seeing in flowing: Adapting CLIP for action recognition with motion prompts learning," Proceedings of the 31st ACM International Conference on Multimedia (ACM), 5339–5347 (2023).
16. Z. Bai, Y. Li, Z. Ru, S. Li, W. Li, and F. Duan, "Rats Behavior Recognition Based on Improved YOLOv8," in 2024 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 1082–1086 (2024).
17. K. M. Hatton-Jones, C. Christie, T. A. Griffith, A. G. Smith, S. Naghipour, K. Robertson, J. S. Russell, J. N. Peart, J. P. Headrick, A. J. Cox, and E. F. Du Toit, "A YOLO based software for automated detection and analysis of rodent behaviour in the open field arena," Computers in Biology and Medicine 134, 104474 (2021).

18. N. G. Nguyen, D. Phan, F. R. Lumbanraja, M. R. Faisal, B. Abapihi, B. Purnama, M. K. Delimayanti, K. R. Mahmudah, M. Kubo, and K. Satou, "Applying Deep Learning Models to Mouse Behavior Recognition," Journal of Biomedical Science and Engineering 12(02), 183–196 (2019).
19. J. Carreira, A. Zisserman, "Quo Vadis, action recognition? A new model and the Kinetics dataset," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4724–4733 (2017).
20. D. I. Krasnov, M. A. Volynsky, and A. A. Gusev, "Preprocessing of skeletal keypoints trajectories in the task of laboratory animal behavior recording automation," Scientific and Technical Journal of Information Technologies, Mechanics and Optics 25(2), 295–302 (2025). [in Russian]
21. D. I. Krasnov, M. A. Volynsky, and A. A. Gusev, "State estimation of an agent with stochastic dynamics using recurrent filters," Informatics and Automation 24(5), 1355–1378 (2025). [in Russian]
22. N. L. Goodwin, J. J. Choong, S. Hwang, K. Pitts, L. Bloom, A. Islam, Y. Y. Zhang, E. R. Szelenyi, X. Tong, E. L. Newman, K. Miczek, H. R. Wright, R. J. McLaughlin, Z. C. Norville, N. Eshel, M. Heshmati, S. R. O. Nilsson, and S. A. Golden, "Simple Behavioral Analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience," Nature Neuroscience 27(7), 1411–1424 (2024).
23. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (2016).
24. K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3D residual networks for action recognition," Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), 3154–3160 (2017).
25. G. Farnebäck, "Two-Frame Motion Estimation Based on Polynomial Expansion," in Image Analysis, J. Bigun and T. Gustavsson (Eds.), Springer Berlin Heidelberg, 2749, 363–370 (2003).
26. C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "RTMDet: An empirical study of designing real-time object detectors," arXiv:2212.07784 (2022).
27. J. Huang, Z. Zhu, F. Guo, and G. Huang, "The devil is in the details: Delving into unbiased data processing for human pose estimation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5699–5708 (2020).