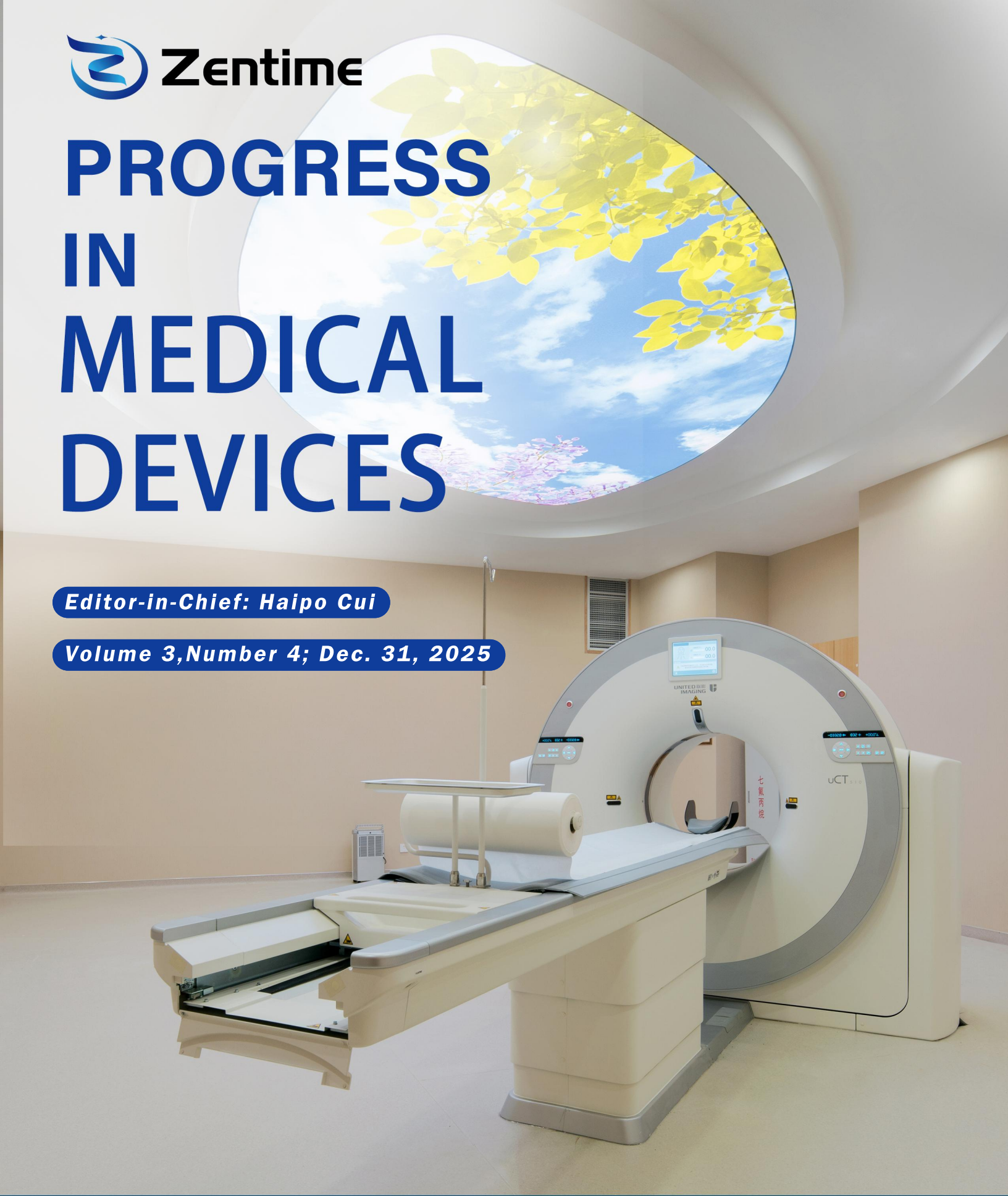
Zentime

PROGRESS IN MEDICAL DEVICES



Editor-in-Chief: Haipo Cui

Volume 3,Number 4; Dec. 31, 2025

ISSN: 2957-5478

Published by ZENTIME PUBLISHING CORPORATION

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/017119gmeubf

Detection of traumatic brain injury using eddy current damping technology

Wenjing Du, Rongguo Yan, Tingting Shi, Ke Wang, Shoucheng Chen

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Corresponding author: Rongguo Yan.

Declaration of conflict of interest: None.

Received February 21, 2025; Accepted April 16, 2025; Published December 31, 2025

Highlights

- This paper presents a new method for brain injury detection, which uses eddy current damping technology. This method is more convenient and faster than the traditional detection method.
- This paper simulates the method in COMSOL software and then verifies it through experiments. The feasibility of the method is proved through the combination of simulation and experiments.

Abstract

Objectives: To investigate the feasibility of using eddy current damping (ECD) technology as a portable and rapid alternative for detecting the location of cerebral hemorrhage. **Methods:** This study leveraged the significant conductivity difference between hemorrhagic and normal brain tissue. ECD technology involved the application of alternating current to a coil, generating a time-varying magnetic field that induces eddy currents in conductive tissues such as hemorrhagic regions. These eddy currents produced a secondary magnetic field that opposes the original, resulting in measurable changes in the ECD signal. Using COMSOL simulations, detailed models of the head, brain, hemorrhage, and coil were constructed to analyze electromagnetic coupling. Changes in coil resistance and inductance due to hemorrhagic tissue were quantified to evaluate the detectability of ECD signals. **Results:** Simulation results confirmed that cerebral hemorrhages generate measurable eddy current responses, which alter the resistance and inductance of the detection coil. These changes provide quantifiable ECD signals that can potentially indicate both the location and volume of hemorrhagic lesions. **Conclusions:** The findings demonstrate the feasibility of using ECD technology as a portable, rapid, and non-invasive method for detecting cerebral hemorrhage. Pending clinical validation, this approach could serve as a practical alternative to conventional imaging, particularly in emergency settings where speed and mobility are essential.

Keywords: Eddy current damping, traumatic brain injury, COMSOL multiphysics, finite element model of the human head

Introduction

Traumatic brain injury (TBI) refers to damage to brain tissue caused by direct or indirect external force to the head and is classified as a mechanical injury. In clinical practice, TBI not only affects brain tissue but can also lead to dysfunction of the nervous and endocrine systems, resulting in complications such as limb impairment, cognitive decline, and psychological disorders [1]. Studies have shown that accurate assessment of TBI severity plays a crucial role in clinical decision-making. Early identification of the extent of brain injury and the location of intracranial hemorrhage (ICH) allows for timely and targeted treatment interventions [2]. Therefore, rapid and accurate assessment is essential to improving survival

Address correspondence to: Rongguo Yan, Department of Biomedical Engineering, School of Health Science and Engineering, University of Shanghai for Science and Technology, No. 516 Jungong Road, Yangpu, Shanghai 200093, P. R. China. E-mail: yanrongguo@usst.edu.cn.

rates and prognosis in patients with acute brain injuries.

Magnetic resonance imaging (MRI) and computed tomography (CT) are currently the most widely used imaging modalities for brain injury diagnosis [3]. MRI utilizes the hydrogen nucleus as the imaging basis, capitalizing on the varying water content across different tissues. These differences in hydrogen concentration result in distinct MRI signals, allowing differentiation between tissue types [4]. Brain injuries such as edema and hemorrhage alter the water content and thus can be detected using MRI [5]. However, MRI is contraindicated for patients with magnetic metal implants or pacemakers due to the presence of a strong magnetic field [6]. Additionally, MRI is time-consuming, generates significant acoustic noise, and produces large amounts of data, limiting its use in portable or emergency settings [7].

CT imaging works by scanning a specific area of the body with X-rays over a defined period [8]. The captured data is processed by a computer based on differences in tissue absorption and attenuation of X-rays [9]. This produces cross-sectional and 3D images that help identify pathological changes [10]. Despite its widespread clinical use, CT has several drawbacks. It relies on ionizing radiation, posing risks to certain populations, especially pregnant women [11]. Additionally, CT scanners are expensive, and the cost of each scan is relatively high. CT also has limitations in qualitative assessments [12]. More importantly, both CT and MRI lack portability and are unsuitable for time-critical situations [13]. For patients with acute brain injury, the time required to reach a hospital, undergo scanning, and receive a diagnosis may result in missed treatment windows, leading to adverse outcomes [14]. Thus, there is an urgent need for a portable device capable of quickly localizing intracranial injuries.

The basic principle of eddy current damping (ECD) technology is that when an alternating current is applied to a conductor, it generates a time-varying magnetic field, according to Ampère's law [15]. This changing magnetic field induces an electric field as described by Faraday's law of electromagnetic induction, generating circulating currents—known as eddy currents—within conductive materials [16]. These eddy currents, in turn, produce a secondary magnetic field that opposes the original one, resulting in a damping effect. Several international research institutions have explored the application of ECD in brain injury detection, demonstrating its feasibility through simulations, system prototyping, and experimental validation against conventional imaging techniques [17-19]. However, research in this area remains limited in China, and many existing simulation studies lack depth or precision.

In this study, we apply ECD technology to simulate brain injury detection using COMSOL Multiphysics. The basic concept involves generating a magnetic field from an alternating current (AC)-driven copper coil, with the brain modeled as a conductive medium. When this field interacts with hemorrhagic brain tissue, eddy currents are induced, leading to detectable changes in the coil's inductance and resistance. These changes are quantifiable and can be used to infer the injury's location. A simplified physical model of the human head was constructed, and sensor measurements confirmed that the equivalent resistance and inductance varied as expected in response to hemorrhagic lesions.

Compared with conventional imaging methods such as CT and MRI, this approach offers significant advantages. First, the portable design allows rapid detection in community clinics, at home, or even in pre-hospital emergency settings, breaking the reliance on fixed hospital equipment. Second, the electromagnetic-based detection system does not require expensive imaging devices or contrast agents, thereby reducing overall medical costs. These features highlight the high clinical value and potential of ECD technology for brain injury assessment.

Materials and methods

COMSOL simulation

Theoretical analysis

In the COMSOL Multiphysics environment, a numerical model was developed to simulate ECD for cerebral hemorrhage detection. An anatomically accurate head model was constructed and integrated with an electromagnetic coil. Tissue-specific bioelectromagnetic parameters were assigned to different anatomical regions. Finite element analysis was used to examine the spatial distribution and

phase shift of eddy currents induced by the alternating magnetic field in hemorrhagic regions. Quantitative analysis of the perturbations in the electromagnetic field relative to hematoma volume confirmed the theoretical feasibility of ECD as a non-invasive method for detecting cerebral hemorrhage.

Normal brain tissue primarily consists of high-fat myelin, with a conductivity of approximately 0.2 S/m, while blood, rich in dissolved ions and charged proteins, has a conductivity of about 0.65 S/m [20]. Due to this significant difference, blood is considerably more conductive than brain tissue. When cerebral hemorrhage occurs, the localized change in conductivity leads to detectable eddy current phenomena under an applied magnetic field. These eddy currents generate secondary magnetic fields that oppose the original field, resulting in increased coil resistance and decreased inductance. By comparing the coil's resistance and inductance before and after measurement, the ECD signal can be quantified to identify the location of the hemorrhage.

Currently, magnetic scanners capable of operating at the micro-Tesla level are not widely used in clinical settings [21]. If it can be demonstrated that a magnetic field at this intensity induces a detectable eddy current in hemorrhagic tissue, it would support the feasibility of using ECD for precise localization of brain injuries at low field strengths.

The AC excitation frequency applied to the coil is set to 1 MHz. This frequency was selected based on several key considerations. First, the LDC1101 inductance-to-digital converter chip used in this study operates within the 500 kHz to 10 MHz range [19]. A frequency of 1 MHz lies near the midpoint of this range, allowing for stable and optimal performance. Furthermore, Shahristani et al. demonstrated that sensor coils resonate around 1 MHz when detecting brain tissue, suggesting this frequency aligns with the brain's natural electromagnetic resonance [19]. Operating at this resonant frequency maximizes the amplitude of the ECD signal, thereby improving detection sensitivity.

Second, the choice of frequency also balances two competing factors: spatial resolution and magnetic field penetration depth. Higher frequencies improve spatial resolution but reduce penetration, while lower frequencies allow deeper penetration at the cost of resolution. Through theoretical analysis and empirical validation, 1 MHz was found to provide an optimal trade-off between these parameters for cerebral hemorrhage detection.

In the simulation phase, the coil excitation current was set to 27 mA AC to ensure sufficient magnetic field intensity. For the physical experiment, however, power was supplied via a USB 2.0 interface, which allows a maximum current of 500 mA per port. To maintain a safe design margin and ensure stable operation, the experimental current was limited to 2.7 mA. This constraint was fully accounted for in the signal processing circuit design and component selection, ensuring system reliability and operational safety during real-time data acquisition and transmission.

Model construction

The simulation model was constructed using COMSOL Multiphysics. A realistic human head model was created using MRI coordinate files provided by the COMSOL library [22]. The model has an approximate height of 20 cm and width of 10 cm (**Figure 1A**). An ellipsoid (axes: 0.055 m, 0.065 m, 0.045 m) was inserted to represent the brain parenchyma, positioned at Y=-0.01 m and Z=0.07 m. A smaller ellipsoid (axes: 0.016 m, 0.018 m, 0.01 m) was added to simulate a cerebral hemorrhage, located at Y=-0.01 m and Z=0.1 m, ensuring it lies within the brain parenchyma (**Figure 1B**).

For the size and shape of the cerebral hemorrhage, the parameters were selected based on previous studies indicating that local brain stimulation may lead to an increase in local blood volume of approximately 1.5 mL. To enhance the accuracy of the COMSOL simulation, the modeled hemorrhage volume must closely approximate actual physiological values. Using a simplified formula for estimating intracerebral hemorrhage volume—defined as half the product of the three ellipsoidal semi-axes—the volume V is calculated as:

$$V = \frac{a \times b \times c}{2} \quad (1)$$

Based on this formula, and the selected semi-axis lengths a, b, and c, the calculated volume of the ellipsoid is 1.44 mL. After iterative adjustments and validation in COMSOL, these dimensions were confirmed to yield a volume

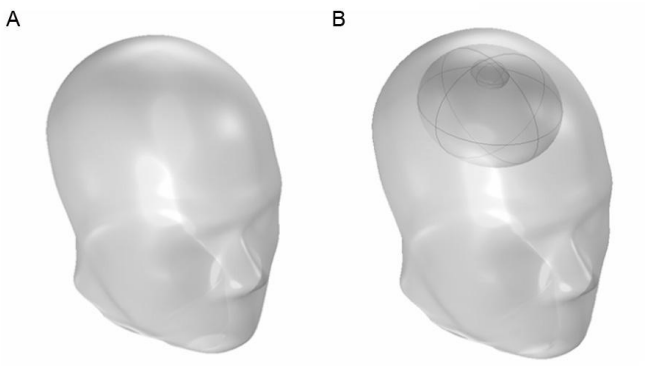


Figure 1. Head model construction process. (A) After the skull model lofting is completed. The Boolean operation segments the curve to ensure the successful lofting of the model, and the model generated after lofting is relatively smooth. (B) The brain parenchyma and cerebral hemorrhage models were constructed. It can be seen that the brain parenchyma and cerebral hemorrhage are in a reasonable position in the skull, with reasonable volume.

Figure 2. The coil model is constructed. The spatial position of the coil is above the overall model.

Figure 3. The setting of coil current. The red arrow represents the direction of the input current in the coil. In this direction, the magnetic field generated by the coil is downward to ensure that it can pass through the cerebral hemorrhage area.

closest to 1.5 mL, aligning with experimental requirements.

For the inductive coil design, a copper coil with an outer diameter of 8.8 cm is used in the simulation. The coil has the following parameters: a peak current of 2.7 mA, 6 turns, and a length of 6 cm. It is positioned directly above the head model, as shown in **Figure 2** [19].

Material parameters were assigned as follows:

Skull: conductivity 0.4 S/m, density 2000 kg/m^3, relative permittivity 827, relative permeability 1.

Brain parenchyma: conductivity 0.2 S/m, density 1000 kg/m^3, relative permittivity 827, relative permeability 1.

Cerebral hemorrhage: conductivity 0.65 S/m, density 1060 kg/m^3, relative permittivity 3030, relative permeability 1.

Coil (copper): conductivity 5.998 S/m, density 8960 kg/m^3, relative permittivity 1, relative permeability 1 [23].

These values simulate the physical properties of human tissues and the coil within the model.

The simulation involves applying alternating current to the coil to generate a magnetic field, which induces eddy currents in conductive tissue such as hemorrhagic areas. The magnetic field was modeled using the AC/DC module, focusing on flux density, current, inductance, and resistance. Ampère's law was used to define the magnetic field behavior, with magnetic insulation applied to all boundaries except where inapplicable. Initial values for the magnetic vector potential in x, y, and z directions were set to 0 Wb/m.

A uniform multi-turn coil was defined with numerical current excitation, and the direction of current input was aligned axially to ensure that the magnetic field penetrates downward

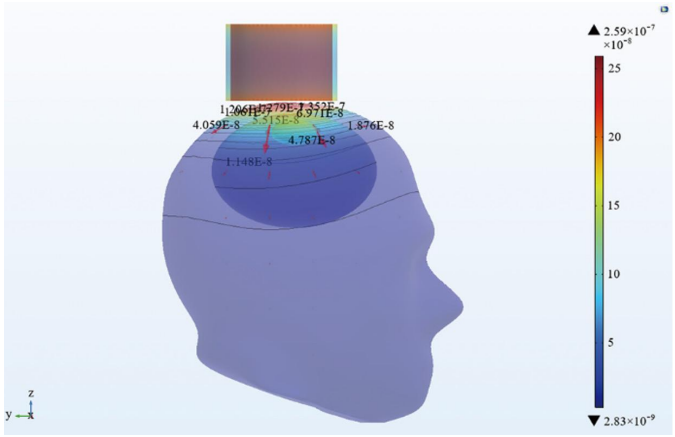


Figure 4. Magnetic field distribution of the whole model. The line on the model represents the magnetic field contour, and the number on the contour represents the value of the magnetic field on the contour. The red arrow represents the direction of the magnetic field. The larger the arrow, the stronger the magnetic field in this area.

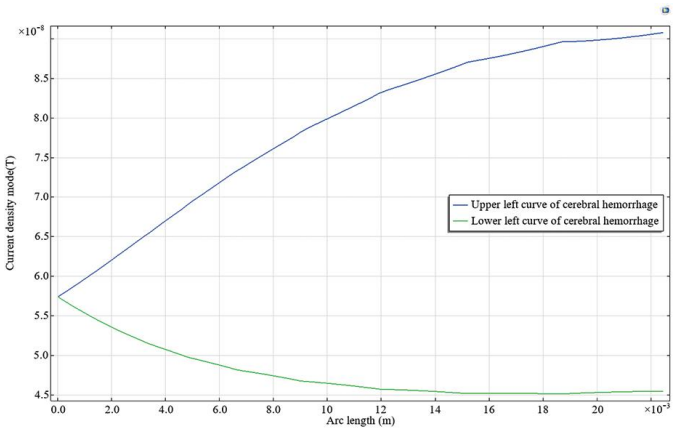


Figure 5. The variation curve of magnetic flux density modulus with arc length in intracerebral hemorrhage. The blue curve represents the upper left curve of cerebral hemorrhage, and the green curve represents the sitting curve of cerebral hemorrhage.

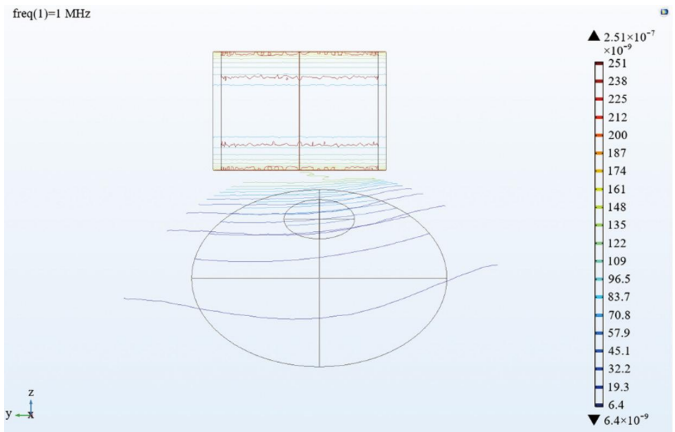


Figure 6. Contour map of cerebral parenchyma and cerebral hemorrhage area. The magnetic field generated by the coil can pass through the brain parenchyma and the area of cerebral hemorrhage.

through the hemorrhagic area (**Figure 3**). This completes the physical field setup for the magnetic field simulation.

Simulation results

To utilize ECD technology for precise localization of cerebral hemorrhage, it is essential to verify that the alternating magnetic field generated by the coil can penetrate the hemorrhagic area. This penetration induces eddy currents within the hemorrhage, enabling its localization. The simulated magnetic field distribution is shown in **Figure 4**. The red arrow indicates the direction of the magnetic field, which passes downward through the head. The contour lines reveal that the magnetic flux density near the hemorrhage is approximately 4.78×10^{-8} T, confirming that the magnetic field effectively penetrates the hemorrhagic region.

To illustrate the spatial variation of magnetic flux density within the hemorrhagic area, two line plots were generated along different sagittal cross-sections. The x-axis represents arc length (m), and the y-axis shows magnetic flux density (T). As shown in **Figure 5**, the maximum flux density within the hemorrhage reaches 9.2×10^{-8} T, while the lowest value, measured at the region furthest from the coil, is about 4.5×10^{-8} T.

Additionally, a 3D visualization (**Figure 6**) was generated by adding contour lines and sectional views, illustrating the spatial distribution of the magnetic field. The contours clearly pass through the hemorrhagic region and gradually diminish in intensity with increasing distance from the coil. These visualizations confirm that the magnetic field is strong enough to reach the hemorrhagic site.

Once magnetic field penetration is confirmed, the next step is to detect hemorrhagic lesions by observing parameter changes in the coil. Eddy currents induced in the hemorrhagic region produce a secondary magnetic field that opposes the original field

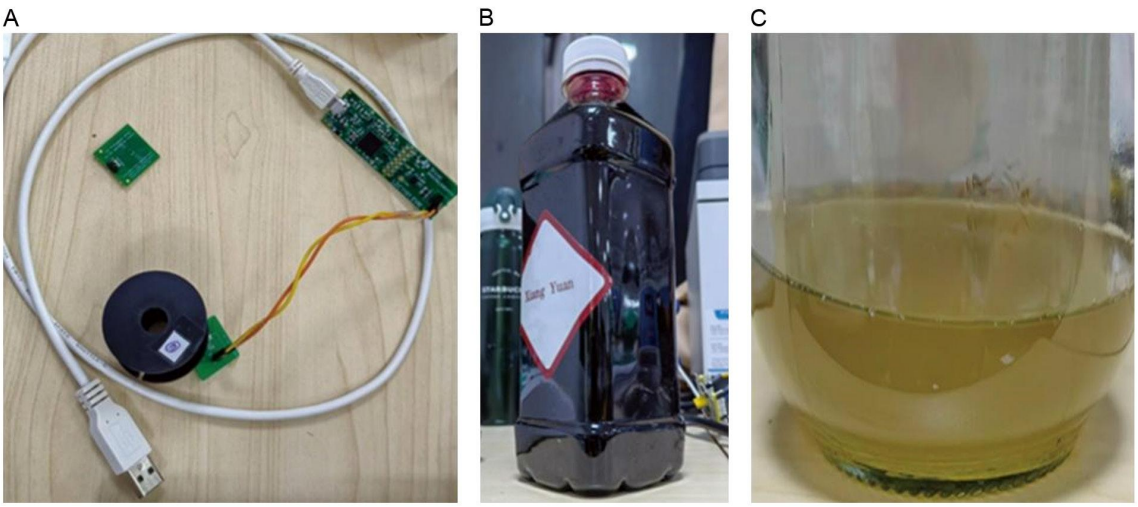


Figure 7. Materials used in the experiment. (A) The figure shows the coil sensor, processing circuit, and USB interface for data transmission with the computer. (B) Fresh pig blood. (C) Gelatin material.

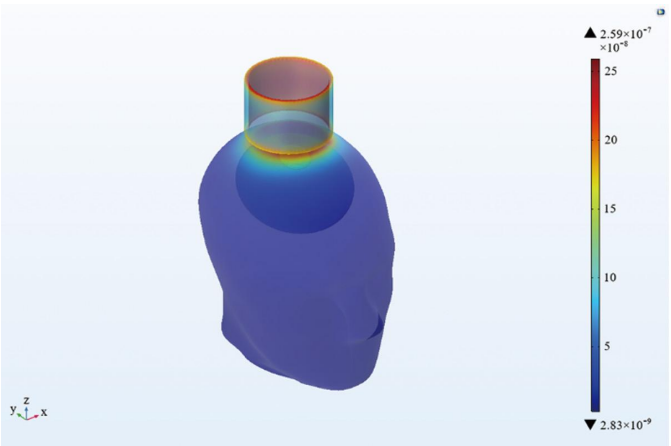


Figure 8. Magnetic field distribution cloud diagram of the whole model.

from the coil. This interaction decreases the coil's inductance and increases its AC resistance due to energy loss. Therefore, variations in coil inductance and resistance can be used as indicators to detect cerebral hemorrhage.

Experimental validation

Simulation results suggest that hemorrhagic regions exhibit a damping effect on the coil's magnetic field, and that the field can effectively penetrate such regions. This forms the basis for quantifying ECD signals.

These signals can be measured using an LC resonant circuit, where a magnetic coil is coupled with a capacitor. The change in parallel resistance (Rp) of the system reflects variations in coil AC resistance. According to Equation (2):

$$Rp = \frac{L}{RC} \quad (2)$$

where Rp is the parallel resistance, L is inductance, R is resistance, and C is capacitance. This equation shows that Rp is inversely proportional to the AC resistance R. Hence, when the coil's AC resistance increases due to eddy current damping, the measured Rp decreases.

The experimental setup is shown in **Figure 7A**, and includes an inductive coil, oscillator, LDC1101 inductance-to-digital converter chip, and an MCU for data processing. Data are transmitted to a computer via USB, where inductance and resistance values are displayed in real time.

Figure 7B shows the fresh pig blood used in the experiment, and **Figure 7C** shows the gelatin phantom (conductivity: 0.2 S/m) used to simulate brain parenchyma. The experiment was conducted at room temperature and initiated immediately after gelatin solidification to prevent dissolution. A 10 mL volume of pig blood was injected into the gelatin to simulate a cerebral hemorrhage. The inductive coil was gradually brought closer to the phantom, and changes in inductance and resistance were monitored in real time via the host computer interface to evaluate detectability.

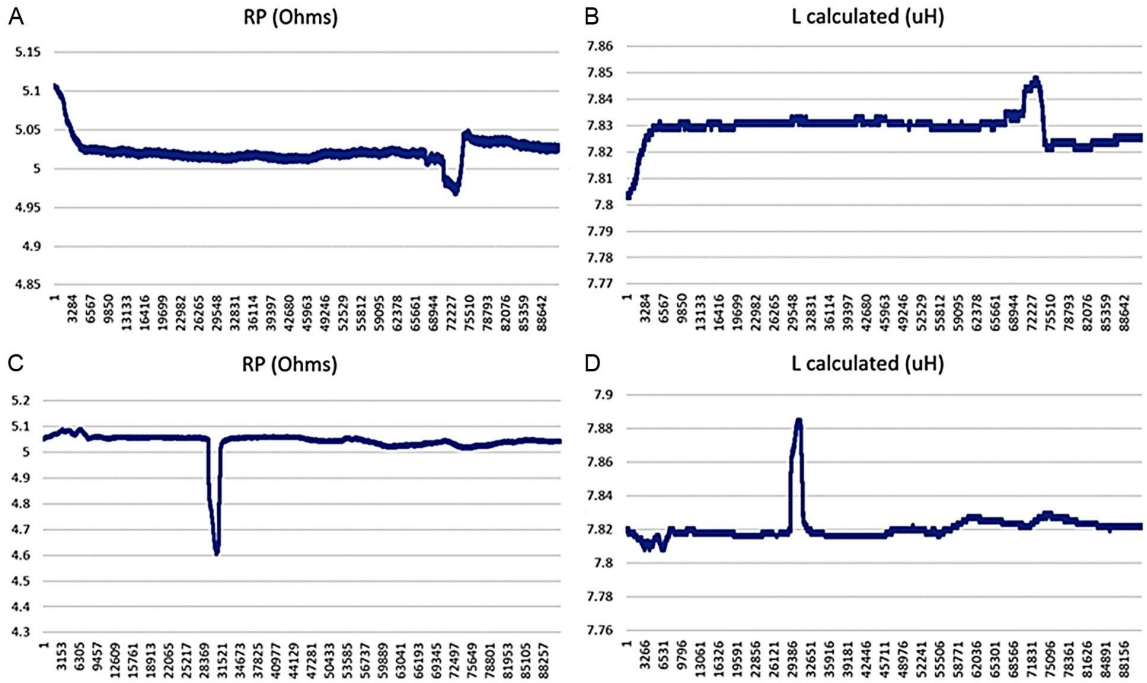


Figure 9. The change curves of circuit resistance and inductance are obtained by the upper computer. (A and B) The volume of pig blood is 40 ml. (C and D) The volume of pig blood is 60 ml.

Results

The magnetic field contour plots of the simulation model confirm that the field generated by the coil can penetrate the cerebral hemorrhagic region, inducing eddy currents that produce a damping effect, thereby influencing the coil's electromagnetic characteristics. **Figure 8** illustrates the magnetic field distribution from the simulation results.

To assess the impact of hemorrhage on coil parameters, a reference model consisting only of the coil and air domain was established, keeping other boundary conditions constant. By comparing the inductance and AC resistance values between the two models, we observed a decrease in inductance and an increase in resistance in the presence of hemorrhage. These trends validate the feasibility of using ECD technology for detecting cerebral hemorrhage.

Figure 9 presents experimental data collected on a computer. **Figure 9A** and **9B** display changes in resistance and inductance when 40 mL of pig blood is used as the detection target, while **Figure 9C** and **9D** show the corresponding data for 60 mL. The curves indicate that, as the detection probe approaches the blood sample, the resistance increases and the inductance decreases—consistent with ECD principles.

Discussion

A variety of techniques have been explored for brain injury detection, including microwave imaging, volume impedance phase shift spectroscopy, near-infrared spectroscopy, EEG, transcranial Doppler ultrasound, and ECD [3, 24]. Among these, ECD stands out for its short detection time and high accuracy.

Given the importance of detecting changes in coil impedance caused by cerebral hemorrhage, a simplified coil-only model was constructed in COMSOL. This allowed for clear visualization of resistance and inductance changes under different conditions. **Figure 10** presents the coil-only model, and **Figure 11** shows the magnetic field distribution generated by the coil. The arrows indicate the magnetic field direction, while the color map reflects its magnitude.

In the coil-only model, the measured inductance is 2.4293±0.0015 μH, while in the full model (including the head), it decreases slightly to 2.4273±0.0018 μH. Meanwhile, the AC resistance increases from 2.12±0.05 mΩ (pure coil)

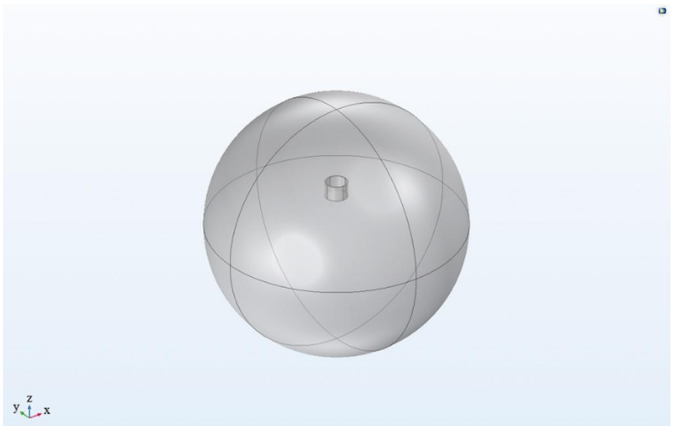


Figure 10. The model with only coil and air domain.

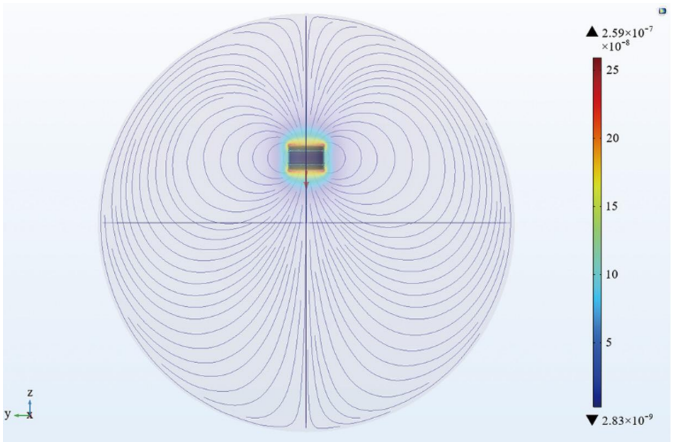


Figure 11. Magnetic field distribution with coil model only. The streamline represents the distribution of the magnetic field in the air domain.

to 3.19±0.07 mΩ (full model). These measurements account for an estimated ±0.5% instrument error and ±0.1% environmental variation. The observed decrease in inductance and increase in resistance support the predicted ECD effects.

This study integrates COMSOL simulations and experimental validation to assess the feasibility of ECD-based detection of cerebral hemorrhage. The electromagnetic model, which incorporates brain tissue and hemorrhagic regions, demonstrates that eddy currents generated in hemorrhagic areas affect coil impedance. Experiments using pig blood and custom detection circuitry revealed a significant correlation between blood volume and coil parameters, aligning with simulation predictions.

Although the current model assumes uniform brain tissue conductivity—neglecting regional (e.g., gray vs. white matter) and temporal variations that may affect quantitative accuracy—the consistent simulation and experimental trends confirm the fundamental feasibility of this approach. Future research will improve model accuracy by integrating tissue-specific conductivity and constructing multi-layered phantoms to better simulate brain structures. These results provide foundational evidence supporting the potential clinical application of ECD technology.

Conclusion

This study uses finite element simulations in COMSOL to verify the feasibility of detecting cerebral hemorrhage via ECD technology. Simulation analysis demonstrates that hemorrhagic regions affect coil inductance and resistance, supporting the use of ECD for detection. A physical model was constructed, and experimental tests using a coil sensor confirmed that inductance and resistance change as expected in the presence of hemorrhage.

Since these changes are numerical differences, they are necessarily influenced by parameters such as hemorrhage volume and distance between the coil and hemorrhagic site. Therefore, this method enables qualitative assessment of hemorrhage by quantifying impedance changes.

Future efforts will aim to further quantify ECD signals induced by cerebral hemorrhage by analyzing precise changes in coil impedance. Optimizing coil geometry and parameters will also be explored to enhance detection performance. If cerebral hemorrhage volume and location can be accurately measured using ECD sensors, this technology could complement or even provide an alternative to current imaging modalities such as CT and MRI. ECD offers a fast, portable, and potentially cost-effective method for detecting intracerebral hemorrhage, overcoming the limitations of conventional techniques. We anticipate continued validation and refinement of this method in future studies.

Author contributions: Wenjing Du performed the data analyses and wrote the manuscript. Wenjing Du and Tingting Shi performed the experiment. Wenjing Du, Tingting Shi, Ke Wang, Shoucheng Chen and Rongguo Yan contributed significantly to the analysis and manuscript preparation.

References

[1] Ghajar J. Traumatic brain injury. Lancet 2000; 356(9233):923-929.

[2] Maas AIR, Menon DK, Manley GT, et al. Traumatic brain injury: progress and challenges in prevention, clinical care, and research. Lancet Neurol 2022;21(11):1004-1060.

[3] Krausz AD, Korley FK, Burns MA. The Current State of Traumatic Brain Injury Biomarker Measurement Methods. Biosensors (Basel) 2021; 11(9):319.

[4] Kessler BA, Goh JL, Pajer HB, et al. Rapid-sequence MRI for evaluation of pediatric traumatic brain injury: a systematic review. J Neurosurg Pediatr 2021;28(3):278-286.

[5] Lunkova E, Guberman GI, Ptito A, et al. Noninvasive magnetic resonance imaging techniques in mild traumatic brain injury research and diagnosis. Hum Brain Mapp 2021;42(16): 5477-5494.

[6] Wald LL, McDaniel PC, Witzel T, et al. Low-cost and portable MRI. J Magn Reson Imaging 2020;52(3):686-696.

[7] Yang RJ, Yan HL, Han DM. Diagnostic Value of Brain Imaging of Multi Slice Spiral CT and MRI Examination for Cerebrovascular Diseases. Chinese Journal of CT and MRI 2021;19(7):15-17.

[8] Wang J, Hu Y, Wu P. Risk factors for positive brain CT scan in children with traumatic brain injury and GCS = 15: A retrospective study. Medicine (Baltimore) 2021;100(4):e24543.

[9] Moen TR, Chen B, Holmes DR 3rd, et al. Low-dose CT image and projection dataset. Med Phys 2021;48(2):902-911.

[10] Aide N, Lasnon C, Desmonts C, et al. Advances in PET/CT technology: an update. Semin Nucl Med 2022;52(3):286-301.

[11] Rehani MM, Yang K, Melick ER, et al. Patients undergoing recurrent CT scans: assessing the magnitude. Eur Radiol 2020;30(4):1828-1836.

[12] Miller GF, DePadilla L, Xu L. Costs of nonfatal traumatic brain injury in the United States, 2016. Med Care 2021;59(5):451-455.

[13] Alouani AT, Elfouly T. Traumatic brain injury (TBI) detection: past, present, and future. Biomedicines 2022;10(10):2472.

[14] Sheth KN, Mazurek MH, Yuen MM, et al. Assessment of brain injury using portable, low-field magnetic resonance imaging at the bedside of critically ill patients. JAMA Neurol 2021;78(1):41-47.

[15] Shahrestani S, Strickland BA, Bakhsheshian J, et al. Transcranial eddy current damping sensors for detection and imaging of hemorrhagic stroke: feasibility in benchtop experimentation. Neurosurg Focus 2021;51(1):E15.

[16] Lee SK, Hua Y. Theory and mitigation of motional eddy current in high-field eddy current shielding. J Appl Phys 2024;136(2):024504.

[17] Kimberly WT, Sorby-Adams AJ, Webb AG, et al. Brain imaging with portable low-field MRI. Nat Rev Bioeng 2023;1(9):617-630.

[18] Shahrestani S, Zada G, Chou TC, et al. Noninvasive transcranial classification of stroke using a portable eddy current damping sensor. Sci Rep 2021;11(1):10297.

[19] Shahrestani S, Zada G, Tai YC. Development of computational models for microtesla-level magnetic brain scanning: a novel avenue for device development. BMC Biomed Eng 2022; 4(1):1.

[20] Gabriel C, Peyman A, Grant EH, et al. Electrical conductivity of tissue at frequencies below 1 MHz. Phys Med Biol 2009;54(16):4863-4878.

[21] Zhang P, Ran Y, Han L, et al. Nanomaterial technologies for precision diagnosis and treatment of brain hemorrhage. Biomaterials 2025;321:123269.

[22] Simulating Irregular Shapes: How to Import Curve Data and Layout into Solids. Available from: https://cn.comsol.com/blogs/modeling-irregular-shapes-how-to-import-curve-data-and-loft-a-solid/

[23] Gabriel C, Gabriel S. Compilation of the dielectric properties of body tissues at RF and microwave frequencies. King's College, London, Department of Physics 1996.

[24] Shahrestani S, Wishart D, Han SMJ, et al. A systematic review of next-generation point-of-care stroke diagnostic technologies. Neurosurg Focus 2021;51(1):E11.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/517419bnpxap

Research on knee osteoarthritis grading based on multidimensional feature fusion

He Ren*, Yina Zhang*, Yutong Xie*, Anqi Wu, Xianglun Kong, Chenxiao Bai, Miao Yu, Yimeng Wang, Ping Li

Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China.
*The authors contribute equally.

Corresponding author: Ping Li.

Declaration of conflict of interest: None.

Declaration of ethics: This study uses an open-source dataset, which has been processed in accordance with relevant ethical guidelines during its collection and release. The data provider has completed the necessary ethical review procedures to ensure that data collection complies with internationally recognized ethical norms such as the Declaration of Helsinki. Strict de-identification processing has been applied to information involving individuals to fully protect the privacy and rights of data providers. Since this study does not directly contact original individuals and only analyzes de-identified open-source data, it does not involve additional risks to human participants, and thus there is no need to apply for ethical review again.

Declaration of patient consent: As the open-source data used in this study has been de-identified and cannot be traced back to specific patients, there is no risk of violating patients' privacy. Therefore, it is not necessary to obtain informed consent from patients.

Received July 17, 2025; Accepted September 29, 2025; Published December 31, 2025

Highlights

- Based on 8,110 OAI X-ray images, 18 radiomic features were processed and selected, with SMOTE combined with category weight balancing employed to address data imbalance.
- Among eight machine learning models, the SVM achieved the best performance.
- SHAP and LIME analyses enhanced model interpretability by identifying key radiomic features influencing predictions.

Abstract

Objective: This study aimed to apply machine learning approaches to the Kellgren-Lawrence (KL) grading of knee osteoarthritis, develop an effective automatic KL grading technique, and provide a methodological reference for clinical diagnosis and research. **Methods:** Data were obtained from the Osteoarthritis Initiative (OAI) knee X-ray image dataset, comprising 8,110 images from the folders of *auto*_test, *train*, and *val*. All images were first subjected to inversion processing, followed by extraction of two-dimensional radiomic features. Feature selection was then conducted using a combination of variance thresholding and analysis of variance (ANOVA), yielding 18 key features. To address class imbalance in the original dataset, this synthetic minority over-sampling technique (SMOTE) and class weight balancing were jointly applied. Eight machine learning models—Decision Trees (DT), Logistic Regression (LR), Random Forests (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Light Gradient Boosting Machine (LightGBM)—were trained for KL grading of knee osteoarthritis. Model performance was evaluated using accuracy, precision, recall, F1-score, and the area under the curve (AUC). For the optimal SVM model, global and local interpretability analy-

Address correspondence to: Ping Li, Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, No. 279, Zhouzhu Highway, Pudong New Area, Shanghai 201318, China. Tel: +86-13764055848. E-mail: lip@sumhs.edu.cn.

ses were further conducted using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) to identify the key factors influencing model decisions. **Results:** The support vector machine (SVM) model achieves the best performance. **Conclusion:** This study establishes an effective machine learning-based method for automatic KL grading of knee osteoarthritis, providing valuable support for clinical diagnosis and research applications.

Keywords: Osteoarthritis, knee osteoarthritis, Kellgren-Lawrence (KL) grading, machine learning, interpretability

Introduction

Arthritis is a category of diseases primarily characterized by joint pain, swelling, stiffness, deformity, and functional impairment. Its major types include osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, and gouty arthritis [1-5]. Among these, osteoarthritis is the most prevalent form. Knee osteoarthritis, a common subtype, is characterized by degeneration of knee cartilage and osteophyte formation, predominantly affecting middle-aged and elderly individuals. Due to pain, many patients with knee osteoarthritis reduce their physical activity, leading to impaired lower-limb circulation [6, 7]. The release of inflammatory factor further compromises vascular endothelial function, causing a hypercoagulable state and an increased risk of lower-limb venous thrombosis [8]. Once detached, thrombi may travel to the lungs via the bloodstream, potentially causing pulmonary embolism and fatal complications [9, 10].

Conventional diagnostic methods for knee osteoarthritis, including physical examination, X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and arthrography, each have limitations. Physical examination outcomes rely on physician experience and subjective judgment, lacking standardized criteria. X-ray and CT are limited in detecting early cartilage wear, bone changes, or soft-tissue lesions [11-13]. MRI, though sensitive, is costly, contraindicated in patients with metal implants, and prone to metal artifacts that degrade image quality. Arthrography is invasive, technically demanding, and carries risks such as infection. Additionally, the high prevalence of depression among knee osteoarthritis patients contributes to substantial socioeconomic burdens [14]. The Kellgren-Lawrence (KL) grading system is a classic radiological tool for assessing the severity of osteoarthritis. However, its accuracy is highly dependent on the subjective interpretation of radiologists, potentially affecting diagnostic consistency [15]. Artificial intelligence (AI) models can reduce subjectivity by enabling standardized feature extraction. Yet, conventional deep learning models often lack interpretability, limiting their acceptance as diagnostic evidence in clinical practice [16].

Machine learning, value for its interpretability, has been increasingly used in disease diagnosis and classification. However, most existing studies suffer from small sample sizes, leading to a higher risk of model overfitting and limited generalization. Additionally, in studies utilizing multiple classifiers, the best classification accuracies were 85.6% and 82.61% [14, 17]. These results fall short of the clinical goal of supporting consistent diagnostic decision-making.

Building upon prior research, this study aims to develop a novel diagnostic method with improved diagnostic accuracy and cost-effectiveness to advance osteoarthritis diagnosis. Specifically, quantitative radiomic features were extracted from knee X-ray images using PyRadiomics to establish a database for model training. Eight machine learning classifiers, including Support Vector Machine (SVM) and AdaBoost, were applied for KL grading. Furthermore, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were employed to interpret the model outputs from both global and local perspectives, enhancing the transparency and clinical interpretability for the results.

Materials and methods

Datasets

Data for this study were sourced from the Osteoarthritis Initiative (OAI) knee osteoarthritis dataset, which includes X-ray images annotated with corresponding KL severity grades. The dataset contains 9,786 knee X-ray images used for knee joint detection and KL grading, organized into four folders: auto_test, test, train, and val. Each folder contains images classified into five KL grades: Grade 0 (Normal), Grade 1 (Suspicious), Grade 2 (Mild), Grade 3 (Moderate),

Table 1. Kellgren-Lawrence (KL) grading and corresponding imaging features for knee osteoarthritis

Class	Kellgren-Lawrence Grading	Imaging Features	Number of X-ray images
Grade 0	Normal	Normal joint space, no osteophytes, no subchondral sclerosis.	3212
Grade 1	Doubtful	Possible joint space narrowing, with possible osteophyte formation in a lip - like shape.	1474
Grade 2	Minimal	Definite osteophyte formation, possible joint space narrowing.	2117
Grade 3	Moderate	Multiple osteophytes, definite joint space narrowing, mild subchondral sclerosis.	1063
Grade 4	Severe	Large osteophytes, marked joint space narrowing, severe subchondral sclerosis.	244

and Grade 4 (Severe). This study selected 8,110 X-ray images from the auto_test, train, and val folders for analysis (**Table 1**).

Data preprocessing

Image processing

Color X-ray images were first converted to grayscale. Image inversion was applied to enhance bone structures for threshold-based segmentation. Subsequently, Otsu's thresholding method was applied to segment the image into foreground (bone) and background regions. Only foreground pixel values were retained, with background pixels being set to 0 for subsequent refinement. A second Otsu thresholding was performed on the foreground to eliminate low-intensity noise and preserve the high-intensity bone core area. A binary mask was generated to locate the Region of Interest (ROI). The upper and lower boundaries of the knee joint were determined using row-pixel statistics combined with a sliding-window search. A quadrilateral ROI consistent with the knee's anatomical structure was delineated through endpoint positioning and spacing adjustment. Finally, polygon filling was applied to generate the mask and define the effective area for feature extraction.

Feature extraction

2D radiomic features were extracted from the preprocessed knee X-ray images using the *Pyradiomics* library. Initially, the ROI was delineated through inversion and binarization of the original image, followed by connected component analysis to remove noise and retain the largest bone-connected region. Subsequently, vertical artifacts were excluded based on column-pixel statistics, generating a rectangular mask that encompassed the core knee joint structure (**Figure 1**). The preprocessed grayscale images and ROI masks were then converted to the *SimpleITK* format to ensure spatial alignment. Using a predefined parameter configuration, 18 first-order statistical features, 14 shape-based features, and 48 texture features were extracted, totaling 80 quantitative features. Finally, feature extraction was performed on each ROI, generating a dataset with feature values, image filenames, and diagnostic labels for subsequent analysis.

Feature selection

To select features with significant influence on the target variable, a two-step selection process combining variance thresholding and analysis of variance (ANOVA) were used. Features with a variance greater than 0.01 were first retained. Subsequently, the SelectKBest function with the *f_classif* criterion was applied to determine features most strongly correlated with the target variable. To further enhance the accuracy and reliability of feature selection, this study integrated variance thresholding and correlation analysis. Specifically, features meeting the variance threshold and showing strong correlation with the target variable were selected, yielding 18 optimal features, which served as the final feature set for subsequent modeling.

Data balancing

The dataset used in this study exhibits class imbalance, with varying sample sizes across categories. This imbalance can bias the model toward majority classes, resulting in suboptimal performance on minority classes. To address this issue, synthetic minority over-sampling technique (SMOTE) was applied to augment the

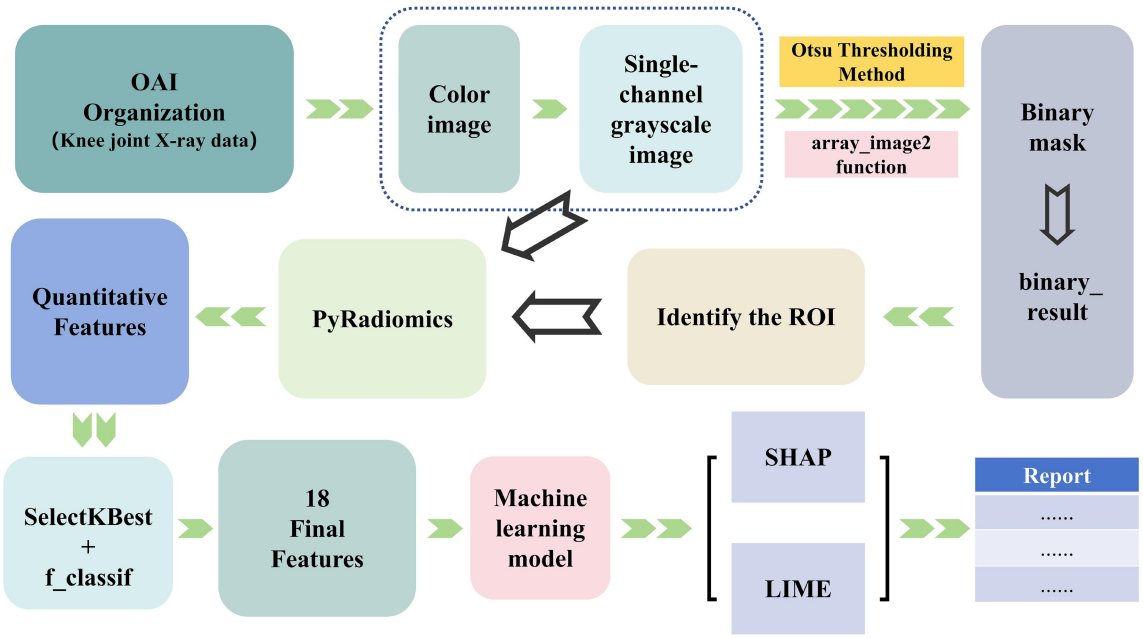


Figure 1. A comprehensive analytical pipeline integrating image processing, radiomic feature extraction, machine-learning model training, and model interpretation. The workflow is designed to extract meaningful information from medical imaging data and facilitate disease diagnosis.

minority class samples, as illustrated in **Figure 2**. SMOTE generates synthetic samples between existing minority - class samples to balance the dataset. Additionally, during model training, different weights were assigned to each class to make the model focus more on the minority - class samples.

To avoid data leakage and ensure the model's generalization to real-world scenarios, the SMOTE algorithm was exclusively applied to the training set to generate synthetic minority samples, while the validation and test sets retained their original class distributions. This strategy mitigated overfitting by preventing the model from learning synthetic patterns that do not exist in real data, enabling an unbiased evaluation of model performance on naturally imbalanced clinical datasets.

Model construction

Support vector machine (SVM)

SVM is characterized by strong generalization ability. It can accurately classify patients based on limited data. Also, its robustness allows tolerance to a certain degree of outliers and noise, contributing to the development of a stable five–class classification model. Thus, SVM performs well in this study. As a supervised learning algorithm based on statistical learning theory, SVM distinguishes samples of different classes by finding a hyperplane that maximizes the margin. Its fundamental principle involves mapping data into a high-dimensional space to identify an optimal hyperplane. This hyperplane separates samples of different classes with the maximum margin.

Other models

In addition to SVM, several other machine learning algorithms were evaluated in this study. The Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) models demonstrated outstanding performance. LightGBM uses a histogram-based algorithm and a leaf-wise growth strategy to enhance training efficiency. It can automatically handle missing values and assess feature importance, thereby assisting in feature selection. Its precise predictions, strong generalization ability, and effectiveness in addressing data imbalance make it particularly suitable for this research.

XGBoost, a gradient boosting decision tree algorithm, incorporates parallel computing and optimization, delivering superior prediction accuracy and automatically handling missing data. These features aid in understanding the rela-

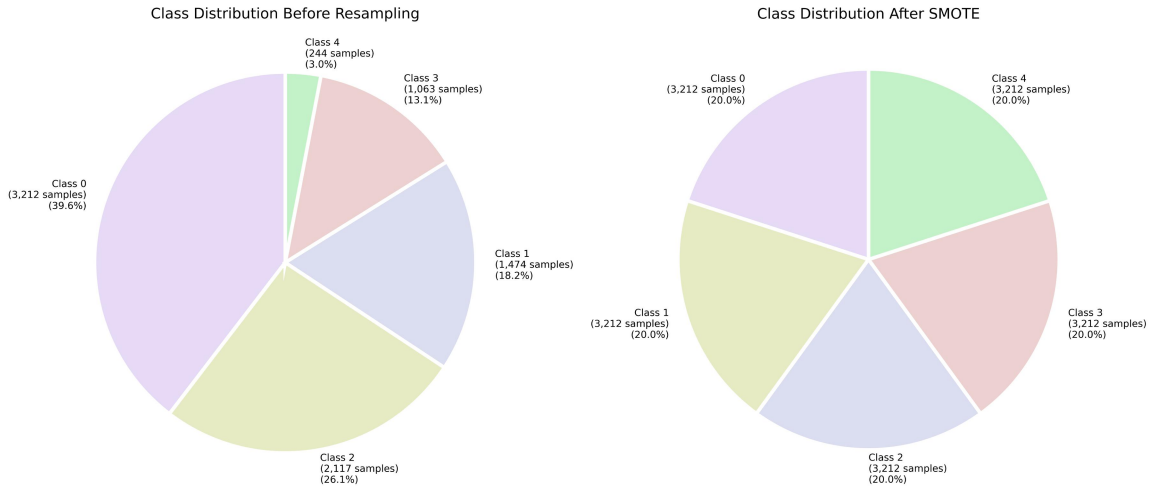


Figure 2. Comparison of the class distribution in the dataset before and after balancing.

tionship between features and model outputs, making them suitable for the five-class classification of knee osteoarthritis.

The K-nearest neighbors (KNN) is simple and intuitive but sensitive to high-dimensional data and noise. Categorical Boosting (CatBoost) performs well in handling imbalanced data by automatically managing class weights and mitigating overfitting. Decision Tree (DT) models offer good interpretability but are prone to overfitting. Random Forest (RF) enhances stability by integrating multiple decision trees but is less efficient in handling imbalanced data. Logistic Regression (LR), a linear model, struggles to capture nonlinear relationships, leading to insufficient classification accuracy. However, it offers good interpretability, simplicity, and ease of use for clinical applications and decision support. Additionally, it provides probability predictions for richer diagnostic information. These advantages make it valuable in specific contexts.

Model evaluation

Model performance was evaluated using five standard metrics: accuracy, precision, recall, F1 score, and area under the curve (AUC). Accuracy represents the proportion of correctly classified instances, indicating overall model correctness. High accuracy indicates that the model performs well in classification. Precision reflects the proportion of true positives among all predicted positives, where higher precision means fewer false positives and improved diagnostic reliability. Recall (sensitivity) evaluates a model's ability to correctly identify actual positive samples; high recall indicates good classification performance. The F1 score balances precision and recall, helping to identify the optimal model. AUC measures a model's ability to distinguish between classes. A higher AUC value indicates better performance and is not affected by class imbalance. These metrics provide a comprehensive evaluation of model performance for the five-class knee osteoarthritis classification task.

Model explainability

To enhance model transparency and clinical trust, interpretability analyses were conducted using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). SHAP quantifies feature contributions to model predictions based on Shapley values. It assigns importance values to features, where positive values indicate a promotion effect and negative values indicate an inhibitory effect on the prediction. This reflects the influence of each feature on the model's output.

LIME, on the other hand, explains individual predictions by constructing local surrogate linear models around sample points, simplifying complex models and visually demonstrating feature impacts. In LIME analysis, high - weight features significantly influence predictions of specific arthritis types. These techniques help doctors understand model reasoning, increasing diagnostic credibility and patient acceptance. They also improve the reliability and clinical applicability of models in diagnosing knee osteoarthritis.

Table 2. Comparison of model performance

Model	Accuracy	Precision	Recall	F1-score	AUC
SVM	0.9284	0.9271	0.9282	0.9270	0.9906
DT	0.7905	0.7872	0.7900	0.7870	0.8872
XGBoost	0.9128	0.9116	0.9125	0.9118	0.9874
KNN	0.8776	0.8840	0.8761	0.8626	0.9558
CatBoost	0.8238	0.8202	0.8234	0.8193	0.9596
LightGBM	0.9141	0.9138	0.9141	0.9139	0.9894
LR	0.4038	0.3819	0.4056	0.3856	0.7183
RF	0.5682	0.5587	0.5687	0.5579	0.8460

Results

Comparison of model performance

In this study, eight machine learning models were applied to the five-class classification task of knee osteoarthritis. The models were compared based on accuracy, precision, recall, F1 score, and AUC. Additionally, receiver operating characteristic (ROC) and Precision-Recall (PR) curves were plotted for visualization. SHAP and LIME were used to illustrate the model's interpretability through multiple plots, providing a comprehensive assessment of the models' performance.

As shown in **Table 2**, SVM outperformed other models across multiple metrics. It achieved an accuracy of 0.9284, with precision and recall at 0.9271 and 0.9282, respectively, indicating excellent classification consistency. The SVM model also yielded an F1 score of 0.9270 and an AUC of 0.9906, further demonstrating its robustness and discriminative capability in the medical grading task.

XGBoost and LightGBM, as representatives of gradient boosting frameworks, also performed well. XGBoost achieved an accuracy of 0.9128, while LightGBM achieved an accuracy of 0.9141, both with AUC values exceeding 0.98. LightGBM achieved an F1 score of 0.9139, approaching that of SVM, indicating its robustness in handling complex feature interactions.

The KNN algorithm demonstrated acceptable overall performance but had a relatively lower F1 score (0.8626), showing slight model imbalance. CatBoost and DT models achieved accuracies of 0.8238 and 0.7905, respectively, with AUC values between 0.88 and 0.96. CatBoost, despite its advantages in handling categorical features, performed moderately in this task. DT was prone to overfitting. In contrast, RF and LR had the lowest accuracy at 0.5682 and 0.4038.

The performance of the eight models was further evaluated by plotting their ROC and PR curves. The ROC curve reflects each model's discriminative ability across different classification thresholds. **Figure 3A** shows that the curves of SVM, LightGBM, and XGBoost are closer to the upper left corner, indicating superior ability to accurately identifying positive samples while minimizing false-positive predictions. In contrast, the ROC curves of the other models are closer to the diagonal, indicating weaker discriminative ability. PR curves display a model's precision across varying recall levels. **Figure 3B** shows that SVM, LightGBM, and XGBoost maintain high precision across most recall ranges, indicating a higher probability that identified positive samples are truly positive. By contrast, the PR curves for LR and RF showed a faster decline in precision as recall increases. This suggests these models have more misclassifications when identifying positive samples, resulting in relatively poorer performance.

Figure 4 and **Table 3** show that the balancing strategy used in this study effectively improved recognition performance for minority classes. Grade 4 achieved a recall rate of 99.84%, with no false-negative predictions in the confusion matrix, indicating the model's excellent ability to identify severe lesions. For Grade 1 and Grade 3, the recall rates were 96.77% and 97.13%, respectively, indicating reliable identification of both early and moderate disease stages while minimizing missed or misdiagnosed cases. Grade 0 has a relatively lower recall rate of 78.67%, and most misclassifications were attributed to normal samples being misclassified as Grade 1 (35 cases) or Grade 2

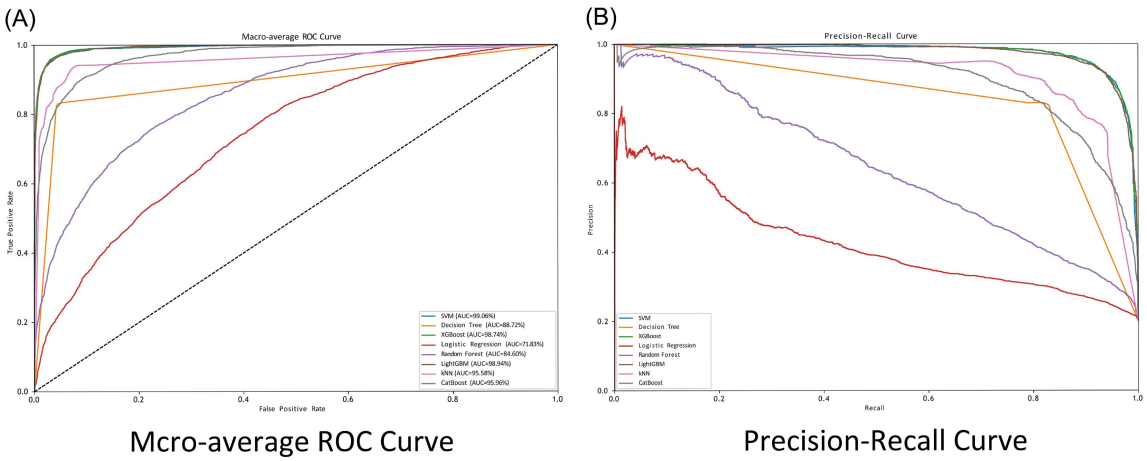


Figure 3. Receiver operating characteristic (ROC) and precision-recall (PR) curves. (A) Macro-averaged ROC curves illustrating the classification performance of eight models across different thresholds; (B) PR curves illustrating the relationship between precision and recall, reflecting the accuracy and completeness of model predictions at varying thresholds.

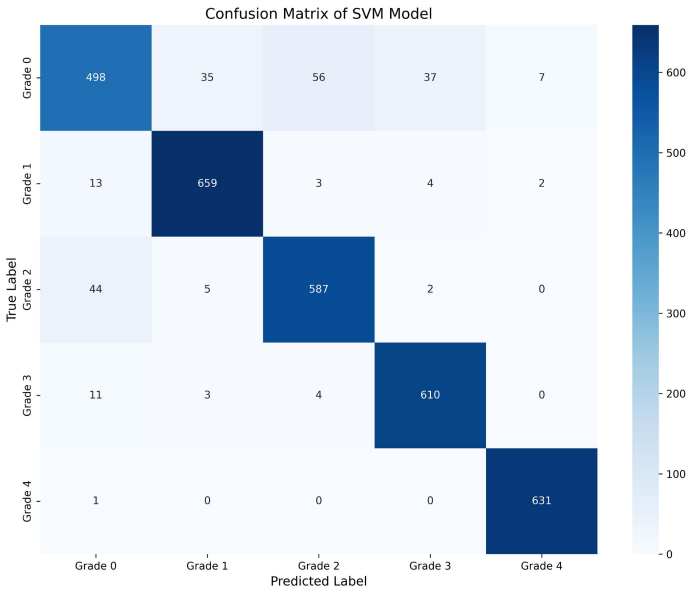


Figure 4. Confusion matrix of the support vector machine (SVM) model for classifying different Kellgren-Lawrence (KL) grades of knee osteoarthritis. The color gradient represents the number of samples, with darker blue indicating a higher count of samples in that cell.

Table 3. Recall (sensitivity) of the support vector machine (SVM) model for each Kellgren-Lawrence (KL) grade of knee osteoarthritis

KL Grade	Recall (Sensitivity)
0	0.7867
1	0.9677
2	0.9201
3	0.9713
4	0.9984

(56 cases). These misclassifications, considered "conservative errors" in clinical practice, can be corrected through further imaging reviews without leading to serious diagnostic risks. Overall, the balancing strategy significantly improves the detection capability for minority classes, especially Grade 4, without inflating overall performance metrics, aligning with clinical diagnosis requirements.

Results of model interpretation

A total of 18 features were ultimately selected for KL grading diagnosis of knee osteoarthritis. Features were extracted using different processing methods and statistical combinations, as well as from diverse data sources. **Table 4** summarizes various key features used in image modeling and analysis, including first-order statistical features (e.g., Feature 04, 18) and features derived from mathematical transformations such as mean absolute deviation, robust mean absolute deviation, entropy, and kurtosis (e.g., Feature 08, 11, 15).

SHAP-based local interpretability analysis

SHAP was used to interpret feature contributions and model decisions. SHAP, based on Shapley values from cooperative game theory,

Table 4. Abbreviated feature identifiers and their corresponding definitions

Features	Corresponding ID
square_firstorder_InterquartileRange	Feature 01
squareroot_firstorder_InterquartileRange	Feature 02
square_firstorder_MeanAbsoluteDeviation	Feature 03
diagnostics_Image-original_Minimum	Feature 04
wavelet-LL_firstorder_InterquartileRange	Feature 05
original_firstorder_InterquartileRange	Feature 06
square_firstorder_Entropy	Feature 07
gradient_firstorder_Skewness	Feature 08
original_firstorder_Uniformity	Feature 09
wavelet-LL_glrlm_RunEntropy	Feature 10
square_firstorder_RobustMeanAbsoluteDeviation	Feature 11
wavelet-LL_firstorder_MeanAbsoluteDeviation	Feature 12
wavelet-LL_firstorder_RobustMeanAbsoluteDeviation	Feature 13
square_firstorder_Kurtosis	Feature 14
squareroot_firstorder_RobustMeanAbsoluteDeviation	Feature 15
exponential_firstorder_InterquartileRange	Feature 16
exponential_firstorder_RobustMeanAbsoluteDeviation	Feature 17
original_firstorder_RobustMeanAbsoluteDeviation	Feature 18

Note: Interquartile Range (IQR) represents the distribution range of the middle 50% of pixel values.

calculates each feature's contribution to predictions, offering an intuitive understanding of how features affect model outputs. **Figure 5** presents the feature importance distribution for the SVM model in the grading task. Overall, Feature 07 and Feature 12 significantly impacted model's predictive performance. A high SHAP value of Feature 07 indicates its core role in assessing lesion severity. Feature 07, a metric quantifying the disorder of image gray value distribution, directly links to key radiographic signs of knee X-rays across KL grades. For KL Grade 0-1, low entropy aligns with the uniform gray distribution of normal joint spaces or mild lip-like osteophytes in Grade 1, as bone tissue maintains high-density uniformity with minimal gray fluctuations. For Grades 2-3, significantly increased entropy correlates with density heterogeneity from definite osteophytes and patchy subchondral sclerosis—osteophyte proliferation enhances local gray-level fluctuations, while contrast between sclerotic and normal bone further boosts entropy. For Grade 4, peak entropy matches the complex structure of large osteophytes and marked joint space narrowing; interleaved bone destruction-repair regions cause the highest gray distribution disorder, consistent with the clinical radiographic feature of 'density inhomogeneity in severe lesions'. This mapping confirms entropy's clinical value as a texture metric and provides an interpretable radiological basis for the model's decisions. Subplots (B-F) display the direction and intensity of feature contributions through swarm plots. The x-axis of the SHAP summary plot represents SHAP values, where positive values indicate that the feature drives predictions toward a specific class, while negative values suggest the opposite. For class 0, Feature 04 significantly impacts model predictions, and a broader distribution of feature values may lead to variations in model predictions. In class 1, Feature 15 and Feature 18 negatively impact model predictions. For class 2, Feature 07 and Feature 12 show significant positive impacts. In class 3, Feature 07 has a strong positive influence. For class 4, Feature 14 positively affects predictions but has a wide distribution, indicating variable impacts on the model output.

This study employed random selection of 10 samples from each class to generate the SHAP decision plot presented in **Figure 6**, illustrating the feature influences on the SVM model across different classes. The horizontal axis represents the SHAP values, where positive values denote that a feature propels the prediction towards the positive class, while negative values indicate a propensity towards the negative class. Features are ranked along the vertical axis in descending order of their importance to the model's output. It can be observed that

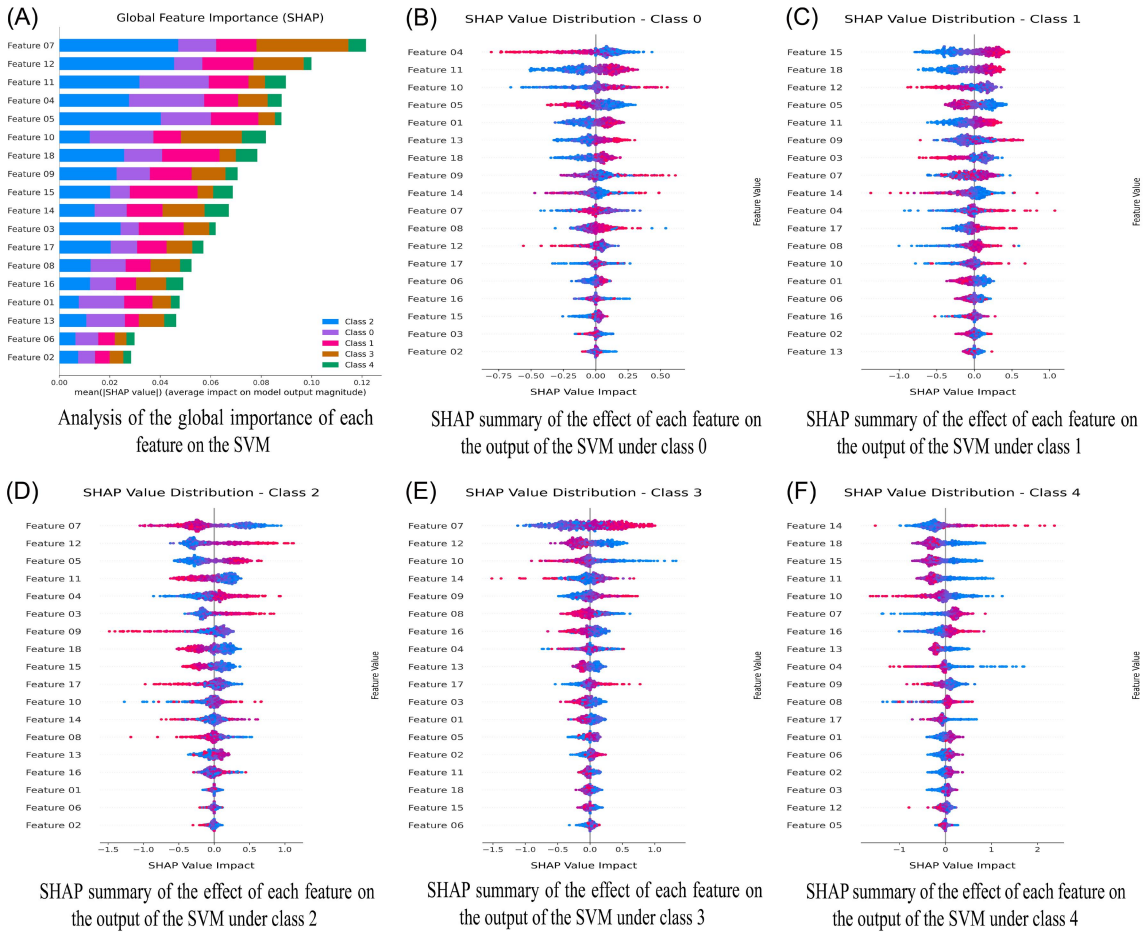


Figure 5. Summary of global feature importance and class-specific feature impact analysis of the support vector machine (SVM) model using SHapley Additive exPlanations (SHAP). (A) Ranking of features by their average SHAP values across all categories; (B-F) Feature's impact plots for classes 0-4, showing each feature's influence on the SVM output. Each point represents an individual sample's SHAP value, with colors indicating feature magnitude.

within Class 0, Feature 04 exerts a positive contribution to the prediction outcome. Features in Class 1 exhibited greater variability in their SHAP values, suggesting a potential interdependence between Feature 12 and Feature 04. Feature 12 in Class 2 demonstrated a distinct negative influence on the model's predictive results. In Class 3, Feature 07 and Feature 10 exhibited analogous patterns, with higher values associated with positive contributions and lower values correlated with negative contributions. Finally, elevated values of Feature 14 and Feature 18 in Class 4 were found to positively impact the model's prediction results.

LIME-based local interpretability analysis

In this study, LIME was used to visualize the feature importance of the SVM model from the perspective of individual prediction samples. As shown in **Figure 7**, features such as Feature 03 and Feature 09 exerted significant positive impacts on positive predictions, indicating a strong positive correlation with the target variable. Feature 16 exerted a notable negative influence, suppressing model predictions. Overall, features with positive impacts predominate.

Discussion

This study utilized a portion of knee X-ray images from the OAI organization. After inversion processing and 2D radiomics feature extraction, 18 key features were selected using variance thresholding and ANOVA. To address data imbalance, SMOTE sampling and class weight balancing were applied. In the model evaluation, eight machine learning models were compared, with the SVM model performing the best.

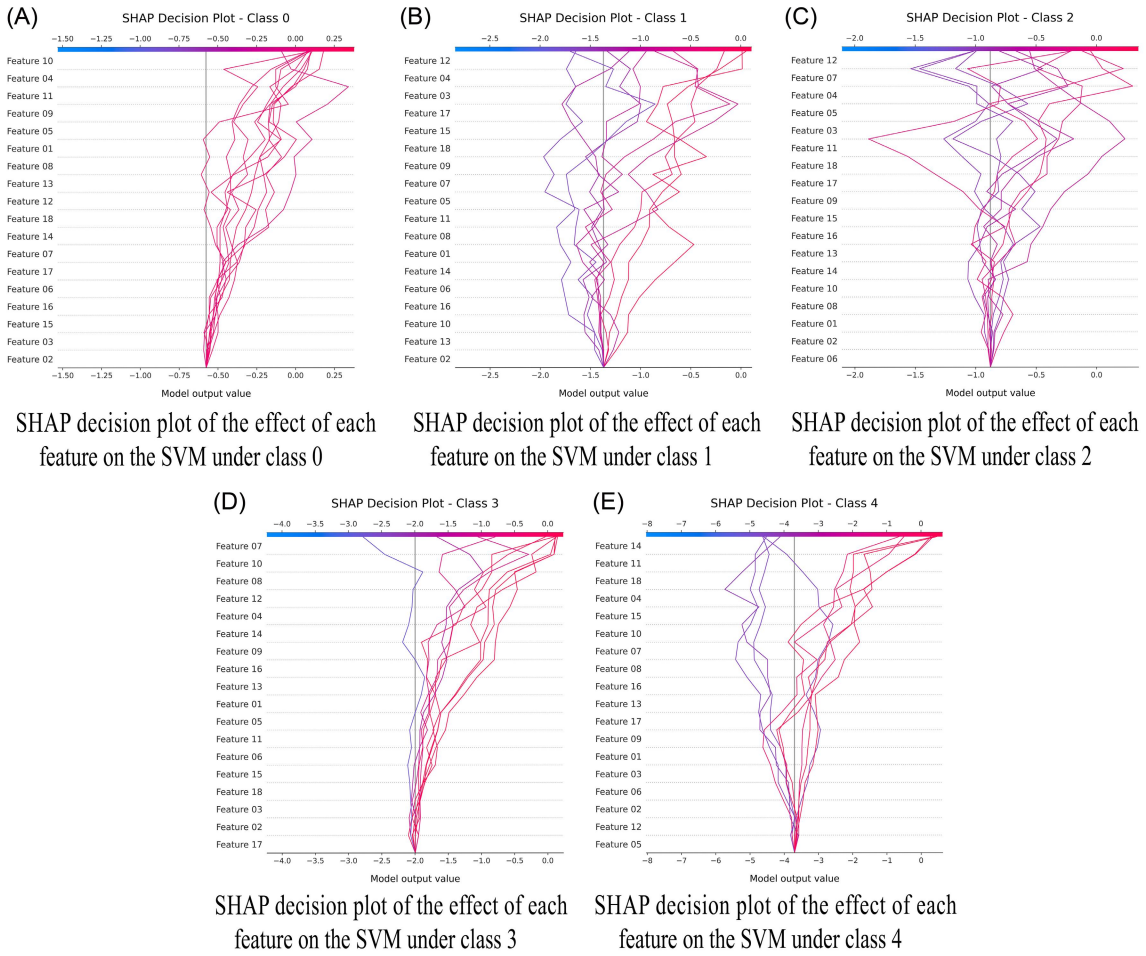


Figure 6. SHapley Additive exPlanations (SHAP)-based decision plot illustrating feature influence on the support vector machine (SVM) model. (A-E) Visualization of the decision-making process of the SVM model for Classes 0 to 4. The influence of each feature is represented by line segments, with the distance from the central axis being proportional to the magnitude of its impact on the model output. Feature values are color-coded, with blue indicating lower values and red indicating higher values.

The model's decision-making process was analyzed using SHAP and LIME to identify key features. This work provides a reliable reference for clinical practice, enhancing the model's credibility and practicality.

Prior studies mainly focused on the binary classification of knee OA, aiming only to determine disease presence. For instance, study of G.B. Joseph et al. achieved an AUC of 0.772, and Xu Wang et al. obtained an AUC of 96.5% and an F1 score of 93.5% [18, 19]. Despite technical advancements in these studies, they fell short in model interpretability and lacked in-depth analysis of feature contributions to predictions. This study advances to a multi-classification framework using the KL scale and various ML models, enabling more precise OA severity prediction. Additionally, SHAP and LIME were employed to enhance the interpretability of the models. As shown in **Table 4** and **Figure 5A**, Feature 07, by quantifying image texture complexity, helps the model detect subtle changes in diseased areas [20, 21]. This is crucial for early diagnosis and monitoring disease progression. Feature 12 reflects the uniformity of low-frequency image components, providing valuable information on joint structural integrity and aiding in distinguishing between normal and pathological joints. **Figures 5** and **6** reveal the impact of features on different KL grades, underscoring their role in model predictions. In classes 0 and 1, Feature 04 helps standardize image processing and distinguish pathological from non-pathological low signals by assessing the discreteness of the lowest signal intensity areas in the patellar image, thus improving grading accuracy. In classes 1 and 2, Feature 12 shows high importance with significant fluctuations, aligning with the global feature impor-

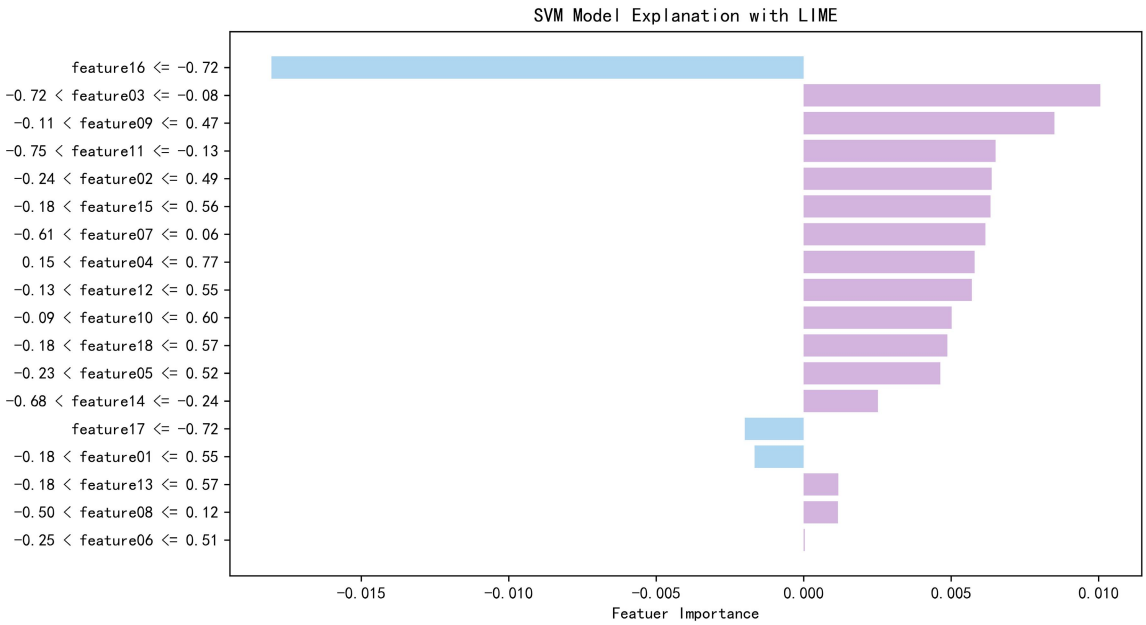


Figure 7. Local feature importance for individual samples interpreted by Local Interpretable Model-agnostic Explanations (LIME). The figure illustrates the influence of each feature on model predictions, ranked by importance. The length of bars represents feature importance, with negative values indicating a negative impact on predictions and positive values indicating a positive impact. The inequality next to the feature name shows the feature value's threshold range.

tance analysis. In class 3, Feature 07 exhibits a clear directional impact on predictions. The varying SHAP values of these features influence model predictions differently, highlighting the transparent and visual nature of the decision-making process. A transparent decision-making process enhances clinicians' trust in model outputs and improves diagnostic precision. LIME further validates the dynamic role of features across different samples, providing clinicians with a clear basis for decision-making.

Nevertheless, this study still has several limitations. X-ray image data were exclusively sourced from the OAI database. The absence of external data from other institutions or imaging modalities might restrict the model's generalizability across diverse clinical scenarios. Additionally, the dataset suffered from significant class imbalance. Future work will incorporate multimodal data from multiple institutions to evaluate the model's generalizability and combine radiomics with deep learning features to enhance sensitivity in early diagnosis.

Conclusion

This study investigated the five-class classification of knee osteoarthritis using multiple machine learning models and demonstrated that SVM exceled in classification accuracy. The model's decision-making process was further elucidated through SHAP and LIME analyses, providing valuable insights into how key features influenced predictions.

Author contributions: He Ren, Yina Zhang, Yutong Xie contributed to the conception of the study. Anqi Wu, Xianglun Kong, Chenxiao Bai performed the experiment and contributed significantly to analysis and manuscript preparation. Miao Yu, Yimeng Wang helped perform the analysis with constructive discussions. Ping Li oversaw the project, offered key revisions to the manuscript, and gave approval for the final version.

References

[1] Hall M, van der Esch M, Hinman RS, et al. How does hip osteoarthritis differ from knee osteoarthritis? Osteoarthritis Cartilage 2022;30(1):32-41.

[2] Gelber AC. Knee Osteoarthritis. Ann Intern Med 2024;177(9):ITC129-ITC144.

[3] Kloppenburg M, Namane M, Cicuttini F. Osteoarthritis. Lancet 2025;405(10472):71-85.

[4] Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. Lancet 2023;402(10416):2019-2033.
[5] Deng P, Wang S, Sun X, et al. Global Trends in Research of Gouty Arthritis Over Past Decade: A Bibliometric Analysis. Front Immunol 2022;13:910400.
[6] Park SY, Wooden TK, Pekas EJ, et al. Effects of passive and active leg movements to interrupt sitting in mild hypercapnia on cardiovascular function in healthy adults. J Appl Physiol (1985) 2022;132(3):874-887.
[7] Harden JE, David Branch J, Reynolds LJ. Impacts of Physical Inactivity Models on Endothelial Function: A Systematic Review. Sports Med 2025;55(8):1937-1952.
[8] Kaya AT, Akman B. Relationship of the Novel Scoring System for Lower Extremity Venous Thrombosis with Pulmonary Embolism. Acad Radiol 2024;31(9):3811-3824.
[9] Baranga L, Khanuja S, Scott JA, et al. In Situ Pulmonary Arterial Thrombosis: Literature Review and Clinical Significance of a Distinct Entity. AJR Am J Roentgenol 2023;221(1):57-68.
[10] Kahn Susan R, de Wit K. Pulmonary Embolism. N Engl J Med 2022;387(1):45-57.
[11] Ou J, Zhang J, Alswadeh M, et al. Advancing osteoarthritis research: the role of AI in clinical, imaging and omics fields. Bone Res 2025;13(1):48.
[12] Li X, Roemer FW, Cicuttini F, et al. Early knee OA definition-what do we know at this stage? An imaging perspective. Ther Adv Musculoskelet Dis 2023;15:1759720X231158204.
[13] Ehmig J, Engel G, Lotz J, et al. MR-Imaging in Osteoarthritis: Current Standard of Practice and Future Outlook. Diagnostics (Basel) 2023; 13(15):2586.
[14] Li D, Lu H, Wu J, et al. Development of machine learning models for predicting depressive symptoms in knee osteoarthritis patients. Sci Rep 2024;14(1):28603.
[15] Pabinger C, Lothaller H, Kobinia GS. Intra-articular injection of bone marrow aspirate concentrate (mesenchymal stem cells) in KL grade III and IV knee osteoarthritis: 4 year results of 37 knees. Sci Rep 2024;14(1):2665.
[16] Farooq MU, Ullah Z, Khan A, et al. DC-AAE: Dual channel adversarial autoencoder with multi-task learning for KL-grade classification in knee radiographs. Comput Biol Med 2023; 167:107570.
[17] Li G, Li S, Xie J, et al. Identifying changes in dynamic plantar pressure associated with radiological knee osteoarthritis based on machine learning and wearable devices. J Neuroeng Rehabil 2024;21(1):45.
[18] Joseph GB, McCulloch CE, Nevitt MC, et al. Machine learning to predict incident radiographic knee osteoarthritis over 8 Years using combined MR imaging features, demographics, and clinical factors: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2022;30(2):270-279.
[19] Wang X, Wang T, Su Z. Attention-enhanced deep learning and machine learning framework for knee osteoarthritis severity detection in football players using X-ray images. Journal of Radiation Research and Applied Sciences 2025;18(2):101428.
[20] Spotorno N, Strandberg O, Stomrud E, et al. Diffusion MRI tracks cortical microstructural changes during the early stages of Alzheimer's disease. Brain 2024;147(3):961-969.
[21] Kelly BS, Mathur P, McGuinness G, et al. A Radiomic "Warning Sign" of Progression on Brain MRI in Individuals with MS. AJNR Am J Neuroradiol 2024;45(2):236-243.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/446813bjkhvg

Multi-method fusion for image segmentation in skin disease analysis

Siqi Wang*, Danhong Li*, Yina Zhang*, Yu Wang, Linrong Yuan, Miao Yu, Jianghui Li, Yimeng Wang, Ping Li

Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China.
*These authors are co-first authors.

Corresponding author: Ping Li.

Declaration of conflict of interest: None.

Declaration of ethics: The study was approved by the Ethics Review Committees of the Medical University of Vienna and the University of Queensland.

Declaration of patient consent: The dataset underwent automated screening using neural networks, followed by multiple manual reviews. All EXIF metadata were removed to eliminate any potentially identifiable information. Therefore, the data are considered anonymized to the best of our knowledge.

Received July 26, 2025; Accepted November 4, 2025; Published December 31, 2025

Highlights

- For the first time, the advantages of two deep learning architectures—SegNet and U-Net—were integrated by averaging their prediction outputs. This ensemble approach overcame the limitations of single models and substantially improved segmentation accuracy.
- The proposed Ensemble Model outperformed both SegNet and U-Net across all major evaluation metrics, including the Intersection over Union (IoU, 93.73%), Dice coefficient (84.85%), precision (93.93%), and loss (0.63), confirming the effectiveness of multi-method fusion.
- Considering the complex morphology and indistinct lesion boundaries characteristic of skin diseases, a standardized preprocessing and data augmentation pipeline was developed to enhance the model's robustness in handling diverse lesion patterns.

Abstract

Objective: The morphological complexity of dermatologic diseases poses considerable challenges to clinical diagnosis. Conventional manual interpretation of skin images is time-consuming and influenced by subjective variability, which limits diagnostic accuracy. Hence, developing advanced medical image segmentation techniques through multi-method fusion is of particular importance. **Methods:** A comprehensive dataset of dermatologic images was utilized and rigorously preprocessed to ensure reliability and consistency. Two representative deep learning models, SegNet and U-Net, were optimized to achieve precise delineation of lesion areas. Building upon their complementary strengths, a novel fusion-based image segmentation framework was proposed, integrating both models to enhance performance through synergistic learning. The effectiveness of the ensemble strategy was validated through extensive experiments using standard evaluation metrics, including the Dice coefficient and Intersection over Union. **Results:** Compared with each individual model, the Ensemble Model yielded consistent improvements across all evaluation indices, with a notable reduction in loss values. These enhancements indicate markedly better learning efficiency and generalization in dermatologic image segmentation tasks. **Conclusion:** By integrating multiple deep learning algorithms, this fusion technique solves the misclassification and omission

Address correspondence to: Ping Li, Faculty of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, 279 Zhouzhu Highway, Pudong New Area, Shanghai 201318, China. Tel: +86-13764055848. E-mail: lip@sumhs.edu.cn.

issues observed in single-model segmentation. It significantly improves overall segmentation accuracy and demonstrates superior performance, particularly in edge detection of complex skin lesions.

Keywords: Skin cancer, deep learning, image segmentation, multi-method fusion

Introduction

Dermatological conditions are prevalent clinical disorders characterized by diverse morphological manifestations, which pose significant challenges to accurate diagnosis and effective treatment [1]. Traditional manual image review is not only inefficient but also prone to subjective bias, thereby limiting diagnostic accuracy. In recent years, advancements in deep learning and computer vision technologies have facilitated the automated recognition and segmentation of dermatological images. However, it is evident that single image processing methods or deep learning techniques alone are insufficient to address challenges such as dataset imbalance, noise interference, and the morphological diversity of lesions [2-4].

Multi-method fusion techniques for medical image segmentation effectively overcome the limitations of single-model approaches by integrating multiple deep learning algorithms. This strategy substantially enhances segmentation accuracy and demonstrates particularly robust performance in edge detection tasks [5-7]. Moreover, the approach provides high flexibility and broad applicability, allowing the selection and integration of the most appropriate segmentation algorithms according to the characteristics of different medical images. Such adaptability enables clinicians to formulate individualized treatment plans, thereby enhancing therapeutic outcomes and improving patients' quality of life [8, 9].

In the field of medical image processing, the development of deep learning can be traced back to the early 1990s. In 1993, convolutional neural networks (CNNs) were first applied to lung nodule detection, representing one of the earliest and most influential uses of deep learning in medical image analysis. Two years later, in 1995, CNNs were employed to identify microcalcifications in mammograms, further demonstrating their potential in medical image recognition. Around 1998, Yann LeCun introduced the well-known LeNet-5 model, an early and successful implementation of CNNs for image recognition tasks, which laid a solid foundation for subsequent advances in medical image analysis [10]. A major breakthrough occurred in 2012, when the AlexNet achieved remarkable success in the ImageNet competition, catalyzing a worldwide surge of research in deep learning and profoundly influencing medical image processing [11]. In 2015, the introduction of the U-Net model marked another milestone in the field. This CNN architecture, specifically designed for biomedical image segmentation, greatly enhanced the accuracy and efficiency of medical image analysis [12]. In 2017, the advent of Residual Pyramid Networks and SegNet further propelled progress in medical image segmentation, introducing new architectures and methodologies that have since shaped modern research directions [13, 14].

This study primarily investigates deep learning-assisted segmentation techniques for dermatological images, as well as the design and validation of strategies for combining multiple methods to achieve segmentation. The research comprises four main components, as illustrated in **Figure 1**: first, data collection and preprocessing are conducted; second, the SegNet and U-Net models are studied and improved; third, a multi-method fusion segmentation strategy is proposed; and finally, the effectiveness and advantages of the proposed approach are validated through experiments and performance evaluations, such as the Dice coefficient and Intersection over Union (IoU) metrics. The objective of this study is to establish a multi-method fusion image segmentation framework that enhances the accuracy and stability of dermatological image segmentation, thereby providing more reliable auxiliary support for clinical diagnosis and treatment.

Materials and methods

Data source

The HAM10000 dataset employed in this study is a well-established, large-scale collection of dermatological images widely utilized in skin disease research [15, 16]. It was publicly released in 2018 by a research team from the Medical University of Vienna. Designed to support the development of deep learning and machine learning tools for early skin cancer

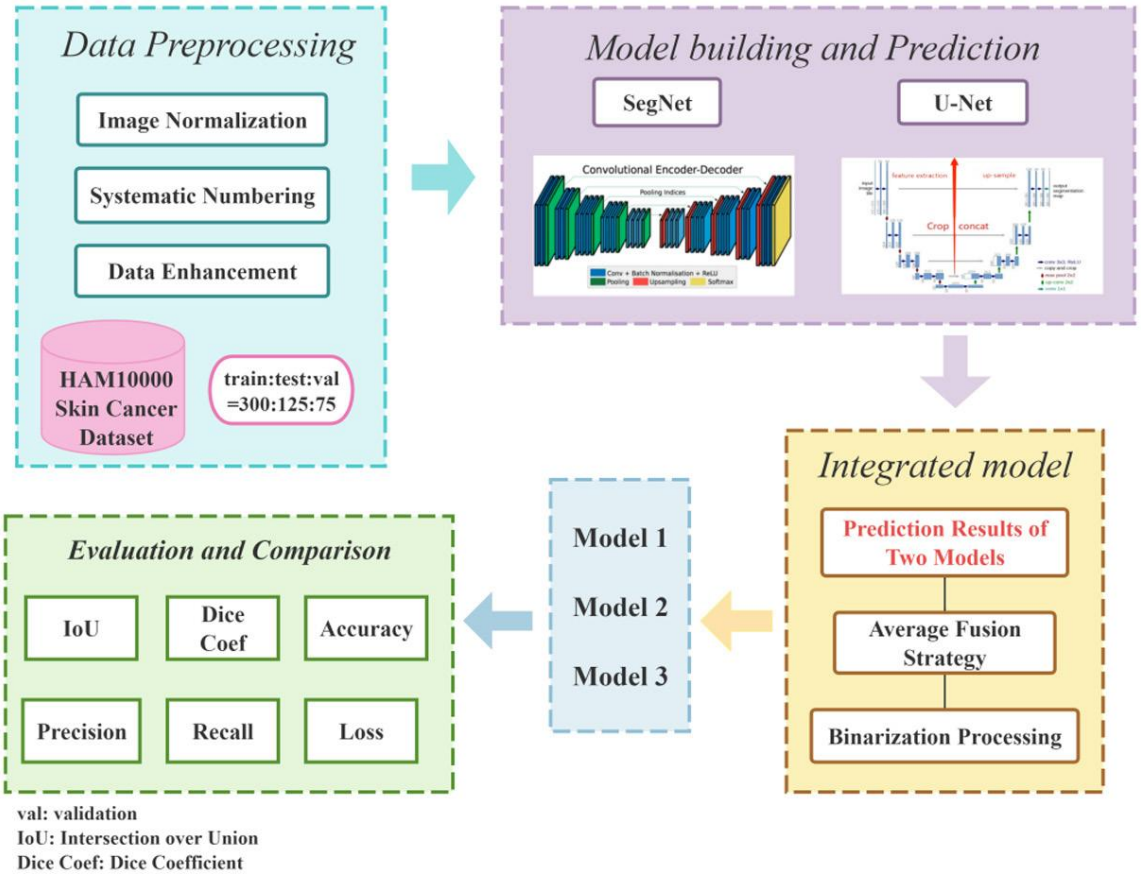


Figure 1. Flowchart.

diagnosis and treatment assistance, this dataset includes more than 10,000 high-resolution images of skin lesions. Each image has been carefully reviewed and annotated with pathological classification labels by board-certified dermatologists. The dataset covers various types of skin lesions, including malignant conditions such as melanoma, basal cell carcinoma, and squamous cell carcinoma, along with a substantial number of benign cases.

Data processing

Data processing is essential for skin lesion image segmentation tasks. In this study, we utilized the HAM10000 skin cancer dataset and performed preprocessing, augmentation, and partitioning steps. First, all images from the HAM10000 dataset were standardized to a uniform resolution of 192×256 pixels to avoid analytical bias caused by dimensional inconsistencies and to improve model learning efficiency. Gaussian filtering was employed to reduce image noise and enhance meaningful features. Pixel values were also normalized to the range [0, 1] to facilitate faster and more stable model convergence. To increase sample diversity and simulate variations in real imaging conditions, data augmentation techniques—including random rotation, horizontal flipping, and color adjustment—were employed. This process helps alleviate overfitting caused by limited data and enhances the model's adaptability and robustness in real clinical environments. Finally, the dataset was divided into a training set (300 images), a test set (125 images), and a validation set (75 images). These subsets were used for model training, generalization assessment, and hyperparameter optimization, respectively. This segmentation approach aims to identify the optimal model architecture and parameters, ultimately enhancing the accuracy and stability of segmentation outcomes.

Model architecture

This study focuses on two deep learning-based image segmentation models: SegNet and U-Net. These architectures have been widely applied in medical image segmentation and

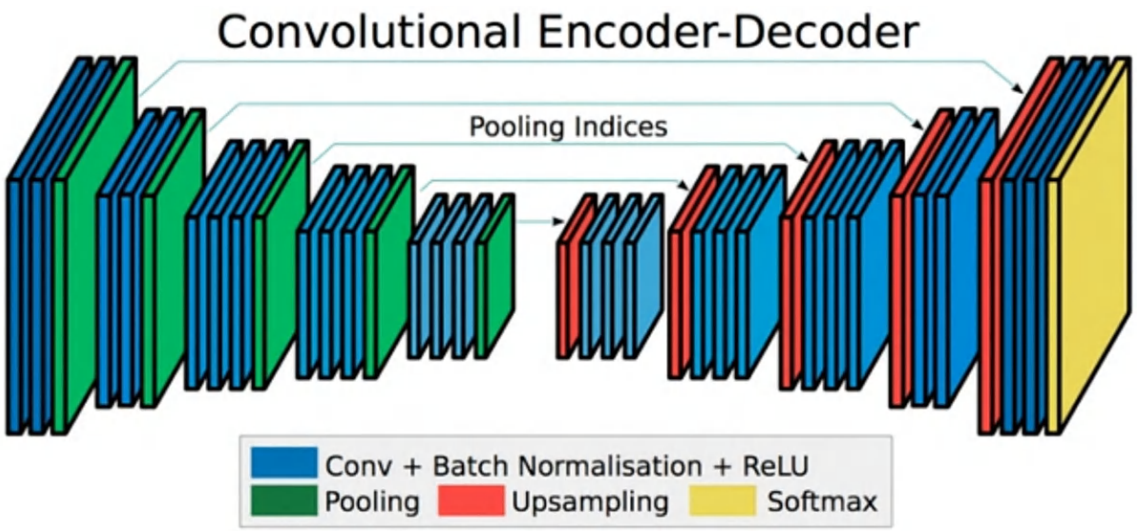


Figure 2. SegNet model architecture diagram, illustrating the encoder-decoder structure of the model. Skip connections enhance the propagation of feature information, optimizing memory and computational efficiency, which makes SegNet well-suited for resource-constrained environments.

have demonstrated strong performance in dermatological image processing [17, 18].

SegNet has established a prominent position in the field of image segmentation due to its efficient and sophisticated design. As illustrated in the model architecture diagram in **Figure 2**, its decoder adopts an innovative spatial information reconstruction strategy. Specifically, it reuses the pooling indices generated during max-pooling in the encoder to perform non-linear upsampling. This method effectively preserves spatial structure information from the original image without distortion and contributes significantly to the accurate restoration and localization of object edges. As a result, the model is able to capture the contour boundaries of target objects clearly and precisely. Furthermore, SegNet has been extensively optimized for computational efficiency and memory usage. Its lightweight architecture enables stable performance even in resource-constrained real-world environments.

U-Net, also known as the U-shaped network, is a pioneering deep learning architecture for medical image segmentation. It has garnered significant attention due to its unique encoder-decoder structure. The core of U-Net's design is its progressively layered convolutional processing pipeline. In the encoder path, image features are extracted and the spatial resolution is reduced through a series of convolutional and downsampling layers. Conversely, the decoder path gradually restores resolution using upsampling operations (e.g., transposed convolutions) to reconstruct detailed segmentation masks. A key feature of U-Net is its skip connections, which concatenate feature maps from the encoder with corresponding maps in the decoder, thereby preserving more contextual information when reconstructing image details. This mechanism enables U-Net to deliver outstanding segmentation accuracy and generalization capabilities even with small-scale datasets.

This study involved the detailed implementation and optimization of the SegNet and U-Net models. The SegNet model takes preprocessed Red, Green, Blue images with a resolution of 192×256 pixels as input. Following processing through its encoder-decoder architecture, it produces binary segmentation maps. In contrast, the U-Net model leverages its distinctive skip-connections to achieve superior segmentation accuracy and finer detail preservation. For training, both models were optimized using the binary cross-entropy loss function and the Adam optimizer, augmented with a learning rate decay strategy to enhance performance.

Evaluation indicators

This study employs five key metrics to thoroughly evaluate the model's segmentation performance.

Jaccard index (IoU)

The Jaccard index, often referred to as IoU, quantifies the overlap between the model's predicted segmentation region and the manually annotated ground truth. It is calculated using Formula 1:

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

In this formula, True Positives (TP) denotes pixels correctly identified as the target class; False Positives (FP) represents pixels incorrectly labeled as the target class; and False Negatives (FN) refers to pixels that belong to the target class but are mistakenly predicted as background. A higher IoU value, especially one approaching 1, indicates more accurate segmentation results.

Dice coefficient

As another metric for assessing segmentation accuracy, the Dice coefficient is particularly useful in binary classification scenarios involving overlapping regions. It is calculated using Formula 2:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (2)$$

Although the Dice coefficient is similar to the IoU, it places greater emphasis on the size of the intersection between the predicted and ground truth regions. As with IoU, a Dice value closer to 1 indicates a more precise segmentation outcome.

Precision

Precision measures the fraction of samples predicted as positive that are correctly identified. It is computed using Formula 3:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

A higher precision value indicates that the model makes fewer false positive errors, meaning its positive predictions are more reliable.

Recall

Recall assesses a model's effectiveness in identifying all relevant positive instances within a dataset. It measures the proportion of actual positives that are correctly identified, as calculated by Formula 4:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

A higher recall value indicates that the model misses fewer positive samples, resulting in a more comprehensive detection of the positive class.

Accuracy

Accuracy measures the overall correctness of a model by calculating the proportion of correctly predicted samples out of the total. It is computed using Formula 5:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Here, True Negatives (TN) represents samples correctly classified as the negative class. This ratio offers a clear indication of the model's general classification performance.

Together, these metrics allow a comprehensive assessment of the model, providing a scientific basis for its optimization and selection. Among them, the Jaccard index (IoU) and Dice coefficient are particularly notable for their effectiveness in evaluating overlap quality in segmentation tasks. In contrast, precision, recall, and accuracy function as general-purpose metrics that are applicable not only to classification but also to a wide range of other evaluation scenarios.

Results

This study conducted a systematic evaluation of SegNet and U-Net for dermatological image segmentation. The assessment encompassed key performance metrics, including IoU, Dice coefficient, precision, recall, and accuracy, alongside an analysis of the training dynamics. Both models demonstrated steady and robust performance throughout the experiments. Notably, U-Net exhibited superior generalization capabilities and achieved higher accuracy, a finding supported by the empirical data in **Table 1** and the visual comparisons in **Figure 3**.

Building upon these initial results, we developed a lightweight fusion strategy. The core of

Table 1. Performance comparison of SegNet and U-Net across different datasets

Dataset	IoU (%)	Dice (%)	Precision (%)	Recall (%)	Accuracy (%)
SegNet (Train)	96.64	79.67	94.63	95.13	97.14
SegNet (Val)	91.77	73.51	85.00	85.85	91.61
SegNet (Test)	91.31	70.80	85.77	80.41	91.18
U-Net (Train)	95.38	87.36	93.26	85.76	94.52
U-Net (Val)	94.17	85.78	91.57	85.07	93.41
U-Net (Test)	93.53	84.29	90.98	81.47	92.78

Note: IoU, Intersection over Union; Dice, Dice coefficient; Train, training set; Val, validation set; Test, test set.

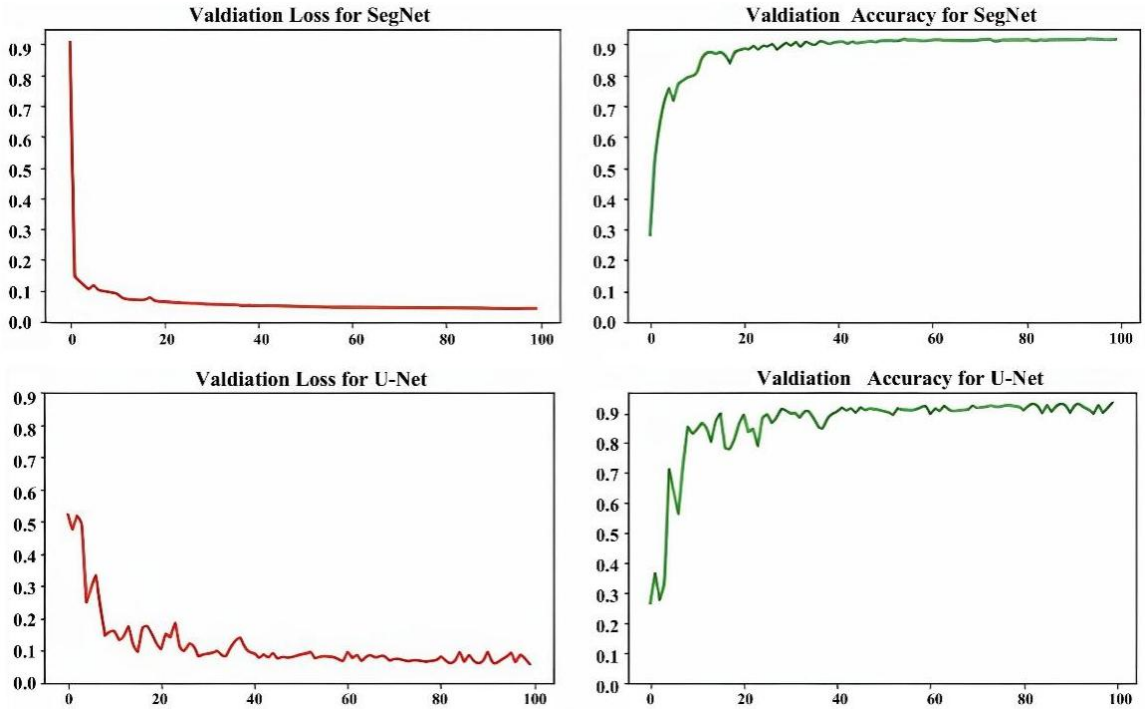


Figure 3. Training statistics of SegNet and U-Net on the validation set, illustrating the trends in loss and accuracy for both models. These curves reflect the training dynamics and generalization performance of the two architectures.

this strategy is a two-step process: first, averaging the probability outputs from the two pretrained (and fixed) models; second, applying a threshold of 0.7 to binarize the averaged results. This approach is notable for enhancing segmentation accuracy without requiring additional model retraining.

To validate the effectiveness of the fused model, we refer to **Table 2** and **Figure 4**. **Table 2** presents numerical results demonstrating the model's advantages across all evaluated metrics—IoU, Dice score, precision, recall, accuracy, and loss. **Figure 4** offers complementary visual evidenc. Together, these results confirm that the Ensemble Model consistently outperforms the individual SegNet and U-Net models. This comprehensive performance advantage provides a solid basis for the subsequent in-depth analysis.

Model performance evaluation

This study conducted a thorough evaluation of two deep learning models, SegNet and U-Net, for dermatological image segmentation tasks (detailed numerical results are shown in **Table 1**). The U-Net model outperformed SegNet across multiple evaluation metrics, especially in terms of generalization capability, segmentation accuracy, and recognition completeness. U-Net achieved higher IoU values on the training, validation, and test datasets, indicating stronger adaptability to diverse data distribu-

Table 2. Performance comparison of the three models on the test set

Model	IoU (%)	Dice (%)	Precision (%)	Recall (%)	Accuracy (%)	Loss
SegNet	91.31	70.80	85.77	80.41	91.18	24.09
U-Net	93.53	84.29	90.98	81.47	92.78	6.47
Ensemble	93.73	84.85	93.93	87.38	92.82	0.63

Note: IoU, Intersection over Union; Dice, Dice coefficient; Loss, cross-entropy loss.

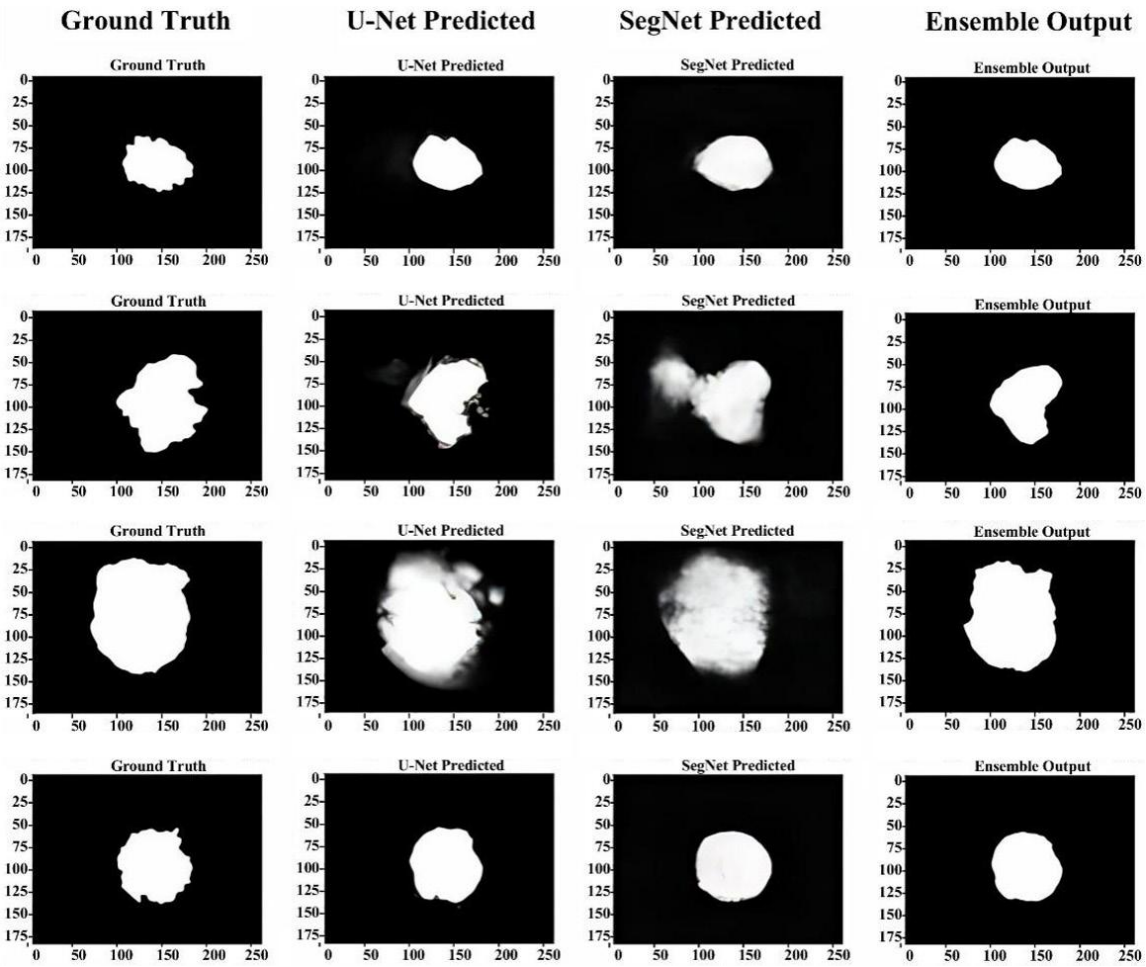


Figure 4. Comparison of segmentation results from different models on the validation set. The Ensemble Model demonstrates superior performance in capturing edge details.

tions and a greater ability to deliver stable and accurate segmentation results. Furthermore, in metrics such as the Dice coefficient, precision, and recall, U-Net either surpassed or matched SegNet, demonstrating superior performance in both precise localization of target regions and comprehensive coverage of annotated pixels. In contrast, though SegNet has higher loss function values, it showed specific advantages in training efficiency and recall. This suggests that SegNet may learn and capture fundamental image features more rapidly during the initial training phase, leading to faster convergence compared to U-Net. Additionally, SegNet's higher recall indicates its superior ability to identify all positive-class instances, enabling more comprehensive detection of target regions.

U-Net's strengths primarily derive from its encoder-decoder architecture and skip-connections, which effectively preserve edge and detail information in medical images – a crucial factor for precise segmentation. Furthermore, U-Net's parameter efficiency and generalization capability enable it to achieve favorable training adaptability and robust performance even with

limited training data. In contrast, SegNet's advantages lie in its spatial information reconstruction techniques and memory optimization, granting it distinct value in computational resource-limited scenarios. Overall, while U-Net exhibits superior segmentation performance in this study, SegNet maintains competitiveness for specific practical applications.

Figure 3 presents the performance of the SegNet and U-Net models on the validation set. The top-left panel shows the validation loss of SegNet over training iterations, which decreases rapidly in the initial phase before stabilizing, indicating effective learning. The top-right panel shows that SegNet's validation accuracy rises quickly and stabilizes at a high level. In the lower-left panel, U-Net's validation loss also declines rapidly and stabilizes, yet remains consistently lower than that of SegNet, reflecting its superior data-fitting ability. The bottom-right panel demonstrates that U-Net's validation accuracy increases rapidly and maintains a higher value than SegNet. These results indicate that both models effectively capture patterns from the training data without overfitting risk. Furthermore, U-Net outperforms SegNet on the validation set in terms of generalization capability and segmentation accuracy, which is consistent with the quantitative metrics reported earlier.

Ensemble model performance evaluation

In this study, a straightforward yet effective fusion strategy was employed to combine predictions from the SegNet and U-Net models for enhanced dermatological image segmentation accuracy. Specifically, pre-trained SegNet and U-Net models were first used to generate independent predictions on the validation set images, producing probability values or real-valued outputs for each pixel category. The prediction outputs from both models were then converted into one-dimensional vectors by flattening the original 2D or 3D pixel matrices, resulting in two 1D arrays designated as sub1 and sub2 respectively.

To integrate the predictions from both models, we constructed a new vector sub by computing the element-wise average of the corresponding results in sub1 and sub2. averaging strategy enhances the final prediction accuracy by leveraging the complementary strengths of both models. Subsequently, a binarization strategy was applied to each element in the sub vector: values exceeding the optimal threshold of 0.7 (determined through grid search) were set to 1, while others were set to 0. The 0.7 threshold was selected to convert continuous prediction values into discrete category labels. In this binary segmentation task, pixels with values above 0.7 are classified as the target category (assigned 1), while those below are classified as background (assigned 0). This approach enables precise pixel-level segmentation by effectively distinguishing foreground from background regions.

Figure 4 compares the segmentation outputs of SegNet, U-Net, and the Ensemble Model against the ground truth on the validation set. Although both U-Net and SegNet demonstrate effective lesion identification in certain cases, they exhibit noticeable differences in handling edge details. In contrast, the Ensemble Model shows superior alignment with the ground truth, particularly in capturing fine edge details within complex image regions where it more accurately delineates lesion boundaries. This result confirms that by integrating the strengths of both base models, the Ensemble Model effectively enhances segmentation accuracy and robustness.

This study conducted a comprehensive evaluation of three models. The test set images first underwent quality enhancement and data augmentation through the "enhance()" function. The trained deep learning models were then employed to generate predictions on the processed test data. For the binary classification task, a threshold of 0.7 was established to convert multi-class probability outputs into binary predictions. During metric computation, we utilized TensorFlow to define and calculate multiple evaluation metrics, including Jaccard index, Dice coefficient, precision, and recall. These metrics quantitatively assessed the spatial alignment between predicted and ground-truth regions, as well as classification accuracy (detailed numerical results are shown in **Table 2**).

The Ensemble Model exhibited superior performance across multiple key evaluation metrics. It achieved an IoU of 93.73%, slightly higher than both SegNet (91.31%) and U-Net (93.53%). In terms of the Dice coefficient, the Ensemble

Model led with 84.85%, indicating superior performance in segmentation fineness. For precision, it attained the highest value of 93.93%, reflecting more reliable identification of target pixels. Similarly, it achieved the top recall rate at 87.38%, showing a greater ability to comprehensively detect positive-class pixels. Regarding accuracy, the Ensemble Model achieved 92.82%, slightly exceeding U-Net's 92.78% and SegNet's 91.18%. The most notable advantage was observed in the cross-entropy loss metric, where the Ensemble Model achieved a value of only 0.63 on the test set, substantially lower than U-Net's 6.47 and SegNet's 24.09. This result highlights its exceptional efficiency in parameter optimization and data fitting. By integrating the strengths of SegNet and U-Net, the Ensemble Model achieves significant performance improvements in image segmentation tasks, particularly in segmentation accuracy and generalization capability. This fusion strategy not only bolsters model generalization and robustness but also improves segmentation precision and completeness, thereby offering stronger competitiveness for practical image segmentation applications.

Discussion

The superior performance of the proposed average fusion strategy primarily stems from the complementary error patterns exhibited by SegNet and U-Net. Pixel-level error analysis on the validation set shows that while SegNet demonstrates high recall at lesion margins, it frequently misclassifies textured background regions as foreground, leading to reduced precision. Conversely, U-Net achieves high Dice scores within homogeneous lesion areas but exhibits decreased detection sensitivity in regions with faint boundaries, leading to lower recall rates. After implementing the fusion strategy, the "edge sensitivity" of SegNet and the "regional consistency" of U-Net demonstrate a negative correlation at the pixel level ($r=-0.68$, $p<0.01$). The simple averaging approach effectively mitigates individual model errors through complementary weighting, elevating the overall Dice score from 70.80% (SegNet) to 84.85% while substantially reducing the cross-entropy loss from 6.47 to 0.63. These results suggests the performance gains are primarily attributable to error complementarity rather than increased model capacity.

Further investigation into the effectiveness of weight averaging reveals its particular suitability for dermatological image segmentation. In such medical images, foreground pixels typically account for less than 10% of the total image, creating a significant class imbalance. Subsequent experiments with pixel-level weighted fusion combined with conditional random field post-processing revealed only a marginal 0.1% increase in lesion-specific Dice score, while the global Dice score dropped by 0.4%. This performance degradation can be attributed to the high-variance noise introduced during the weight estimation process, which offset any potential improvements. Therefore, under the dual constraints of imbalanced data distribution and computational efficiency, fixed-weight average fusion represents an optimal point on the simplicity-effectiveness Pareto frontier. This approach achieves robust performance enhancement while avoiding the introduction of additional hyperparameters typically required by complex post-processing methods.

The contribution of the training strategies is equally important. Batch normalization significantly stabilized gradient flow, allowing both sub-models to converge within 40 epochs—about 30% faster than training without batch normalization. Meanwhile, data augmentation—including random rotation, color dithering, and elastic deformation—reduced the Dice score variance on the validation set by 18%, indicating that the models learned more generalized and diverse features. Collectively, these strategies produced base predictions with low bias and low variance for the fusion stage, thereby amplifying the overall ensemble gains.

In contrast to previous studies, the present work further highlights the advantages of the proposed image segmentation framework. For example, Wu et al. introduced an improved Residual Attention Deconvolution U-Net model that surpasses conventional SegNet and U-Net in segmenting lung nodules from computed tomography images [19]. However, its adaptability and stability tend to be limited when confronted with complex image structures and diverse segmentation requirements. Han et al. presented a multi-channel medical image segmentation network, the Hierarchically Weighted Attention SegNet (HWA-SegNet), to address challenges in delineating irregular skin lesion boundaries [20]. While HWA-SegNet achieved

a segmentation accuracy of 96.33% on the public International Skin Imaging Collaboration 2018 dataset, its Jaccard index remained at 85.90%. The current study also focuses on segmenting diverse skin lesion images from the HAM10000 dataset, particularly those exhibiting fine structural details and ambiguous boundaries in localized regions. The proposed average fusion strategy achieves a Jaccard index of 93.73% on the current dataset, surpassing that of HWA-SegNet by approximately 7.8 percentage points, while maintaining an accuracy of 92.82%. This result reflects the method's effectiveness in balancing segmentation completeness with precision.

While the fusion strategy proposed in this study demonstrates superior performance over individual models, its potential remains limited when compared to more sophisticated ensemble techniques. For instance, Sanga et al. and Wang et al. introduced post-processing methods based on conditional random fields or weighted voting mechanisms to combine model predictions, achieving notable improvements in segmentation accuracy and robustness [21, 22]. Nevertheless, such approaches relying on fixed weights or predetermined rules often lack the flexibility to adapt to varying image characteristics. Although the average fusion strategy used in this study is straightforward and effective, it does not fully account for the spatially varying performance of the two constituent models across different image regions. As a result, a more adaptive fusion method that dynamically adjusts weights based on localized model performance could further enhance segmentation outcomes.

Despite the demonstrated effectiveness of the proposed strategy, two main limitations warrant consideration. First, the training data were obtained from a single center, highlighting the need for validation on larger, multi-center datasets acquired with diverse imaging devices to further assess its generalization capability. Second, the model contains a moderate number of parameters and has not undergone pruning or quantization, limiting its immediate applicability in mobile or edge computing scenarios. Future work will focus on developing dynamic weighted fusion mechanisms guided by uncertainty estimation and exploring lightweight network architectures, with the goal of achieving real-time, high-accuracy dermatological lesion segmentation in resource-limited clinical environments.

Conclusion

This study implemented a simple averaging strategy to integrate SegNet and U-Net, leading to measurable improvements in dermatological image segmentation accuracy and generalization capability. Experimental results show that the Ensemble Model achieves enhanced performance across multiple evaluation metrics, with a particularly notable reduction in loss value, confirming its improved model efficacy. These advancements contribute to more precise delineation of lesion boundaries and detailed characterization of internal textures, underscoring the methodological value and practical potential of the fusion approach in medical imaging-assisted diagnosis.

Author contributions: Siqi Wang, Danhong Li, and Yina Zhang jointly conceived the study design and experimental protocol, implemented the SegNet, U-Net, and Ensemble Models, and wrote the initial manuscript. Yu Wang, Linrong Yuan, Miao Yu, Jianghui Li, and Yimeng Wang performed data preprocessing, model training and evaluation, analyzed the results, and contributed to manuscript writing. Ping Li provided critical revisions and supervised the entire project.

References

[1] Ray A, Sarkar S, Schwenker F, et al. Decoding skin cancer classification: perspectives, insights, and advances through researchers' lens. Sci Rep 2024;14(1):30542.

[2] Haque S, Ahmad F, Singh V, et al. Skin Cancer Detection Using Deep Learning Approaches. Cancer Biother Radiopharm 2025;40(5):301-312.

[3] Lee K, Cavalcanti TC, Kim S, et al. Multi-Task and Few-Shot Learning-Based Fully Automatic Deep Learning Platform for Mobile Diagnosis of Skin Diseases. IEEE J Biomed Health Inform 2023;27(1):176-187.

[4] Sarwar N, Irshad A, Naith QH, et al. Skin lesion segmentation using deep learning algorithm with ant colony optimization. BMC Med Inform Decis Mak 2024;24(1):265.

[5] Akram A, Rashid J, Jaffar MA, et al. Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of

Medical Things. Skin Res Technol 2023; 29(11):e13524.
[6] Verma N, Ranvijay, Yadav DK. Hybrid of Deep Feature Extraction and Machine Learning Ensembles for Imbalanced Skin Cancer Datasets. Exp Dermatol 2024;33(12):e70020.
[7] Juan CK, Su YH, Wu CY, et al. Deep convolutional neural network with fusion strategy for skin cancer recognition: model development and validation. Sci Rep 2023;13(1):17087.
[8] Zhang M, Zhang Y, Liu S, et al. Dual-attention transformer-based hybrid network for multimodal medical image segmentation. Sci Rep 2024;14(1):25704.
[9] Liu T, Bai Q, Torigian DA, et al. VSmTrans: A hybrid paradigm integrating self-attention and convolution for 3D medical image segmentation. Med Image Anal 2024;98:103295.
[10] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:2278-2324.
[11] Singh SP, Wang L, Gupta S, et al. 3D Deep Learning on Medical Images: A Review. Sensors (Basel) 2020; 20(18):5097.
[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 2015;234-241.
[13] Yuan J. Brain tumor image segmentation method using hybrid attention module and improved mask RCNN. Sci Rep 2024;14(1):20615.
[14] Badrinarayanan V, Kendall A,Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2017; 39(12):2481-2495.
[15] Skin Cancer MNIST: HAM10000. Available at: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
[16] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5:180161.
[17] Azad R, Aghdam EK, Rauland A, et al. Medical Image Segmentation Review: The Success of U-Net. IEEE Trans Pattern Anal Mach Intell 2024;46(12):10076-10095.
[18] Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Med Imaging 2021; 21(1):19.
[19] Wu Z, Li X, Zuo J. RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning. Front Oncol 2023;13:1084096.
[20] Han Q, Wang H, Hou M, et al. HWA-SegNet: Multi-channel skin lesion image segmentation network with hierarchical analysis and weight adjustment. Comput Biol Med 2023;152: 106343.
[21] Sanga P, Singh J, Dubey AK, et al. DermAI 1.0: A Robust, Generalized, and Novel Attention-Enabled Ensemble-Based Transfer Learning Paradigm for Multiclass Classification of Skin Lesion Images. Diagnostics (Basel) 2023;13(19): 3159.
[22] Wang Y, Wen Z, Su L, et al. Improved brain metastases segmentation using generative adversarial network and conditional random field optimization mask R-CNN. Med Phys 2024; 51(9):5990-6001.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Review Article https://doi.org/10.61189/925623shkhmp

Research progress on vascular anastomosis technology

Wanwen Yang, Lin Mao, Yadan Yang, Chengli Song

Shanghai Institute for Minimally Invasive Therapy, School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China.

Corresponding author: Lin Mao.

Declaration of conflict of interest: None.

Received March 5, 2025; Accepted April 10, 2025; Published December 31, 2025

Highlights

- Systematic classification of anastomosis techniques, including manual sutures, robotic assistance, biomedical adhesives, energy welding, and stapling devices.
- Critical analysis of biodegradable materials in addressing foreign body reactions and balancing degradation and mechanical performance.
- Future directions emphasizing intelligent material design, multimodal technology fusion, and specialized device development.

Abstract

Vascular anastomosis, as one of the core surgical techniques, directly determines clinical efficacy in trauma repair, organ transplantation, and vascular reconstruction. This paper systematically reviews the development and current research status of vascular anastomosis techniques such as traditional manual sutures, robotic-assisted technologies, biomedical adhesives, and energy welding. Traditional manual sutures, regarded as the gold standard of vascular anastomosis, have achieved ultra-precise anastomosis at the 0.1-mm level through advancements in microsurgical techniques, making them the most clinically prevalent method. However, they are limited by issues such as foreign body retention, high demands on surgeon, prolonged operative times, and high postoperative stenosis rates. Robotic-assisted systems offer enhanced precision in complex anatomical regions, achieving submillimeter accuracy. However, their widespread adoption is constrained by high costs, reliance on suturing, and steep learning curves. Biomedical adhesives and energy welding techniques significantly reduce operative time but are not yet clinically applicable due to insufficient anastomotic strength. Although the GEM Coupler stapling devices have been clinically applied, other stapling technologies remain limited in scope, with ongoing research in structural designs and biodegradable materials. Future advances in vascular anastomosis are expected to focus on three directions: material innovation, technological breakthroughs, and clinical translation.

Keywords: Vascular anastomosis, manual sutures, stapling devices, biodegradable materials

I Introduction

Vascular anastomosis, as a core technique in surgical practice, involves the precise reconnection of severed vessel ends to restore vascular continuity, ensuring unimpeded blood flow and maintaining normal physiological functions of tissues and organs. Based on vascular injury patterns, three primary anastomotic configurations are employed: end-to-end anastomosis, end-to-side anastomosis, and side-to-side anastomosis [1]. End-to-end anastomosis is suitable for transverse injuries requiring precise matching of luminal diameters and axial alignment; End-to-side anastomosis is commonly used in bypass grafting or for connecting vessels with significant diameter discrepancies; and Side-to-side anastomosis is applied in vas-

Address correspondence to: Lin Mao, Shanghai Institute for Minimally Invasive Therapy, School of Health Science and Engineering, University of Shanghai for Science and Technology, No. 516 Jungong Road, Yangpu District, Shanghai 200093, China. Tel: +86-21-55572159. E-mail: linmao@usst.edu.cn.

cular shunting and dialysis access creation. These techniques are extensively utilized in trauma management, vascular pathology reconstruction, plastic surgery, and organ transplantation [2-5]. The continuous advancement of various surgical techniques and the increasing complexity of surgical procedures have put forward higher requirements for vascular anastomosis technology, demanding higher precision, efficiency and reduced complications.

Although traditional manual sutures are cost-effective and adaptable, they are prone to mechanical endothelial damage and are associated with high postoperative restenosis rates [6]. Notably, breakthroughs in super microsurgical techniques have enabled manual sutures to achieve 0.1-mm level precision in vascular anastomosis [7]. However, super microsurgical operations require extensive microsurgical expertise, with complex anatomical sites such as intracranial vessels requiring up to 30 minutes for anastomosis, significantly increasing the risk of ischemic complications [8].

In recent years, vascular anastomosis techniques have undergone a transformative shift from manual sutures toward minimally invasive and intelligent approaches. Robotic-assisted systems, leveraging multi-degree-of-freedom precise manipulation, achieve submillimeter-level anastomotic precision [9]. Non-suture techniques, such as laser welding, biomedical adhesives, and stapling devices, reduce operative time to 5-10 minutes [10-14]. Furthermore, the integration of novel biodegradable materials into stapling devices has significantly mitigated the risk of foreign body reaction [15].

In conclusion, vascular anastomosis, as a critical surgical technique, significantly enhances patients' quality of life and recovery rates by restoring or optimizing blood perfusion. Clinicians and researchers worldwide are continuously advancing research in anastomotic techniques, device design, and material science to improve the precision, safety, and procedural efficiency of vascular anastomosis, ultimately delivering superior therapeutic outcomes and rehabilitation. This review aims to systematically evaluate the advancements, limitations, and clinical applicability of both traditional and emerging vascular anastomosis techniques, providing evidence-based insights for future research directions.

II Review progress

Vascular anastomosis methods

Manual sutures

Manual sutures involve using sutures and surgical needles to approximate blood vessels. This technique remains the most clinically prevalent anastomotic method and is widely recognized as the gold standard for vascular anastomosis. The history of vascular surgery has been subject to debate, with no unified consensus. In 1889, Alexander Jassinowsky proposed the feasibility of vascular suturing and successfully performed an arterial anastomosis two years later [16]. In 1897, J.B. Murphy expanded upon Jassinowsky's work by introducing a novel framework for vascular anastomosis [17]. However, both pioneers erroneously concluded that endothelial disruption would inevitably trigger coagulation, hindering successful implementation. In 1902, French surgeon Alexis Carrel introduced a groundbreaking suturing method: placing three equidistant sutures at the vascular rupture site to stabilize the wound into a triangular configuration, followed by everting the vessel ends, precisely aligning them, and suturing with sterilized silk threads [18]. This approach ensured a smooth intraluminal surface and unobstructed blood flow. Carrel further established four ideal principles for vascular anastomosis: (1) Prevention of luminal stenosis at the anastomotic site; (2) Maintenance of intimal smoothness; (3) Tight intimal apposition between vascular ends; (4) Avoidance of direct contact between sutures and blood [19]. The advent of microsurgical manual sutures in 1960 marked the beginning of microvascular suturing techniques under microscopy. However, super-microsurgical operations demand exceptional microsurgical expertise, with complex anatomical sites such as intracranial vessels requiring 30 minutes for anastomosis [8]. Persistent challenges, including operator hand tremor and visual fatigue, substantially elevate the risks of ischemic complications.

Furthermore, the properties of suture materials critically influence vascular healing outcomes. Surgical sutures must exhibit optimal biocompatibility, sufficient tensile strength, and ease of knotting. Sutures are broadly classified into absorbable and non-absorbable categories. Absorbable sutures, such as polyglycolic acid

(PGA), polydioxanone (PDS), polyhydroxybutyrate (PHB), and polylactic acid (PLA), are rarely used in vascular anastomosis due to their prolonged retention within vessels, which poses a high infection risk [20]. In clinical vascular suturing, non-absorbable sutures are commonly employed. These sutures, made from non-degradable materials such as nylon, polyester, polytetrafluoroethylene (PTFE), and polypropylene, offer superior mechanical strength, with polypropylene being the most widely used in vascular surgery due to its excellent strength.

Manual sutures, despite their advantages of stable clinical efficacy, cost-effectiveness, and ease of implementation, are associated with significant limitations, including operator-dependent outcomes, prolonged operative times, and suture-induced foreign body retention. The latter can trigger inflammatory responses, platelet aggregation, endothelial dysfunction, and intimal hyperplasia, ultimately leading to anastomotic stenosis. To address these shortcomings, alternative vascular anastomosis techniques have been developed to mitigate the inherent limitations of manual sutures.

Robotic-assisted anastomosis

With the rapid advancement of automation control and computer vision technology, robotic systems have increasingly been integrated into medical practice. In 1985, robots were first applied in healthcare as positioning assistance devices during neurosurgical procedures [21]. Subsequently, robotic technology has been applied across diverse surgical disciplines, including neurosurgery, minimally invasive surgery, and microsurgery. Robotic-assisted vascular anastomosis emerged as an extension of medical robotics development. In this approach, the surgeon, seated at a console, manipulates robotic arms and instruments via control interfaces, such as trackballs and control wheels. This approach combines the benefits of minimally invasive approaches with robotic precision, enabling submillimeter-level accuracy in vascular anastomosis.

In 1995, Intuitive Surgical developed the prototype of the da Vinci Surgical System, named “Lenny”, which required fixation to the operating table and lacked interchangeable instruments [22]. In 2000, the da Vinci system received approval from the U.S. Food and Drug Administration (FDA) and began tis widespread application in general surgery. In 2002, Wisselink et al. first reported two successful cases of robotic-assisted laparoscopic aortofemoral bypass procedures [23]. Subsequently, robotic surgical systems have evolved toward diversification and specialization: The da Vinci platform has undergone iterative upgrades, expanding to complex procedures such as single-port surgery; Emerging systems like Hugo of Medtronic and Versius of CMR utilize modular and lightweight designs to reduce costs; and domestic Chinese systems, including “Tianji” and “Tuomai”, have overcome technical barriers and entered clinical use. Furthermore, robotic-assisted vascular anastomosis has been validated in broader applications, such as renal transplantation and pancreaticoduodenectomy, demonstrating enhanced surgical efficacy and safety profiles [9, 24, 25].

Compared to traditional manual sutures, medical robotic systems offer distinct advantages: reduced operator fatigue, enhanced maneuverability, improved surgical precision, and the ability to perform anastomose in narrow surgical fields or with submillimeter diameters [26]. However, robotic-assisted vascular anastomosis presents several limitations: (1) Prolonged operative time, with no significant reduction compared to manual sutures; (2) Persistent reliance on sutures, inheriting suture-related complications; (3) High costs and maintenance demands. For instance, the da Vinci Surgical System is priced at up to $ 3.5 million, with disment and regular maintenance. This economic burden restricts effective surgeon training, potentially increasing procedural risks; (4) Immature haptic feedback technology. While current systems integrate force feedback, vibration sensing, and visual cues to simulate tactile perception, yet their force resolution and dynamic response rates remain inadequate for complex vascular anastomosis; (5) Technically demanding operations with a steep learning curve [27].

Biomedical adhesives

Biomedical adhesives refer to vascular anastomosis methods utilizing biocompatible adhesive materials, as shown in **Figure 1**. Based on their composition, they are classified into natural adhesives, semi-synthetic adhesives, and synthetic adhesives. Early biomedical adhesives

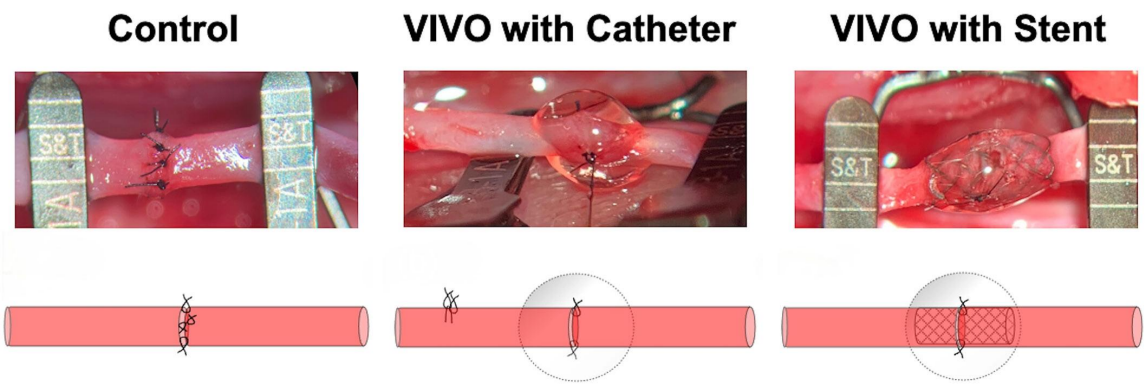


Figure 1. Biomedical adhesives: Sealants combined with sutures and stents [58].

were primarily natural derivatives, offering excellent biocompatibility and degradability but limited adhesive strength and durability, thus rarely applied in vascular anastomosis. In 1960, Nathan H. pioneered the use of synthetic adhesives for vascular repair [28]. Despite significant advancements in material formulations and delivery systems in recent decades, the core functionalities of these adhesives in vascular anastomosis remain consistent: (1) Hemostatic agents to promote wound healing; (2) Sealants, often combined with sutures or stents, to enhance anastomotic integrity and prevent postoperative leaks; (3) Standalone adhesives applied directly to vascular ends for sutureless anastomosis [29-31].

Biomedical adhesives offer several advantages, including minimal vascular wall trauma, procedural simplicity, rapid anastomosis, and adaptability to diverse vascular configurations. Representative tissue adhesives include fibrin glue, cyanoacrylate adhesives, and BioGlue biologic sealant. An ideal vascular adhesive should possess the following characteristics: (1) High mechanical strength to securely bond vascular ends and withstand hemodynamic forces; (2) Excellent biocompatibility, without inducing immune responses or thrombogenesis; (3) Controlled degradability to allow gradual absorption while minimizing long-term foreign body reactions; (4) Ease of preparation and application; (5) Non-interference with tissue function and mobility [10]. Despite their potential, biomedical adhesives face critical challenges in vascular applications, including insufficient long-term durability under cyclic mechanical stress, incomplete vascular wall compatibility leading to interfacial failure, and instability in dynamic hemodynamic environments. Consequently, researchers are actively developing novel materials and engineering strategies to enhance adhesive performance and safety. While various bioadhesives have demonstrated efficacy in preclinical studies, none have yet achieved successful clinical translation for vascular anastomosis.

Energy welding

Energy welding achieves vascular anastomosis by converting optical or electrical energy into thermal energy, which heats the anastomotic ends to induce partial denaturation and unwinding of the triple-helical collagen structure into randomly coiled polypeptide chains. This process facilitates synergistic aggregation of the denatured proteins, thereby bonding the vascular ends. Currently, the predominant energy sources in research are laser and high-frequency electric current [11, 32].

Laser welding

Laser light is a coherent light beam amplified by stimulated radiation. The active medium includes CO_2, Ar, Nd:YAG, He-Ne, and infrared diodes. In the 1970s, laser welding was gradually adopted for biological tissue repair, with the Nd:YAG laser being the first applied to vascular anastomosis [33, 34]. Over the past four decades, advancements in laser technologies have expanded their applications in vascular anastomosis. Researchers have identified key factors influencing the efficacy of laser-welded vascular anastomosis, including laser parameters (Wavelength, power output, and pulse width), tissue penetration depth, and welding duration. Optimal thermal effects for achieving robust anastomosis require a pulse width of 1 μs/min during vascular anastomosis [35]. Laser

parameters vary significantly across different laser systems, and their interactions with specific biological tissues determine the penetration depth within the tissue. Typically, CO_2 lasers exhibit a penetration depth of 10-50 μm in vascular tissue, which is notably shallower compared to other laser types.

Laser welding offers advantages such as high precision, short operative time, absence of foreign body reactions, and a reduced risk of postoperative aneurysms. However, the high collagen fiber and hemoglobin content in vascular tissues result in elevated laser absorption coefficients and penetration depths, which may cause thermal damage to the vessel wall, compromise anastomotic strength, and impair surgical outcomes [36]. To address these limitations, researchers have employed reinforcement materials such as albumin, polylactic-co-glycolic acid (PLGA), and indocyanine green (ICG) to augment laser-mediated anastomosis [12, 37, 38]. Despite these advancements, laser-welded vascular anastomosis has not yet been adopted clinically, with improving initial anastomotic strength remaining the critical challenge for its widespread implementation.

High-frequency electric welding

High-frequency electric welding utilizes heat generated by high-frequency alternating current to thermally fuse tissues, as shown in **Figure 2**. The current is delivered via electrodes to achieve tissue cutting, coagulation, or closure. In 1962, Sigel and Acevedo successfully sealed linear vascular incisions using high-frequency current transmitted through bipolar electrocautery forceps [39]. During the 1990s, high-frequency electric welding technology advanced rapidly. Barry et al. used radiofrequency current combined with mechanical pressure to weld the carotid artery in dogs and investigated the effect of radiofrequency current on the mechanical and histological properties of the arterial wall [40]. However, this method has notable limitations: (1) Electrode-tissue adhesion: Removal of electrodes may disrupt the anastomotic seal; (2) Uncontrolled parameters: Difficulty in optimizing current intensity and anastomosis duration compromises outcomes [41, 42].

In summary, energy welding techniques face critical challenges in clinical applications, including imprecise thermal control, excessive thermal damage, and insufficient anastomotic strength, necessitating continuous technological refinement and innovation.

Photochemical tissue bonding

Photochemical tissue bonding (PTB) is a sutureless anastomosis technique that achieves tissue adhesion through laser-activated photochemical reactions. This method induces electron transfer between photosensitizing chemicals and biological tissues, leading to covalent crosslinking of collagen molecules at the anastomotic site. In 2007, O'Neill and Winograd from Harvard University pioneered this technique, successfully anastomosing rat femoral arteries by applying photosensitizers to both vascular ends and irradiating them with a 532 nm laser (power density: 0.5 W/cm^2, energy density: 150 J/cm^2) [43]. During the 30-second irradiation on each side, an intraluminal catheter was inserted to maintain vascular patency.

Unlike energy welding, which depends on thermal energy, PTB relies on photochemical crosslinking, eliminating thermal damage to surrounding tissues. The technique employs 532 nm laser light combined with photosensitizers to initiate collagen covalent bonding, enabling rapid and robust tissue adhesion. Commonly used photosensitizers include Rose Bengal, Riboflavin-5-Phosphate (R5P), Methylene Blue (MB), and N-2-Hydroxypyridine (N-HTP) [44, 45].

As an advancement over laser welding, PTB addresses critical limitations such as thermal injury and high-energy laser requirements. However, challenges persist, including high operational costs and difficulties in laser parameter optimization, necessitating further research for clinical translation.

Microvascular stapler anastomosis

Microvascular stapler anastomosis refers to a vascular anastomosis method that replaces sutures with specialized devices, primarily categorized into staples, magnets and rings based on vascular eversion fixation mechanisms, as shown in **Figure 3**.

Staples

Staple-based anastomosis achieves vascular connection by mechanically pressing surgical staples (metal or polymeric) into both vascular

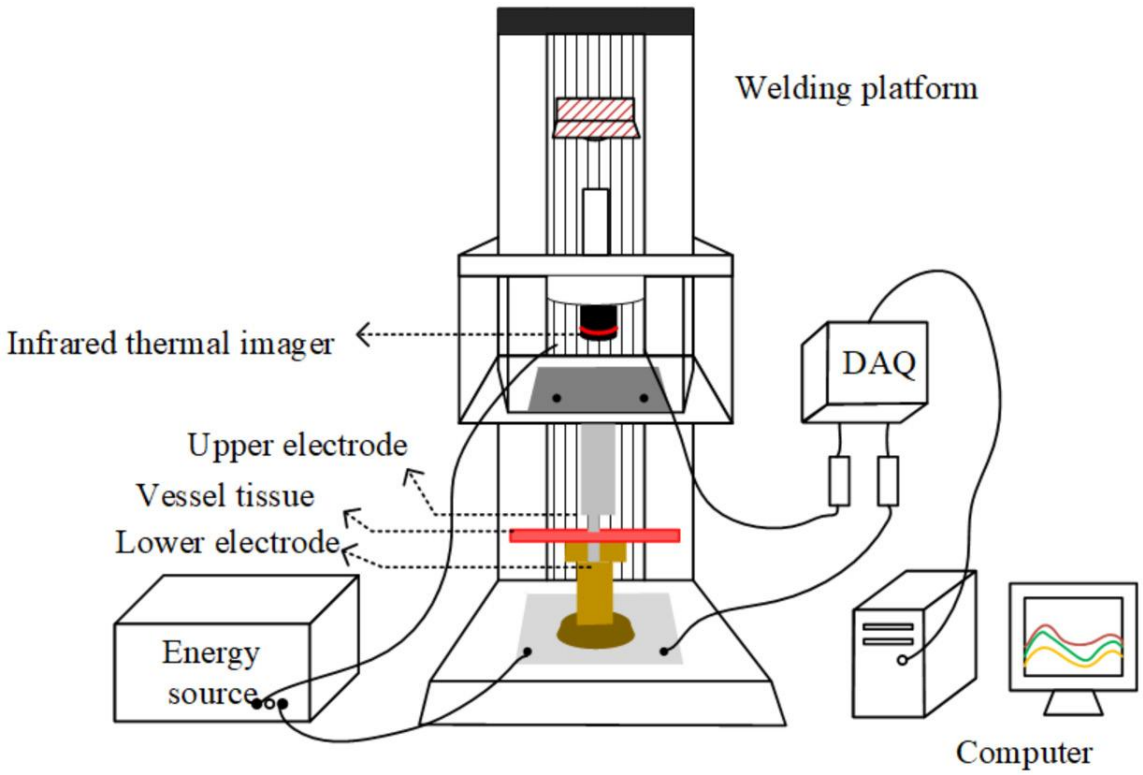


Figure 2. High-frequency electric welding: The treated vessel tissue is placed between the upper and lower electrodes, and the upper and lower electrodes deliver energy to the vessel tissue. The infrared thermal imager begins to record the tissue temperature until the end of the welding experiment [42].

ends. These stapling devices typically comprise a handle, anvil head, cartridge, and staples. During anastomosis, staples are deployed in a staggered arrangement to seal potential leaks. In 1908, Hungarian surgeon Hümér Hültl innovated the first surgical stapler, inspired by regular staplers, modifying the anvil to create B-shaped staples upon compression [46]. This design ensures secure vascular fixation while preserving blood supply and nutrient delivery to stapled tissue margins.

Compared to sutures, staples offer lower infection rates and shorter operative times, making them widely used in gastrointestinal anastomosis [47]. However, there are several factors that limit their application in vascular anastomosis [48]. Firstly, the vascular needs to be valgus 180°, which increases the risks of anastomotic stenosis; Secondly, permanent staples may induce chronic inflammation or foreign body reactions; Lastly, improper staple placement can cause impaired healing or functional compromise.

Magnets

Magnetic anastomosis achieves vascular reconstruction by implanting magnetic devices at both vascular ends, enabling instantaneous coupling via magnetic attraction. In 1978, Obora et al. first performed end-to-end venous anastomosis in rats using magnetic rings [49]. In 2004, Erdmann et al. successfully applied magnetic rings for side-to-side arterial and venous anastomosis in foxhounds. Since then, magnetic anastomosis has significantly evolved [50]. To date, it has been validated in animal models for end-to-end anastomosis, side-to-side anastomosis, and artificial vascular grafting [15, 50, 51].

Magnetic Anastomosis has several advantages, including simplified operation, shortened operative time, and reduced complications, such as anastomotic leakage. However, its limitation include challenges in repositioning once the magnetic rings are coupled. Suboptimal alignment necessitates device removal and replacement, potentially increasing surgical trauma.

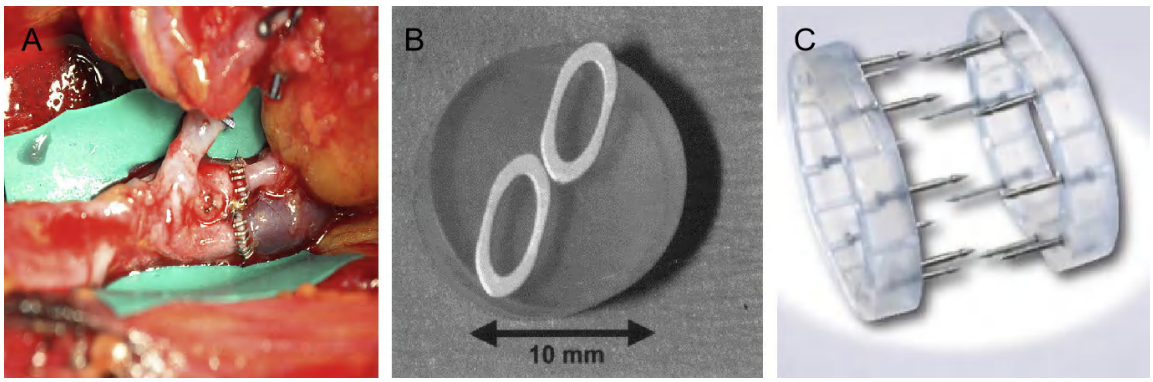


Figure 3. Typical anastomosis devices of microvascular stapler anastomosis [50, 59]: (A) Staples; (B) Magnetic rings; (C) GEM Coupler.

Rings

Ring-type staplers primarily consist of two coupling rings. During anastomosis, vessels are threaded through the rings, everted, and fixed, followed by ring apposition for end-to-end vascular connection. Based on eversion fixation methods, these staplers are classified into needle-ring and needleless types.

Needle-ring staplers feature alternating needles and holes on each ring. Vessels are everted and secured onto the needles, achieving anastomosis through needle-hole coupling. In 1900, Erwin Payr designed a magnesium ring to assist anastomosis by threading the proximal vessel through the ring, everting its margin, and inserting it into the dilated distal vessel [52]. In 1960, Holt and Lewis first proposed paired needle-perforated rings [53]. Subsequently, Ostrup and Berggren refined this design in 1986 by alternating needle-hole distribution, developing the Unilink system—the precursor to the clinically established GEM Coupler [54]. The GEM Coupler offers advantages, including ease of operation, precise intimal apposition, absence of intraluminal foreign bodies, and reduced anastomotic leakage. Although initially deemed suitable only for venous anastomosis due to arterial wall elasticity and thickness, ongoing research explores its arterial applicability.

Needleless staplers employ inner and outer rings: vessels are threaded through the inner ring, everted, and secured via outer ring fixation. Li et al. developed a novel device with two connecting rings and two posterior rings, addressing efficiency and reliability limitations of manual sutures [14]. Although needleless designs eliminate the use of metal needles, reducing the damage to the vascular intima, their clinical potential remains unvalidated due to insufficient long-term animal studies and human trials.

From the perspective of materials, the three types of stapling devices predominantly utilize non-degradable materials such as polyethylene, stainless steel, and titanium. Their long-term retention impairs physiological vascular function and restricts remodeling at the anastomotic sites. Consequently, researchers are exploring biodegradable materials for vascular staplers. These materials provide short-term mechanical support during anastomosis and are gradually absorbed post-healing, effectively preventing long-term implant-related complications. Additionally, their degradation releases bioactive substances that promote vascular regeneration and repair. Needleless designs eliminate metallic components but lack validation in long-term animal and clinical studies. Zhang Qiong and Xu Shuogui developed biodegradable shape memory sleeve-needle ring system for anastomosis (PLA needle rings and polycaprolactone (PCL) sleeves), achieving successful rabbit jugular vein anastomosis [55]. Li Zhuoqun et al. fabricated a 3D-printed polypropylene-based biodegradable stapler, evaluating its hypertension tolerance and clinical potential in rabbit models. Ryan Brewster et al. designed a biodegradable stapler composed of polylactic acid (PLA) and copolymers with dual inner and outer rings, successfully anastomosing porcine carotid arteries [57].

However, these biodegradable vascular staplers still face multiple technical challenges. Firstly, mismatched degradation kinetics and vascular healing timelines remain a critical

issue. For instance, PCL exhibits a degradation period of 12-24 months, far exceeding the typical 3-6 months required for vascular remodeling. This discrepancy leads to premature mechanical failure or delayed material absorption, compromising anastomotic integrity. Secondly, long-term safety validation remains insufficient. Degradation byproducts may impair vascular repair through mechanisms including local toxicity, metabolic burden, and immune reactions. Consequently, systematic long-term biocompatibility evaluations are imperative before clinical adoption, particularly to assess chronic risks like ectopic calcification and systemic metabolic disturbances.

III Conclusion

Vascular anastomosis, as a cornerstone of surgical practice, has undergone over a century of evolution. The emergence of robotic-assisted anastomosis, biomedical adhesives, energy welding, and stapling devices has driven transformative technological breakthroughs. Concurrently, advancements in biodegradable materials promise to accelerate the refinement and clinical translation of these techniques.

Traditional manual sutures, recognized as the gold standard, have achieved 0.1 mm-level precision through microsurgical innovations and remain the most widely used clinical method. However, manual sutures remain limited by operator dependency, prolonged operative durations, and high postoperative stenosis rates, which necessitate urgent solutions. The rise of surgical robotics has mitigated operator-related variability, enhancing both precision and procedural complexity. Sutureless techniques—including biomedical adhesives, energy welding, and stapling devices—have reduced anastomotic time to 5-10 minutes. Furthermore, the integration of biodegradable materials into sutures and staplers has emerged as a pivotal research direction to address foreign body retention. In addition, some researchers have explored combining various anastomosis methods, such as manual sutures, biomedical adhesives, magnetic anastomosis and energy welding to optimize the joint application of multiple anastomosis technologies.

Despite their unique advantages, all anastomotic techniques face distinct challenges. Surgical robotic systems, while achieving high precision, are hindered by prohibitive costs, lack of haptic feedback, and steep learning curves. Energy welding methods remain clinically unvalidated due to thermal collateral damage and insufficient anastomotic strength. Non-degradable materials such as staples and magnetic rings provoke chronic inflammation upon prolonged retention, while biodegradable alternatives require optimization to balance mechanical performance with degradation kinetics. Specifically, the degradation rates of biodegradable materials need to be regulated to match the healing cycle of blood vessels (3-6 months). Energy welding for immediate anastomotic sealing can be combined with biomedical adhesives for thermal protection, achieving both superior mechanical strength and minimized thermal injury to the anastomosis. Extensive animal experiments and clinical trials are needed to verify the long-term safety and effectiveness of these methods.

In conclusion, the future development of vascular anastomosis technology should focus on solving current technical bottlenecks, offering more efficient and safe solutions for the fields of cardiovascular disease, trauma treatment, and organ transplantation through three key directions of material innovation, technology integration, and clinical transformation.

Author contributions: Lin Mao, Chengli Song and Wanwen Yang and Yadan Yang conceived this project. Wanwen Yang and Yadan Yang collected and organized the literature. All authors contributed to the writing of the manuscript.

References

[1] Werker PM, Kon M. Review of facilitated approaches to vascular anastomosis surgery. Ann Thorac Surg 1997;63(6 Suppl):S122-S127.

[2] Hoballah JJ, Bechara CF. Vascular reconstructions: anatomy, exposures and techniques. Springer Nature, 2021.

[3] Woo SJ, Koo HT, Park SO, et al. Evolution of anatomical studies on the arterial, venous, and lymphatic system in plastic surgery. Arch Plast Surg 2022;49(6):773-781.

[4] Girardi LN, Krieger KH, Lee LY, et al. Management strategies for type A dissection complicated by peripheral vascular malperfusion. Ann Thorac Surg 2004;77(4):1309-1314.

[5] Siddique A, Bose AK, Özalp F, et al. Vascular anastomotic complications in lung transplantation: a single institution's experience. Interact Cardiovasc Thorac Surg 2013;17(4):625-631.

[6] Zeebregts C, Heijmen R, Van den Dungen J, et al. Non-suture methods of vascular anastomosis. Br J Surg 2003;90(3):261-271.
[7] Ratoiu VA, Cretu A, Hodea FV, et al. Supermicrosurgical Vascular Anastomosis—A Comparative Study of Lumen-Enhancing Visibility Techniques. J Clin Med 2025;14(2):555.
[8] Ramanathan D, Hegazy A, Mukherjee SK, et al. Intracranial in situ side-to-side microvascular anastomosis: principles, operative technique, and applications. World Neurosurg 2010; 73(4):317-325.
[9] Tiong HY, Goh BYS, Chiong E, et al. Robotic kidney autotransplantation in a porcine model: a procedure-specific training platform for the simulation of robotic intracorporeal vascular anastomosis. J Robot Surg 2018;12(4):693-698.
[10] Bal-Ozturk A, Cecen B, Avci-Adali M, et al. Tissue adhesives: From research to clinical translation. Nano Today 2021;36:101049.
[11] Gomes DF, Galvão I, Loja MAR. Overview on the Evolution of Laser Welding of Vascular and Nervous Tissues. Appl Sci 2019;9(10):2157.
[12] Schöni DS, Bogni S, Bregy A, et al. Nanoshell assisted laser soldering of vascular tissue. Lasers Surg Med 2011;43(10):975-983.
[13] Bahmani A, Arthanari S, Shin KS. Formulation of corrosion rate of magnesium alloys using microstructural parameters. Journal of Magnesium and Alloys 2020;8(1):134-149.
[14] Li H, Gale BK, Sant H, et al. A novel vascular coupling system for end-to-end anastomosis. Cardiovasc Eng Technol 2015;6(3):294-302.
[15] Liu P, Liu X, Yang L, et al. Enhanced hemocompatibility and rapid magnetic anastomosis of electrospun small-diameter artificial vascular grafts. Front Bioeng Biotechnol 2024;12: 1331078.
[16] Barker WF. A Century's Worth of Arterial Sutures. Ann Vasc Surg 1988;2(1):85-91.
[17] Murphy JB. Resection of Arteries and Veins Injured in Continuity--End-to-End Suture: Experimental and Clinical Research. Publishers Printing Company, 1897.
[18] Carrel A. Anastomose bout à bout de la jugulaire et de la carotide primitive. Lyon méd 1902;99:114-116.
[19] Aida L. Alexis Carrel (1873-1944): Visionary vascular surgeon and pioneer in organ transplantation. J Med Biogr 2014;22(3):172-175.
[20] D'Cunha P, Pande B, Kathalagiri MS, et al. Absorbable sutures: chronicles and applications. Int Surg J 2022;9(7):1383-1394.
[21] Stadler P. Robot-Assisted Vascular Surgery. Robotic Surgery 2021;1365-1381.
[22] Palep JH. Robotic assisted minimally invasive surgery. J Minim Access Surg 2009;5(1):1-7.
[23] Wisselink W, Cuesta MA, Gracia C, et al. Robot-assisted laparoscopic aortobifemoral bypass for aortoiliac occlusive disease: a report of two cases. J Vasc Surg 2002;36(5):1079-1082.
[24] Campi R, Pecoraro A, Vignolini G, et al. The first entirely 3D-printed training model for robot-assisted kidney transplantation: the RAKT box. Eur Urol Open Sci 2023;53:98-105.
[25] Menichini G, Malzone G, Tamburello S, et al. Safety and efficacy of Symani robotic-assisted microsurgery: Assessment of vascular anastomosis patency, thrombus, and stenosis in a randomized preclinical study. J Plast Reconstr Aesthet Surg 2024;96:1-10.
[26] Malzone G, Menichini G, Innocenti M, et al. Microsurgical robotic system enables the performance of microvascular anastomoses: a randomized in vivo preclinical trial. Sci Rep 2023;13(1):14003.
[27] Lengyel BC, Chinnadurai P, Corr SJ, et al. Robot-assisted vascular surgery: literature review, clinical applications, and future perspectives. J Robot Surg 2024;18(1):328.
[28] Healey JE, Gallager HS, Moore EB, et al. Experiences with plastic adhesive in the non-suture repair of body tissues. Am J Surg 1965;109:416-423.
[29] Heitzer M, Brockhaus J, Kniha K, et al. Mechanical strength and hydrostatic testing of VIVO adhesive in sutureless microsurgical anastomoses: an ex vivo study. Sci Rep 2021;11(1):13598.
[30] Annabi N, Zhang YN, Assmann A, et al. Engineering a highly elastic human protein-based sealant for surgical applications. Sci Transl Med 2017;9(410):eaai7466.
[31] Annabi N, Tamayol A, Shin SR, et al. Surgical materials: Current challenges and nano-enabled solutions. Nano Today 2014;9(5):574-589.
[32] Moritz WR, Raman S, Pessin S, et al. The History and Innovations of Blood Vessel Anastomosis. Bioengineering (Basel) 2022; 9(2):75.
[33] White RA, Patrick Abergel R, Lyons R, et al. Biological effects of laser welding on vascular healing. Lasers Surg Med 1986;6(2):137-141.
[34] Okada M, Shimizu K, Ikuta H, et al. An alternative method of vascular anastomosis by laser: experimental and clinical study. Lasers Surg Med 1987;7(3):240-248.
[35] Bregy A, Bogni S, Bernau VJ, et al. Solder doped polycaprolactone scaffold enables reproducible laser tissue soldering. Lasers Surg Med 2008;40(10):716-725.

[36] Li C, Wang K. Effect of welding temperature and protein denaturation on strength of laser biological tissues welding. Optics & Laser Technology 2021;138:106862.
[37] McNally KM, Sorg BS, Welch AJ, et al. Photothermal effects of laser tissue soldering. Phys Med Biol 1999;44(4):983-1002; discussion 2 pages follow.
[38] Sorg BS, Welch AJ. Laser-tissue soldering with biodegradable polymer films in vitro: Film surface morphology and hydration effects. Lasers Surg Med 2001;28(4):297-306.
[39] Sigel B, Acevedo FJ. Electrocoaptive union of blood vessels: a preliminary experimental study. J Surg Res 1963;3(2):90-96.
[40] Barry KJ, Kaplan J, Connolly RJ, et al. The effect of radiofrequency-generated thermal energy on the mechanical and histologic characteristics of the arterial wall in vivo: implications for radiofrequency angioplasty. Am Heart J 1989;117(2):332-341.
[41] Muzychenko P, Chernyak V, Savosko S, et al. Morphological assessment of vascular anastomoses performed using high frequency electric welding. Int J Morphol 2021;39(4):1183-1189.
[42] Wang H, Yang X, Madeniyeti N, et al. Temperature distribution of vessel tissue by high frequency electric welding with combination optical measure and simulation. Biosensors (Basel) 2022;12(4):209.
[43] O'Neill AC, Winograd JM, Zeballos JL, et al. Microvascular anastomosis using a photochemical tissue bonding technique. Lasers Surg Med 2007;39(9):716-722.
[44] Ark M, Cosman PH, Boughton P, et al. Photochemical Tissue Bonding (PTB) methods for sutureless tissue adhesion. Int J Adhes Adhes 2016;71:87-98.
[45] Scott BB, Randolph MA, Guastaldi FP, et al. Light-activated vascular anastomosis. Surg Innov 2023;30(2):143-149.
[46] Akopov A, Artioukh DY, Molnar TF. Surgical staplers: the history of conception and adoption. Ann Thorac Surg 2021;112(5):1716-1721.
[47] Ghosh S, More N, Kapusetti G. Surgical staples: Current state-of-the-art and future prospective. Med Nov Technol Devices 2022;16:100166.
[48] Ribaudo JG, He K, Madira S, et al. Sutureless vascular anastomotic approaches and their potential impacts. Bioact Mater 2024;38:73-94.
[49] Obora Y, Tamaki N, Matsumoto S. Nonsuture microvascular anastomosis using magnet rings: preliminary report. Surg Neurol 1978; 9(2):117-120.
[50] Erdmann D, Sweis R, Heitmann C, et al. Side-to-side sutureless vascular anastomosis with magnets. J Vasc Surg 2004;40(3):505-511.
[51] Lu Q, Liu K, Zhang W, et al. End-to-end vascular anastomosis using a novel magnetic compression device in rabbits: a preliminary study. Sci Rep 2020;10(1):5981.
[52] Payr E. Ueber Verwendung von Magnesium zur Behandlung von Blutgefässerkrankungen. Deutsche Zeitschrift für Chirurgie 1902;63: 503-511.
[53] Holt MH, Macalalad F, Lewis FJ. Autogenous Oxygenation With Cardian Bypass, Hypothermia, and an Atrioventricular Clamp. J Thorac Cardiovasc Surg 1960;40:536-548.
[54] Östrup LT, Berggren A. The UNILINK instrument system for fast and safe microvascular anastomosis. Ann Plast Surg 1986;17(6):521-525.
[55] Zhang Q, Xu SG. Biodynamic properties of vascular anastomosis using a new-type degradable anastomosis clip. Chinese Journal of Tissue Engineering Research 2014; 18(52): 8487-8490.
[56] Li Z, Zhang Y, Li Y, et al. 3D-Printed Biodegradable Stapler for Vascular Anastomosis: A Feasibility Study in Rabbits. J Am Coll Surg 2020;231(4 Suppl 1):S297.
[57] Brewster R, Gale BK, Sant HJ, et al. A biodegradable vascular coupling device for end-to-end anastomosis. J Med Biol Eng 2018;38:715-723.
[58] Heitzer M, Ooms M, Katz MS, et al. Evaluation of the long-term results of vascular anastomosis using polyurethane adhesive and shape-memory stent in the rat carotid artery model. Microsurgery 2022;42(5):480-489.
[59] Reddy C, Pennington D, Stern H. Microvascular anastomosis using the vascular closure device in free flap reconstructive surgery: A 13-year experience. J Plast Reconstr Aesthet Surg 2012;65(2):195-200.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Review Article https://doi.org/10.61189/880658qzgyhf

A review of low thermal damage technologies in electrosurgery

Yuxiang Luo, Yong Wang, Jiuzhou Zhao, Xiangzhou Meng, Yanan Hou, Yao Zheng, Yu Zhou

School of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China.

Corresponding author: Yu Zhou.

Declaration of conflict of interest: None.

Received April 21, 2025; Accepted August 8, 2025; Published December 31, 2025

Highlights

- Comprehensive Technology Review – It systematically compares three major techniques for reducing thermal damage in electrosurgery.
- Balanced Evaluation Framework – Each method is evaluated across multiple dimensions, including effectiveness, applicability, complexity, and cost.
- Forward-Looking Insight – It discusses future trends such as AI integration and the development of advanced materials for smarter surgical systems.

Abstract

Electrosurgery is widely applied for precise tissue cutting and coagulation through high-frequency electrical energy, yet it carries the risk of collateral thermal injury. This review examines emerging strategies to mitigate such damage, including cooling systems, real-time temperature monitoring, pulsed cutting, and laser- or ultrasound-assisted techniques. By improving temperature regulation, enhancing feedback accuracy, and enabling adaptive power control, these innovations reduce heat diffusion and tissue charring, thereby enhancing surgical safety and outcomes.

Keywords: Electrosurgery, thermal damage, adaptive control

Introduction

In early surgical practice, tissue separation, tumor excision, and necrotic tissue removal were primarily performed with surgical blades. However, the sharpness of the blade often caused rupture of adjacent blood vessels, leading to bleeding that required hemostasis by manual pressure, ligation, or the use of hot cautery. With advances in medical technology and the discovery of electromagnetic induction, surgery gradually transitioned toward electronic methods. In the early 20th century, Walter reported that high-frequency currents could generate spark-like effects between electrodes and tissue at a short distance, which were applied to treat skin tumors [1]. In 1911, Clark employed very high voltage and extremely low current with a single electrode to destroy tissue, producing marked desiccation at the contact site [2]. Subsequently, with support from the Liebel-Flarsheim Company, Bovie developed the first electrosurgical instruments. By the mid-1980s, ionization devices using high-frequency alternating current to generate localized high heat with a single electrode were introduced, enabling both tissue cutting and coagulation [3]. Since then, electrosurgery has become a widely adopted clinical technique [4].

Address correspondence to: Yu Zhou, School of Health Sciences and Engineering, University of Shanghai for Science and Technology, No. 516 Jungong Road, Shanghai 200093, China. Tel: +86-18021042556. E-mail: zhouyu@usst.edu.cn.

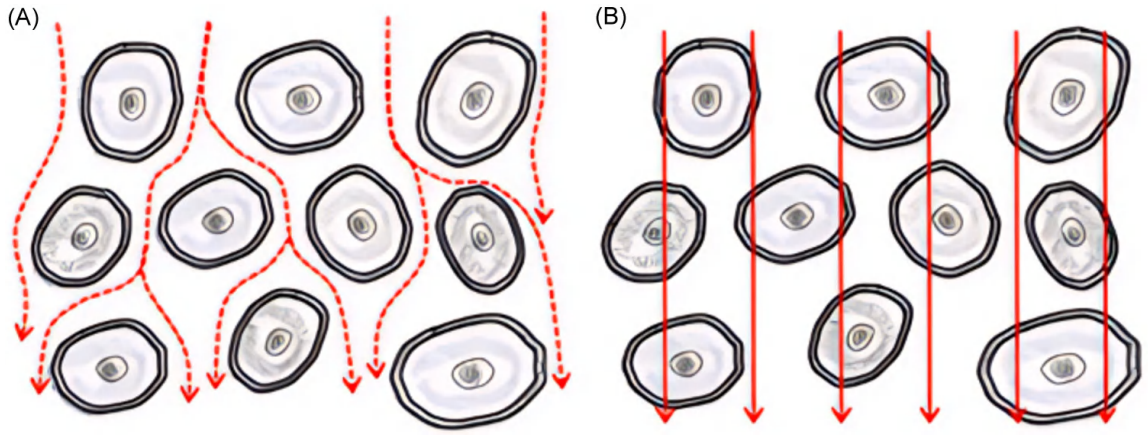


Figure 1. The internal current path of the organization at different frequencies. (A) Low-frequency current path within the tissue; (B) High-frequency current path within the tissue [8].

Electrosurgical devices broadly refer to instruments that utilize high-frequency electrical energy (typically 0.2-5 MHz) for tissue cutting, dissection, coagulation, and related procedures. These include monopolar and bipolar electrosurgical devices, radiofrequency knives, ion knives, argon plasma coagulators, and laser scalpels. A standard electrosurgical system consists of a high-frequency power source and return electrode pads. Current passes through the patient's body, and heat is generated at the tissue–electrode interface, enabling precise cutting and coagulation. With advances in material science and surgical technology, a variety of electrosurgical knives have been developed to meet the needs of different surgical scenarios. However, the high temperatures produced during cutting inevitably cause thermal diffusion, which can damage surrounding normal tissue and even result in charring or necrosis. Therefore, reducing thermal injury has become a major focus in the development of electrosurgical devices. This review will discuss the basic principles of high-frequency electrosurgery, mechanisms of thermal damage, and emerging technologies aimed at minimizing such effects.

Operating principle of electrosurgery

Electrosurgery applies high-frequency electrical energy to biological tissues, offering two key clinical advantages.

First, because biological tissues are composed of numerous cells with membranes that behave as capacitors, the current pathway is frequency-dependent [5]. Direct current is blocked, and low-frequency alternating current encounters high impedance, so current primarily flows through extracellular fluid (**Figure 1A**). In contrast, high-frequency alternating current can traverse both extracellular and intracellular fluids (**Figure 1B**). When the amplitude is sufficiently high, Joule heating raises the temperature of intracellular fluid to its boiling point. Vapor expansion induces microscopic cell rupture, producing the macroscopic cutting effect [6]. In an experiment conducted by Burcu Tunc, a sudden temperature drop of 0.5-18 °C was observed during the initial ablation phase, indirectly confirming gas formation during electrosurgical cutting and supporting this mechanism [7].

Second, the frequency range of electrosurgical currents (0.2-5 MHz) is far higher than physiological signals in the human body [8]. This minimizes interference with both the patient's and the surgeon's physiological functions during the procedure.

As shown in **Figure 2**, progressive heating leads to cellular dehydration above 80 °C, followed by vaporization above 100 °C. This sequence underlies the mechanism of tissue cutting in high-frequency electrosurgery [9].

Mechanisms of thermal effects in electrosurgery and analysis of tissue thermal damage

Definition and classification of thermal damage

Definition of thermal damage

Thermal damage refers to the structural and functional changes in cells or tissues caused

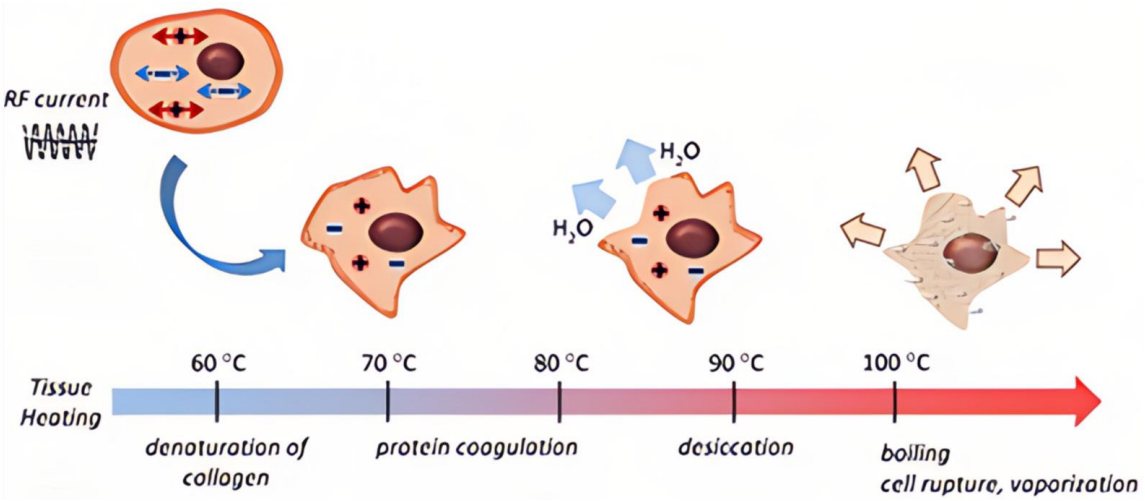


Figure 2. State changes of cells at different temperatures.

by exposure to excessive temperatures, leading to damage. High temperatures can cause water evaporation within tissues or protein denaturation, ultimately resulting in tissue death. Thermal damage typically occurs when tissues are exposed to temperatures exceeding their normal tolerance threshold [10].

Classification of thermal damage

Thermal damage is commonly classified based on temperature and exposure time. The primary classifications are as follows:

(1) Mild Thermal Damage (Low-Temperature Damage): This occurs when tissues are exposed to temperatures between 40 °C and 60 °C for short durations. Mild damage may result in temporary loss of cellular function, but tissue recovery is generally expected. This type of damage is often seen in localized high-temperature procedures, such as hot compresses or light electrosurgical cutting.

(2) Moderate Thermal Damage (Coagulation Damage): Damage occurs when tissues are exposed to temperatures between 60 °C and 80 °C for longer durations. In moderate thermal damage, cellular proteins begin to denature, leading to functional impairments. This damage can often be managed surgically, with the affected area typically showing signs of coagulation. The damaged cells may not recover.

(3) Severe Thermal Damage (Vaporization and Carbonization Damage): Exposure to temperatures above 80 °C for extended periods results in the evaporation of water, vaporization, carbonization, or complete charring of the tissue. Cellular death becomes irreversible, and the tissue cannot recover. Severe thermal damage is commonly observed with devices like high-frequency electrosurgical knives, and it is the primary mechanism behind the cutting action in electrosurgery.

Common mechanisms of thermal damage caused by high-frequency electrosurgical knives

Power and heat generation

Electrosurgery generates heat within tissue via high-frequency currents, causing rapid temperature increases at the point of contact. When electrosurgical power is too high, the current generates significant heat as it passes through the tissue. For a given current, higher power results in more heat generation and a faster temperature rise at the local tissue site. Excessive heat accumulation can surpass the tissue's tolerance threshold, leading to thermal damage.

During surgery, the electrosurgical power must be adjusted according to the specific requirements of the procedure. If the power is too high, heat may not dissipate quickly enough within the tissue, causing overheating. Conversely, if the power is too low, the desired cutting or coagulation effect may not be achieved, resulting in incomplete results. Therefore, precise power adjustment based on tissue types and

surgical needs is crucial in clinical practice to avoid excessive heat accumulation.

Heat propagation and tissue thermal conductivity

Heat conduction refers to the transfer of heat from high-temperature regions to cooler areas through molecular vibration. Different tissues have varying thermal conductivities, which influence how heat is distributed within them. Tissues with higher water content, such as muscles and blood, generally have better thermal conductivity, allowing heat to spread quickly and reducing the risk of localized overheating. In contrast, tissues with lower water content, such as fat, have poor thermal conductivity, meaning heat spreads slowly and can accumulate, increasing the risk of thermal damage.

During electrosurgery, heat is generated as current passes through tissue. When power output is too high, tissue temperature rises rapidly. In low-conductivity tissues like fat, heat cannot dissipate efficiently, leading to heat accumulation and potentially severe damage such as charring. In contrast, tissues with higher conductivity, such as muscle, spread heat more effectively, which slows the rise in temperature and helps reduce the risk of damage.

The thermal conductivity of tissue is closely related to the power settings of the electrosurgery. When power is too high, heat accumulates rapidly, especially in low-conductivity tissues like fat, where heat dissipation is inefficient. This causes local temperatures to rise quickly, leading to protein denaturation, tissue necrosis, and even carbonization. On the other hand, in tissues with higher thermal conductivity, such as muscle, heat is spread more effectively, reducing the rate of temperature increase and lowering the risk of thermal damage. Therefore, adjusting the power settings based on tissue thermal conductivity is essential for minimizing thermal damage during surgery.

Common mechanisms of thermal damage caused by high-frequency electrosurgical knives

Thermal damage caused by high-frequency electrosurgical knives involves two primary mechanisms:

(1) Power Setting and Impedance Variability: When electrosurgery is used for cutting or coagulation, the impedance of the tissue it contacts varies depending on the type of tissue and its pathological state. Since the power setting is often fixed by the surgeon during the procedure, any imbalance between the power output and the tissue's impedance can result in inefficient energy transfer. If the power is too high, the heat generated by the current may exceed the tissue's capacity to dissipate it, causing heat to accumulate and leading to extensive and severe thermal damage. Conversely, if the power is too low, the tissue may not receive sufficient heat to achieve the desired effects, such as incomplete coagulation or inadequate cutting, resulting in suboptimal surgical outcomes.

(2) Thermal Injury to Surrounding Healthy Tissue: The fundamental operation of electrosurgery involves heating tissue to achieve cutting or coagulation. However, this process inevitably affects surrounding healthy tissues, causing "collateral damage". This is especially noticeable in tissues with low thermal conductivity, such as fat, where heat tends to concentrate in localized areas. These tissues' inability to dissipate heat efficiently increases the risk of thermal damage. Even with precise control of the electrosurgery's power, heat will still propagate and affect surrounding healthy tissues. Therefore, careful management of both power settings and tissue characteristics is critical to minimizing this risk during surgery.

Technologies for reducing thermal damage

While high-frequency electrosurgical knives are widely used in surgical procedures, thermal damage to surrounding tissues due to their high temperatures remains a significant concern. In recent years, several technical strategies have been developed to address this issue. These strategies can be broadly classified into three categories: real-time feedback systems for intelligent power regulation, physical cooling mechanisms to actively reduce temperature at the surgical site, and waveform modulation techniques to interrupt continuous heat delivery. This section discusses representative techniques from each category, along with their respective mechanisms, advantages, and challenges.

Temperature feedback and intelligent power control

The first approach focuses on controlling heat generation at the source through intelligent power regulation based on real-time feedback. Among various feedback signals, temperature and impedance are commonly used to assess tissue conditions and adjust power output accordingly.

With advancements in electrosurgical equipment and its application by surgeons, numerous methods have emerged to evaluate the performance of electrosurgical devices, including power exposure, accuracy, and more [10]. Initially, these methods were used only for evaluation, but later, researchers proposed using the numerical indicators obtained from these evaluations as feedback to the electrosurgical device for intelligent cutting. Devices capable of such intelligent cutting are referred to as adaptive electrosurgical knives, and the algorithm that processes the feedback information and adjusts the electrosurgery's power output in real-time is called the adaptive algorithm (also referred to as intelligent power control).

In the implementation of adaptive algorithms, key reference indicators include tissue surface temperature and impedance. Temperature directly reflects the tissue's damage state, while tissue impedance indirectly indicates the tissue type and whether the heat exposure is within the normal range. This section uses temperature feedback as an example to explain the implementation of power adaptive algorithms.

A real-time temperature monitoring system can continuously track both the electrode tip's and surrounding tissue's temperature during high-frequency electrosurgery. Using temperature sensors, the surgeon remains informed about the temperature changes in the tissue, helping avoid thermal damage caused by excessive heat [11]. For example, infrared temperature sensors detect infrared radiation to measure tissue surface temperature, allowing rapid feedback. These sensors offer the advantage of non-contact measurement, making them suitable for dynamic adjustments during surgery. Alternatively, thermocouple sensors, typically in the form of miniature probes, directly contact the tissue for precise temperature measurements. These sensors are ideal for high-precision monitoring, ensuring optimal cutting results while minimizing thermal damage to surrounding tissues.

By integrating these feedback mechanisms, adaptive electrosurgical knives can adjust their power output in real-time, improving cutting accuracy and reducing thermal damage to healthy tissues.

Figure 3 illustrates the use of a traditional monopolar electrosurgery system based on real-time temperature monitorin [12]. The current-sensing transformer and voltage-sensing transformer are represented by CT (Current Transformer) and PT (Power Transformer), respectively. The high-frequency transformer (T) increases the output voltage. The tissue is placed between the electrosurgical electrode (ES) and the return pad, with the tissue surface temperature being monitored in real-time via an infrared thermography system.

The primary advantage of real-time temperature monitoring is its ability to provide early warnings when temperatures rise excessively. The adaptive algorithm helps the surgeon take corrective actions, such as reducing power or temporarily halting the procedure. This monitoring approach specifically addresses the first mechanism of damage discussed in Chapter 2, ensuring that the electrosurgical power output remains within a safe range by precisely controlling temperature. This is particularly beneficial for lengthy surgeries and temperature-sensitive tissues, as it significantly reduces the risk of thermal damage. Additionally, the high accuracy of the temperature monitoring system, coupled with the power adaptation algorithm, minimizes collateral damage, making it an effective strategy for improving surgical outcomes.

However, the implementation of temperature-feedback-based adaptive algorithms still faces several challenges. First, real-time temperature acquisition requires highly sensitive, fast-response sensors that do not interfere with surgical operations. Second, from an algorithmic perspective, modeling the relationship between temperature and power regulation is difficult due to the complex and variable nature of tissue types and surgical environments. Finally, from a mechanical standpoint, integrating sensors seamlessly into electrosurgical equipment

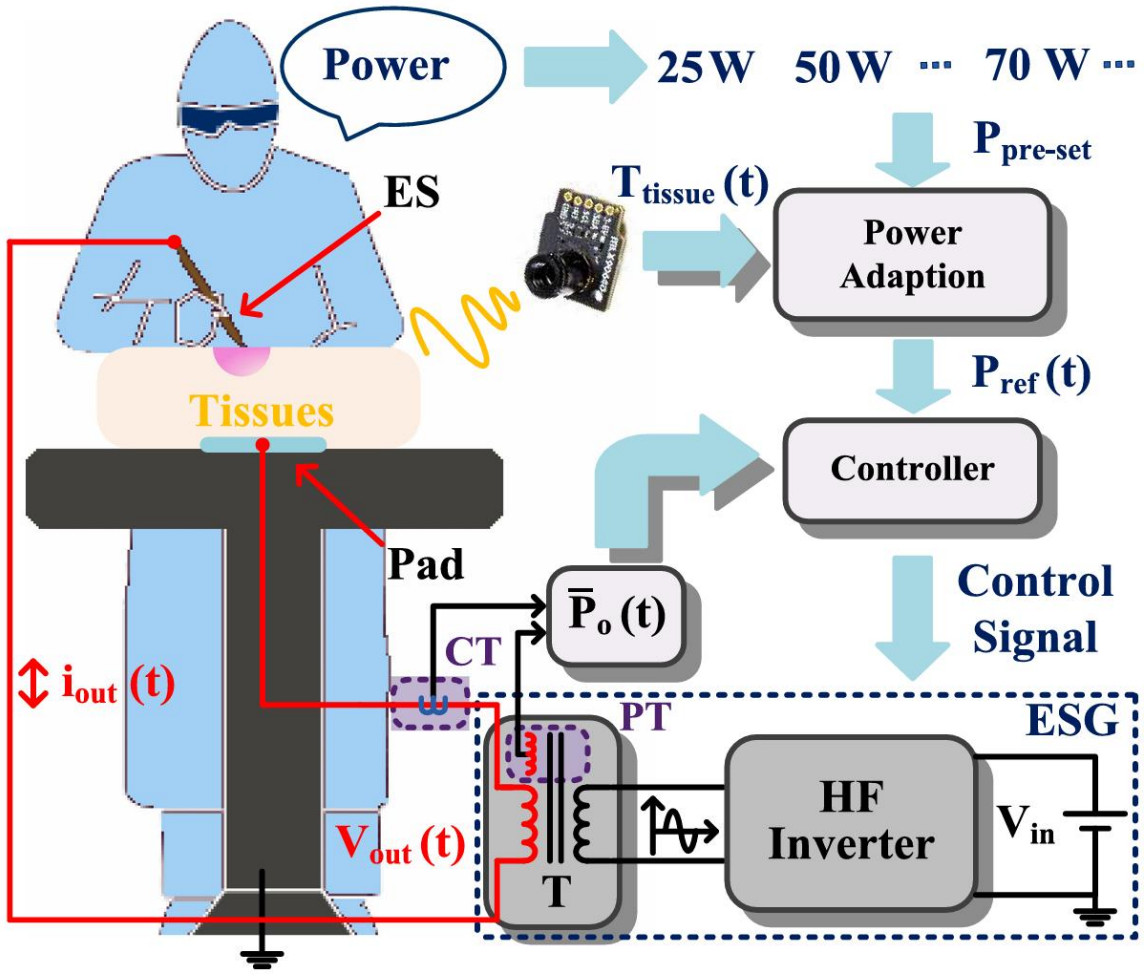


Figure 3. Flowchart of adaptive electrosurgery operation based on real-time temperature monitoring [12].

without compromising ergonomics or functionality remains a significant technical challenge.

Moreover, most current systems employing this technology can only detect surface tissue temperatures or rely on impedance feedback, which fails to accurately reflect temperature variations in deeper tissues. Additionally, sensor latency or measurement errors may cause delayed power adjustments, undermining thermal control. Furthermore, most control algorithms are heuristic in nature and struggle to adapt to the rapidly changing conditions encountered during surgery.

The future development of this technology will largely depend on advancements in temperature sensors. High-precision, interference-resistant miniature sensors are essential, and the lack of such devices is a major limitation preventing this technology from reaching its full potential. Previous studies have also demonstrated that optimizing electrode geometry can significantly minimize lateral thermal spread during radiofrequency procedures [13]. Additionally, Yang et al. designed and simulated a multi-functional radiofrequency electrode structure, enabling integration of temperature feedback with structural optimization [14]. The integration of artificial intelligence (AI) and machine learning can enhance the responsiveness of cutting algorithms. Once these algorithms mature, standardized feedback parameters will be crucial for their consistent implementation.

Active cooling system

Beyond regulating the power source, another strategy to reduce thermal damage involves actively removing heat from the electrosurgical knife or surrounding tissues. This approach is particularly crucial in extended procedures where thermal buildup is unavoidable.

Cooling systems are widely used to reduce thermal damage caused by high-frequency electrosurgical knives, especially in long or large-scale surgeries where the knife tip temperature continuously rises, leading to overheating and unnecessary tissue damage. The cooling system works by lowering the electrosurgical knife tip's temperature, reducing heat accumulation and effectively preventing thermal damage. For instance, Qian et al. achieved precision cooling during radiofrequency ablation under real-time temperature monitoring, with both simulation and experimental results confirming significant temperature reduction [15].

The water cooling system is the most commonly used method. It circulates cooling water through the electrosurgical knife handle and tip, removing excess heat and maintaining the knife tip temperature within an acceptable range. Research has shown that water cooling systems can significantly reduce the temperature of high-frequency electrosurgical knives and mitigate heat diffusion [16].

The air cooling system uses cool air or carbon dioxide (CO_2) to cool the knife tip and surrounding tissue. This method offers a fast response time and does not interfere with the surgical field of view, making it suitable for short-term procedures or small-scale cuts. In the air cooling system, controlling the flow rate and temperature of the gas is crucial for efficiently reducing temperature without disrupting the surgical operation.

In summary, cooling systems primarily address the second thermal damage mechanism discussed in Chapter 2. By providing additional cooling, these systems help dissipate excess heat rapidly, reducing heat exposure to surrounding healthy tissue. Not only do cooling systems significantly reduce the risk of thermal damage, but they also provide a more stable working environment for surgeons, ensuring optimal cutting and coagulation performance.

The main challenges faced by active cooling systems include the integration of the cooling mechanism into electrosurgical devices, ensuring the stability of the cooling effect, and preventing the surgical field from being obstructed by smoke or vapor produced during the cooling process. Additionally, as temperature changes dynamically, fluctuations in the cooling channel's flow rate can pose potential safety concerns that need to be addressed.

Furthermore, incorporating a cooling system increases the structural complexity and weight of the device, which may reduce the surgeon's operational sensitivity. Cooling systems may also lead to overcooling or condensation, potentially impairing tissue visibility or causing cold-related tissue injury.

To address these challenges, future cooling systems should incorporate artificial intelligence technologies to dynamically adjust the cooling intensity based on real-time temperature readings from the electrode tip. Additionally, passive or compact cooling technologies, such as phase-change materials or thermoelectric components, should be explored further.

Pulsed cutting technology

A recent development in electrosurgery is waveform modulation, specifically pulsed cutting technology. Unlike continuous high-frequency current, this technique delivers energy in short bursts, allowing tissues to cool between pulses.

Pulsed cutting technology replaces continuous high-frequency current with intermittent pulse currents, providing cooling periods during which the tissue can dissipate heat naturally. The principle of this technology is to achieve tissue cutting through short, high-frequency pulses, while the gaps between pulses allow for heat dissipation. By adjusting pulse duration, output voltage, and power can be dynamically controlled in real-time.

This method not only minimizes thermal damage to surrounding tissues but also improves cutting precision, as the tissue remains cooler during the procedure. It is especially beneficial for surgeries where maintaining tissue integrity and reducing thermal damage are crucial.

Firstly, pulsed cutting reduces the average temperature, preventing tissue damage such as protein denaturation and necrosis, which result from prolonged exposure to high temperatures. Secondly, the intermittent current reduces the spread of heat to surrounding tissues, effec-

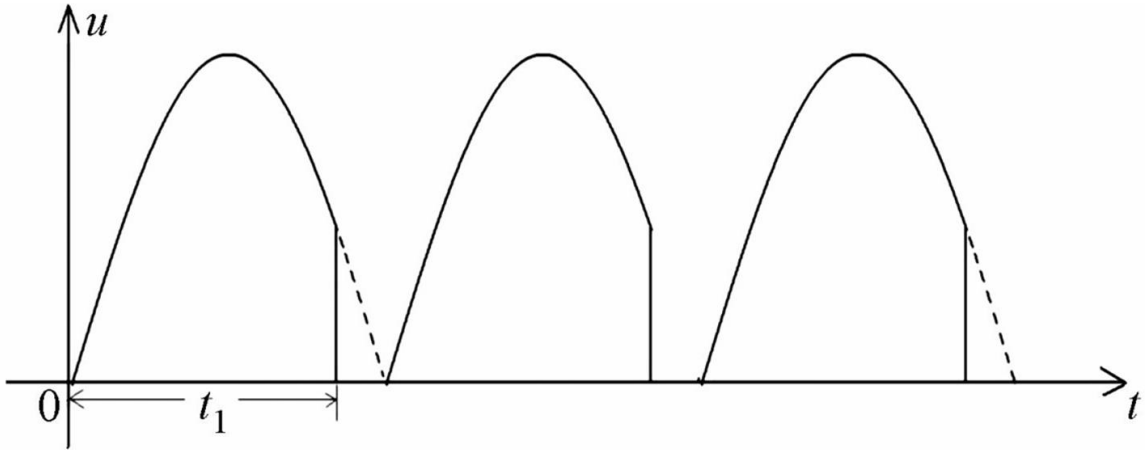


Figure 4. Schematic diagram of pulse cutting power control system output [17].

tively limiting the area affected by thermal damage.

Figure 4 illustrates pulsed cutting technology allows electrosurgical knives to maintain effective cutting performance while significantly reducing thermal damage, making it especially suitable for tissues sensitive to temperature changes. This technique has proven to be an effective, low-thermal-damage surgical method. Experimental results by Bao Congbo et al. demonstrated that without a blank period, widespread heat diffusion and noticeable carbonization occurred [17]. However, with intermittent current blank periods, heat diffusion was greatly minimized, and tissue carbonization was reduced.

Pulsed cutting technology addresses both thermal damage mechanisms discussed in Chapter 2. It achieves power control during electrosurgical output through adaptive algorithms and uses the pulsed method to control excessive heat, effectively reducing the risk of thermal damage.

While pulsed cutting technology is highly effective, its implementation requires precise control of pulse parameters—such as duration, duty cycle, and frequency—to balance efficient cutting with minimal thermal diffusion. Additionally, in regions of dense tissue, cutting efficiency may decrease, and different tissue types respond differently to pulsed energy, making uniform parameter settings difficult.

In summary, the three strategies—intelligent feedback control, active cooling, and pulsed waveform modulation—represent distinct but complementary approaches to minimizing thermal damage in electrosurgery. Their integration into unified, smart surgical platforms may define the next generation of safe and precise electrosurgical devices.

The future development of this technology lies in three main directions: first, the development of adaptive pulse modulation systems based on real-time tissue feedback; second, integrating pulsed cutting with intraoperative image-guided navigation to achieve precise energy targeting; and third, incorporating AI-driven control systems to dynamically adjust pulse output based on tissue response, thereby enhancing cutting precision and surgical safety.

Summary and future outlook

Since the introduction of high-frequency electrosurgical knives, they have revolutionized surgery with their efficient cutting and coagulation abilities. However, the widespread use of these devices has led to thermal damage, which limits their effectiveness. Thermal damage not only affects surgical outcomes and prolongs recovery times but can also lead to surgical failure. Thus, reducing thermal damage caused by high-frequency electrosurgical knives has become a key focus in surgical innovation.

Recent progress has been made in addressing this issue with various technological solutions, such as cooling systems, temperature feedback control, and pulsed cutting modes. Cooling systems effectively lower the knife tip temperature, reducing thermal diffusion from the source. The

Table 1. Comprehensive comparison of three technologies

Evaluation Dimension	Adaptive Control	Active Cooling System	Pulsed Cutting Technology
Effectiveness	Maintains tissue temp <80 °C with real-time feedback	Lowers knife temp by >20 °C; mitigates lateral heat spread	Reduces heat accumulation; carbonization area down ~35%
Applicability	Suitable for most tissues; ideal for long or precise surgeries	Best for large-area or prolonged procedures; limited by structural setup	Ideal for thermally sensitive tissues; protects nerves and vessels
Operational Complexity	High – Requires sensor integration and algorithm control	High – Adds hardware bulk; requires fluid/gas control	Medium – Needs pulse generator and tuning
Cost Impact	+30-50% – Sensors and AI algorithm overhead	+40-60% – Cooling system cost and maintenance	~+10% – Can reuse existing power hardware
Quantitative Evidence	Charring reduced by ~40%; cutting temp stabilized ±2 °C [13]	Tip temp drop ~23 °C; lower lateral spread [12]	Carbonization reduced ~35%; avg temp lower by 10-15 °C [13]
Overall Evaluation	High precision, intelligent; higher threshold for implementation	Effective thermal risk control; best for specific cases	Mature and cost-efficient; suitable for broader adoption

combination of temperature feedback and intelligent power regulation technologies allows for real-time adjustments, minimizing thermal damage to surrounding tissues. Pulsed cutting technology mitigates heat accumulation through intermittent currents, improving the precision and safety of the cut.

These three technologies each have distinct advantages and disadvantages, addressing the reduction of thermal damage through both internal control and external intervention. The outcomes vary in effectiveness depending on the approach, as summarized in **Table 1**.

Despite these advancements, further improvements are still needed, particularly in device complexity, ease of operation, and real-time feedback mechanisms during surgery. Additionally, refining thermal management strategies to accommodate different surgical types, tissue characteristics, and patient variations remains a key focus for future research. The future development of reducing thermal damage in high-frequency electrosurgical knives will likely center around three key areas:

(1) Intelligent Adaptive Control: With the advancement of artificial intelligence (AI) and machine learning, future electrosurgery systems will not only be tools but intelligent assistants [18-20]. By analyzing large amounts of surgical data (such as temperature, impedance, and tissue type), AI will help surgeons adjust the knife's output parameters in real-time, accurately controlling temperature changes during cutting and minimizing thermal damage. For example, deep learning algorithms will allow electrosurgery systems to automatically adjust power, frequency, and pulse modes according to the tissue's real-time status, ensuring optimal cutting and coagulation effects while minimizing thermal damage. Future electrosurgical systems may incorporate additional sensors (e.g., infrared sensors, micro thermocouples, pressure sensors) to monitor multiple data dimensions such as tissue temperature, pressure, and elasticity [20]. Combined with AI for multi-level data analysis, the system could automatically adjust output and issue warnings when tissue temperature nears safety thresholds, ensuring more precise operations and reducing thermal risks.

(2) Higher Precision Temperature Monitoring and Feedback Systems: To further improve cutting precision and reduce thermal damage, future temperature monitoring systems will focus on miniaturization, real-time capabilities, and accuracy. Current temperature sensors (such as infrared or thermocouples) often suffer from large size and slow response times [21]. Future miniaturized sensors will offer more accurate monitoring of small temperature changes in tissue, providing real-time feedback to the electrosurgical system, enabling timely intervention. For example, implantable temperature sensors can be placed directly in the surgical area to monitor temperature changes continuously, providing more accurate data than external sensors. These sensors could wirelessly connect to the surgical equipment, providing

"zero error" temperature feedback and helping avoid tissue damage due to overheating.

(3) Introduction of Innovative Materials with Good Thermal Conductivity and Biocompatibility: Beyond thermal management, innovative surgical materials such as biodegradable staples have been explored for synergistic safety improvement in gastrointestinal anastomosis [22]. With the development of materials science, future high-frequency electrosurgical knives may utilize nanocoating electrodes or biocompatible materials (e.g., the Ti-SS304 layered composite material discovered by Harasani I et al.) to reduce thermal damage [23]. These new electrodes not only improve heat dissipation but also enhance tissue-contact efficiency, reducing carbonization and thermal damage. For example, carbon nanotube coatings can improve electrode thermal conductivity and biocompatibility, stabilizing the electrosurgery during high-temperature cutting and reducing the risk of overheating. Additionally, these materials reduce electrode wear and corrosion, extending the lifespan of the equipment.

Author contributions: Yuxiang Luo was responsible for collecting data, writing the paper, revising the content, and proofreading the paper. Zhou Yu was responsible for providing direction for the paper and guiding the writing. Others was responsible for offering technical support, reviewing relevant literature, and assisting in refining the manuscript.

References

[1] Moak E. Electrosurgical unit safety. The role of the perioperative nurse. AORN J 1991;53(3): 744-746, 748-749, 752.

[2] Massarweh NN, Cosgriff N, Slakey DP. Electrosurgery: history, principles, and current and future uses. J Am Coll Surg 2006;202(3):520-530.

[3] Marrero K, Fingeret A. The innovator of electrosurgery. J Craniofac Surg 2019;30(7):1936-1937.

[4] Kearns S, Connolly E, McNally S, et al. Randomized clinical trial of diathermy versus scalpel incision in elective midline laparotomy. Br J Surg 2001;88(1):41-44.

[5] Martinsen OG, Heiskanen A. Bioimpedance and bioelectricity basics. Elsevier, 2023.

[6] Feldman L, Fuchshuber P, Jones DB. The SAGES manual on the fundamental use of surgical energy (FUSE). Springer, 2012.

[7] Tunc B, Gulsoy M. Tm:fiber laser ablation with real-time temperature monitoring for minimizing collateral thermal damage: ex vivo dosimetry for ovine brain. Lasers Surg Med 2013;45(1):48-56.

[8] Mazumder SK, Bao C, Mehta AI. Power-electronics enabled precision-power electrosurgery. IEEE Power Electronics Magazine 2023;10(4):20-25.

[9] Bao C, Mazumder SK. Output power computation and adaptation strategy of an electrosurgery inverter for reduced collateral tissue damage. IEEE Transactions on Biomedical Engineering 2023;70(6):1729-1740.

[10] Loiola BR, Orlande HR, Dulikravich GS. Thermal damage during ablation of biological tissues. Numerical Heat Transfer, Part A: Applications 2018;73(10):685-701.

[11] Bao C, Mazumder SK. Reduced collateral tissue damage using thermal-feedback-based power adaptation of an electrosurgery inverter. IEEE Transactions on Power Electronics 2022;37(10):11540-11545.

[12] Belik DV, Kustov IN, Bukovskiy MP, et al. Reduced thermal injuries bio-tissue during surgery with the use of electrosurgical devices. 2016 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE) 2016;427-428.

[13] Mao L, Xue H, Hu Z, et al. Optimizing electrode design to minimize thermal spread in radiofrequency-induced colonic anastomosis. Progress in Medical Devices 2023;1(1):42-54.

[14] Yang W, Mao L, Chen Y, et al. Design and simulation of a multi-functional radiofrequency tissue welding electrode. Progress in Medical Devices 2023;1(3):155-162.

[15] Qian R, Lin H, Gu D, et al. Precision cooling radiofrequency ablation under tumor boundary temperature control. Progress in Medical Devices 2023;1(2):120-130.

[16] Donzelli J, Leonetti JP, Wurster RD, et al. Neuroprotection due to irrigation during bipolar cautery. Arch Otolaryngol Head Neck Surg 2000;126(2):149-153.

[17] Bao C, Mazumder SK. MCU-based safer coagulation mode by nonfixed duty cycle for an electrosurgery inverter. IEEE Transactions on Power Electronics 2024;39(7):7938-7951.

[18] Waheed A, Mahmood T. Published- Integrating artificial intelligence (AI) in the identification of pain. Anaesthesia, Pain & Intensive Care 2025;29(1):1-2.

[19] Kim B, Ryan K, Kim JP. Assessing the impact of information on patient attitudes toward artificial intelligence-based clinical decision support

(AI/CDS): a pilot web-based SMART vignette study. J Med Ethics 2025;51(8):541-549.
[20] Li X, Sheng Y, Wu QJ, et al. Clinical commissioning and introduction of an in-house artificial intelligence (AI) platform for automated head and neck intensity modulated radiation therapy (IMRT) treatment planning. J Appl Clin Med Phys 2025;26(1):e14558.
[21] Altuntaş H, Arslan MS. A New Approach for Infrared Temperature Measurement Sensor Systems and Temperature Control for Domestic Induction Hobs. Sensors (Basel) 2025;25(1):235.
[22] Cai X, Mao L, Shen J, et al. Research progress of biodegradable staples in gastrointestinal anastomosis. Progress in Medical Devices 2024;2(1):38-43.
[23] Harasani I, Jamaati R, Aval HJ. Electrochemical aspects and in vitro biocompatibility of Ti-SS304 layered composite. Journal of Materials Research and Technology 2023;26:2847-2862.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/828047gamhgw

A method for predicting the outcome of PD1/PD-L1 inhibitors in non-small cell lung cancer

Wensong Yan[1], Shiju Yan[1], Yunhua Xu[2]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. [2]Department of Oncology, Shanghai Chest Hospital, Shanghai 200030, China.

Corresponding author: Shiju Yan.

Declaration of conflict of interest: None.

Received January 13, 2025; Accepted April 15, 2025; Published December 31, 2025

Highlights

- This study introduces a novel method to predict the efficacy of PD1/PD-L1 inhibitors in non-small cell lung cancer by extracting radiomic features from pre-treatment and post-treatment CT images.
- Integrating biological features and radiomic features enhances predictive performance.
- The newly proposed segmentation model achieved a dice coefficient of 90.09%, enabling accurate lesion segmentation.

Abstract

Objective: To propose a method for predicting immunotherapy outcome in non-small cell lung cancer based on computed tomography images before and after treatment. **Methods:** An improved U-net model incorporating Efficient Channel Attention was used to segment lesions. Radiomic features of lesions were extracted using PyRadiomics package and combined with biological indicators. Feature selection and dimensionality reduction were performed using linear discriminant analysis and Pearson correlation algorithms. A support vector machine was used to establish the predictive model. **Results:** The proposed segmentation model achieved a Dice coefficient of 90.09%, a positive predictive value of 89.23%, and an intersection over union of 82.15%, outperforming mainstream segmentation models. The proposed predictive model achieved an area under the curve of 85.05%, accuracy of 77.59%, specificity of 81.68% and sensitivity of 73.52%, all superior to models based solely on single-time computed tomography images or lacking biological features. **Conclusion:** The proposed method provides an effective approach for predicting the efficacy of immunotherapy in non-small cell lung cancer patients and offers a valuable tool to support clinical decision-making.

Keywords: Immunotherapy, medical image segmentation, non-small cell lung cancer, machine learning, feature engineering

Introduction

Lung cancer is a prevalent malignancy worldwide, with non-small cell lung cancer (NSCLC) accounting for 80%-85% of all lung cancer cases, making its treatment a critical focus in the medical field [1, 2]. Studies have demonstrated that immune checkpoint inhibitors (ICIs) significantly enhance the objective response rate and prolong the survival of lung cancer patients. As a result, PD-L1 testing is clinically recommended for NSCLC patients to identify those who may benefit from immunotherapy [3, 4]. However, even PD-L1-negative patients have shown benefits from ICIs, suggesting that the efficacy of ICIs in NSCLC is not universally predictable [5, 6]. Additionally, PD-L1 testing, being invasive, is not universally suitable [7]. Therefore, developing a model capable of accurately and non-invasively predicting the efficacy of lung cancer immunotherapy is crucial for optimizing treatment strategies and improving patient outcomes. The segmentation of lesion areas and the selection of relevant features are critical components of this process.

The U-Net network proposed by Ronneberger et al. has been widely used in medical image processing due to its superior performance and

Address correspondence to: Shiju Yan, School of Health Sciences and Engineering, University of Shanghai for Science and Technology, No.580 Jungong Road, Shanghai 200093, China. Tel: +86-18217617984. E-mail: yanshiju@usst.edu.cn.

relatively low training sample requirements [8]. The U-Net++ proposed by Zong et al. enhanced feature map connectivity across different resolutions, enabling the capture of higher-level features and improving segmentation performance [9]. Chen et al. proposed the Trans-U-net, which inputs features extracted by a convolutional neural network encoder into a transformer to model global relationships, thereby enhancing the model's contextual understanding capabilities [10]. Cao et al. proposed the Swin-U-net, which incorporates a Hierarchical Self-Attention Mechanism and Shifted Window technique to improve feature extraction and processing efficiency [11]. While U-Net and U-Net++ are convolutional neural network-based models known for their lightweight parameters, they have limitations in capturing global image structure and context. In contrast, transformer architecture-based models, such as Trans-U-Net and Swin-U-net, utilize the transformer's ability to capture global information, making them particularly effective in computing long-distance dependencies. This enables more efficient feature learning and detailed recovery.

He et al. established a model to predict the outcome of ICIs using radiomic biomarkers as features, achieving an Area Under the Curve (AUC) of 0.81 [12]. In another study, He et al. classified the ICI outcomes in NSCLC patients into disease progression and partial response using CT image features, achieving AUC values of 0.78 and 0.81, respectively [13]. Shen et al. extracted ten texture feature sets to predict efficacy, resulting in an AUC of 0.81 [14]. Gong et al. combined 1,118 CT radiomic features to predict ICI efficacy, achieving an AUC of 0.82 [15]. These studies demonstrate the effectiveness of imaging data in predicting treatment outcomes. However, they primarily focused on medical image data from a single time point, overlooking the temporal changes in tumor characteristics. Additionally, these studies did not incorporate patients' biological indicators. Combining imaging data with biological and clinical indicators offers a more comprehensive assessment of a patient's overall health.

This study presents an improved U-Net segmentation model based on Efficient Channel Attention (ECA) to accurately segment the lesion regions of interest in CT images [16]. The proposed prediction model combines radiomic features from chest CT images and biological indicator features collected from two time points, before and after immunotherapy. Feature dimensionality reduction and selection were performed using LDA and Pearson algorithms, and a SVM was employed to establish the prediction model [17-19].

Materials and methods

Materials

This study was approved by the Ethics Committee of Shanghai Chest Hospital (Ethics approval number: KS24074). The research included 427 pairs of pre- and post-treatment CT images and biological markers (12 biomarkers, including white blood cell count and neutrophil count) from 103 NSCLC patients [20, 21]. Lesions were annotated by two senior radiologists, while treatment efficacy was assessed by one senior radiologist and one senior pulmonologist according to the RECIST 1.1 criteria. Efficacy was categorized into three types: partial response, stable disease, and progressive disease [22, 23].

Methods

The overall workflow of this study is illustrated in **Figure 1**. The proposed segmentation model was employed to segment lesions, followed by extraction of various features from CT images. Biological features were also collected from clinical test reports of the patients. Feature selection and dimensionality reduction were then performed to identify the features sensitive to predicting immunotherapy efficacy. Finally, the processed features were input into an SVM classifier to predict the immunotherapy outcome for the patients.

Lesion segmentation model

An improved segmentation model, named the Multi-Size Self-Attention model (MSSA) is proposed in this study. The workflow is illustrated in **Figure 2**. We employed a multi-size self-attention learning module based on the ECA module, which is embedded after every two 3x3 convolutional layers in the encoder. Additionally, a prompt-based cropping module that uses the mouse-clicks as prior knowledge was incorporated into the MSSA.

Multi-scale self-attention module

The multi-scale self-attention module employs a two-layer structure. Initially, the original feature map is divided into sub-feature maps, each with half the original size. These sub-feature maps are then processed using the multi-head self-attention mechanism to compute pixel-wise associations in both local

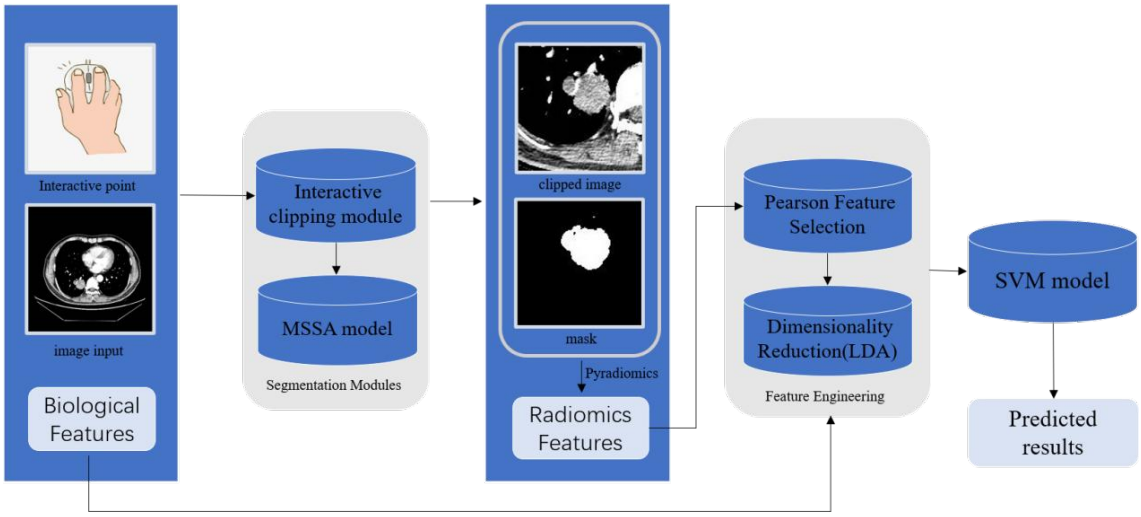


Figure 1. Overview of the prediction framework. The framework consists of three main modules: segmentation, feature engineering, and prediction. The segmentation module segments the lesion regions of interest, which are then input into PyRadiomics to extract radiomic features. These features undergo feature selection and dimensionality reduction, which are combined with biological features for further dimensionality reduction. Finally, the processed features are input into the SVM model.

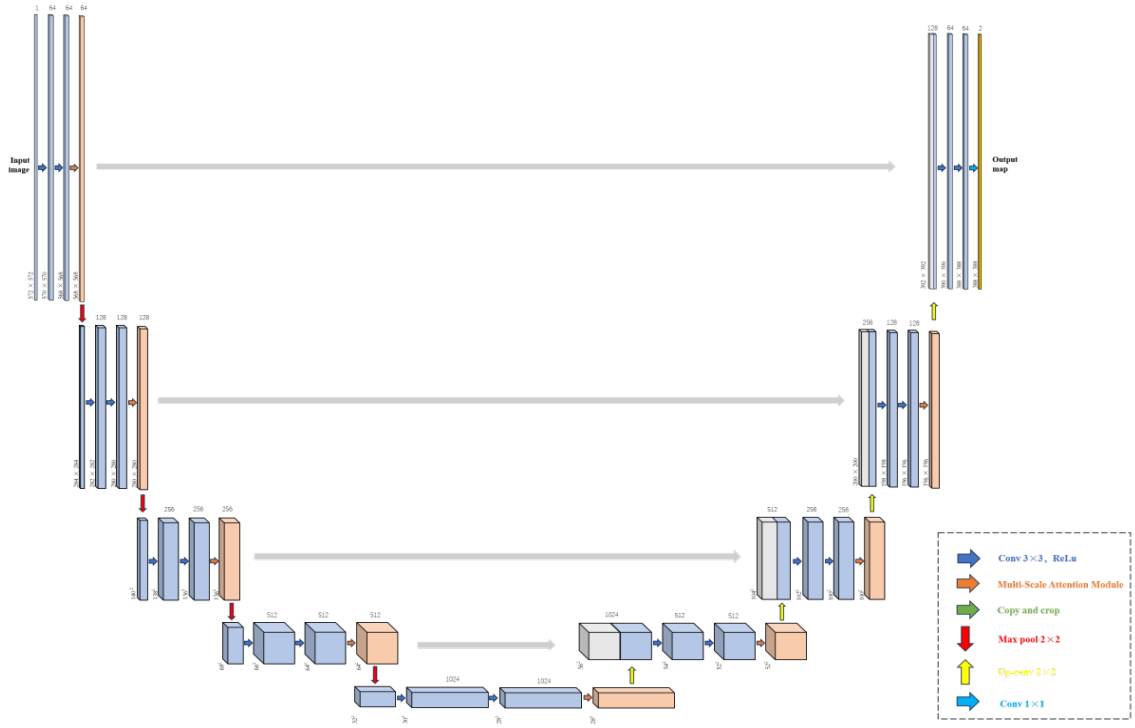


Figure 2. The architecture of the proposed MSSA. Blue boxes correspond to a multi-channel feature maps; Orange boxes refer to feature maps processed by the Multi-Size Self-Attention Learning Block; White boxes represent copied feature maps; Arrows denote different operations.

and global regions. The resulting pixel features are further processed by the ECA module, which enhances key features by weighting each channel. Afterward, the processed sub-feature maps are concatenated back to their original size, producing a weighted feature map. This weighted feature map is then divided into sub-feature maps, each with one-fourth of the original size, and the aforementioned operations are repeated. This method is particularly effective for lesion segmentation in lung CT images, addressing issues such as boundary blurring and feature extraction difficulties, ultimately achieving more precise segmentation. The diagram is shown in **Figure 3**.

The multi-head self-attention mechanism

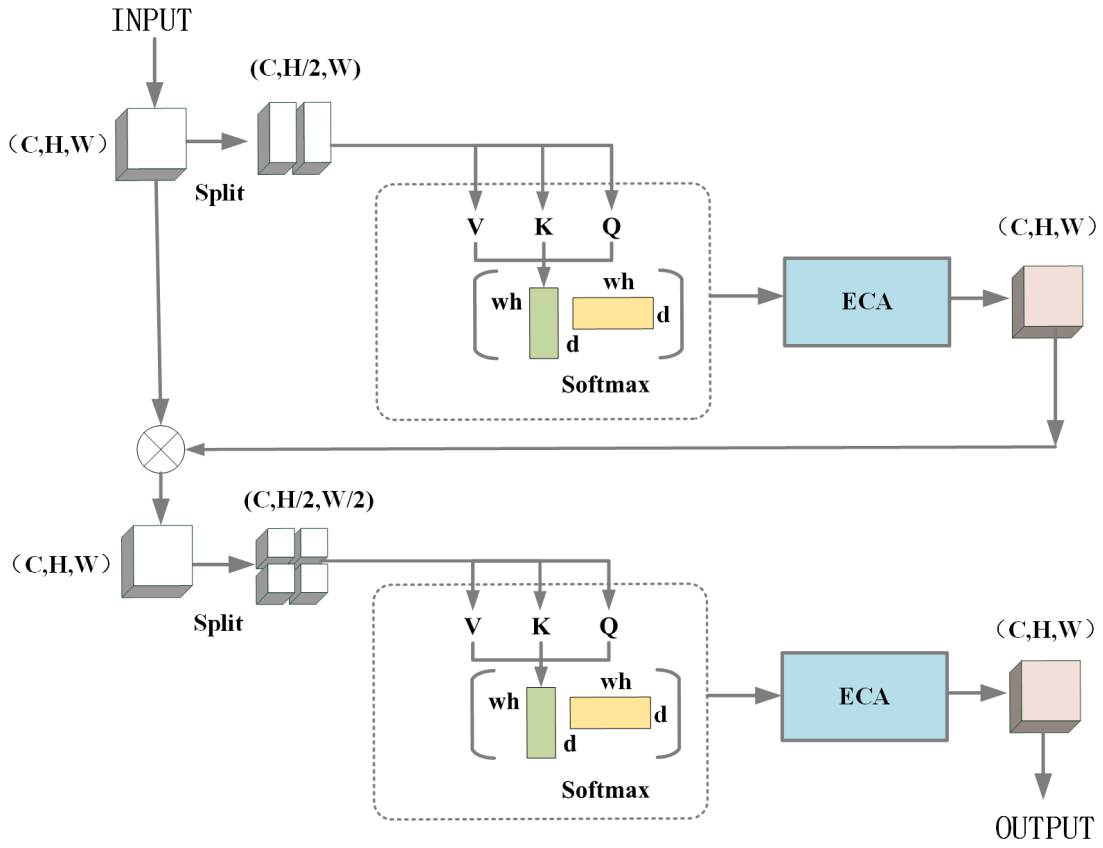


Figure 3. Schematic diagram of the multi-scale self-attention learning module (H, W, and C represent the height, width, and channel dimensions of the feature map, respectively). The input feature map is divided into 2 sub-feature maps of half size and then sequentially fed into the multi-head self-attention module and Efficient Channel Attention (ECA) module. Subsequently, the sub-feature maps are concatenated back to the original size. Following this, the resulting feature map is further divided into 4 sub-feature maps of one-quarter size, and the aforementioned process is repeated.

generates weight matrices W_g, W_k and W_v based on the similarity relationships between pixels [24]. Then, the query matrix (Q), key matrix (K), and value matrix (V) are obtained according to Equation (1):

$$\begin{aligned} Q &= W_g \cdot X \\ K &= W_k \cdot x \\ V &= W_v \cdot x \end{aligned} \quad (1)$$

Thus, the mathematical expression of the self-attention mechanism is:

$$\sum_{i=1}^{N}\left[\mathrm{Softmax}\left(\frac{\exp(s_i)}{\sum_{j=1}^{N}\exp(s_j)}\right)\cdot \boldsymbol{V}_i\right]; s_i = \frac{QK_i^{\mathrm{T}}}{\sqrt{d}} \quad (2)$$

Where: N represents the maximum index value within the matrix; Si denotes the similarity between Q and K; Vi represents the value indexed by i in V; and d is the dimension of Q and K. Following the self-attention mechanism, an ECA module is incorporated. The ECA module utilizes a linear transformation combined with global average pooling, offering lower computational complexity compared to other channel attention mechanisms. **Figure 4** illustrates the corresponding schematic diagram:

First, global average pooling is applied to the feature map, as shown in Equation (3):

$$\mathrm{GAP}(\boldsymbol{U}) = \frac{1}{W\times H}\sum_{i=1,j=1}^{WH} U_{i,j} \quad (3)$$

Then, convolution operations are performed on *GAP(X)* using a dynamic 1D convolutional kernel of size k to learn the importance weights between different channels. The definition of the convolution kernel adaptive function and weights ω∈R1×1×C is as follows:

$$k = \psi(C) = \left|\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{\mathrm{odd}} \quad (4)$$

$$\boldsymbol{\omega} = \mathrm{Sigmoid}\left(\mathrm{ClD_k}\left(\mathrm{GAP(U)}\right)\right) \quad (5)$$

Where, in Equation (4), k represents the convolutional kernel size, $|.|_{\mathrm{odd}}$ ensures that k is an odd number, and γ and b are used to adjust the ratio between C and k; in Equation (5), the vector ω adjusts the attention

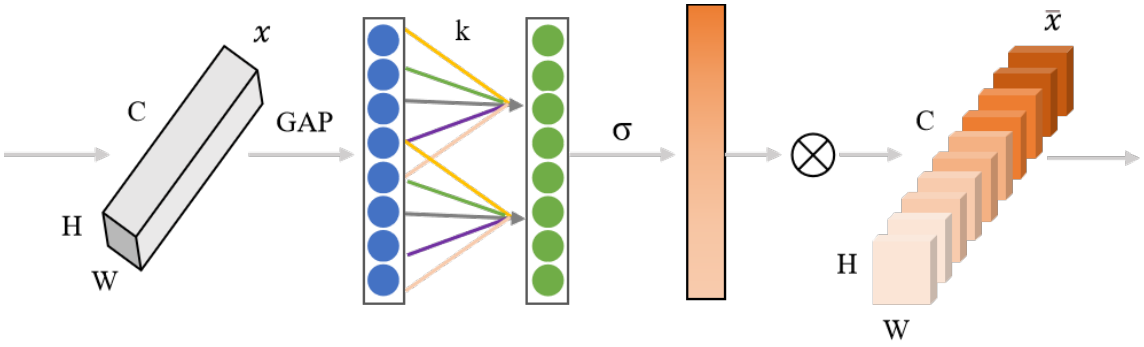


Figure 4. Schematic diagram of the Efficient Channel Attention (ECA) mechanism. X represents the input feature map. First, global average pooling (GAP) is performed, followed by a convolution using a 1D convolutional kernel of size k. σ denotes the Sigmoid activation function, represents element-wise multiplication, and is the weighted feature map. The features of each channel are adjusted according to their importance.

weights of each channel, and the Sigmoid function computes the activation values of the 1D convolution output. Finally, the generated weights are applied to the original feature map through element-wise multiplication, enhancing important channels while suppressing less important ones [25].

Prompt-based cropping module

In this module, a physician is required to perform a single click at the center of the lesion on any slice of the CT scan. The click point serves as the center, for cropping the CT image according to Equation (6), resizing the CT image to (128, 128) pixels, as illustrated in **Figure 4**. When processing 3D CT data, this interactive point is propagated to other CT slices via translation, allowing all CT slices to undergo the cropping operation.

$$\begin{aligned} X1 &= X - \frac{R}{2} \\ X2 &= X + \frac{R}{2} \\ Y1 &= Y - \frac{R}{2} \\ Y2 &= Y + \frac{R}{2} \\ \mathrm{Img} &*= \mathrm{Img}[X1{:}X2, Y1{:}Y2] \end{aligned} \tag{6}$$

Here, (*X, Y*) represents the coordinates of the interactive point provided by the physician, R is the size of the desired ROI to be cropped, and *X1, X2, Y1, Y2* are the horizontal and vertical coordinates after cropping. The brackets [] denote the slicing operation. To save labor, the interactive points in the training set are obtained based on the mask images.

Feature selection

Radiomic features were extracted from the CT images using the PyRadiomics tool. These features were then matched with the patient's data and the timing of the CT examinations. This resulted in a dataset of 427 pairs of 3D CT images from the same patient before and after treatment, comprising over 1,500 different feature differences. Two feature selection methods (Pearson and LASSO) were applied to this dataset to identify the top 21 radiomic features most relevant to treatment efficacy [26-29]. Furthermore, these selected features were combined with 12 biological features for Principal Component Analysis (PCA) and LDA, aiming to optimize the data representation [30, 31].

Outcome prediction

SVM was selected as the classification prediction model due to its strong ability in handling high-dimensional and nonlinear data. By maximizing the margin and utilizing kernel functions, SVM demonstrates excellent generalization performance and robustness against overfitting, making it widely used in the medical field as well as other domains [32-34].

Experiments and results

Experimental configuration

All experiments were conducted on a system equipped with an Intel i12-12700KF processor, Nvidia RTX 3090 GPU, and 16 GB of memory. The software configuration included Python 3.9 and PyTorch 1.8. The dataset was split into training and testing sets in an 8:2 ratio. For the segmentation model, stochastic gradient descent was chosen as the optimizer, with a batch size of 9, an initial learning rate of 0.001, and a total of 100 iterations. The loss function combined binary cross-entropy and Dice loss (BCE-Dice-Loss). Due to the high computational cost, 5-fold cross-validation was employed

Table 1. Comparison of Segmentation Performance of Various Models on the Test Set

Methods/ Evaluation Metrics		IoU (%)	PPV (%)	Dice (%)	HD_95
U-net*		78.92±1.52	85.58±1.31	86.26±1.52	2.35±0.51
U-net++*		78.52±1.12	85.21±1.46	86.34±1.23	2.85±0.56
Swin-U-net*		79.82±1.58	86.27±1.71	87.06±1.12	2.23±0.36
Trans-U-net*		79.12±1.62	85.63±1.43	86.45±1.01	2.09±0.41
MSSA		80.08±1.27	88.86±1.12	88.40±0.95	2.21±0.43
Incorporation of Prompt-based Cropping Module	U-net*	79.85±1.34	87.92±1.01	87.85±1.82	2.33±0.53
	U-net++*	79.23±1.29	87.12±1.89	87.26±1.43	2.62±0.51
	Swin-U-net*	80.23±1.45	88.51±1.27	88.05±1.32	2.35±0.56
	Trans-U-net*	80.28±1.04	88.11±1.81	88.29±1.27	2.07±0.58
	MSSA	82.15±1.02	89.23±1.04	90.09±1.28	2.18±0.31

Note: * Indicates significant difference compared to MSSA model. IoU, Intersection over Union; PPV, Positive Predictive Value.

Table 2. Comparison of Segmentation Performance in Ablation Experiments for Different Models

Baseline Model	Module1	Module2	IoU (%)	PPV (%)	Dice (%)	Hd_95
√			78.92±1.52	85.58±1.31	86.26±1.52	2.15±0.51
√	√		80.08±1.27	88.86±1.12	88.40±0.95	2.86±0.31
√		√	79.85±1.34	87.92±1.01	87.85±1.82	2.33±0.53
√	√	√	82.15±1.02	89.23±1.04	90.09±1.28	2.18±0.43

Note: IoU, Intersection over Union; PPV, Positive Predictive Value.

to evaluate segmentation performance, balancing computational efficiency with evaluation reliability. The classification model SVM was trained and tested using the Scikit-learn package in Python and its performance was evaluated using 10-fold cross-validation, providing fine-grained and robust evaluation with relatively low computational burden [35].

Results

Lesion segmentation model

The upper part of **Table 1** presents the performance of various models on the lung lesion segmentation task. The U-Net network, serving as the baseline model, achieved a Dice coefficient of 86.26%. The best-performing comparison model was the Swin-U-Net, with a Dice coefficient of 87.06%. Our proposed method, MSSA, achieved an intersection over union (IoU) of 80.08%, positive predictive value (PPV) of 88.86%, and Dice coefficient of 88.40%, outperforming all other models.

The lower part of **Table 1** shows the performance of each model after incorporating the prompt-based cropping module. All models exhibited an improvement of over 1% in the dice coefficient. Notably, the MSSA model achieved an outstanding dice coefficient of 90.09%, representing a performance improvement of 6.72% compared to the baseline U-Net model.

Ablation experiments of segmentation model

Table 2 presents a comparison of segmentation performance for models derived from the base U-Net network, incorporating different combinations of two modules (Module 1: Multi-scale Self-attention Learning Module and Module 2: Prompt-based Cropping Module). Incorporating the Multi-scale Self-attention Learning Module improved the model's IoU, PPV, and Dice coefficient by 1.45%, 3.83% and 2.71%, respectively. Adding the Prompt-based Cropping Module further enhanced IoU, PPV and Dice coefficient by 1.19%, 2.73% and 1.84%, respectively. When both modules were integrated, the model's IoU, PPV and Dice coefficient are improved by 4.09%, 4.27% and 4.44%, respectively. These results validate that the combined use of both modules significantly enhances segmentation performance.

Table 3. Outcome Prediction Results with Different Data and Feature Selection Combinations

Feature Data	Feature Engineering	AUC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)
Data-A	Lasso + PCA	49.41±8.22	51.13±3.20	35.32±2.79	42.32±1.25	33.68±17.59
	Lasso + LDA	47.10±11.85	54.80±10.83	77.87±10.06	20.90±7.07	59.47±14.86
	Pearson +PCA	61.31±15.52	61.24±10.60	43.31±21.21	75.52±13.20	56.01±14.22
	Pearson +LDA	58.17±5.78	52.54±5.83	38.46±5.32	69.67±3.84	44.84±15.28
Data-B	Lasso + PCA	67.53±13.87	60.17±7.91	38.49±13.96	81.40±8.57	60.49±17.29
	Lasso + LDA	78.26±10.27	69.41±9.01	80.12±8.97	60.93±16.72	68.90±16.39
	Pearson +PCA	72.26±9.06	63.73±9.40	53.01±13.28	74.83±12.57	58.45±14.39
	Pearson +LDA	78.97±9.39	69.04±7.99	62.24±9.67	78.66±11.21	65.43±13.68
Data-C	Lasso + PCA	74.44±5.23	66.93±6.96	52.00±14.67	78.37±6.89	63.27±11.17
	Lasso + LDA	82.24±6.13	75.85±7.54	66.94±12.88	82.88±4.07	73.91±4.97
	Pearson +PCA	81.83±6.28	75.84±8.71	66.30±14.27	83.51±7.08	75.23±8.72
	Pearson +LDA	85.63±9.67	77.60±9.93	73.52±8.90	81.68±7.29	74.62±9.42

Note: PCA, Principal Component Analysis; LDA, Linear Discriminant Analysis; PPV, Positive Predictive Value.

Feature selection and outcome prediction

To evaluate the performance of the model, this study employed three groups of data:

Data-A: Radiomic features derived solely from the first period CT images of patients.

Data-B: Radiomic features from both the first and second period CT images.

Data-C: Radiomic features from both periods combined with biological features.

For each data group, two feature selection methods (Pearson and LASSO) and two feature dimensionality reduction methods (PCA and LDA) were applied in various combinations. The experimental results are shown in **Table 3**, with all prediction models using SVM for classification. The optimal feature selection scheme was identified as using Pearson and LDA. The final results for specificity, sensitivity, AUC, accuracy, and precision reached 85.63%, 77.60%, 73.52%, 81.68% and 74.62%, respectively. Notably, sensitivity, AUC, and accuracy outperformed other feature combinations.

Comparing Data-A and Data-B, experiments based solely on radiomic features from pre- and post-treatment CT images significantly outperform those based on a single-period CT image across all evaluation metrics. This highlights the superiority of using multi-period CT images. When comparing Data-B and Data-C, it was observed that the combination of radiomic and biological features consistently yielded better predictive performance over radiomic features alone, regardless of the feature engineering technique used. This indicates that combining radiomic and biological features provides a more effective strategy for predicting the efficacy of immunotherapy for lung cancer patients.

Ablation experiments on outcome prediction

Table 4 presents the comparison of classification performance for models based on the baseline SVM, using a dataset containing radiomic features from two-period CT images and biological features. Incorporating the Pearson feature selection method improved the model's AUC and Accuracy by 62.62% and 27.96%, respectively. Adding the LDA feature dimensionality reduction method resulted in an improvement in AUC and accuracy by 64.52% and 23.43%, respectively. When both methods are applied simultaneously, the AUC and accuracy increased by 81.80% and 41.61%, respectively, demonstrating that the incorporation of these feature engineering techniques can significantly enhance classification performance.

Discussion

This study proposes a method for predicting the efficacy of immunotherapy in lung cancer, leveraging pre- and post-treatment CT images.

Table 4. Ablation Experiment Results for Outcome Prediction

Baseline Model	Pearson	LDA	AUC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)
√			47.10±11.85	54.80±10.83	77.87±10.06	20.90±7.07	59.47±14.86
√	√		76.62±6.82	70.12±5.46	63.42±10.72	75.92±7.98	66.18±9.45
√		√	77.49±12.94	67.64±12.96	64.34±11.92	77.56±8.50	63.06±12.83
√	√	√	85.63±9.67	77.60±9.93	73.52±8.90	81.68±7.29	74.62±9.42

Note: LDA, Linear Discriminant Analysis; AUC, Area under the curve; PPV, Positive Predictive Value.

It combines an improved U-Net segmentation model, radiomic feature extraction, and clinical biological indicators to achieve precise predictions of immunotherapy outcomes for NSCLC patients.

In terms of lesion segmentation, we introduced a multi-scale self-attention learning module based on ECA. This module enhances the model's ability to identify lesion areas by performing self-attention computation and channel weighting at multiple scales. The experimental results show that the improved MSSA model outperforms some existing popular segmentation models across four evaluation metrics, including the Dice coefficient. Particularly, after integrating the prompt-based cropping module, the Dice coefficient reached 90.09%, significantly improving the segmentation accuracy. This improvement not only facilitates more precise extraction of radiomic features from lesion areas but also provides a reliable foundation for subsequent efficacy predictions.

In constructing the predictive model, this study incorporated data from pre- and post-treatment CT images, along with relevant clinical biological indicators. Using feature engineering methods like Pearson correlation analysis and LDA dimensionality reduction, we selected the most relevant feature combinations associated with treatment efficacy. The experimental results showed that models combining radiomic features from pre- and post-treatment CT images and biological features outperformed models relying solely on single-period CT images or those without biological features in terms of AUC, accuracy, sensitivity and specificity. This suggests that dynamic information on tumor changes during treatment, coupled with the clinical state of patients, plays a crucial role in predicting treatment efficacy, thereby offering more comprehensive and accurate predictive results.

Despite the achievements of this study, several limitations should be acknowledged. First, the sample size is relatively small and primarily drawn from a single center, which may limit the generalizability of the model. Future studies should validate and refine the model using larger, multi-center datasets. Second, the current research focuses predominantly on NSCLC patients. The applicability of the model to other types of lung cancer or its potential for predicting immunotherapy efficacy in other cancers remains to be explored.

Conclusion

This study proposes a method for predicting the outcome of immunotherapy in lung cancer patients based on pre- and post-treatment CT images. By integrating the proposed segmentation model, radiomic feature extraction, and clinical biological indicators, the method provides accurate predictions of immunotherapy efficacy in NSCLC patients. The proposed segmentation model outperforms existing popular methods, providing an effective solution for CT lesion segmentation in NSCLC. The prediction model achieves specificity, sensitivity, AUC, accuracy, and precision rates of 85.63%, 77.60%, 73.52%, 81.68%, and 74.62%, respectively. These results provide new insights into predicting immunotherapy efficacy in NSCLC, enhancing the scientific basis for medical decision-making. This method shows strong potential for broad application in clinical application, offering oncologists more precise and effective diagnostic and treatment support.

Author contributions: Wensong Yan was responsible for conceptualization, methodology, and writing of the original draft. Shiju Yan provided supervision, offered technical guidance, and contributed to review and editing. Yunhua Xu handled data collection and offered medical guidance. All authors have read and agreed to the final version of the manuscript.

References

[1] Hendriks LE, Remon J, Faivre-Finn C, et al. Non-small-cell lung cancer. Nat Rev Dis Primers 2024;10(1):71
[2] Lahiri A, Maji A, Potdar PD, et al. Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer 2023;22(1):40.
[3] Hsieh K, Dickstein DR, Runnels J, et al. Radiotherapy and immunotherapy in lung cancer. Biomedicines 2023;11(6):1642.
[4] Khadela A, Postwala H, Rana D, et al. A review of recent advances in the novel therapeutic targets and immunotherapy for lung cancer. Med Oncol 2023;40(5):152.
[5] Pan Y, Fu Y, Zeng Y, et al. The key to immunotherapy: how to choose better therapeutic biomarkers for patients with non-small cell lung cancer. Biomark Res 2022;10(1):9.
[6] John T, Sakai H, Ikeda S, et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in advanced non-small cell lung cancer: a subanalysis of Asian patients in CheckMate 9LA. Int J Clin Oncol 2022;27(4):695-706.
[7] Yin X, Liao H, Yun H, et al. Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer. Semin Cancer Biol 2022;86(Pt 2):146-159.
[8] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 2015;234-241.
[9] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support 2018;3-11.
[10] Chen J, Lu Y, Yu Q, et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 2021.
[11] Cao H, Wang Y, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation. European conference on computer vision 2022;205-218.
[12] He B, Dong D, She Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. J Immunother Cancer 2020;8(2):e000550.
[13] He B-X, Zhong Y-F, Zhu Y-B, et al. Deep learning for predicting immunotherapeutic efficacy in advanced non-small cell lung cancer patients: a retrospective study combining progression-free survival risk and overall survival risk. Transl Lung Cancer Res 2022;11(4):670-685.
[14] Shen L, Fu H, Tao G, et al. Pre-immunotherapy contrast-enhanced CT texture-based classification: a useful approach to non-small cell lung cancer immunotherapy efficacy prediction. Front Oncol 2021;11:591106.
[15] Gong J, Bao X, Wang T, et al. A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. Oncoimmunology 2022;11(1):2028962.
[16] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020;11534-11542.
[17] Jelodar H, Wang Y, Yuan C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2019;78:15169-15211.
[18] Cleophas TJ, Zwinderman AH. Bayesian Pearson Correlation Analysis. In: Cleophas TJ, Zwinderman AH, editors. Modern Bayesian Statistics in Clinical Research. Cham (SWITZERLAND). Springer International Publishing, 2018. p. 111-118.
[19] Wang H, Zhou W, Shao Y. A new fast ADMM for kernelless SVM classifier with truncated fraction loss. Knowl-Based Syst 2024;283:111214.
[20] Benzekry S, Grangeon M, Karlsen M, et al. Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data. Cancers 2021;13:6210.
[21] Mariam A, Kamath S, Schveder K, et al. Large-Scale Meta-Analysis of Potential Biomarkers for Treatment Response to Anti-PD-1/PD-L1 Immune Checkpoint Inhibitors. medRxiv 2020;2020.2011.
[22] Sun C, Ma X, Meng F, et al. Tumor microenvironment (TME) and single-source dual-energy CT (ssDECT) on assessment of inconformity between RECIST1. 1 and pathological remission in neoadjuvant immunotherapy of NSCLC. Neoplasia 2024;50:100977.
[23] He LN, Chen T, Fu S, et al. Reducing number of target lesions for RECIST1. 1 to predict survivals in patients with advanced non-small-cell lung cancer undergoing anti-PD1/PD-L1 monotherapy. Lung Cancer 2022;165:10-17.
[24] Zhong Y, Tang Z, Zhang H, et al. A froth image segmentation method via generative adversarial networks with multi-scale self-attention mechanism. Multimed Tools Appl 2024;83:19663-19682.
[25] Tian Q, Wang Z, Cui X. Improved Unet brain tumor image segmentation based on GSConv module and ECA attention mechanism. arXiv preprint arXiv:2409.13626 2024.

[26] Rahadian H, Bandong S, Widyotriatmo A, et al. Image encoding selection based on Pearson correlation coefficient for time series anomaly detection. Alex Eng J 2023;82:304-322.

[27] James G, Witten D, Hastie T, et al. An Introduction to Statistical Learning: with Applications in Python. Cham (SWITZERLAND). Springer International Publishing, 2023. p. 69-134.

[28] Antoniadis A, Lambert-Lacroix S, Poggi J-M. Random forests for global sensitivity analysis: A selective review. Reliab Eng Syst Safe 2021;206:107312.

[29] Ghosh P, Azam S, Jonkman M, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access 2021;9:19304-19326.

[30] Kurita T. Principal Component Analysis (PCA). In: Ikeuchi K, editors. Computer Vision: A Reference Guide. Cham (SWITZERLAND). Springer International Publishing, 2021. p. 1013-1016.

[31] Egwom OJ, Hassan M, Tanimu JJ, et al. An LDA–SVM machine learning model for breast cancer classification. BioMedInformatics 2022;2(3):345-358.

[32] Shanmugam S, Dharmar S. Implementation of a non-linear SVM classification for seizure EEG signal analysis on FPGA. Eng Appl Artif Intell 2024;131:107826.

[33] Alqhtani SM, Soomro TA, Shah AA, et al. Improved Brain Tumor Segmentation and Classification in Brain MRI With FCM-SVM: A Diagnostic Approach. IEEE Access 2024;12:61312-61335.

[34] Hoque R, Billah M, Debnath A, et al. Heart disease prediction using SVM. Int J Sci Res Arch 2024;11:412-420.

[35] Zhang X, Liu C-A. Model averaging prediction by K-fold cross-validation. J Econom 2023;235:280-301.