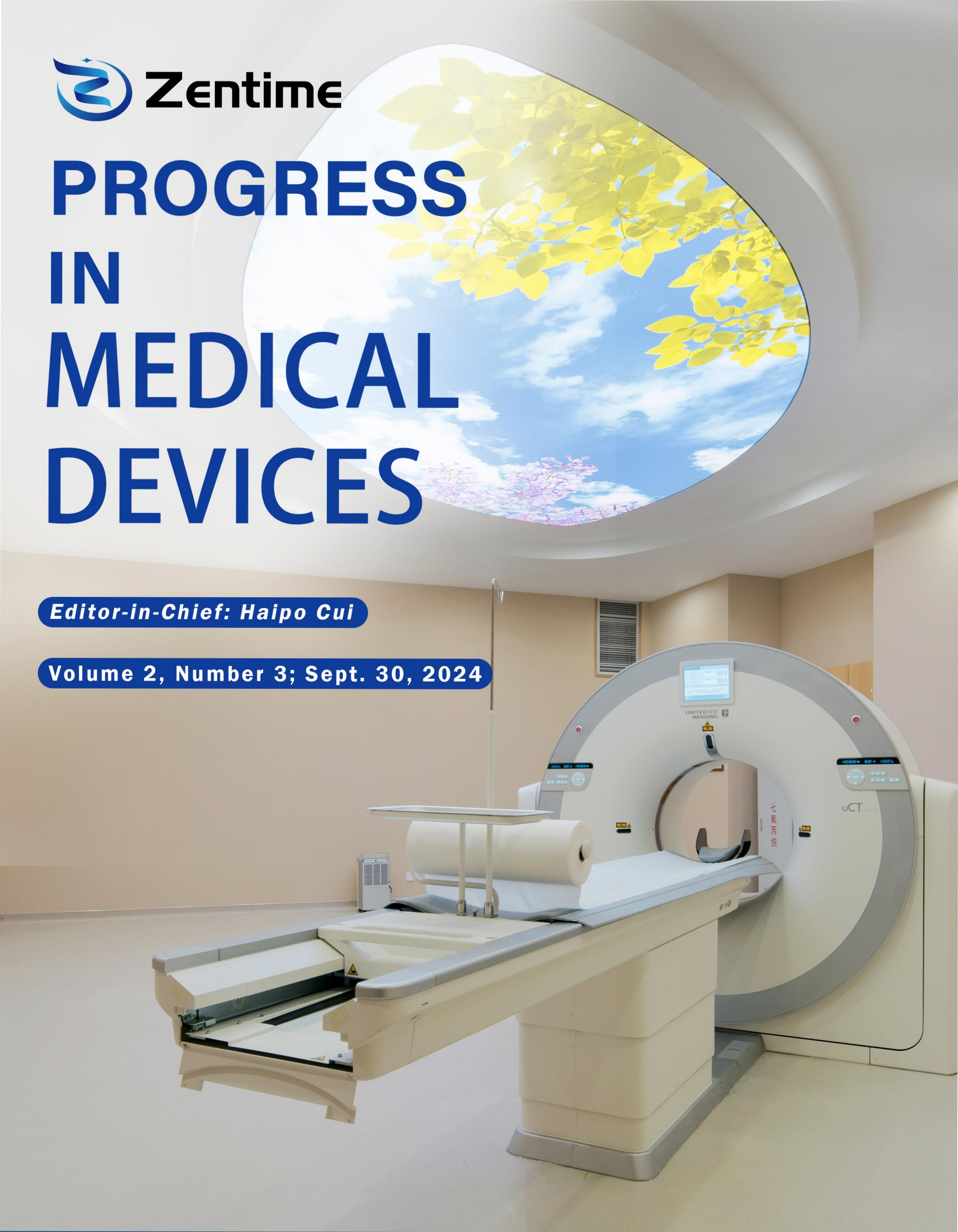
Zentime

PROGRESS IN MEDICAL DEVICES



Editor-in-Chief: Haipo Cui

Volume 2, Number 3; Sept. 30, 2024

ISSN: 2957-5478

Published by ZENTIME PUBLISHING CORPORATION

Progress in Medical Devices
ISSN: 2957-5478

Review Article https://doi.org/10.61189/650204jodubt

Review of methods for detecting electrode-tissue contact status during atrial fibrillation ablation

Mengying Zhan, Jiahao Zhang, Yuqiu Zhou, Qijun Xie, Fangfang Luo, 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 January 29, 2024; Accepted March 25, 2024; Published September 30, 2024

Highlights

- The effect of electrode-tissue contact force on the efficacy and safety of ablation of atrial fibrillation was reviewed in detail.
- The existing contact force sensing catheters on the market are compared and introduced.
- Three impedance-related methods for assessing catheter adherence are introduced.

Abstract

Atrial fibrillation is a common cardiac arrhythmia with an annually increasing global prevalence. Ablation of atrial fibrillation is a minimally invasive procedure that treats atrial fibrillation by using a catheter to deliver radiofrequency energy to heart tissues generating abnormal electrical potentials. The success of this procedure relies significantly on the adhesion between the catheter and the heart tissue, presenting a challenge in accurately assessing the contact force (CF) during surgery. To improve the safety and success rate of surgery, researchers are committed to developing various methods to evaluate or detect catheter-tissue CF. Among these, some studies integrated optical fibers or magnetic elements into the catheter tip to create CF sensing catheters that monitor CF in real time; other studies used impedance measurement, electrical coupling index, local impedance and other methods to evaluate the CF between the catheter and the tissue by measuring changes in electrical signals. These methods have achieved certain success in clinical practice, offering new ways to improve the effectiveness and safety of cardiac radiofrequency ablation surgery.

Keywords: Catheter, atrial fibrillation, bioimpedance, contact force

Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia typically caused by irregular electrical signals generated by the two upper chambers of the heart, causing erratic atrial contractions. This abnormal atrial activity affects the pumping function of heart, and patients may experience symptoms including palpitations, shortness of breath, chest pain, fatigue, dizziness, and other symptoms. AF not only diminishes quality of life but also heightens the risk of severe health problems, including stroke, heart failure, and other arrhythmia diseases. Globally, the incidence of AF is on the rise. Data from the Framingham Heart Study indicate a three-fold increase in the number of AF patients worldwide over the past 50 years [1]. Research by Miyasaka et al. pointed out that by 2050, the number of patients with AF in the United States could increase to 15.9 million [2]; Krijthe et al. pointed out that by 2060, the number of patients with AF in European Union countries could reach 17.9 million [3]. Asia, the most densely populated region in the world, might experience an even steeper increase, potentially reaching 72 million AF cases by 2050, with around 2.9 million of these patients suffering AF-related strokes [4]. Over time, the probability of AF progressing from initial paroxysmal AF to long-term persistent or permanent AF is as high as 15%, with rates increasing to 27-36% over ten years of follow-up [5]. Numerous factors contribute to the onset of AF. Age is the

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.

most important factor affecting AF. The prevalence of AF increases sharply after the age of 65, affecting about 5% of those over 65 [6]. In addition, traditional influencing factors such as diabetes and hypertension also paly roles, with more than one-fifth of AF cases attributed to hypertension and a small proportion to diabetes [7]. Affected by modern life, some mental sub-health states can lead to illness in young people. For example, research by Rosman et al. shows that post-traumatic stress disorder is associated with an increased risk of AF in young people [8]. Some unhealthy lifestyles including alcohol abuse, obesity, smoking, and obstructive sleep apnea also increase the incidence of AF [9].

There are some commonly used methods to treat AF, including medication, cardiac RF ablation, cardiac ablation, electrical cardioversion and pacemaker. Drug therapy is one of the most common treatments for AF. Antiarrhythmic drugs are usually used to maintain a normal heart rate, while anticoagulant drugs help prevent blood clots and reduce the risk of stroke. Compared with other treatments, drug therapy is simpler and carries lower risks, but long-term use of drugs may lead to some drug resistance and side effects. Defibrillation is a medical procedure in which doctors use an electric shock to reset the heart's electrical signals, restoring a normal heart rhythm. While effective in quickly normalizing heart rate, this treatment is not a cure for AF and carries certain risks, especially for patients with other health issues. For patients requiring long-term heart rate control, a cardiac pacemaker is recommended. This small device, implanted in the chest, primarily adjusts heart rate by sending electrical signals to the heart, effectively alleviating symptoms of AF. However, it does not cure the underlying condition. Cardiac ablation is a procedure that involves destroying or removing heart tissue causing AF, aiming to restore a normal heart rhythm. This method has the potential to completely cure AF, but relatively high-risk due to the open-chest surgical approach, which also extends recovery time. RF catheter ablation, including procedures like AF ablation and catheter ablation, is a minimally invasive surgery that utilizes catheters and RF energy to destroy abnormal tissue in the heart atrium that generates irregular electrical signals. This approach is suitable for patients with inadequate responses or adverse reactions to medication. Compared to drug therapy, this treatment method offers a more enduring effect, potentially reduce dependence on medication, and may also lower the stroke risk when combined with anticoagulant therapy. Compared to open-chest surgical procedures like cardiac ablation, RF catheter ablation has a more precise treatment target and involves lower surgical risks. Studies have shown that ablation, as a first-line treatment for AF, can significantly reduce arrhythmia recurrence, improve arrhythmia-related symptoms and improve patients' quality of life, and reduce the incidence of adverse reactions [10].

The incidence of AF is on the rise worldwide, underscoring the important to improve both the effectiveness and safety of AF surgery. AF ablation, a procedure that uses radiofrequency (RF) energy delivered via a catheter to treat affected areas of the heart, relies heavily on the adhesion between the catheter and heart tissue. This article focuses on the relationship between catheter-tissue adhesion and AF ablation, the impact of catheter adhesion on AF ablation outcomes, and the current characterization of catheter-tissue contact. Research indicates that contact force (CF) is the most intuitive factor characterizing the degree of adhesion, as appropriate CF can significantly improve the effectiveness and safety of treatment. The study of catheter-tissue contact methods can be categorized into four main approaches: CF sensing catheter, electrical coupling index (ECI), electrode-tissue interface impedance (IR) and local impedance (LI).

The relationship between catheter adhesion and AF ablation

Cardiac RF ablation is a minimally invasive medical procedure performed using a thin, long, and unidirectional / bidirectional flexible catheter with one or more electrodes at its tip to record signals and deliver energy to tissue. The catheter is inserted through a small incision in the patient's groin and navigated through the femoral vein to the heart, where the tip is positioned against the target tissue. By releasing RF energy, the procedure aims to destroy the tissue responsible for generating abnormal electrical signals [11, 12]. While cardiac RF ablation offers many advantages compared with traditional thoracotomy and drug treatment, it also presents unique challenges. This remote-controlled surgical operation mode makes it impossible for clinicians to perceive the real force between the treatment instrument and the tissue [13]. This loss of tactile and kinesthetic feedback imposes certain limitations: 1. Doctors may find it challenging to accurately locate certain structures or evaluate certain characteristics of tissues due to the inability to palpate [14]; 2. It is hard to properly control of the applied force. Excessive force may increase trauma or damage to surrounding healthy tissues, while insufficient force may compromise the effectiveness of the treatment [13]. Therefore, CF is crucial during cardiac RF ablation.

To improve the safety and success rate of surgery, many scholars have analyzed and optimized cardiac RF ablation based on clinical treatment and the results of some in vitro experiments. Yokoyama et al. conducted experiments using canine thigh muscles, and confirmed that CF is the main determinant of the effect of RF treatment [15]. With a fixed RF power output, increasing the CF increases leads to significantly higher tissue temperature and more thermal damage range, as well as a higher probability of thrombus formation and steam burst. Under the condition of 40 g CF and 30 W RF output, the thermal damage depth was greater than that at 10 g CF with 50 W RF output. Reddy et al. followed 34 patients with AF for 12 months and confirmed that the CF between the catheter and tissue affected the treatment results of RF ablation: lower catheter-tissue CF can lead to a higher AF recurrence rate [16]. When the average CF during ablation was <10 g, the AF recurrence rate was 100%. This rate reduced to 47% when the average CF was between 10-20 g, and further reduced to 20% when the average CF exceeded 20 g. Shurrab et al. followed 1,428 adult patients for 10-53 weeks, comparing those who used CF sensing catheters (n=552) with those who used conventional catheters during surgery [17]. The findings indicated that patients undergoing surgery using CF sensing catheters generally have lower recurrence rates and shorter operative time. These catheters also reduce the need for excessive reliance on fluoroscopy during the procedure [17].

Moreover, the study suggested that the optimal CF for cardiac RF ablation is approximately 17±5 g, which balances the risks of recurrence and complications such as cardiac tamponade or effusion requiring intervention [17]. Nakagawa et al. proposed that during the treatment process, the RF output power can be controlled according to the CF between the catheter tip and the tissue to prevent steam bursts and impedance rises that affect the effectiveness of the treatment [18]. In their study, 18 patients with symptomatic refractory paroxysmal AF were treated with AF ablation using CF-controlled RF output power. Settings were adjusted based on CF: 35-45 W for <10 g, 25-34 W for 11-30 g, 15-24 W for 31-50 g, and 5-14 W for >50 g. A transthoracic echocardiogram was performed on the second day after surgery to rule out pericardial effusion and other complications. As a result, no complications such as pericardial effusion, cardiac tamponade, or stroke were observed, except one case developed atrial tachycardia. There were no steam bursts, rising impedance, or electrode eschar during the treatment.

In summary, CF is a crucial indicator of the degree of catheter-tissue contact during cardiac RF ablation, directly impacting the procedure's effectiveness and safety. Conditions such as steam explosion, impedance rise, and electrode eschar that affect RF energy output will gradually increase as the CF increases. Appropriate CF can reduce the recurrence rate of AF and greatly improve the results of cardiac RF ablation. Therefore, ongoing research is focused on refining methods to characterize the catheter-tissue contact to further improve treatment results.

Related studies characterizing catheter-tissue contact

CF sensing catheter

CF measurement has a positive impact on operation time and postoperative recurrence rate, which has made CF sensing irrigation catheters widely used clinically in recent years [19]. Current electrodes for AF ablation catheters include monopolar catheters (usually 3.5-4.0 mm), with or without a CF catheter; and multipolar catheters designed to encircle the pulmonary veins to deliver RF energy to targeted tissue areas [20]. Operating these catheters is highly challenging, typically requiring the operator to control the catheter end effector on the proximal handle by pushing, pulling, rotating, and bending [21]. In view of the difficulty in performing catheterization procedures, there have been many studies on integrating tactile sensors to the catheter tip. CF sensors in catheters are mainly divided into optical fiber CF sensors and magnetic CF sensors [22].

Common designs for fiber optic CF sensors in catheters include intensity-modulated sensors (sensing the CF by measuring changes in the intensity of the optical signal), phase-modulated sensors via Fabry-Perot interferometry (the CF information is inferred by measuring changes in the phase of the optical signal), and a wavelength-modulated sensor with a fiber Bragg grating (a fiber Bragg grating is a periodic structure that senses CF by modulating the wavelength of light. When the abutment force changes, the periodic structure of the grating undergoes slight deformation, resulting in a

Table 1. Comparison of five types of catheters

Catheter name	Brand	Sensor type	Catheter outer diameter	Force measuring range	Deviation	Electrode description	Additional features
Tacticath™ Quartz	Abbott	Optical fiber	8F	Unknown	<3 g	Unknown	Unknown
AcQBlate®Force	Acutus Medical	Optical fiber	8F	0-60 g	<3 g	Gold electrode	Rinse and cool
ThermoCool SmartTouch®	Johnson&Johnson	Electromagnetic	8F	0-60 g	Unknown	Six electrodes	Embedded temperature sensor
Intellanav Stablepoint™	Boston Scientific	Electromagnetic	7.5F	Unknown	<5 g	Four electrodes	Local impedance combined with CF
ThermoCool SmartTouch® SF	Johnson&Johnson	Electromagnetic	8F	0-70 g	<3 g	Six electrodes	Rinse and cool

Note: CF, contact force; SF, surround flow.

change in wavelength, thereby enabling the measurement of the abutment force). These designs enable precise sensing of both axial and lateral forces at the catheter tip [23]. Magnetic CF sensing catheters are usually equipped with magnetic materials such as magnets or magnetic alloys at the tip. These magnetic elements generate an electric field, while magnetic field sensor located at the proximal end of the catheter detects changes in the magnetic field generated by the catheter tip. When the tip contacts tissue, the magnetic field changes, allowing the system to calculate the CF between the catheter and the tissue, thereby enabling real-time monitoring of the CF between the catheter tip and the tissue.

Tacticath™ Quartz (Abbott, Abbott Park, IL, USA) and AcQBlate® Force (Biotronik, Berlin, Germany), which are frequently used clinically, are optical fiber CF sensing catheters. Tacticath™ Quartz uses the Fabry-Perot interferometry method to measure how a flexible titanium alloy structure deforms upon contact with tissue, from which both the magnitude and direction of the CF can be deduced [22, 24]. The AcQBlate® Force uses a fiber Bragg grating wavelength-modulated sensor that changes its length at the corresponding portion of the fiber when the catheter tip contacts tissue [24]. Magnetic CF sensing catheters include ThermoCool SmartTouch® (Biosense Webster, Diamond Bar, CA, USA), Intellanav Stablepoint™ (Boston Scientific, Marlborough, MA, USA), and Smarttouch® SF (Johnson&Johnson, New Brunswick, NJ, USA). The ThermoCool SmartTouch® is a saline flush catheter with a magnet signal transmitter at the distal end of the spring and three magnetic position sensors near the spring that measure the micro-deflection of the spring to calculate the force magnitude and angle on the tip electrode. The tip of Stablepoint™ is equipped with a precision-machined spring with three inductive sensors made of ferromagnetic cores and coils near the catheter tip. Contact with tissue causes the coils to move inwards, and the axial and transverse forces acting on the tip can be calculated from Hooke's law. The tip of Smarttouch® SF is equipped with a machined precision spring with three magnetic sensors distributed in relatively close parts of the catheter. The sensors detect changes in signal received by the sensor coil when the catheter tip encounters force, enabling calculation of the force and direction based on Hooke's law [24-26]. CF-sensing catheters have been used in clinical procedures, which can provide real-time feedback of the force exerted between the catheter tip and the myocardium, improving the effectiveness and safety of complex ablation procedures [26]. See **Table 1** for details.

The CF between the catheter tip and tissue is crucial to ablation surgery. In addition to CF sensing catheters, researchers are also working on other characterization methods. Some researchers have proposed that impedance changes can be monitored to estimate the physical contact between the catheter tip and the tissue, as well as the effect of RF ablation [27, 28]. Real-time changes in impedance during ablation are a direct physical consequence of actual lesion formation, and therefore, impedance serves as the only commercially available and clinically accessible indicator that directly reflects real-time biophysical status of the tissue during ablation [29].

The method of ECI

The method of assessing electrode-tissue contact by measuring the LI between the catheter tip and the tissue using a three-terminal circuit model was pioneered by Yokoyama et al., which has been tested in animal studies and demonstrates the feasibility of evaluating the

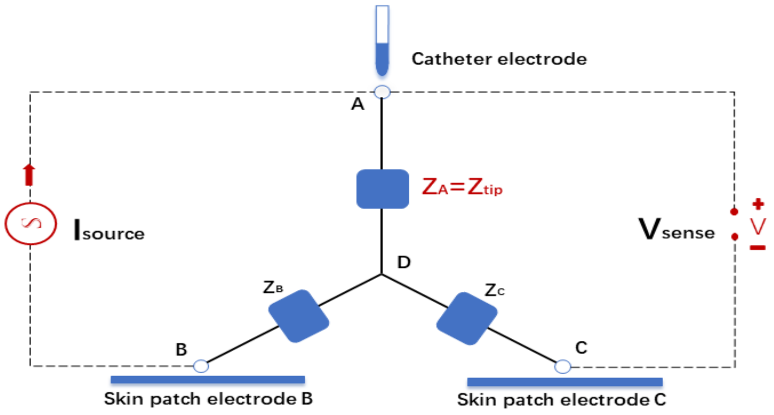


Figure 1. ECI three-terminal impedance measurement circuit model. ECI, electrical coupling index.

degree of contact using LI measurements [15]. Based on this research, Piorkowski et al. employed the EnSite Contact™ system to measure the complex impedance between the catheter tip and the tissue [30]. They mathematically combined the real and imaginary parts of the collected impedance to obtain the ECI. **Figure 1** shows the three-terminal circuit model used by ECI. In a clinical application involving 12 patients undergoing RF ablation for AF, various parameters such as electrocardiogram amplitude, pacing threshold, and the impedance at the catheter-tissue interface were concurrently measured [30]. When the catheter touched the tissue, the ECI changed significantly, from (118±15) to (145±24). The average ECI difference between contact and non-contact sites was (32.7±11.6) ECI units, and the ECI of vascular tissue was significantly higher than that of trabecular tissue and smooth myocardium. This study not only confirmed the practicality of measuring catheter LI at the catheter-tip tissue interface during AF ablation, but also explored the impact of patch location, operator, body mass index (BMI), and type of AF on the ECI calculation results [30].

An N-way analysis of variance showed that these factors had no significant impact on ECI. The research by Deno et al. indicated that the ECI could convey information about the effective delivery of RF ablation energy to the tissue, serving as a useful indicator for monitoring the RF catheter ablation process for AF [31]. In their study, the researchers derived a fitting formula for ECI based on pig experiments. They compared ECI with clinical judgments, pacing thresholds, electrogram amplitudes, ablation lesion depth, and transmural effects, affirming the specificity of catheter-tissue impedance and ECI for assessing tip electrode-tissue contact or coupling. Three-terminal impedance and ECI measurements were taken at non-contact locations during the experiment. The average impedance change was less than 10 Ω over an interval of (6.3±2.2) hours, and the average ECI change was only (5.8±5.6)%. This suggests that ECI measurements are stable over long-term use. Linear regression analysis of the real and imaginary parts of impedance in the experimental data compared to pacing and electrophysiology system yielded an impedance weight of 5.1 times resistance. The relationship between ECI, force, and catheter position (distance and angle) was validated in isolated bovine myocardium. Impedance and ECI were found to be insensitive to the surface and above distances until the electrode tip was approximately 2 mm from the impedance. In the range of 0-20 g, ECI changes were nearly proportional to the applied force and depth (an increase of 1.5 ECI units per gram of force). The results indicate that the catheter-tissue contact angle has no significant impact on ECI outcomes.

The method of IR

ECI is determined by applying a small high-frequency current between the catheter electrode and the first skin patch, while simultaneously measuring the voltage difference between this electrode and another skin patch. This process calculates the impedance, both resistance and reactance, of the electrode-tissue contact, ultimately determining the ECI value. While ECI provides reliable measurements within the left atrium, it can erroneously indicate good contact even when the electrode is floating within the pulmonary veins, likely influenced by the nearby high-impedance lung tissue. Therefore, Van et al. proposed an improved method for measuring the IR, as depicted in **Figure 2** [32]. Unlike ECI, the IR method is suitable for multi-electrode catheters. It involves applying current between the target electrode and adjacent electrodes, simultaneously measuring the voltage between the target electrode and the skin patch to calculate impedance. This method concentrates the driving current in a small area around the target electrode, effectively avoiding the influence of distant tissues on ECI measurements. Researchers compared the IR method with ECI alongside CF sensing catheters in both in vivo and in vitro pig experiments. In vitro experiments used PVC pipes

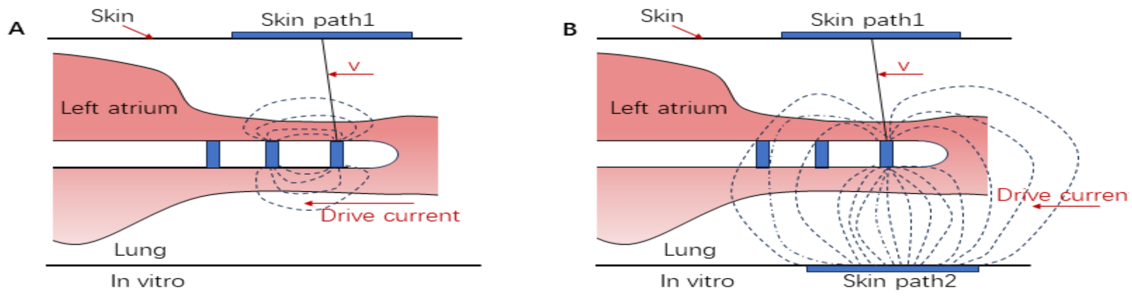


Figure 2. Comparison diagram of IR and ECI. (A) A schematic diagram of the principle of the IR method. It is necessary to apply a current between the target electrode and the adjacent electrode, to measure the voltage between the target electrode and the skin patch electrode; (B) A schematic diagram of the principle of the ECI method. It is necessary to apply a current between the target electrode and the skin patch electrode, to measure the voltage between the target electrode and another skin patch electrode. IR, electrode-tissue interface impedance; ECI, electrical coupling index.

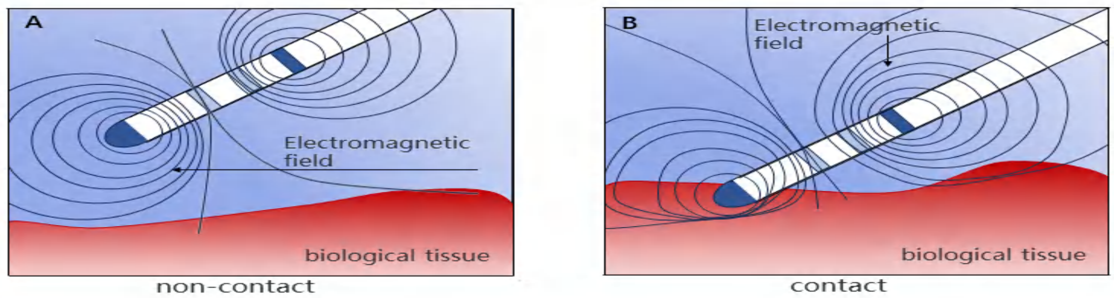


Figure 3. Schematic diagram of local electric field under the action of floating and contact with LI method. (A) A schematic diagram of the electromagnetic field when the electrode is suspended; (B) A schematic diagram of the electromagnetic field when the electrode is in contact with tissue. LI, local impedance.

to simulate high impedance at the distal end. When the catheter was placed inside the pipe, ECI and IR measurements increased by 32.2% and 3.2%, respectively. IR demonstrated a significantly lower sensitivity to distal impedance compared to ECI. In the in vitro experiments, ECI and IR were measured 237 and 288 times, respectively, along with synchronous impedance readings and CF sensor readings. The correlation coefficient (R2) for the fitted power was 0.84, indicating a clear relationship between impedance and force for both measurement methods. The results from in vitro experiments were further validated in in vivo experiments. A total of 373 CF and IR measurements were taken during in vivo experiments, showing a significant correlation (P=0.64) between CF and impedance increases. When the electrode moved in vivo from the inferior vena cava to the right atrium, the ECI value increased by 20%. However, the IR value remained stable when the catheter shifted between the right atrium, inferior vena cava, or superior vena cava. These data suggest that the IR method is convenient for assessing the contact level between circular multi-pole ablation catheters and tissue interfaces, and it is less susceptible to the influence of high-impedance tissues at a distance.

The method of LI

To address the susceptibility of impedance measurements to distant structures such as muscles, lungs, and bones, Sulkin et al. proposed the LI method, which eliminates the need for skin surface patch electrodes [32-34]. This system is applicable to multi-electrode catheters, where the catheter tip is equipped with at least three microelectrodes. The LI method employs a four-electrode impedance measurement technique. A small current (5.0 μA, 14.5 KHz) is applied between the tip electrode and the proximal electrode to generate a local electric field. The microelectrodes and the distal electrode can measure the potential difference generated by nearby heart tissue or high-resistance myocardial tissue contact. The impedance is then calculated by dividing the voltage by the applied current. The schematic diagram in **Figure 3** illustrates the local electric field formed by the small current in the LI method during both floating and contact states. This system rigorously tested through in vivo and in vitro experiments. The results demonstrated that the LI method can effectively distinguish healthy myocardium from blood. Importantly, the measurements provided by LI method are independent of catheter orientation and tissue thickness. The LI method showed high sensitivity to tissues with different resistivities, allowing for a clear differentiation between contact and non-contact states between the catheter and tissue.

Conclusion

Catheter ablation presents significant advantages in treating AF; however, challenges persist during the surgical process, particularly in accurately assessing the degree of contact between the catheter tip and the tissue. This paper views various assessment of catheter tip-tissue contact in AF ablation and introduces existing methods for evaluating contact.

The assessment of contact between the catheter and tissue is crucial in ablation surgery. Appropriate CF can significantly enhance the effectiveness of treatment and reduce recurrence rates. Conversely, excessive CF may increase the risk of thrombus formation and steam pops, while insufficient CF can lead to a higher recurrence rate of AF. This paper primarily delves into the in-depth study of evaluation methods, including CF-sensing catheters and LI, used in current research.

In clinical AF ablation surgeries, operators manipulate the catheter using techniques such as pushing, pulling, rotating, and bending. To improve surgical precision and ease the procedure for physicians, CF sensors are placed at the catheter tip to monitor real-time CF between the catheter and tissue. The sensors employed in CF-sensing catheters, including fiber optic sensors and magnetic CF sensors, such as Tacticath™ Quartz, AcQBlate® Force, and ThermoCool SmartTouch®, are widely used in clinical treatments.

Additionally, three-terminal circuits were utilized to collect LI between the catheter tip and tissue for evaluating CF. The ECI method mentioned in this paper uses a configuration with one treatment electrode and two skin patch electrodes to calculate impedance. The IR method, developed to address the impact of high impedance in pulmonary vein tissue on ECI measurements, uses a configuration with two treatment electrodes and one skin patch electrode to calculate impedance. Some researchers found that, apart from pulmonary vein tissue, other high-impedance tissues in the body, such as bones and lungs, could potentially affect contact assessment. To address this issue, researchers proposed using a configuration with three intracardiac treatment electrodes to calculate impedance in the LI method.

In summary, various methods for assessing catheter-tissue contact can assist catheter operators in performing AF ablation procedures more easily and accurately. Future research and technological developments will further improve the accuracy of evaluations, allowing these methods to gain broader applications in clinical practice.

Author contributions: Mengying Zhan: Collect and summarize materials, write and revise the paper, and draw graphs and tables; Jiahao Zhang: Participate in paper writing, and revise the content of the paper; Yuqiu Zhou: Consult literature, collect data, and draw graphs and tables; Qijun Xie: Proofread the paper; Fangfang Luo: Proofread the paper; Yu Zhou: Provide writing ideas and guide paper writing.

References

[1] Schnabel RB, Yin X, Gona P, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. Lancet 2015;386(9989):154-162.
[2] Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006;114(2):119-125.
[3] Krijthe BP, Kunst A, Benjamin EJ, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J 2013;34(35):2746-2751.
[4] Chiang CE, Wang KL, Lip GY. Stroke prevention in atrial fibrillation: an Asian perspective. Thromb Haemost 2014;112(05):789-797.
[5] Gunawardene MA, Willems S. Atrial fibrillation progression and the importance of early treatment for improving clinical outcomes. Europace 2022;24(Suppl2):ii22-ii28.
[6] Tan AY, Zimetbaum P. Atrial fibrillation and atrial fibrosis. J Cardiovasc Pharmacol 2011;57(6):625-629.
[7] Huxley RR, Lopez FL, Folsom AR, et al. Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011;123(14):1501-1508.
[8] Rosman L, Lampert R, Ramsey CM, et al. Posttraumatic stress disorder and risk for early incident atrial fibrillation: a prospective cohort study of 1.1 million young adults. J Am

Heart Assoc 2019;8(19):e013741.
[9] Kornej J, Börschel CS, Benjamin EJ, et al. Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights. Circ Res 2020;127(1):4-20.
[10] Andrade JG. Ablation as first-line therapy for atrial fibrillation. Eur Cardiol 2023;18:e46.
[11] Padmanabhan D, Rao PSM, Pandya HJ. Force sensing technologies for catheter ablation procedures. Mechatronics 2019;64:102295.
[12] Polygerinos P, Ataollahi A, Schaeffter T, et al. MRI-compatible intensity-modulated force sensor for cardiac catheterization procedures. IEEE Trans Biomed Eng 2010;58(3):721-726.
[13] Trejos AL, Patel RV, Naish MD. Force sensing and its application in minimally invasive surgery and therapy: a survey. Proc Inst Mech Eng, Part C 2010;224(7):1435-1454.
[14] Puangmali P, Althoefer K, Seneviratne LD, et al. State-of-the-art in force and tactile sensing for minimally invasive surgery. IEEE Sens J 2008;8(4):371-381.
[15] Yokoyama K, Nakagawa H, Shah DC, et al. Novel contact force sensor incorporated in irrigated radiofrequency ablation catheter predicts lesion size and incidence of steam pop and thrombus. Circ Arrhythm Electrophysiol 2008;1(5):354-362.
[16] Reddy VY, Shah D, Kautzner J, et al. The relationship between contact force and clinical outcome during radiofrequency catheter ablation of atrial fibrillation in the TOCCATA study. Heart rhythm 2012;9(11):1789-1795.
[17] Shurrab M, Di Biase L, Briceno DF, et al. Impact of contact force technology on atrial fibrillation ablation: a meta-analysis. J Am Heart Assoc 2015;4(9):e002476.
[18] Nakagawa H, Kautzner J, Natale A, et al. Locations of high contact force during left atrial mapping in atrial fibrillation patients: electrogram amplitude and impedance are poor predictors of electrode-tissue contact force for ablation of atrial fibrillation. Circ Arrhythm Electrophysiol 2013;6(4):746-753.
[19] Boersma L, Andrade JG, Betts T, et al. Progress in atrial fibrillation ablation during 25 years of Europace journal. Europace 2023;25(9):euad244.
[20] Hong KL, Borges J, Glover B. Catheter ablation for the management of atrial fibrillation: current technical perspectives. Open Heart 2020;7(1):e001207.
[21] Cheng W, Law PK. Conceptual design and procedure for an autonomous intramyocardial injection catheter. Cell Transplant 2017;26(5):735-751.
[22] Cheng W, Yao M, Zhai B, et al. Contact force sensors in minimally invasive catheters: current and future applications. Expert Rev Med Devices 2021;18(5):445-455.
[23] Xu L, Miller MI, Ge J, et al. Temperature-insensitive fiber-optic contact force sensor for steerable catheters. IEEE Sens J 2016;16(12):4771-4775.
[24] Küffer T, Haeberlin A, Knecht S, et al. Validation of the accuracy of contact force measurement by contemporary force-sensing ablation catheters. J Cardiovasc Electrophysiol 2023;34(2):292-299.
[25] Liang JJ, Santangeli P. Contact force sensing during atrial fibrillation ablation: Clinical experience and effects on outcomes. Expert Rev Cardiovasc Ther 2016;14(6):749-759.
[26] Hoffmayer KS, Gerstenfeld EP. Contact force-sensing catheters. Curr Opin Cardiol 2015;30(1):74-80.
[27] Cao H, Tungjitkusolmun S, Choy YB, et al. Using electrical impedance to predict catheter-endocardial contact during RF cardiac ablation. IEEE Trans Biomed Eng 2002;49(3):247-253.
[28] He DS, Bosnos M, Mays MZ, et al. Assessment of myocardial lesion size during in vitro radio frequency catheter ablation. IEEE Trans Biomed Eng 2003;50(6):768-776.
[29] Chu GS, Calvert P, Futyma P, et al. Local impedance for the optimization of radiofrequency lesion delivery: a review of bench and clinical data. J Cardiovasc Electrophysiol 2022;33(3):389-400.
[30] Piorkowski C, Sih H, Sommer P, et al. First in human validation of impedance-based catheter tip-to-tissue contact assessment in the left atrium. J Cardiovasc Electrophysiol 2009;20(12):1366-1373.
[31] Deno DC, Sih HJ, Miller SP, et al. Measurement of electrical coupling between cardiac ablation catheters and tissue. IEEE Trans Biomed Eng 2013;61(3):765-774.
[32] Van Es R, Hauck J, Van Driel VJ, et al. Novel method for electrode-tissue contact measurement with multi-electrode catheters. Europace 2018;20(1):149-156.
[33] Gaspar T, Sih H, Hindricks G, et al. Use of electrical coupling information in AF catheter ablation: a prospective randomized pilot study. Heart Rhythm 2013;10(2):176-181.
[34] Sulkin MS, Laughner JI, Hilbert S, et al. Novel measure of local impedance predicts catheter–tissue contact and lesion formation. Circ Arrhythm Electrophysiol 2018;11(4):e005831.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Review Article https://doi.org/10.61189/721472czacxf

Integrating Traditional Chinese Medicine massage therapy with machine learning: A new trend in future healthcare

Yichun Shen[1], Shuyi Wang[1], Yuhan Shen[1], Hua Xing[2]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China. [2]Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai 200080, China.

Corresponding author: Shuyi Wang.

Acknowledgement: This work was supported by Shanghai "Science and Technology Innovation Action Plan" Medical Innovation Research Special Program (23Y11921700) and Shanghai Municipal Health Commission Health Industry Clinical Research Special Program (20234Y0077).

Declaration of conflict of interest: None.

Received April 12, 2024; Accepted July 11, 2024; Published September 30, 2024

Highlights

- Machine learning can enhance the individualization of treatment in Chinese massage.
- An intelligent system improves the efficiency of Traditional Chinese Medicine massage therapy.
- The integration of Traditional Chinese Medicine Massage Therapy with machine learning represents a new trend in future healthcare.

Abstract

The growing demand for healthcare has brought Traditional Chinese Medicine (TCM) massage therapy into the spotlight in academic circles. Numerous studies have underscored the effectiveness of TCM massage in health promotion, disease amelioration, and quality of life enhancement. However, the field faces challenges such as inconsistent training and inadequate transfer of experiential knowledge. Recently, machine learning has shown potential in the medical field and its application in TCM massage therapy offers new developmental opportunities. This paper reviews key research areas exploring the synergy between machine learning and Chinese massage therapy, including acupoint localization and identification, massage practice, and personalized treatment plans. It summarizes progress and identifies the challenges in integrating these technologies. Despite potential risks, merging these technologies is poised to be a trend in future healthcare, driven by advances in computer technology and the needs of TCM practitioners.

Keywords: Traditional Chinese medicine, machine learning, massage therapy

Introduction

With the increasing demand for healthcare services, Traditional Chinese Medicine (TCM) massage therapy is receiving heightened attention and recognition within academia. Research and clinical practices have confirmed the significant benefits of TCM massage in health improvement, illness treatment, and life quality enhancement [1-3]. These endeavors have established a robust scientific foundation and provided insights into the therapeutic mechanisms of TCM massage therapy. Despite these advances, the field still grapples with issues such as inadequate knowledge transfer, inconsistent training standards, and unclear criteria for therapeutic outcomes, which impede further development and adoption [4-6]. Additionally, the high doctor-patient ratio increases the workload of practitioners. Addressing these challenges through modern engineering technology could revolutionize TCM massage techniques, making them more scientific and standardized. This approach would enhance the effectiveness of

Address correspondence to: Shuyi Wang, School of Health Science and Engineering, University of Shanghai for Science and Technology, NO.516, Jungong Road, Shanghai 200093, China. E-mail: wangshuyi@usst.edu.cn.

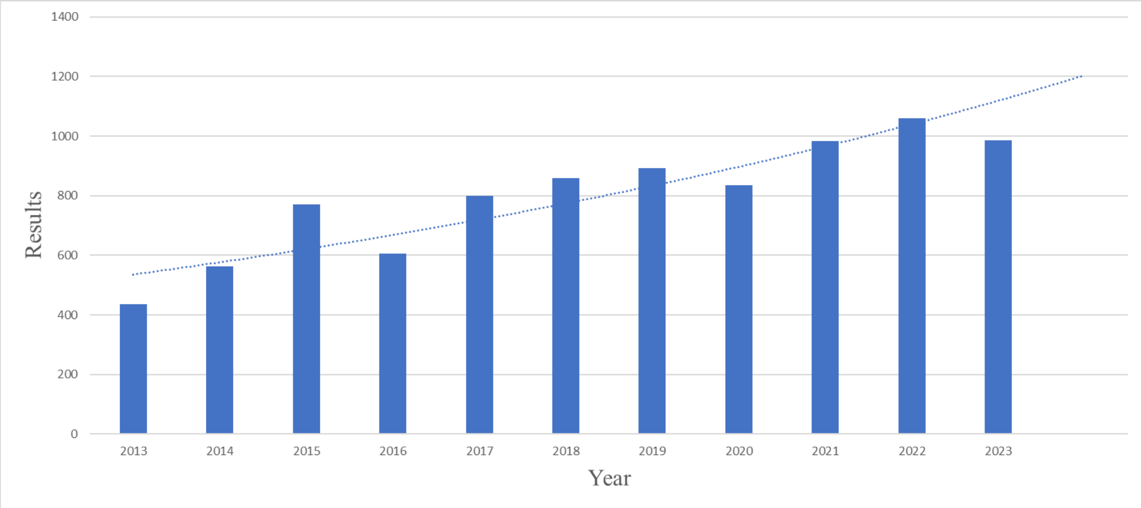


Figure 1. Number of published paper entries in the last ten years.

massage therapy and facilitate the delivery of superior healthcare services.
In recent years, the evolution of computer technology, particularly in artificial intelligence with a focus on machine learning, has become a prominent subject within engineering and technology disciplines [7]. Machine learning involves computer systems that learn by analyzing and recognizing patterns in datasets, enabling predictions, decisions, or task executions based on this data. Characterized by automation, generalization, iterative optimization, data-centricity, and adaptability, machine learning is rapidly expanding across various sectors [1, 2, 8, 9].

In healthcare, machine learning is increasingly used for image recognition, recommendation systems, and medical diagnostics, enhancing the development of intelligent technologies. Notably, the application of machine learning in TCM has gained momentum in academia, with significant publications emerging. For example, Xu et al. developed a dynamic treatment strategy for rectal cancer using a survival cost-sensitive classification learning algorithm, which demonstrated clinical effectiveness [10]. Chen et al. utilized a Convolutional Neural Network (CNN) for symptomatic text classification in TCM, achieving accurate symptom categorization [11]. Moreover, Yang et al. introduced the “Zhongjing” large language model tailored for Chinese medicine, refined through expert feedback, marking a significant advancement in language model applications within the field [12].

The potential for integrating machine learning with Chinese massage therapy (Tui-na) is further highlighted by the increasing number of academic publications over the past decade, as illustrated in **Figure 1**, derived from Google Scholar searches using keywords “Chinese massage therapy”, “machine learning”, and “Chinese Tui-na”.

Despite the growing evidence supporting the efficacy of machine learning in TCM massage therapy, a comprehensive review of this integration is still lacking. This study aims to critically evaluate the risks and benefits by analyzing existing literature, which will not only inform clinical decisions but also inspire further research.

Application of machine learning in Chinese massage therapy practice

TCM massage, a cornerstone of Chinese medicine, plays an essential role in regulating meridians, qi, and blood circulation, aiming to prevent and treat diseases, promote health, and restore bodily balance [13]. It employs a theoretical framework that includes concepts of qi, blood, meridian systems, and meridian points. Following evidence-based TCM principles, massage techniques are tailored to the individual symptoms and physical characteristics of patients. TCM massage is proven effective in clinical settings for treating a range of conditions such as musculoskeletal pain, neurological disorders, respiratory issues, and circulatory disorders [14-16].

The advent of machine learning has marked significant advancements in traditional Chinese massage therapy, enhancing the integration of engineering technology with alternative massage systems. Currently, the application of machine learning for quantitative and objective assessments of massage therapy outcomes is expanding. Scholars globally focus on three primary areas where machine learning intersects with Chinese massage therapy:

• Automated acupoint localization and identification: Machine learning is applied to automate the identification and localization of human acupoints through image processing and pattern recognition. Advances in deep learning algorithms, utilizing extensive acupoint datasets for model training, have significantly improved the precision and efficiency of acupoint positioning, enhancing the practice of TCM massage.

• Intelligent massage assistance system: Integrating machine learning with intelligent sensing technologies, this advanced system provides real-time monitoring and guidance for TCM massage. Analyzing the effects of massage therapy and patient feedback, the system supports data-driven decisions, aiding practitioners in their therapeutic approaches.

• Personalized massage treatment plan: Machine learning facilitates the creation of customized treatment plans by integrating individual health data and physiological parameters. This system adapts massage techniques and procedures to the patient's specific conditions and feedback, enabling dynamic adjustments for more accurate and effective treatments.

Application of machine learning to automated acupoint localization and identification

Acupoints are integral to Chinese medicine, facilitating the flow of qi and blood, alleviating pain, and regulating internal organs, thus playing a vital role in maintaining physical health [17, 18]. Accurate identification and application of acupoints are critical in massage therapy. Traditionally, locating acupoints has relied on the experience and expertise of practitioners, a process that can be subjective and time-consuming [19]. In contrast, automated acupoint localization and recognition using machine learning presents an effective solution to overcome these limitations.

A common strategy involves the use of vision devices for image recognition, where machine learning techniques, particularly deep learning algorithms like CNN, are applied to automatically recognize acupoint images. This automation enhances the accuracy and efficiency of acupoint localization. Research has shown that leveraging bony joint feature points significantly improves acupoint localization [20, 21]. Some studies use neural networks to capture human joint data and apply proportional geometric measurements to understand acupoint geometric relationships, combining this with three-dimensional information for precise acupoint positioning. These methods have achieved an average localization accuracy of 2.36 cm and real-time performance of up to 14 frames per second, demonstrating adaptability across various body types, environments, and postures [22].

Further increasing accuracy, Wang et al. employed an Artificial Neural Network for more refined acupoint identification. Their approach includes image preprocessing to segment the region of interest on the sole, mapping the Foot Acupoint Profile to the foot image to accurately locate reference massage points [23]. Additionally, Zhang et al. developed an integrated deep confidence network that merges manually labeled and rule-matched data, reducing manual effort in acupoint localization and enhancing model accuracy [24].

There are also advancements in acupoint visualization techniques. Zheng et al. explored an Augmented Reality overlay method for facial virtual acupoints, where real-time computed facial acupoint coordinates are superimposed onto the actual scene, providing physicians with an intuitive means of locating acupoints [25].

Ongoing research continues to explore new methods utilizing machine learning in acupoint localization. One study applied temperature specificity detection and edge detection algorithms to automatically pinpoint major chest organs and localize acupoints based on the physiological structures of the chest and upper limbs, achieving a localization accuracy over 90.12% [26]. Another innovative approach involves the fusion of RGB and depth images in a deep CNN to create a novel 3D acupoint localization method, as outlined in **Table 1** [27].

Application of machine learning to intelligent massage assistance systems

Intelligent massage assistance systems combine artificial intelligence technologies with traditional massage expertise to analyze data, recognize and classify massage techniques, conduct personalized health data analyses, and provide real-time monitoring and feedback. These systems employ machine learning, deep learning, and sensor technology to enhance the accuracy, personalization, and efficiency of traditional massage practices [28, 29]. They assist practitioners in selecting appropriate techniques, adjusting treatment plans, and delivering more scientifically informed and efficient therapy, thus contributing to the preser-

Table 1. Comparative analysis of techniques related to acupoint localization in traditional Chinese medicine massage

Localization approach	Machine learning models	Advantages	Weaknesses
Traditional machine learning [26, 27]	BP neural networks	Robustness	Gradient vanishing problem, Overfitting problem
Deep learning [20]	Convolutional Neural Network	Strong generalization ability	Data quality limitations
Machine learning based on temperature specificity [24]	Support Vector Machine	Leverage individual physiological features	Data acquisition complexity
Deep learning based on RGB recognition [25]	Convolutional Neural Network	3D localization	Data quality limitations

Note: BP, Backpropagation; RGB, red-green-blue.

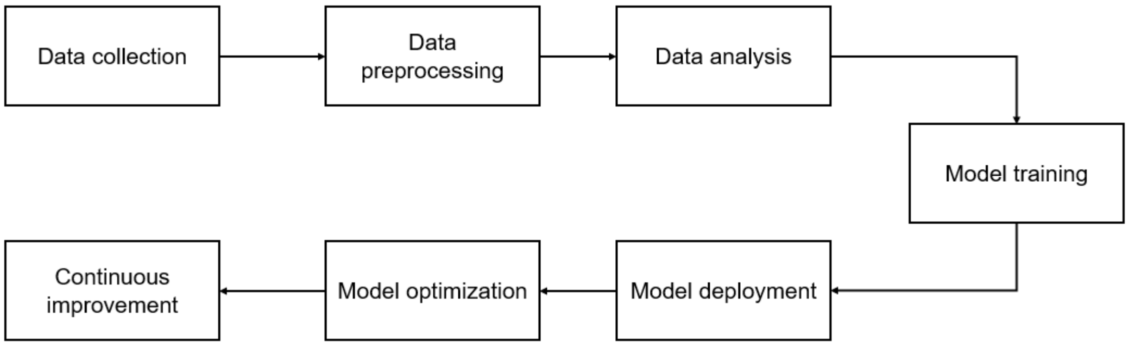


Figure 2. General framework for creating an intelligent massage assistance system.

vation and advancement of traditional Chinese massage technology [30].

A crucial component in developing these systems is the precise acquisition of massage techniques. Zhu et al. developed a manipulation acquisition system using tactile sensors, employing CNNs, long short-term memory networks, and attention mechanisms to identify massage techniques with a 100% classification accuracy [31]. This system deviates from traditional methods that focus only on force distribution across different hand regions of a practitioner. Instead, it utilizes a triaxial force sensor to detect the magnitude and direction of spatial forces, crucial for quantifying forces in TCM physiotherapy practices like massage. Support Vector Regression is then applied to translate strain signals into force components, allowing for the continuous recognition of varied force magnitudes and directions [32].

Beyond mechanical data analysis, Jia et al. explored capturing real-time image feature data of patients' joints and gestures during autonomous massage sessions [33]. They used a CNN to calculate human postures, providing real-time cues for autonomous massage and improving the accuracy of massage procedures.

The overarching goal in developing intelligent massage assistance systems is to quantify and evaluate TCM massage techniques. Li et al. created a learning framework that combines deep learning with variable impedance control technology [34]. This framework determines the necessary impedance gain and angle for each joint of a massage robot, enabling autonomous decision-making during massage operations [32]. Additionally, these systems often utilize various physiological signals. For example, Long et al. developed muscle fatigue state classifiers using surface Electromyography signals to assess massage effects, and Zheng et al. used the circular Hough transform for iris segmentation from periocular images, evaluating the effectiveness of periocular massage [35, 36].

These advancements illustrate a shared process in developing intelligent massage assistance systems that leverage machine learning technologies, as depicted in **Figure 2**.

Application of machine learning to personalized massage treatment plans

TCM Personalized Massage Therapy tailors treatments based on TCM theories and individual requirements, emphasizing the creation of customized regimens. These plans consider the patient's constitution, condition, symptoms, and specific needs, incorporating TCM meridian

Table 2. The currently accessible LLMs for TCM

Name	Release Time	Service Area	Supported Languages	Data Entry
MedChatZH [42]	2024	Medical guidance	Chinese, English	764,000
HuaTuo [43]	2023	Comprehensive Diagnosis	English	
Qibo [44]	2024	Corpus	English	
TCM-GPT [45]	2023	Comprehensive Diagnosis	English	7 billion
Zhongjing [12]	2024	Medical guidance	English	70,000

Note: LLMs, large language models; TCM, traditional Chinese medicine; GPT, generative pre-trained transformer.

theory and diagnostic techniques such as observation, olfaction, inquiry, and pulse diagnosis [37].

Although research in this area is still emerging, several scholars have applied machine learning algorithms to personalize recommendations for comprehensive treatment plans that include massage and acupuncture. For instance, Sun et al. utilized Support Vector Machine algorithms to select clinical acupoints for treating functional constipation, with comparative experiments showing a total effective rate of 96.56% in the experimental group, significantly higher than the control group's 75.02% [38]. Chae et al. explored artificial intelligence methods to quantify the target disease for each acupoint, aiming to improve the specificity of acupoint indications [39].

Additionally, Large Language Models (LLMs) are increasingly recognized as a valuable tool in this domain. LLMs, which use deep learning to generate coherent natural language texts, are particularly useful for providing insights and recommendations on TCM diagnosis, treatment strategies, medication, and disease prevention [40]. However, there are currently no LLMs specifically designed for TCM massage therapy, indicating a gap in the application of this technology. This paper highlights the need for such models and provides an overview of existing LLMs in TCM, as summarized in **Table 2** [41].

Key challenges in integrating machine learning with TCM massage and innovative solutions

Key challenges

Data access and labeling issues

TCM massage encompasses a wealth of clinical experience and theories, representing a vast pool of unstructured knowledge passed down through teaching and practice. This knowledge, which includes a diverse range of manual techniques, diagnostic methods, and treatment concepts, often lacks the structured data format crucial for analysis by machine learning algorithms. The primary challenge lies in effectively using existing data to refine labels for machine learning applications. Labeling in machine learning refers to the assignment of correct labels or categories to data for training and testing algorithmic models, a critical process in supervised learning where models rely on existing labels to make predictions and classifications. The task of efficiently gathering, organizing, and annotating massage movement data to create suitable datasets for machine learning applications requires the integration of expert knowledge, modern technology, and advanced data processing methods. Addressing this challenge is essential for integrating TCM massage principles with machine learning, propelling the scientific and modern evolution of TCM massage practices.

Model training and optimization issues

The diversity and variability inherent in TCM massage contribute to differences in techniques among patients, conditions, and practitioners. Constructing a model capable of accommodating such diversity is a significant challenge. The model must account for individual variances and unique circumstances to ensure the efficacy and safety of the therapy. Numerous studies have focused on developing a versatile and adaptable massage model through continuous exploration and optimization efforts.

Clinical validation and credibility issues

Implementing machine learning techniques in TCM massage necessitates clinical validation to verify the therapeutic efficacy and safety of the interventions. Clinical validation assesses the model's performance in real-world settings, testing its adaptability and effectiveness across different scenarios. Additionally, verifying the model's credibility is crucial to ensure the accuracy and reliability of the results produced by the machine learning algorithm. Through rigorous clinical validation and credibility verification processes, the effectiveness and safety of massage therapy can be guaranteed, advancing the sustainable development and application of machine learning technology within the realm of TCM massage.

Innovative solutions

Standardized TCM massage techniques

To create a standardized dataset for TCM massage, expert knowledge and practitioner experience have been extensively utilized. Numerous studies aim to catalog massage actions, detailing techniques, intensity, and targeted areas [46, 47]. Advanced technologies such as deep learning and computer vision are employed to process images and videos of massage movements, extracting key features and annotating the data accordingly. Additionally, natural language processing is applied to analyze TCM massage-related literature semantically, further refining and enhancing data annotation [31, 33]. Integrating these technologies facilitates the systematization and standardization of TCM massage knowledge, laying a solid foundation for incorporating machine learning within the field.

Selection of highly adaptive models

The integration of deep learning has emerged as a prominent academic solution, enabling the design of complex neural network structures and extensive data training to meet diverse massage needs. Transfer learning techniques are also employed, allowing knowledge transfer from one domain to another to enhance model generalization and adaptability. Furthermore, reinforcement learning is used to allow the model to continuously optimize its strategies through interaction with the environment, thus adapting to varying massage requirements [24]. By leveraging these advanced methodologies, a more intelligent and adaptive massage model can be developed, enhancing the potential for advancement and application in TCM massage practices.

Building cooperation mechanisms

A collaborative mechanism is proposed to integrate machine learning technology with clinical practice, aligning with the health promotion principles of the Healthy China Policy [48]. This involves creating evidence-based experimental protocols and evaluation criteria to validate the effectiveness and safety of combining TCM massage with machine learning models [49]. This collaborative approach aims to modernize TCM massage, elevate therapy standards, and support the overarching goal of promoting health and well-being.

Conclusion and outlook

Reflecting on the research findings and current literature, integrating TCM massage with machine learning technology holds significant potential and promises notable advancements. Machine learning introduces novel perspectives and methods to TCM massage therapy, enabling personalized treatment, efficient diagnosis, and optimization of treatment strategies. This technology facilitates the analysis and collection of clinical data for practitioners, identifies individual treatment factors, and automates the generation of treatment plans. Personalized treatment, a highlight of machine learning, involves constructing models that predict patient conditions and recommend tailored treatment plans, thereby enhancing the relevance and effectiveness of therapies. Additionally, the digitization of traditional medicine represents a significant shift for TCM, where databases created for machine learning become valuable training and benchmark data, supporting the inheritance and innovation of TCM Tuina techniques.

However, the integration of machine learning with TCM massage is not without risks. One major concern is the potential for model misdiagnosis. Machine learning models, which make predictions based on training data, may not be perfectly adapted to every individual, leading to incorrect diagnostic and treatment decisions that could impact patient health. Additionally, TCM practitioners require further training and technical support to collaborate effectively with engineers in dataset construction. Social acceptance is another crucial factor, especially for new health-related technologies, which often face public skepticism and distrust initially; thus, increasing public acceptance of machine learning technologies is essential.

In conclusion, integrating TCM massage with machine learning involves various risks and challenges that necessitate a comprehensive consideration of technology, law, and ethics. It requires gradual development and careful application to achieve optimal clinical outcomes while ensuring patient safety and therapeutic efficacy. Enhanced collaboration between academic institutions and healthcare facilities is vital to advance the application of machine learning in TCM massage, continuously improving the quality and efficacy of therapy.

Looking ahead, as machine learning technology further advances in the medical field, the integration of TCM massage and machine learning is expected to deepen and become more refined. Through collaborative efforts, more comprehensive datasets and models can be developed to better support and guide TCM massage therapy. This collaborative approach will also promote in-depth research and preservation of TCM massage techniques, cultivate a new generation of professionals skilled in both medical practices and machine learning, and inject new vitality into the evolution of massage therapy. Ultimately, this integration aims to enhance the widespread adoption of TCM massage in healthcare, offering safer and more effective treatment options to a broader patient base and benefiting the community as a whole.

Author Contributions: Yichun Shen and Yuhan Shen performed the data analyses and wrote the manuscript; Shuyi Wang and Hua Xing helped perform the analysis with constructive discussions.

References

[1] Lawler SP, Cameron LD. A randomized, controlled trial of massage therapy as a treatment for migraine. Ann Behav Med 2006;32(1):50-59.

[2] Sagar S, Dryden T, Wong RK. Massage therapy for cancer patients: a reciprocal relationship between body and mind. Curr Oncol 2007;14(2):45-56.

[3] Walach H, Güthlin C, König M. Efficacy of massage therapy in chronic pain: a pragmatic randomized trial. J Altern Complement Med 2003;9(6):837-846.

[4] Wang GC. Reflections on the standardized development of traditional Chinese medicine massage. J Tradit Chin Med Manage 2017;25(15):16-17.

[5] Wang ZG, Fan YC, Liu DL. A brief discussion on the discipline development of rehabilitation massage therapy. Chin J Rehabil Theory Pract 2011;17:796-797.

[6] An GH. Reflections on the standardized development of traditional Chinese medicine massage. China Association of Traditional Chinese Medicine Massage Branch. Proceedings of the 9th annual academic conference of the China Association of Traditional Chinese Medicine Massage Branch and the Zhejiang Province Traditional Chinese Medicine Massage Branch Continuing Education Project 2006;2.

[7] Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. J Phys: Conf Ser 2018;1142:012012.

[8] Liakos KG, Busato P, Moshou D, et al. Machine learning in agriculture: A review. Sensors 2018;18(8):2674.

[9] Tarca AL, Carey VJ, Chen XW, et al. Machine learning and its applications to biology. PLoS Comput Biol 2007;3(6):e116.

[10] Xu YY, Li QY, Yi DH, et al. Dynamic Treatment Strategy of Chinese Medicine for Metastatic Colorectal Cancer Based on Machine Learning Algorithm. Chin J Integr Med. 2024; 1-8.

[11] Chen Z, Zhang D, Liu C, et al. Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning. Integr Med Res 2024;13(1):101019.

[12] Yang S, Zhao H, Zhu S, et al. Zhongjing: Enhancing the Chinese Medical Capabilities of Large Language Model through Expert Feedback and Real-world Multi-turn Dialogue. AAAI Conference on Artificial Intelligence 2023.

[13] Mercati M. The handbook of Chinese massage: tui na techniques to awaken body and mind. Vermont (US). Healing Arts Press, 2000.

[14] Wei X, Wang S, Li L, et al. Clinical evidence of chinese massage therapy (Tui Na) for cervical radiculopathy: a systematic review and meta-analysis. Evid Based Complement Alternat Med 2017;2017:9519285.

[15] Yang X, Zhao H, Wang J. Chinese massage (Tuina) for the treatment of essential hypertension: a systematic review and meta-analysis. Complement Ther Med 2014;22(3):541-548.

[16] Zhang X, Cao D, Yan M, et al. The feasibility of Chinese massage as an auxiliary way of replacing or reducing drugs in the clinical treatment of adult type 2 diabetes: a systematic review and meta-analysis. Medicine (Baltimore) 2020;99(34):e21894.

[17] Li F, He T, Xu Q, et al. What is the Acupoint? A preliminary review of Acupoints. Pain Med 2015;16(10):1905-1915.

[18] Felhendler D, Lisander B. Pressure on acupoints decreases postoperative pain. The Clin J Pain 1996;12(4):326-329.

[19] Bäumler PI, Simang M, Kramer S, et al. Acupuncture point localization varies among acupuncturists. Forsch Komplementmed 2012;19(1):31-37.

[20] Kang Y, Park J, Lee T, et al. Analysis of the Bone Proportional Method for Determining Acupoints in the Upper and Lower Abdominal Region in Males and Females. J Acupunct Res 2019;36(4):264-271.

[21] Regional Office for the Western Pacific. WHO standard acupuncture point locations in the Western Pacific Region. Available at: https://iris.who.int/handle/10665/353407

[22] Hu W, Sheng Q, Sheng X. A novel realtime vision-based acupoint estimation for TCM massage robot. 2021 27th International Conference on Mechatronics and Machine

Vision in Practice (M2VIP) 2021;771-776.
[23] Wang CC. Artificial Neural Network Based Image Detection Applied to Foot Massage Machines. J Rob, Networking Artif Life 2023;10(1):39-42.
[24] Zhang C, Wu Q, Wang J, et al. Massage Acupoint Positioning Method of Human Body Images Based on Transfer Learning. Trait du Signal 2023;40(6):2761-2768.
[25] Zheng CL, Feng Y. Design and Implementation of Face Acupoint Recognition System Based on Augmented Reality Technology. Comput Knowl Technol 2022;18(17):89-91.
[26] Shu J, Ding R, Jin A, et al. Acupoint Selection for Autonomous Massage Based on Infrared Thermography. Trait du Signal 2022;39:355-362.
[27] Masood D, Qi J. 3D Localization of hand acupoints using hand geometry and landmark points based on RGB-D CNN fusion. Ann Biomed Eng 2022;50(9):1103-1115.
[28] Zhang QY, Yi Y, Li C. Application of Genetic Algorithm Optimized BP Neural Network in Traditional Chinese Massage Robot. Appl Sci Technol 2017;44(2):73-77.
[29] Yang XP, Wu YD. Design of human acupoint positioning system based on PSO-BP neural network. Embed Technol 2018;44(09):75-78.
[30] Ramírez-Fernández C, Hernández-Capuchin I, Meza-Sánchez M, et al. RehWave: a real-time tele-rehabilitation haptic device for massage therapy. Proceedings of the 7th Mexican Conference on Human-Computer Interaction 2018;1-4.
[31] Zhu S, Lei J, Chen D. Recognition method of massage techniques based on attention mechanism and convolutional long short-term memory neural network. Sensors (Basel) 2022;22(15):5632.
[32] Liu GT, Yu PS, Tao Y, et al. Hybrid 3D printed three-axis force sensor aided by machine learning decoupling. Int J Smart Nano Mater 2024;15(2);261-278.
[33] Jia YK, Ding RT, Ren W, et al. Gesture Recognition of Somatosensory Interactive Acupoint Massage Based on Image Feature Deep Learning Model. Trait du Signal 2021;38(3):565-572.
[34] Li Z, Zeng C, Deng Z, et al. Learning Variable Impedance Control for Robotic Massage With Deep Reinforcement Learning: A Novel Learning Framework. IEEE Systs, Man, Cybern Mag 2024;10(1):17-27.
[35] Long D, Wang X, Tian M, et al. Estimation of Fatigue Status by sEMG Signal Using SVM Algorithm in Massage Assessment. 2019 IEEE International Conference on Mechatronics and Automation (ICMA) 2019;1316-1320.
[36] Zheng XB, Ling BW, Zeng ZT. Evaluation of effectiveness of eye massage therapy via classification of periocular images. Multimed Tools Appl 2022;81(4):5743-5760.
[37] Wang Q. Individualized medicine, health medicine, and constitutional theory in Chinese medicine. Front Med 2012;6(1):1-7.
[38] Sun F, Liu Z, Zhang W. Clinical Acupoint Selection for the Treatment of Functional Constipation by Massage and Acupuncture Based on Smart Medical Big Data Analysis. J Healthc Eng 2023;2023:9798104.
[39] Chae YB. Suggestions for the study of acupoint indications in the era of artificial intelligence. J Physiol Pathol Korean Med 2021;35(5):132-138.
[40] Alberts IL, Mercolli L, Pyka T, et al. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? Eur J Nucl Med Mol Imaging 2023;50(6):1549-1552.
[41] Kim JK, Chua M, Rickard M, et al. ChatGPT and large language model (LLM) chatbots: The current state of acceptability and a proposal for guidelines on utilization in academic medicine. J Pediatr Urol 2023;19(5):598-604.
[42] Tan Y, Zhang Z, Li M, et al. MedChatZH: A tuning LLM for traditional Chinese medicine consultations. Comput Biol Med 2024;172:108290.
[43] Zhang HB, Chen JY, Jiang F, et al. Huatuogpt, towards taming language model to be a doctor. (preprint) arXiv:2305.15075 2023.
[44] Zhang H, Wang X, Meng Z, et al. Qibo: A Large Language Model for Traditional Chinese Medicine. (preprint) arXiv:2403.16056 2024.
[45] Yang G, Liu X, Shi J, et al. TCM-GPT: Efficient pre-training of large language models for domain adaptation in Traditional Chinese Medicine. Comput Methods Programs Biomed Update 2024;6:100158.
[46] Ding LJ, Lv J, Liao YH. Research Progress on the Biomechanics of Traditional Chinese Medicine Massage Techniques. J Med Biomech 2023;38:1051-1056.
[47] Song PF, Xing H, Ma BF, et al. Construction of a Data Collection Platform for “Sitting Knee Adjustment Method” in the Treatment of Knee Osteoarthritis and Preliminary Exploration of Manual Operation Parameters. Beijing J Tradit Chin Mede 2023;42(4):442-445.
[48] Wang HF. Healthy China: Policy System and Development Path. People’s Forum 2024; 14-19.
[49] Dang L, Shi Q. Research on Chinese traditional medical massage robotic products usability design process. J Phys: Conf Ser 2020;1650(2):022014.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/883654uegazz

Optimization design and performance study of magnesium alloy vascular clamp

Weiwei Fan, Lin Mao, Bojun Liu, Chengli Song

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

Corresponding author: Lin Mao.

Declaration of conflict of interest: None.

Received March 21, 2024; Accepted June 5, 2024; Published September 30, 2024

Highlights

- A V-shaped vascular clamp featuring a locking mechanism and transverse teeth has been developed.
- Comparative analysis of clamps with various inner diameters reveals optimal closure with specific configurations.
- The designed clamp presents superior stress-strain response, robust clamping force, and consistent corrosion resistance.

Abstract

Purpose: This study investigates the effects of varying inner diameter of a vascular clamp made from an Mg–Nd–Zn–Zr alloy on its functional performance. The primary objectives are to optimize the clamp's structure, assess its performance across different inner diameters, and ultimately determine the optimal configuration. **Methods**: We developed a V-shaped vascular clamp equipped with a locking mechanism and transverse teeth. The study involved comparing vascular clamps with various inner diameters (0.35 mm, 0.4 mm, 0.5 mm and 0.6 mm, denoted as R0.35, R0.4, R0.5 and R0.6, respectively), achieved by modifying the clamp design. Finite element analysis simulated the closure process of these clamps, both with and without blood vessels, to analyze stress and strain distribution. Subsequently, we manufactured a clamp with the optimized design and conducted performance evaluations, including a closing strength test and an in vitro immersion test. **Results**: Among the tested vascular clamps, the R0.5 clamp demonstrated the lowest strain (0.50798) and minimal stress on blood vessels (0.7629 MPa). Notably, the R0.5 clamp remained intact during clamping fracture experiments and demonstrated a maximum closing force of 334.98±15.4 mmHg. Regarding corrosion resistance, the clamped position showed a higher corrosion rate (0.179±0.00551 mg.cm^{-2}.day^{-1}) compared to the open clamp (0.161±0.00306 mg.cm^{-2}.day^{-1}). **Conclusion**: The R0.5 clamp demonstrated superior performance in finite element analysis, showing effective vascular closure, strong clamping force, and uniform corrosion behavior. Overall, these results highlight its potential as an effective tool for vascular closure.

Keywords: Degradable magnesium alloy, hemostatic clip laparoscopic, surgery structural design, finite element analysis, corrosion behavior

Introduction

The vascular clamp, a pivotal advancement in medical technology, plays an essential role in surgical clamping and closing blood vessels, and it can also serve as a tool for secure attachment [1]. Currently, pure titanium (T) or titanium alloy vascular clips are extensively used in abdominal surgery, favored for their high strength, excellent ductility, and biocompatibility, which ensure secure and complete vascular occlusion. However, the use of titanium alloy clips can present certain drawbacks.

A significant drawback of titanium alloys is their slow degradation rate in the human body, leading to permanent implantation. This can potentially cause long-term allergic reactions [2, 3]. Moreover, the X-ray absorption coefficient of

Address correspondence to: Lin Mao, 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.

titanium alloys is substantially higher than that of human tissues, posing risks of interference with imaging exams and possible misdiagnoses. Postoperative complications associated with titanium clips include loosening, displacement, and premature detachment. These problems can lead to medical complications such as biliary leakage or bleeding, requiring a second surgery for clip removal [3].

In response to these challenges, the development of biodegradable Lapro-Clips, made from polymers, represents a significant innovation. These clips naturally degrade and are absorbed by the body after fulfilling their hemostatic role, thereby eliminating the risks associated with permanent clips. However, these polymer clips have a significantly lower strength compared to metal clips. To compensate for their reduced elastic modulus and strength while enhancing clamping force, the size of Lapro-Clips needs to be increased [4]. Additionally, the complexity of their structure can hinder their widespread application and implantation in the human body.

Magnesium and its alloys increasingly employed as biomedical metals in medical devices such as bone nails, cardiovascular stents, anastomosis nails, and vascular clamps. Magnesium alloys are also renowned as revolutionary metallic implant materials [5, 6]. Vascular clamps made from magnesium and its alloys provide promising biodegradability, biocompatibility, mechanical compatibility, and safe and reliable clamping performance, making them a low-cost and easy-to-use option, potentially eliminating the need for secondary operations [7].

Compared to traditional vascular clamp materials, magnesium alloy stands out with its lower density, higher specific strength and stiffness, and an elastic modulus similar to human bone [8-11]. However, due to the low corrosion resistance of magnesium alloys, apical cracks and excessive corrosion often occur during the use, resulting in poor sealing effects [12]. Additionally, the poor malleability of magnesium alloy makes them prone to deformation and excessive stress and strain, resulting in the rupture of the vascular clamp [13, 14].

To overcome these drawbacks, ongoing research efforts aim to refine the properties of magnesium alloy through innovative structural designs and advanced processing techniques. For example, Ikeo N et al. recently developed a Mg-Zn-Ca vascular clamp with improved ductility via casting, extrusion, blanking, and heat treatments [15]. Additionally, Ding et al. designed a circular, "V" shaped clamp using Mg-2Zn-0.2Ca and Mg-4Zn-0.2Ca materials, specifically preventing fracture during the clamping process [4].

Magnesium alloys are at the forefront of research as promising implant materials, with various alloys like AZ91, AZ31, WE43, WE54, and the biodegradable Mg-Nd-Zn-Zr being extensively studied [16]. Commercial magnesium alloys like AZ31 and WE43 were primarily designed for structural applications, emphasizing mechanical properties such as strength and toughness. However, when used as degradable biomaterials, they exhibit notable limitations. Furthermore, most AZ series magnesium alloys contain aluminum (Al), which, in physiological conditions, can act as a neurotoxin, potentially leading to Alzheimer's disease or muscle fiber damage [17, 18].

In physiological environments, the main issue with magnesium and magnesium alloys is their poor corrosion resistance, which may result in insufficient mechanical performance of implants during tissue healing, hindering their clinical application [19]. In contrast, Mg-Nd-Zn-Zr alloy (abbr. JDBM), designed specifically as a degradable biomaterial, offers better corrosion resistance and superior mechanical properties and biocompatibility compared to commercial alloys [20].

This research builds on prior studies, employing finite element analysis to optimize the clamp's structure for effective blood vessel occlusion. Subsequently, we employed a laser cutting process to fabricate the clamp according to the optimized design. We performed tests to measure the clamping force and assess the likelihood of clamp rupture. Additionally, we conducted in vitro immersion tests to evaluate the corrosion behavior of the magnesium alloy vascular clamp in artificial plasma, enhancing our understanding of the corrosion process involved.

Materials and methods

Structural design and optimization

The magnesium alloy vascular clamp studied here was meticulously represented using SolidWorks, as shown in **Figures 1A-B**, highlighting a 0.4 mm radius. **Figure 1C** displays the clamp's structural dimensions, while **Figure 1D** shows the tooth profile

Table 1. Size parameters of transverse tooth parameter

Transverse tooth parameter	Tooth length		Tooth height	Tooth pitch
	L1 (mm)	L2 (mm)	h (mm)	p (mm)
	0.1	0.2	0.1	0.1

Table 2. Mooney-Rivlin 5 Model Parameters

	Input parameters
Mooney-Rivlin 5 model	C10 = 1.89×10^{4} Pa
	C01 = 2.75×10^{3} Pa
	C20 = 8.572×10^{4} Pa
	C11 = 5.9043×10^{5} Pa
	C02 = 0 Pa
	Incompressible parameters D1 = 5×10^{-5} Pa^{-1}

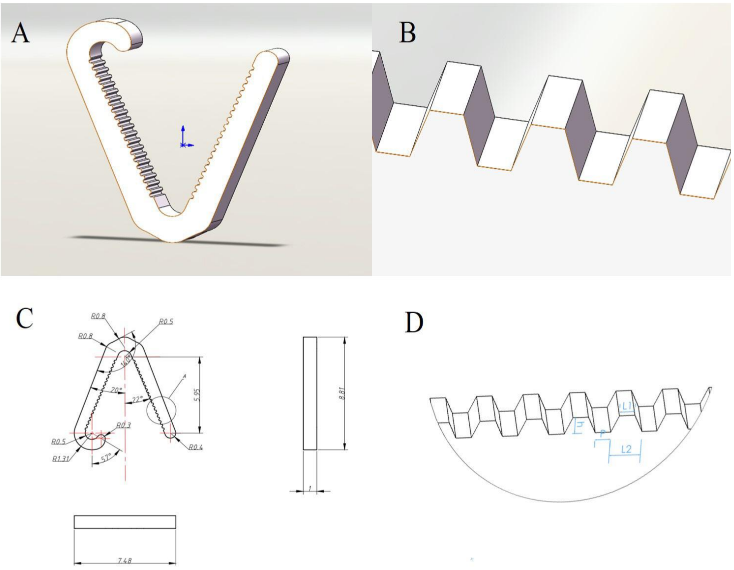


Figure 1. Modeling and structural dimensions of the designed vascular clamp. (A) Geometric model of the vascular clamp; (B) Lateral tooth details of the vascular clamp; (C) Schematic diagram of vascular clamp and structure size; (D) Transverse tooth structure dimension diagram.

parameters: L2 is the lower tooth length and L1 is the upper tooth length (L2=2L1); h is the tooth height; and p is the tooth pitch. Detailed parameters are provided in **Table 1.**

The clamp wall is intentionally designed with a gradually increasing thickness from the end to the top, effectively preventing breakage at the top during clamp usage. The top of the clamp is rounded and features a smaller opening, making it easier for the clamp to snugly fit against the blood vessel wall and clamp the vessel. To enhance friction between the clamp and the blood vessel wall and minimize clamp slippage, small transverse teeth are incorporated within the clamp design [21]. Additionally, a locking structure is introduced at the upper end of the left arm of the vascular clamp, improving its closure performance and preventing potential shifting or detachment during later stages of use.

Finite element simulation analysis

During blood vessel clamping, the mechanical properties of the vascular clamp undergo changes. We employed the finite element analysis software ANSYS Workbench to assess the stress and strain experienced by the vascular clamp under two conditions: non-blood vessel closure and blood vessel closure. Vascular clamps were fabricated using Mg-Nd-Zn-Zr alloy. The chosen intermediate venous segment had an internal diameter of 2.5 mm, a wall thickness of 0.25 mm, and a length of 10 mm. The blood vessel material was modeled using the Mooney-Rivlin 5 model with rubber-like properties. The parameters of this model are detailed in **Table 2**.

In the simulation, the assembly model of the magnesium alloy clamp and the blood vessel was imported into ANSYS Workbench 19.2 for finite element analysis and mesh generation.

Both blood vessel closure and non-blood vessel closure simulation processes employed hexahedral grids, with grid refinement applied to regions prone to stress concentration. The presence of transverse teeth led to non-uniform and relatively sparse grid generation on the transverse tooth cross-sections, necessitating grid refinement on these two lateral surfaces to achieve grid uniformity. The sweep method was utilized for further optimization of the vascular clamp grid, with the outer cross-sections serving as source faces and the entire vascular clamp as the target face. To ensure convergence of the calculation for closing the vascular clamp around the vessel, further grid refinement of the vessel is necessary. The transverse sections on both sides of the vessel

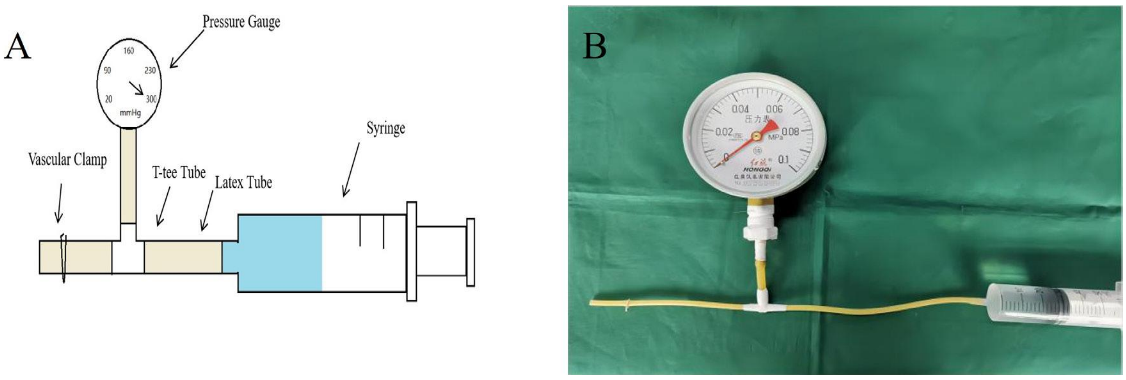


Figure 2. Test device for mechanical properties. (A) Theoretical diagram of mechanical properties test device; (B) Actual mechanical properties test device. Figure 2B is from the Research Center for Shanghai Institute for Minimally Invasive Therapy, School of Health Science and Engineering, University of Shanghai for Science and Technology.

Table 3. Simulated body fluids (artificial plasma) parameters

	Reagent name	Capacity(g/L)
Simulated body fluids (artificial plasma)	NaCl	6.8
	$NaHCO_3$	2.2
	KCl	0.4
	Na_2HPO_4	0.126
	$CaCl_2$	0.2
	$MgSO_4$	0.1
	NaH_2PO_4	0.026

are selected for Face Meshing, which generated mapped grids. The internal grid layers of the selected faces were adjusted to refine the grid count of the vessel. Meanwhile, mapped grids, known for their precision, facilitate faster solution speeds compared to free grids.

In finite element analysis, appropriate boundary conditions were established to reflect the real-world behavior of the vascular clamp during the vessel clamping process. The model's boundary conditions included applying opposite and moving displacements to the two arms of the clamp and securing the outer arc surface with Fixed Support to prevent shifting. To handle simulated deformation and hyperelastic material incorporation, the Large Deflection switch was activated. Additionally, the Analysis Settings were adjusted to include an automatic load step with additional manual load steps configured. To simulate blood vessel closure, the cross-section of one side was stabilized with Fix Support, and Nonlinear Adaptive Region was used to address convergence difficulties caused by blood vessel mesh deformation.

Closing strength property

To simulate and test the mechanical properties of the vascular clamp in vitro, we utilized a testing device as illustrated in **Figure 2A**. This device consists of a pressure gauge, a latex tube with an inner diameter of 2 mm and an outer diameter of 4 mm, a T-tee tube, and a syringe. The images of the mechanical property testing device are depicted in **Figure 2B**. Here, the latex tube serves as a blood vessel substitute. The T-tee tube facilitates the connection between the pressure gauge and the latex tube. One side of the latex tube is clamped using the clamp to simulate the closure of the blood vessel, while the other side is connected to a syringe containing simulated body fluid. Details of the simulated body fluid are provided in **Table 3**.

The procedure begins by securing the left side of the latex tube with the vascular clamp. Next, simulated body fluid was then gradually injected into the tube via the syringe. As the clamp fails to fully occlude the latex tube, the pressure gauge needle will move, indicating leakage from the rubber tube. The maximum pressure noted on the gauge at this point was recorded. Finally, the experiment was repeated thrice to calculate the average maximum closing force exerted by the vascular clamp.

In vitro corrosion performance

The corrosion behavior and degradation patterns of the magnesium alloy vascular clamp were meticulously examined through an in vitro immersion test using artificial plasma, with parameters outlined in **Table 3**. Both open-

Table 4. Comparison of tip deformation, maximum equivalent plastic strain and maximum stress of vascular clamp with different inner diameters

	Vascular clamp diameter		
	R0.35	R0.4	R0.5
Tip deformation (mm)	3.2398	3.2354	3.0889
Maximum equivalent plastic strain	0.51065	0.56558	0.49335
Maximum equivalent stress (MPa)	221.05	221.65	213.5

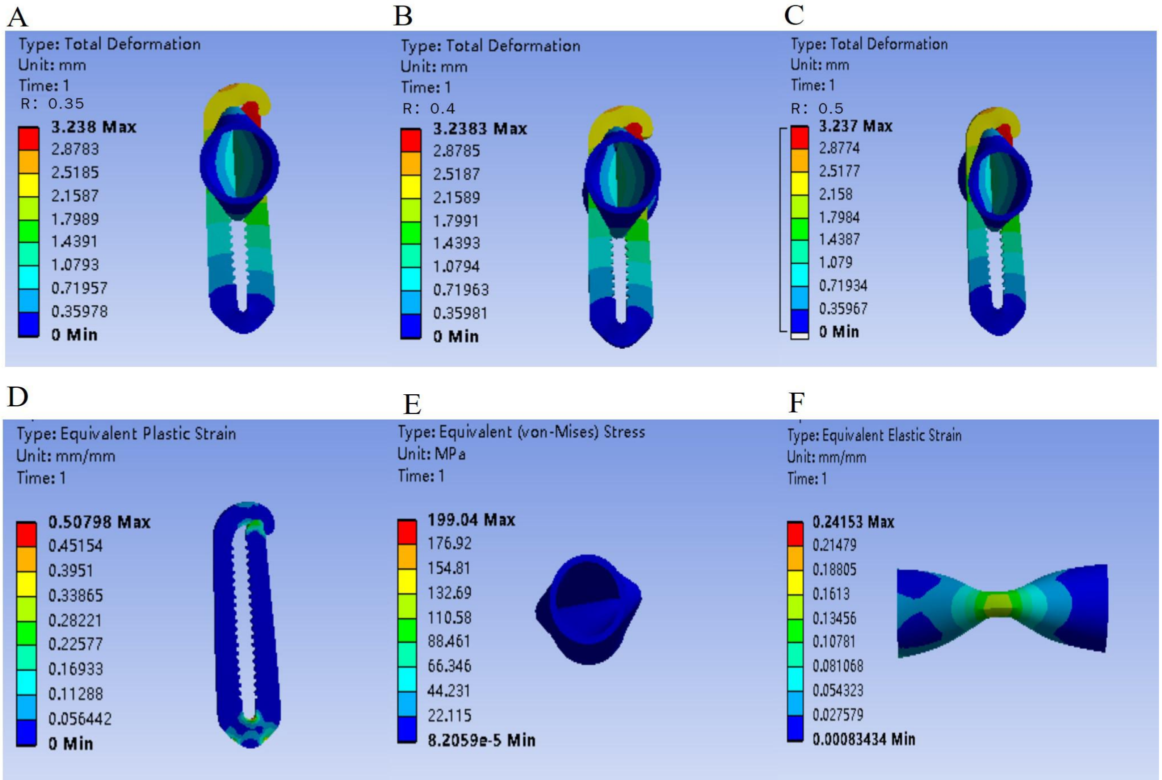


Figure 3. Mechanical simulation cloud image of vascular clamp (R0.35, R0.4, R0.5). (A-C) Deformation cloud images of the clamps (R0.35, R0.4, R0.5); (D) Equivalent plastic strain distribution of R0.5 clamp after vascular closure; (E) Diagram of stress distribution of blood vessels clamped by R0.5 clamp; (F) Strain distribution of blood vessels clamped by R0.5 clamp.

clamp and closed-latex tube clamps were tested simultaneously.

The immersion test followed the ASTM-G31-72 standard. The ratio of artificial plasma volume (ml) to the vascular clamp surface area (cm^2) was 1.8 ml/mm^2. The solution was maintained in a thermostatic water bath at 37±0.5 °C and changed every 24 hours to simulate a dynamic physiological environment. After soaking for 24 hours, the hydrogen gas produced from the degradation of the magnesium alloy vascular clamp was measured using an inverted scale pipette.

The pH value of the solution after 24 hours of immersion was measured using a pH meter. After the designated period, the sample was removed, rinsed, dried, and cleaned with a standard chromic acid solution. The clamps were then weighed. The difference in weight before and after immersion represents the weight loss.

Three sets of parallel samples were taken for each group. The corrosion rate (CR) of the magnesium alloy vascular clamp in artificial plasma was calculated using the formula:

$$CR = (8.76 \times 104 \times W) / (A \times T \times D)$$

Where:

W (g) is the weight change of the clamp.

A (cm^2) is the surface area of the clamp.

T (hours) is the immersion time.

D (g/cm^3) is the density of the magnesium alloy.

Results

Mechanical simulation results and analysis

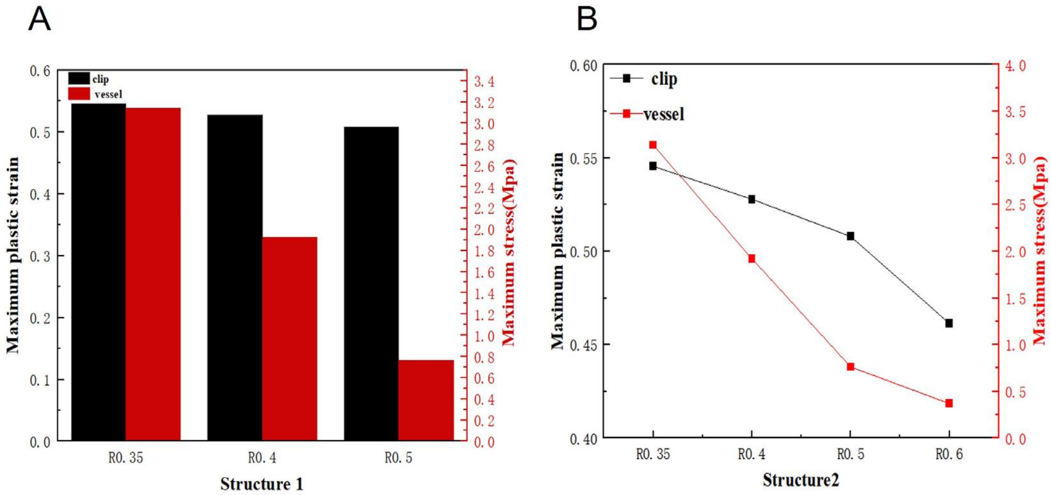


Figure 4. The maximum equivalent plastic strain and the maximum equivalent stress of blood vessel with different inner diameter clamp. (A) the finite element analysis results of three types of vascular clamps (R0.35, R0.4, and R0.5); (B) the decreasing trends of maximum strain and stress in vessels with different clamp diameters (R0.35, R0.4, R0.5, R0.6).

This section presents the results of finite element simulation aimed at elucidating the impact of various inner diameters on vascular occlusion in the absence of blood vessels. As detailed in **Table 4**, a key finding is the inverse relationship between the inner diameter of the vascular clamp and the maximum equivalent plastic strain. Specifically, the strain reached its lowest, 0.49335, at an inner diameter of 0.5 mm, indicating the most favorable outcome among the tested diameters.

Among the three inner diameters, the R0.4 vascular clamp exhibited the highest maximum equivalent stress at 221.65 MPa. Conversely, the R0.5 vascular clamp demonstrated a slightly lower maximum stress of 213.5 MPa, representing a 3.68% reduction compared to the R0.4 vascular clamp. The differences in tip deformation across the three vascular clamps were relatively minor, yet the R0.35 clamp exhibitsed more deformation, whereas the R0.5 clamp showed the least.

Based on these results, it can be concluded that the R0.5 scheme is optimal for minimizing tip deformation, achieving the lowest maximum plastic strain, and generating the smallest maximum equivalent stress.

Figures 3A-C illustrate the deformation cloud images of the vascular clamps for blood vessels with different inner diameters (R0.35, R0.4, R0.5). Notably, the tip deformation across all three clamp sizes was consistent, approximately 3.2 mm. These images clearly demonstrate the clamps' effective closure mechanism, highlighting their ability to completely occlude blood vessels without leaving any gaps, thus meeting surgical requirements.

Figure 3D displays the distribution of equivalent plastic strain in the R0.5 vascular clamp after vessel closure. The maximum plastic strain occurred at the inner arc of the top section of the clamp, aligning with the area of the highest stress concentration. Examining the stress distribution cloud diagram in **Figure 3E**, the maximum stress recorded is 199.04 MPa. **Figure 3F** presents the strain cloud diagram of a blood vessel closed by the R0.5 clamp, showcasing the maximum strain of 0.24153 experienced by the blood vessel. Despite the inevitable local stress concentration after the vessel clamping, a well-designed structure can minimize stress within the blood vessels, achieving the smallest maximum plastic strain of the vascular clamp and reducing the maximum equivalent stress on the particular blood vessel segment.

Figure 4A presents the finite element analysis results of three types of vascular clamps: R0.35, R0.4, and R0.5. The maximum equivalent plastic strain was 0.54551 for the R0.4 clamp, 0.5279 for the R0.35 clamp, and 0.50798 for the R0.5 clamp, with R0.5 clamp exhibiting the lowest strain. As the inner diameter of the vascular clamp increases, the maximum strain gradually decreases. The maximum stress was 3.1388 MPa for the R0.35 clamp, 1.9215 MPa for the R0.4 clamp, and 0.7629 MPa for the R0.5 clamp, representing a 75.7% decrease from the R0.35 clamp and a 54.9% decrease from the R0.4 clamp, indicating superior performance.

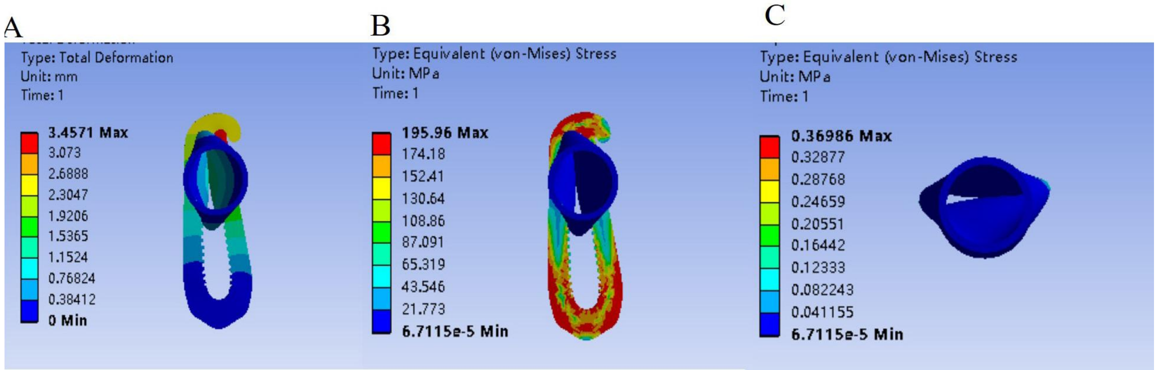


Figure 5. Mechanical simulation cloud image of R0.6 vascular clamp. (A) Deformation cloud image of the R0.6 clamp; (B) Stress distribution cloud map of the clamp; (C) Stress distribution diagram of blood vessels clamped by R0.6.

Table 5. Maximum closing force of magnesium alloy vascular clamp.

	R0.5 magnesium alloy vascular clamp
Sample 1	43 kPa
Sample 2	46 kPa
Sample 3	45 kPa
Mean pressure (kPa)/ (mmHg)	44.66±2.05 kPa / 334.98±15.4 mmHg

Figure 4B illustrates the decreasing trends of maximum strain and stress in vessels with increased clamp diameters. However, subsequent analysis with the inclusion of the R0.6 clamp closure reveals that although it further reduces stress and strain, it fails to meet the requirements for vessel closure performance due to its inability to fully close the vessels.

The finite element simulation results clearly demonstrate the superior performance of the R0.5 vascular clamp over its counterparts with smaller inner diameters. To rigorously assess the structural efficacy of the R0.5 clamp, the study introduces the R0.6 vascular clamp, featuring an increased internal diameter. Vascular clamping simulations were conducted using blood vessels, and the outcomes are presented in **Figure 5**.

The results indicate that, despite the smaller strain after blood vessel closure (0.46138) and lower stress on blood vessels (0.36986 MPa) exhibited by the R0.6 clamp compared to the R0.5 clamp (strain: 0.50798, stress: 0.7629 MPa), there are noticeable shortcomings. The tip deformation of the R0.6 clamp (3.4571) exceeded that of the R0.5 clamp (3.237). Moreover, during the vascular clamping simulation, it was observed that there were significant gaps between the two arms of the clamp, resulting in incomplete closure of the blood vessels. This phenomenon hinders effective hemostasis and may lead to postoperative bleeding, ultimately compromising the success of the procedure. Consequently, the structural optimality of the R0.5 clamp is re-affirmed.

Test result of closure strength

The experimental results on closure strength are presented in **Table 5**. The average pressure of the R0.5 magnesium alloy vascular clamp was measured to be 44.66±2.05 kPa, corresponding to 334.98±15.4 mmHg. Normal systolic blood pressure in humans ranges from 90-140 mmHg, while individuals with grade III hypertension typically have a systolic blood pressure of around 180 mmHg (24 kPa), which is considerably lower than the pressure gauge reading of 300 mmHg. Therefore, it can be inferred that the maximum average pressure value (334.98±15.4 mmHg) measured by the experiment for the vascular clamp during actual closure exceeds the normal blood pressure range for humans and even surpasses the systolic blood pressure of hypertensive individuals, indicating a notable capacity to tolerate high pressures.

In vitro corrosion performance

The corrosion characteristics of Mg-Nd-Zn-Zr vascular clamps were investigated through in vitro immersion tests, assessing both hydrogen evolution and pH value measurements. **Figure 6A** illustrates the daily hydrogen evolution volume of Mg-Nd-Zn-Zr vascular clamps immersed in artificial plasma, indicating a volume below 0.3 ml/cm^2. Over the 10-day

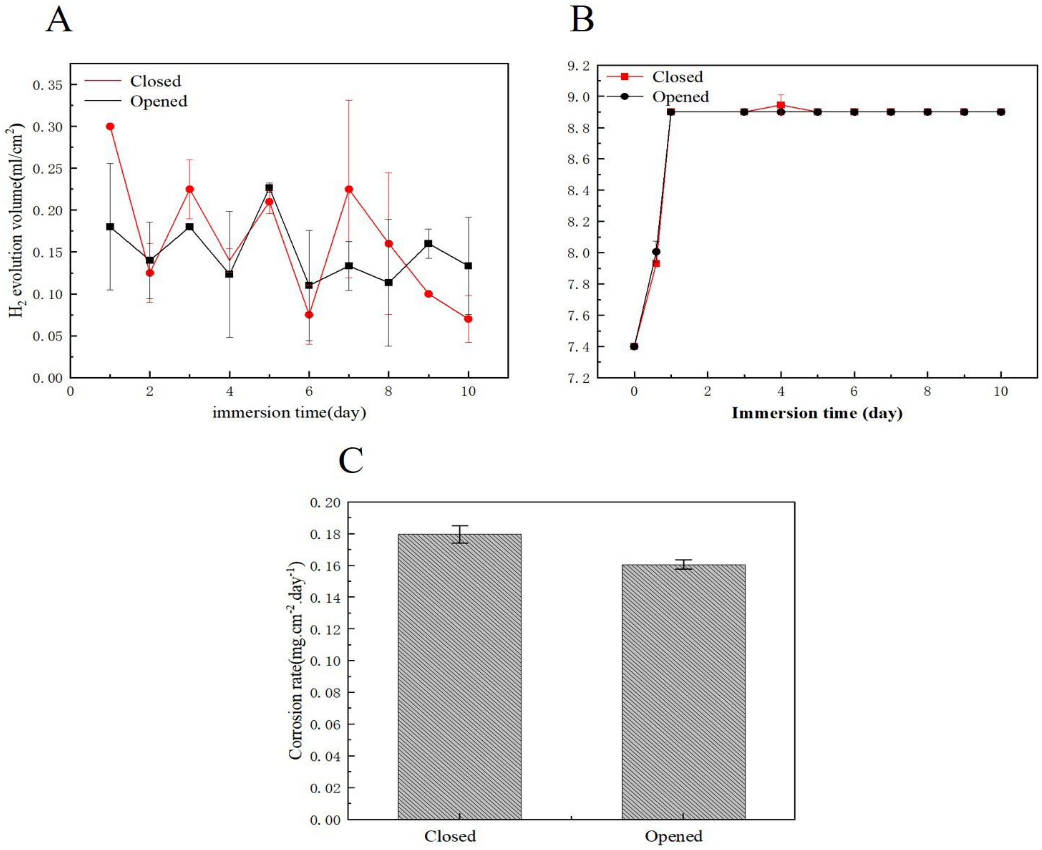


Figure 6. Results of in vitro immersion tests for magnesium alloy vascular clamp. (A) Daily hydrogen evolution of R0.5 vascular clamp in artificial plasma solution; (B) PH value of artificial plasma for the immersion of R0.5 magnesium alloy vascular clamp; (C) Corrosion rate of magnesium alloy vascular clamp.

experimental period, the closed clamp exhibited a higher average hydrogen evolution sum (1.63±0.1539 ml/cm^2) compared to the open clamp (1.58±0.1334 ml/cm^2), as demonstrated by experimental data analysis.

Figure 6B depicts the pH dynamics of the artificial plasma over the immersion time for the Mg-Nd-Zn-Zr vascular clamp. It is evident from **Figure 6C** that within the initial 24 hours of immersion, the pH value of the of the immersion medium for Mg-Nd-Zn-Zr vascular clamp increased rapidly from 7.4 to approximately 8.9. Subsequently, the pH changes gradually leveled off, stabilizing around 8.9 after the second day of immersion.

The corrosion rate of the Mg-Nd-Zn-Zr vascular clamp was determined using weight loss experiments, and the corresponding bar graph is displayed in **Figure 6C**. The average corrosion rate for the closed magnesium alloy vascular clamp was calculated as 0.179±0.00551 mg·cm^{-2}·day^{-1}, while the average corrosion rate for the open magnesium alloy vascular clamp was 0.161±0.00306 mg·cm^{-2}·day^{-1}. These results align with the corrosion rate derived from weight loss, indicating a higher corrosion rate for the closed vascular clamp compared to the open vascular clamp in the artificial plasma solution.

Discussion

Laparoscopic surgery often encounters challenges related to bleeding and the need for timely hemostasis. The development of the vascular clamp represents a significant advancement in medical hemostatic tools. Traditional hemostatic tools, such as surgical sutures and anastomosis nails, have their limitations. Sutures may not always provide reliable hemostasis or secure closure [22]. Anastomosis nails, on the other hand, often require multiple nails, which may increase the risk of blood leakage [23]. In comparison, the vascular clamp is smaller and more flexible, making it particularly advantageous for application in narrow tissue depths and areas that are challenging to access with standard hemostatic clamps. The vascular clamp offers improved maneuverability and precision in achieving effective hemostasis during laparoscopic surgery.

Previous studies on magnesium alloy vascular clamps primarily focused on the material's biodegradability and biocompatibility, while our design also emphasizes its mechanical reliability and ease of use [24]. The introduction of a locking mechanism to the upper end of the clamp's left arm is an essential improvement, addressing the issue of clamp displacement and detachment that was prevalent in earlier designs [25]. This feature ensures a more secure and stable closure, reducing the risk of post-surgical complications.

In laparoscopic surgery, the circular inner diameter opening at the top of the vascular clamp is pivotal in successfully closing blood vessels. The optimal size of this opening is crucial as it ensures a snug fit against the vessel wall, thereby facilitating a successful closure. However, it is also necessary to consider the structural integrity of the clamp [26].

A smaller diameter often results in a tighter curvature at the bend between the clamp's arms, leading to significant change in thickness across these components. This leads to a phenomenon called stress concentration, where the maximum stress in a localized area is much higher than the average stress the structure endures. Stress concentration can cause fracture in brittle materials and even induce fatigue cracks in ductile materials [27].

The top inner part of the vascular clamp is particularly susceptible to stress concentration due to the sudden change in shape during the clamping action. Therefore, optimizing the design of this area is a key focus of this research. We aim to find a balance between minimizing the circular inner diameter opening for efficient vessel closure and ensuring structural integrity to avoid stress concentration and potential fractures.

Based on the existing design of vascular clamp, we propose an optimized design by changing the inner diameter of the bottom side of the clamp. In addition to the existing vascular clamp of 0.4 mm, two other inner diameters were designed, 0.5 mm and 0.35 mm. The stress and strain conditions of the three inner diameters were compared to select the best structure of the vascular clamp. At the same time, the optimal performance of the vascular clamp was verified by further increasing the inner diameter of the vascular clamp.

One of the primary objectives of the vascular clamp design is to ensure complete closure of the blood vessel, thus guaranteeing its safety and stability. Through finite element simulations, our study has shown that the vascular clamp with an inner diameter of 0.5 mm exhibits the most favorable stress-strain performance during complete blood vessel closure. It has been demonstrated that clamps with larger inner diameters possess superior stress-strain characteristics during complete closure of the blood vessel. We observed that as the inner diameter increases, the stress concentration at the bend of the vascular clamp gradually diminishes, leading to a gradual reduction in stress strain. However, at an inner diameter of 0.6 mm, the contact area between the blood vessel and the clamp decreases, preventing the clamp from adequately fitting the blood vessel. Consequently, an inner diameter of 0.6 mm fails to meet the practical requirements for closing blood vessels effectively during clinical surgeries.

Conclusion

Based on the findings from two sets of simulation analyses comparing vascular clamps with varying inner diameters, the vascular clamp with an inner diameter of 0.5 mm demonstrated the most favorable mechanical properties among the three variations tested. Subsequently, an optimized Mg-Nd-Zn-Zr alloy vascular clamp was fabricated using a laser cutting method.

In the gripping force test experiment, the R0.5 vascular clamp exhibited a maximum closing force of 334.98±15.4mmHg, surpassing both the normal blood pressure value of the human body and the systolic blood pressure value of the hypertensive group (180 mmHg). This indicates that the clamp possesses strong gripping capability. Moreover, corrosion rate analysis revealed that the closed clamp (0.179±0.00551 mg.cm^{-2}.day^{-1}) exhibited a slightly faster corrosion rate compared to the open clamp (0.161±0.00306 mg.cm^{-2}.day^{-1}).

Consequently, the newly developed Mg-Nd-Zn-Zr alloy vascular clamp holds significant promise as an effective instrument for closing blood vessels during laparoscopic surgery.

There are several limitations to our study that should be acknowledged. Firstly, the accuracy of our results may be influenced by the assumptions and simplifications inherent in the simulation model. Moreover, the simulation analysis was conducted solely under static conditions, whereas the mechanical behavior of the vascular clamp in real-life situations is influenced by additional factors, such as blood flow.

Further investigation into the biodegradability, a critical attribute for reducing long-term complications, are recommended. Conducting animal experiments to examine the corrosion behavior of the degradable magnesium alloy vascular clamp within the body could help simulate real surgical scenarios more accurately. Such studies would ideally include evaluations of the vascular clamp's impact through blood biochemical analysis and histological examinations, further validating its clinical applicability and safety.

Author Contributions: Lin Mao, and Weiwei Fan conceived this project. Lin Mao and Weiwei Fan carried out structural design and finite element simulation. Lin Mao and Chengli Song designed a closing strength experiment and an in vitro immersion experiment. Weiwei Fan, Bojun Liu, Xin Zheng and Yujie Zhou performed experiments and collected the relevant data and analyzed. All the authors contributed to the writing of manuscript.

References

[1] Ooi BP, Hassan MR, Kiew KK, et al. Case report of a hemostatic clip being retained for 2 years after deployment. Gastrointest Endosc 2010;72(6):1315-1316.

[2] Javed F, Al-Hezaimi K, Almas K, et al. Is titanium sensitivity associated with allergic reactions in patients with dental implants? A systematic review. Clin Implant Dent Relat Res 2013;15(1):47-52.

[3] Fage SW, Muris J, Jakobsen SS, et al. Titanium: a review on exposure, release, penetration, allergy, epidemiology, and clinical reactivity. Contact Dermatitis 2016;74(6):323-345.

[4] Ding PF, Liu YC, He XH, et al. In vitro and in vivo biocompatibility of Mg-Zn-Ca alloy operative clip. Bioact Mater 2019;4:236-244.

[5] Radha R, Sreekanth D. Insight of magnesium alloys and composites for orthopedic implant applications-a review. J Magnesium Alloys 2017;5(3):286-312.

[6] Zhao D, Witte F, Lu F, et al. Current status on clinical applications of magnesium-based orthopaedic implants: A review from clinical translational perspective. Biomaterials 2017;112:287-302.

[7] Zhang RF. Preparation and properties of biodegradable Mg-Zn-Ca alloy vascular clamp for laparoscopic surgery. Tianjin University of Technology 2017.

[8] Witte F, Kaese V, Haferkamp H, et al. In vivo corrosion of four magnesium alloys and the associated bone response. Biomaterials 2005;26(17):3557-3563.

[9] Yufeng Z, Bin L, Xuannan G. Research progress of biodegradable medical metal materials Materials Review 2009;23(1):6.

[10] Fischer J, Crostack H, Windhagen, et al. Synchrotron-radiation based microtomography for in-vivo corrosion measurements of magnesium alloys. ORS Transactions 2005.

[11] Witte F. The history of biodegradable magnesium implants: A review. Acta Biomater 2010;6(5):1680-1692.

[12] Wang RX, Hong LX, Zhang XB. Research progress in corrosion resistance of biomedical magnesium alloys. J Mater Eng 2021;49(12):14-27.

[13] Ren D, Zeng F, Liu L, et al. Bending Properties of Mg Alloy Tailored Arc-Heat-Treated Blanks. Materials 2019;12(6):977.

[14] Wang Z, Cao H, Zheng H, et al. Effect of aging treatment on the mechanical property and precipitation transformation of Mg-Co-Y alloy. Mater Sci Eng: A 2023;881:145322.

[15] Ikeo N, Nakamura R, Naka K, et al. Fabrication of a magnesium alloy with excellent ductility for biodegradable clips. Acta Biomater 2016;29:468-476.

[16] Bai H. Preparation, microstructure and in vivo and in vitro properties of magnesium alloy vascular clamp. Tianjin University of Technology 2018.

[17] Zhang Y, Li J, Li J. Effects of microstructure transformation on mechanical properties, corrosion behaviors of Mg-Zn-Mn-Ca alloys in simulated body fluid. J Mech Behav Biomed Mater 2018;80:246-257.

[18] Liu H, Zhang W, Fang Y, et al. Neurotoxicity of aluminum oxide nanoparticles and their mechanistic role in dopaminergic neuron injury involving p53-related pathways. J Hazard Mater 2020;392:122312.

[19] Abdel-Gawad SA, Shoeib MA. Corrosion studies and microstructure of Mg−Zn−Ca alloys for biomedical applications. Surf Interfaces 2019;14:108-116.

[20] Mao L, Zhou H, Chen L, et al. Enhanced biocompatibility and long-term durability in vivo of Mg-Nd-Zn-Zr alloy for vascular stent application. J Alloys Compd 2017;720:245-253.

[21] Mao L, Zheng X, Tian Y, et al. Structural Optimization, Fabrication, and Corrosion Behaviors of Biodegradable Mg-Nd-Zn-Zr Alloy Hemostatic Clip. Metals 2022;12(11):1979.

[22] Wang XC, Wu QT, Zheng ZZ, et al. Research progress of medical sutures. J Anhui Eng Univ 2020;35(05):1-11.

[23] Lei J. Preparation and research of degradable

magnesium alloy for anastomosis nail. Lanzhou University of Technology 2022.
[24] Mochizuki A, Kaneda H. Study on the blood compatibility and biodegradation properties of magnesium alloys. Mater Sci Eng C Mater Biol Appl 2015;47:204-210.
[25] Sun R, Ren LZ, Zhang ZP, et al. Pilot study of minimum occlusive force of vascular clamps on arterial vessels in rats. Sci Rep 2021;11(1):6042.
[26] Chen HY, Einstein DR, Chen K, et al. Computational Modeling of Vascular Clamping: A Step Toward Simulating Surgery. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference 2005;302-303.
[27] Wang YN, Oyen D, Guo W, et al. StressNet - Deep learning to predict stress with fracture propagation in brittle materials. npj Mater Degrad 2021;5(1):6.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/846307fkxccq

Analysis of urinary non-formed components at home based on machine learning algorithms

Yifei Bai, Rongguo Yan, Yuqing Yang, Chengang Mao

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

Corresponding author: Rongguo Yan.

Declaration of conflict of interest: None.

Received April 12, 2024; Accepted July 11, 2024; Published September 30, 2024

Highlights

- The study evaluated five machine learning algorithms in analyzing urinary non-formed components. Among them, the Random Forests model demonstrated the highest accuracy, precision, recall, and F1 score, suggesting its effectiveness in analyzing urinary non-formed components.
- A technological innovation is introduced for home urinalysis, offering the potential to enhance medical efficiency and patient experience.

Abstract

Objective: Machine learning can automatically extract valuable insights from vast datasets, predict and classify diseases, and evaluate drug efficacy. To assess the effectiveness of machine learning algorithms in analyzing non-formed components in urine, real medical data were processed and annotated. **Methods:** Five models, including K-Nearest Neighbors, Decision Trees, Random Forests, Support Vector Machines, and Gaussian distributions, were constructed to quantitatively analyze 12 non-formed urine components, such as vitamin C, white blood cells, and urinary bilirubin. The efficacy of these models was then compared. **Results:** It was found that the Random Forest model outperformed others, achieving the lowest mean squared error, high recall rate, accuracy, and area under the curve. **Conclusions:** These findings indicate that machine learning offers significant potential for studying non-formed urine components, potentially enhancing the precision and effectiveness of disease detection and providing valuable support for clinical decision-making.

Keywords: Home urine component analysis, machine learning models, quantitative results

Introduction

Urine composition analysis is a widely utilized medical diagnostic technique that accurately reflects various physiological characteristics, such as bilirubin, protein, and glucose levels [1]. Beyond its application in diagnosing and monitoring diseases of the kidneys and urinary tract, urine dry chemical analysis is crucial for assessing individual health, drug usage, auxiliary diagnosis, and monitoring occupational diseases. It also plays a significant role in diagnosing conditions affecting other organs. However, traditional urine analysis often proves costly and impractical for home use due to the need for specific tools and laboratory settings [2].

The medical industry has extensively leveraged machine learning, particularly in areas such as wearable sensors, medicinal chemistry, brain and cancer research, and medical imaging. The unique strength of machine learning algorithms lies in their ability to analyze large datasets to uncover patterns, elucidate underlying phenomena, and identify correlations, which can significantly aid medical professionals in better understanding and diagnosing patients [3]. For many diseases, machine learning-based approaches have shown potential to enhance the performance, accuracy, predictability, and reliability of diagnostic systems.

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

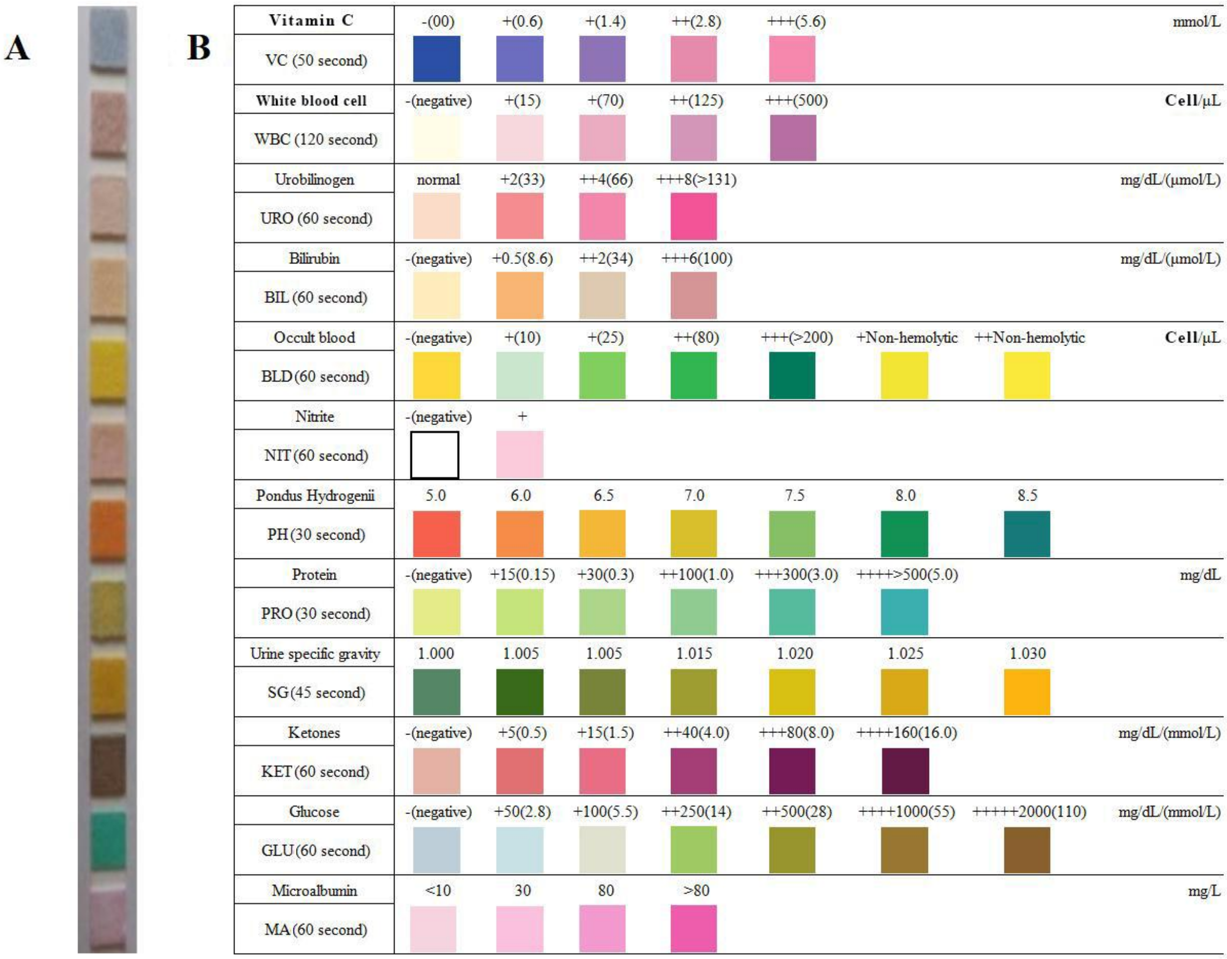


Figure 1. URIT 12-pads Test strips (A) and its color charts (B).

The field of urine analysis and related domains offer a broad array of applications for machine learning. Technological advancements in color recognition, sensor technology, and related fields have significantly enhanced the analysis of urine components. Moreover, research efforts aimed at improving user experience, portability, and the development of intelligent analytical systems hold potential to make urine detection technology more practical and widely applicable [4-6].

This study processed and annotated real medical data to develop models based on K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Support Vector Machines (SVM), and Gaussian distributions. These models were used for the quantitative analysis of 12 non-formed components in urine, including vitamin C, white blood cells, urobilinogen, etc. The primary objective was to investigate the efficacy of machine learning algorithms in analyzing urinary non-formed components. Additionally, the study compared the recognition capabilities of the five models.

Material and methods

Principles of urine dry chemistry test

Urine analysis using the dry-chemistry strip method is the most prevalent detection technique, as depicted in **Figure 1**. In this method, a urine sample is directly applied to a dry test strip [7, 8]. The water content of the urine acts as the solvent, causing specific elements in the urine to chemically react with materials in the test strip pad, resulting in a color change. Darker shades indicate higher concentrations of the tested components, establishing a correlation between color intensity and component concentration [9]. Quantitative analysis of these non-formed components was conducted by comparing the color-changed test strip with a urine analysis color chart, as illustrated in **Figure 2**. The simplicity and rapidity of dry chemistry urine analysis make it a favored option for home use, as well as in clinics and other medical settings.

Principles of machine learning algorithm

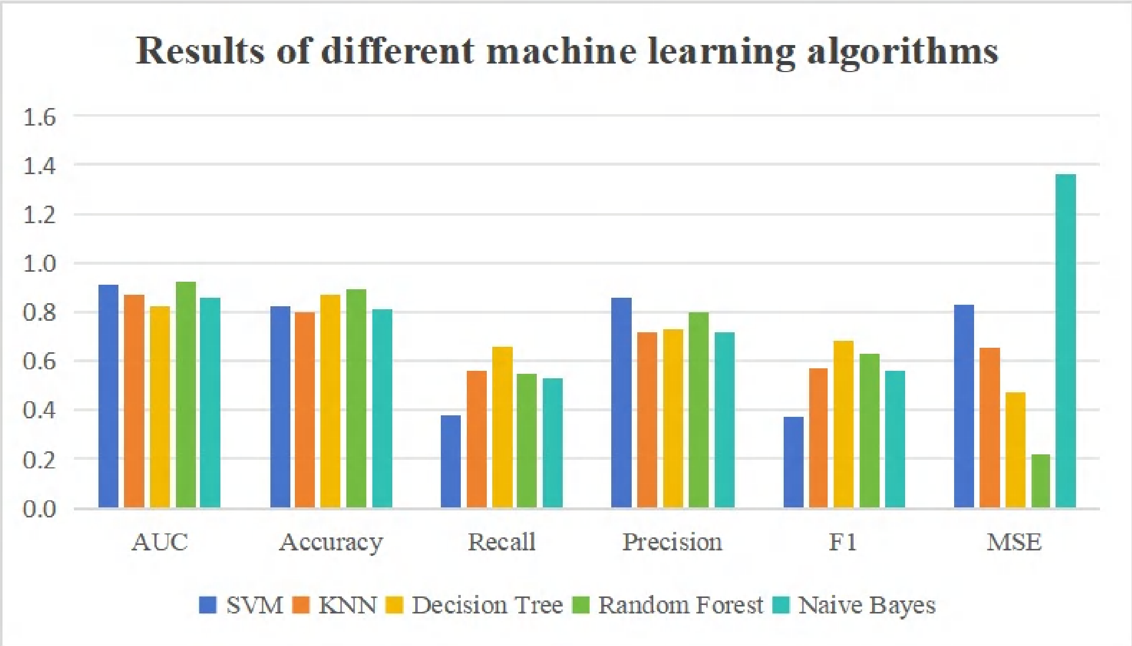


Figure 2. Comparison of the results of different Machine Learning algorithm. SVM, Support Vector Machines; KNN, K-Nearest Neighbors; AUC, Area Under the Curve; MSE, Mean Squared Error.

KNN

The KNN technique is a widely employed method for both regression and classification tasks. Central to KNN is the premise that a sample is likely to belong to a certain class if the majority of its KNN in the feature space are members of that same class [10].

The KNN algorithm operates through several steps. Initially, it constructs a feature space model using labeled training examples. Then, it identifies the k training samples closest to the new, unlabeled sample within this feature space. Based on the labels of these k neighbors, the algorithm predicts the classification or regression outcome for the new sample [11].

The KNN algorithm is valued for its simplicity and ease of understanding, along with its effectiveness in handling non-linear problems. However, it has some drawbacks, particularly when applied to large datasets. The algorithm is sensitive to noise and outliers in the feature space, which can adversely affect its accuracy. Additionally, it is computationally demanding, as it involves calculating the distance between samples for each query, which increases the computational complexity.

Decision tree

A popular method for constructing decision trees is known as Classification and Regression Trees. This approach uses a greedy algorithm to split data at each node, aiming to maximize information gain or minimize Gini impurity [12]. The algorithm continues to divide the data into two subsets until a stopping condition is met, such as reaching the maximum tree depth or having only a minimal number of samples at a node. The result of this process is a binary tree structure, where each leaf node represents a class label or prediction value, and each internal node corresponds to a feature.

Classification and Regression Trees algorithms are highly effective due to their ability to handle both numerical and categorical data, and they provide a straightforward method for visualizing and interpreting the correlations between features and outcomes. However, they are prone to overfitting. This issue can be mitigated with techniques such as pruning, which helps to simplify the model and improve its generalizability by removing sections of the tree that provide little predictive power.

SVM

SVM is a robust supervised learning method effectively used for both binary and multiclass classification challenges. Based on the principles of the perceptron model, SVM seeks to find a line (in two-dimensional space) or a hyperplane (in three-dimensional or higher-dimensional spaces) that optimally separates data points of different classes [13]. The ideal hyperplane is the one that maximizes the margin between these classes, with the most significant data points, known as support vectors, defining this decision boundary.

For multi-class classification tasks, SVM can be implemented using two primary approaches: the "one-versus-rest" (also known as "one-versus-all") and the "one-versus-one" (also referred to as "pairwise") methods [14, 15]. In the "one-versus-rest" approach, an SVM model is trained for each class, where the specific class is treated as the positive class and all others are combined into a negative class. This requires training multiple SVM models—one for each class. On the other hand, the "one-versus-one" method involves training an SVM model for every possible pair of classes, resulting in n(n-1)/2 models for n classes. To ascertain the final classification, a voting or scoring mechanism is used, where the class that receives the highest score from the SVM models is selected.

In this study, we utilized 8,036 sets of color data from urinalysis multi-test strips to train an SVM-based color classification algorithm employing the "one-versus-rest" approach. We designated 1,004 sets of color data for testing and an additional 1,004 sets for validation of the method. The polynomial kernel function was chosen for the SVM model to optimize the classification process.

Random forest

Random Forest is an ensemble learning technique that constructs multiple decision trees and aggregates their predictions through voting or averaging. Each tree is developed using bootstrap samples from the original dataset, involving random sampling with replacement [16]. Additionally, during the construction of each tree, a random subset of features is selected for node splitting. This means that each node decision is made based on a limited subset of features, rather than the entire feature set. This strategy not only increases the diversity among the trees but also helps reduce overfitting, thereby enhancing the robustness and overall performance of the model. Random Forest is particularly effective for both classification and regression tasks, capable of managing high-dimensional data and providing valuable insights into feature importance [17, 18].

Gaussian naive bayes

Gaussian Naive Bayes is a specialized form of the Naive Bayes classifier designed to handle continuous data, assuming that the distribution of features follows a Gaussian (normal) distribution. This forms part of the broader Naive Bayes algorithmic framework, which posits that each feature is independent given the class label [19, 20]. Although this independence assumption may not always align with real-world data complexities, it simplifies the computational process significantly. Despite its simplicity, Gaussian Naive Bayes can perform effectively across a variety of practical applications.

Analysis of urinary non-formed components based on machine learning algorithms

Data collection

In this study, the standard urine poly-reagent band was photographed, and its color was extracted using a smartphone. Due to the limited amount of standard data available, the color data of the collected standard urine poly-reagent band was expanded using the interpolation method, resulting in a comprehensive standard color dataset consisting of 1908 entries. Additionally, actual urine color information from 935 patients was obtained from the hospital to further enhance the model's functionality.

Data cleaning

In the data cleaning process, several key steps need to be considered:

• Handling Missing Values: Various techniques can be employed to address missing values, depending on the rate of missing data and its significance. These techniques include eliminating fields, imputing missing values, or recollecting data [21]. Approaches for filling in missing information can utilize the mean, median, quantiles, mode, random values, interpolation, and other methods. Additionally, business knowledge or experience can guide this process. Virtual variables that map to high-dimensional space can also be introduced, or a model for missing data prediction can be constructed.

• Format and Content Processing: Ensuring the correctness and integrity of data formats and content is crucial for data processing. This involves transforming, purifying, and standardizing the data in an appropriate manner to facilitate further analysis and modeling.

• Duplicate Data Removal: Removing redundant entries helps prevent their influence on the model. This can be achieved by identifying and eliminating duplicate data from the dataset through the comparison of content or record identifiers.

• Noise Data Handling: Anomalous or erroneous data points are referred to as noise data. An excessive amount of noisy data can hinder the model's ability to generalize, while an appropriate amount of noise data can help avoid overfitting. When working with noisy data, it is important to assess the situation and handle noise in data by data cleaning, outlier detection, and smoothing.

Feature extraction

This study selected 11 features, including brightness, contrast, saturation, color distance etc., and performed feature extraction on the dataset [22-25]. Some of the formulas used are as follows:

$$\text{brightness} = \frac{R+G+B}{3} \quad (1)$$

Relative brightness:

$$l1 = 0.2126 * \text{R} + 0.7152 * \text{G} + 0.0722 * \text{B} \quad (2)$$

$$l2 = 0.2126 + 0.7152 + 0.0722 \quad (3)$$

$$\text{contrast} = \frac{l1 + 0.05}{l2 + 0.05} \quad (4)$$

$$\text{saturation} = \frac{X}{R+B+G} \quad (5)$$

X is the maximum value among R, G, and B.

$$\text{color temperature} = \frac{R}{B}(B \neq 0) \quad (6)$$

$$\text{color distance} = \sqrt{(1-R)^2 + (1-G)^2 + (1-B)^2} \quad (7)$$

Labeling

The data were labeled based on the detection outcomes corresponding to the input data using Arabic numerals to ensure compliance with the required data format for machine learning model training. The correlation between the label numbers and the respective detection outcomes was recorded for future reference.

Training, validation, and testing sets

This study employed a random sampling approach to allocate the data into training, validation, and testing sets. Specifically, 80% of the data are designated as training samples, 10% as validation samples, and the remaining 10% as test samples.

Results

Assessment methods

Seven metrics were calculated using the validation set to assess the models' performance:

• Area Under the Curve (AUC): The accuracy of the classifier in multi-class classification issues was indicated by the area under the ROC curve. Higher AUC values indicate greater classifier performance [26].

• Accuracy: The ratio of the number of samples correctly predicted by the model to the total number of samples. The accuracy was calculated as follows [27]:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

The number of positive examples that the model correctly predicted is represented by TP (True Positive), the number of negative examples that the model correctly predicted is represented by TN (True Negative), the number of negative examples that the model incorrectly predicted as positive examples is represented by FP (False Positive), and the number of positive examples that the model incorrectly predicted as negative examples is represented by FN (False Negative).

• Mean Squared Error (MSE): The average of the squared differences between the model's predictions and the true values [28].

• Confusion Matrix: A matrix used to evaluate the accuracy of a classification model, containing the metrics of TP, TN, FP, and FN [29].

• Recall Rate (R): The proportion of positive examples correctly predicted by the model among all positive examples. The Recall Rate was calculated as follows [30]:

$$R = \frac{TP}{TP+FN} \quad (9)$$

• Precision Rate (P): The proportion of true positive examples among all samples predicted as positive by the model. The Precision Rate was calculated as follows [31]:

$$P = \frac{TP}{TP+FP} \quad (10)$$

• F1 Score: The harmonic average of the recall rate R and the precision rate P, providing a weighted average that combines the accuracy of the model. The F1 Score was calculated as follows [32, 33]:

Table 1. Results of different Machine Learning algorithms

	AUC	Accuracy	Recall	Precision	F1	MSE
SVM	0.91	0.82	0.38	0.86	0.37	0.83
KNN	0.87	0.80	0.56	0.72	0.57	0.65
Decision Tree	0.82	0.87	0.66	0.73	0.68	0.47
Random Forest	0.92	0.89	0.55	0.80	0.63	0.22
Naive Bayes	0.86	0.81	0.53	0.72	0.56	1.36

Note: SVM, Support Vector Machines; KNN, K-Nearest Neighbors; AUC, Area Under the Curve; MSE, Mean Squared Error.

$$F1 = \frac{2 \times P \times R}{P + R} \quad (11)$$

Discussion

Comparative Analysis of Results

The performance of the five Machine Learning algorithms employed in this study are compared based on the results presented in **Table 1**.

• AUC: The Random Forest and SVM models exhibit superior AUC scores of 0.92 and 0.91, respectively, outperforming the KNN and Gaussian Naive Bayes models, which scored 0.87 and 0.86, respectively. The Decision Tree model has the lowest AUC score of 0.82.

• Accuracy Rate: The Random Forest model achieves the highest accuracy of 0.89, followed by the Decision Tree and SVM models with 0.87 and 0.82 accuracy, respectively. The Gaussian Naive Bayes and KNN models have lower accuracy rates, both around 0.80.

• Recall Rate: The Decision Tree model has the highest recall rate of 0.66, followed by the KNN and Random Forest models with 0.56 and 0.55, respectively. The Gaussian Naive Bayes and SVM models have recall rates of 0.53 and 0.38, respectively.

• Precision Rate: The SVM model attains the highest precision score of 0.86, followed by the Decision Tree and Random Forest models with 0.80 and 0.73, respectively. The Gaussian Naive Bayes and KNN models have relatively lower precision scores, both around 0.72.

• F1 Score: The Decision Tree and Random Forest models demonstrate better F1 scores of 0.68 and 0.63, respectively, while the KNN and Gaussian Naive Bayes models have lower F1 scores of 0.57 and 0.56, respectively. The SVM model has the lowest F1 score of 0.37.

• MSE: The Random Forest model has the lowest MSE of 0.22, indicating the best performance. The Decision Tree and KNN models have MSEs of 0.47 and 0.65, respectively, while the SVM and Gaussian Naive Bayes models have higher MSEs of 0.83 and 1.36, respectively.

Error analysis of the test results

The Random Forest model exhibited the strongest performance on the validation set, achieving high accuracy and reliability in detecting non-formed components in urine samples. This success can be attributed to the model's ability to handle complex relationships within the data and reduce variance through ensemble learning techniques [34].

The SVM model demonstrated good precision and AUC scores but had relatively lower recall and accuracy. This could be due to the sensitivity of SVM performance to the choice of kernel function and the optimization of hyperparameters, which can be challenging when dealing with nonlinear data [35].

The KNN model performed comparably to the SVM model, but offered a slight advantage in terms of recall and F1 Score. The number of neighbors and the distance measure used in the KNN algorithm can significantly influence its performance, depending on the characteristics of the dataset [36].

The Decision Tree model outperformed the Random Forest and SVM models in terms of recall and F1 Score, while also demonstrating favorable accuracy and MSE. The performance of the Decision Tree model can be influenced by factors such as feature selection and the presence of noise in the dataset [37].

The Gaussian Naive Bayes model showed the poorest performance, particularly in terms of recall and precision. This can be attributed to the model's underlying assumption of feature independence, which may not hold true for the complex relationships present in the urine data [38].

Conclusions

To summarize the findings, in conjunction with the insights from **Figure 2**, the following conclusions can be drawn:

• The Random Forest model demonstrated distinct advantages in terms of prediction performance and generalization capacity. It achieved the best results across several key metrics, including AUC, accuracy, and MSE, indicating its strong overall predictive ability.

• The Decision Tree model exhibited strong performance in terms of recall rate and F1 score, suggesting that it is particularly effective at capturing and leveraging important information within the data.

• The SVM model achieved the highest precision among the tested algorithms. However, it performed relatively poorly in terms of recall and F1 score, indicating that further parameter tuning and optimization may be necessary to enhance its overall effectiveness.

• While the Gaussian Naive Bayes and KNN models did not demonstrate exceptional performance, they exhibited relatively steady and consistent results across a range of evaluation metrics.

• These observed differences in model performance can be primarily attributed to the varying adaptability of the different algorithms to the specific characteristics and complexity of the data.

Author Contributions: Yifei Bai performed data acquisition, processing, analysis and model building and wrote the manuscript. Yifei Bai, Rongguo Yan, Yuqing Yang, and Chengang Mao made substantial contributions to data analysis and manuscript preparation.

References

[1] Yan RG, Wei RF, Shen JK, et al. Research progress on color determination methods of reagent strips in routine urinalysis. Int J Biomed Eng 2019;42:532-536.

[2] Ryan D, Robards K, Prenzler PD, et al. Recent and potential developments in the analysis of urine: a review. Anal Chim Acta 2011;684(1-2):8-20.

[3] Shehab M, Abualigah L, Shambour Q, et al. Machine learning in medical applications: A review of state-of-the-art methods. Comput Biol Med 2022;145:105458.

[4] Wang JW. Design of a portable urine analyzer. Chongqing University 2020.

[5] Wang F. Research and design of urine analyzer based on color sensor. Jilin University 2014.

[6] Liu G, Hu N, Ma Z, et al. A portable analyzer based on a novel optical structure for urine dry-chemistry analysis. J Instrum 2018;13(07):T07002.

[7] Cui LS. Research on intelligent analysis system for home urine detection. Southeast University 2022.

[8] Li XY. Clinical value of combined use of urine dry chemistry test method and urine sediment test method for routine urine testing. Chin Mod Drug Appl 2019;13(15):78-80.

[9] Li YY, Yan GR, Qiu XY, et al. Application of FIA algorithm in urine component detection. Software 2018;39(06):40-45.

[10] Liu XN, Wang K, Wang CD, et al. Research on identification of college students with difficulties based on KNN algorithm. Electron. Compon Inf Technol 2023;7(08):88-92.

[11] Jawthari M, Stoffova V. Predicting students' academic performance using a modified kNN algorithm. Pollack Periodica 2021;16(3):20-26.

[12] Ju MX, Zhou YX, Tang WN, et al. CART Decision Tree power consumption anomaly detection method based on data imbalance. Jilin Electr Power 2024;52(01):53-56.

[13] Li ZT. Development history and application of Support Vector Machines. Inf Syst Eng 2024;(03):124-126.

[14] Xu JH. An extended one-versus-rest support vector machine for multi-label classification. Neurocomputing 2011;74(17):3114-3124.

[15] Guo RN. Ceramic raw material component identification based on Support Vector Machine. Sci Technol Innovation 2024;(02):32-35.

[16] Guo ZY, Lao ZK, Chen YF, et al. Application of Random Forest algorithm in prediction of droplet diameter in electrospray printing. Mech Des Manuf 2024;(03):153-157.

[17] Liu HX, Li X, Ding JG, et al. Strip thickness prediction based on Random Forest feature selection and genetic algorithm optimization. Steel Rolling 1-16.

[18] Tin Kam H, Hull JJ, Srihari SN. Decision combination in multiple classifier systems. IEEE Trans Pattern Anal Mach Intell 1994;16(1):66-

75.

[19] Wang SH, Gao R, Du RJ. Dependency extension of naive Bayesian classifiers based on Gaussian kernel function. Control Decis 2015;30(12):2280-2284.

[20] Bao QQ, Sun CR. An improved naive Bayes method applied to intelligent triage. Mod Hosp 2024;24(03):424-427.

[21] Li JF, Zhang JY, Li LK, et al. Discussion on processing methods of missing values in clinical trials. J Peking Union Med Coll 1-15.

[22] Shi KW. Neural basis of color brightness at different time frequencies. Shanghai University of Sport 2023.

[23] Chen LP, Zhou H, Zhang NY, et al. Color image segmentation method based on improved convex hull and color contrast. J Hebei Univ, Nat Sci Ed 2018;38(05):543-548.

[24] Ibraheem NA, Hasan MM, Khan RZ, et al. Understanding Color Models: A Review. ARPN J Sci Technol 2012.

[25] Guo JJ, Li QW, Cheng HL, et al. Tree image segmentation algorithm based on Lab color distance and GMM. Inf Technol 2016;(02):1-4+9.

[26] Liu BW, Bu Y, Zou DH, et al. An improved tooth color classification model based on residual network. Software Eng 2024;27(03):52-57.

[27] Liu JQ, Zhong SH, Li SZ, et al. patite source rock identification method based on Machine Learning. Geotectonics Miner 1-14.

[28] Liu AL. Research on obesity level estimation based on Machine Learning. Shaoxing University 2023.

[29] Jin M. Research on pig behavior identification and classification methods based on Machine Learning. Shaoxing University of Arts and Sciences 2021.

[30] Zhang DS. Image classification algorithm based on parameterized quantum circuits and Machine Learning. Nanchang University 2023.

[31] Hu T, Wu MT, Hu W, et al. Classification prediction of fly ash activity based on Machine Learning. J. Cent South Univ (Sci Technol) 2023;54(10):3829-3839.

[32] Huang KL. Research on voice fraud-related behavior based on Machine Learning. University of International Business and Economics 2017.

[33] Wang ZG, Zhang HY, Miao DQ. Automatic selection method of non-maximum suppression threshold based on F1 valu. J Intell Syst 2020;15(05):1006-1012.

[34] Cheng L, Chen XW, De Vos J, et al. Applying a random forest method approach to model travel mode choice behavior. Travel Behav Soc 2019;14:1-10.

[35] Azadeh A, Saberi M, Kazem A, et al. A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. Applied Soft Computing 2013;13(3):1478-1485.

[36] Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. Big Data 2019;7(4):221-248.

[37] Galathiya A, Ganatra A, Bhensdadia C. Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning. Int J Comput Sci Inf Technol 2012;3(2):3427-3431.

[38] Khalid H, Khan A, Zahid Khan M, et al. Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease. Comput Intell Neurosci 2023;2023:9266889.

Progress in Medical Devices
ISSN: 2957-5478

Zentime

Research Article https://doi.org/10.61189/568091unpkqk

Construction and comparative analysis of an early screening prediction model for fatty liver in elderly patients based on machine learning

Xiaolei Cai[1*], Qi Sun[2*], Cen Qiu[2*], Zhenyu Xie[1], Jiahao He[2], Mengting Tu[3], Xinran Zhang[2], Yang Liu[2], Zhaojun Tan[2], Yutong Xie[2], Xixuan He[1], Yujing Ren[1], Chunhong Xue[1], Siqi Wang[2], Linrong Yuan[2], Miao Yu[2], Xuelin Cheng[4], Xiaopan Li[4], Sunfang Jiang[4], Huirong Zhu[1]

[1]Tangqiao Community Health Service Center, Shanghai 200127, China. [2]Shanghai University of Medicine and Health Sciences, Shanghai 201318, China. [3]Shanghai DianJi University, Shanghai 201306, China. [4]Health Management Center, Zhongshan Hospital Affiliated to Fudan University, Shanghai 200032, China.
*The authors contribute equally.

Corresponding authors: Sunfang Jiang and Huirong Zhu.

Acknowledgement: None.

Declaration of conflict of interest: None.

Declaration of Ethics: This article does not contain any studies with human participants or animals performed by any of the authors. Ethical approval for this study was obtained from Tangqiao Community Health Service Center.

Received May 11, 2024; Accepted July 16, 2024; Published September 30, 2024

Highlights

- This study collected three years of physical examination data from older adults in the Tangqiao community of Shanghai, which is more regionally representative.
- The most suitable model for this study was selected from six machine learning models to construct a fatty liver risk prediction model for the elderly.
- This study combines six feature selection algorithms with varying performance to screen the features most rele vant to fatty liver.

Abstract

Objective: To construct a prediction model for fatty liver disease (FLD) among elderly residents in community using machine learning (ML) algorithms and evaluate its effectiveness. **Methods:** The physical examination data of 4989 elderly people (aged over 60 years) in a street of Shanghai from 2019 to 2023 were collected. The subjects were divided into a training set and a testing set in a 7:3 ratio. Using feature selection and importance sorting methods, eight indicators were selected, including high-density lipoprotein cholesterol, body mass index, uric acid, triglycerides, albumin, red blood cell, white blood cell, and alanine aminotransferase. Six ML models, including Categorical Features Gradient Boosting, eXtreme Gradient Boosting, Light Gradient Boosting Machine, Random Forest, Decision Tree, and Logistic Regression, were constricted, and their predictive performances were compared via accuracy, precision, recall, F1 score, and Area Under Receiver Operating Characteristic Curve. **Results:** Among the six ML models, the Categorical Features Gradient Boosting model demonstrated the highest prediction accuracy of 0.74 for FLD in elderly community population, along with a precision of 0.70, a recall of 0.73, a F1 score of 0.71, and an area under the curve of 0.74. **Conclusions:** In the context of rapid development of artificial intelligence, a community-based elderly FLD prediction model constructed using ML algorithms aid family general practitioners in the early diagnosis, early treatment, and health management of local FLD patients.

Keywords: Fatty liver, machine learning models, disease screening, health management, community diagnosis

Address correspondence to: Sunfang Jiang, Health Management Center, Zhongshan Hospital Affiliated to Fudan University, Gate 5 East Campus, No. 179 Fenglin Road, Xuhui District, Shanghai 200032, China. Email: jiang.sunfang@zs-hospital.sh.cn. **Huirong Zhu,** Tangqiao Community Health Service Center, No.131 Pujian Road, Pudong New District, Shanghai 200127, China. Email: rachel1022@126.com.

Introduction

Fatty Liver Disease (FLD) is a globally prevalent disease. With changes in lifestyle and population demographics, as well as factors such as obesity and excessive alcohol consumption, FLD has become the leading chronic liver disease in China, and its prevalence is on the rise, posing a significant threat to people's health, particularly that of the elderly [1]. FLD can be diagnosed through liver biopsy or imaging techniques. Liver biopsy is the gold standard for diagnosing steatosis but associated with invasiveness and bleeding risk [2]. Meanwhile, imaging methods such as computed tomography and magnetic resonance imaging are time-consuming and expensive and are not always available in remote areas. Early diagnosis of FLD based on risk factors helps clinicians identify adverse events related to FLD and take more lifestyle intervention measures for its prevention [3]. Machine learning (ML) is a promising technique for medical data analysis and disease prediction, making significant contributions to the field of healthcare. For instance, Jiang et al. used ML to comprehensively describe the overall decline in intrinsic capabilities among elderly people in communities [4]; Islam et al. developed an efficient, interpretable ML framework for predicting the risk of hypertension in Ethiopian patients [5]; Abnoosian et al. developed a novel ML framework to enhance the performance of diabetes prediction models and address various challenges [6]. It is evident that ML is playing an increasingly important role in various fields, particularly in healthcare. By analyzing big data, building models, and making automated decisions, ML offers more efficient and accurate solutions for the healthcare industry. This study aims to propose a ML model to predict FLD, helping doctors to preliminarily screen high-risk populations and make early predictions for FLD.

Materials and methods

Data preparation

The overall workflow of this study is shown in **Figure 1**. The dataset used was provided by a community health service center on a street in Shanghai, containing health examination data for the elderly over three years (2019, 2022, 2023), with a total of 4,989 entries. All individuals were aged 60 and above, and the data was anonymized. The dataset includes 24 variables such as gender, blood biochemical indicators, and age, with the distribution of 23 continuous variables shown in **Figure 2**. To preserve the authenticity of the data, 1,144 entries with missing values were removed. In this study, FLD diagnosis was based on ultrasound results. Ultimately, out of the 3,845 participants in the study, 2,074 were potentially diagnosed with FLD through ultrasound.

Statistical analysis

Data analysis was conducted using Python 3.11.4 software. The continuous variables were expressed as mean ± SD, and the comparison between the FLD group and non-FLD group was conducted using an independent sample t-test. A p-value less than 0.05 were considered statistically significant.

Data processing

In ML, most algorithms, such as Logistic regression (LR), Support Vector Machines, and K-nearest neighbors, can only process numerical data and are unable to handle text. In such cases, to make data compatible with algorithms and libraries, encoding is performed, which involves converting textual data into numerical data.

In this study, the dataset consisted of 25 columns, with data types including only numerical and categorical variables. For numerical variables, since the collected data came with units, these units were removed, and the data type was converted to float. For categorical variables, one hot encoding was used for gender, transforming the feature into dummy variables, with "female" marked as 1 and "male" marked as 0; for ultrasound results, label encoding was used, with "fatty liver" marked as 1 and "non-fatty liver" marked as 0.

Feature selection

The dataset included 24 clinical features: age, sex, body mass index (BMI), white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb), serum creatinine (SCR), mean corpuscular volume (MCV), platelets (PLT), triglycerides (TG), mean corpuscular hemoglobin concentration (MCHC), direct bilirubin (DBIL), aspartate aminotransferase (AST), alanine aminotransferase (ALT), AST/ALT, albumin-globulin ratio (A/G), total cholesterol (TC), uric acid (UA), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), albumin (ALB), globulin (GLB), total protein (TP), and total bilirubin (TBIL). Feature selection aims to remove irrelevant or redundant features while retaining other original features to obtain a subset that better describes the given problem with minimal performance loss. This study employed

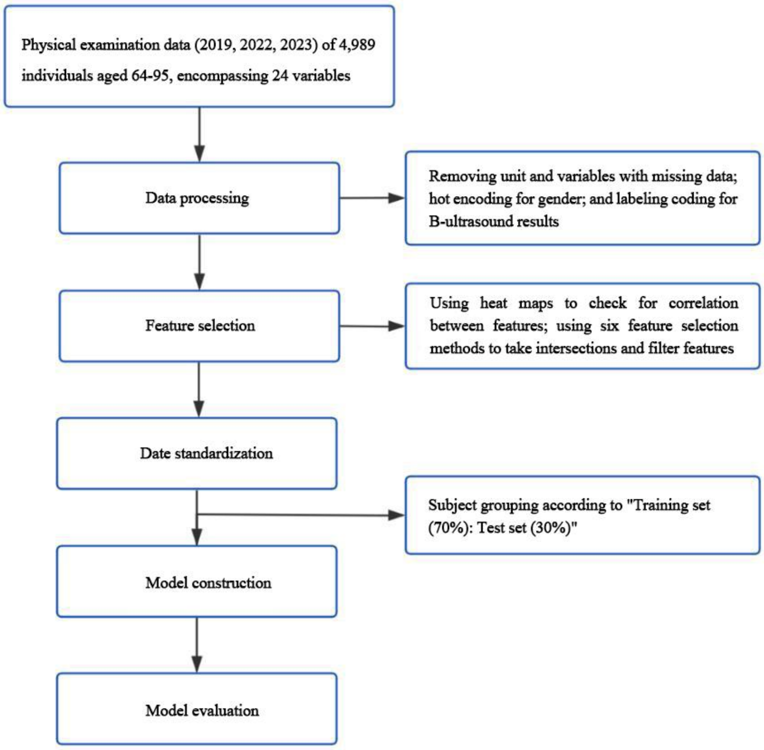


Figure 1. Workflow of the ML-based FLD prediction method proposed in this study. ML, machine learning; FLD, fatty liver disease.

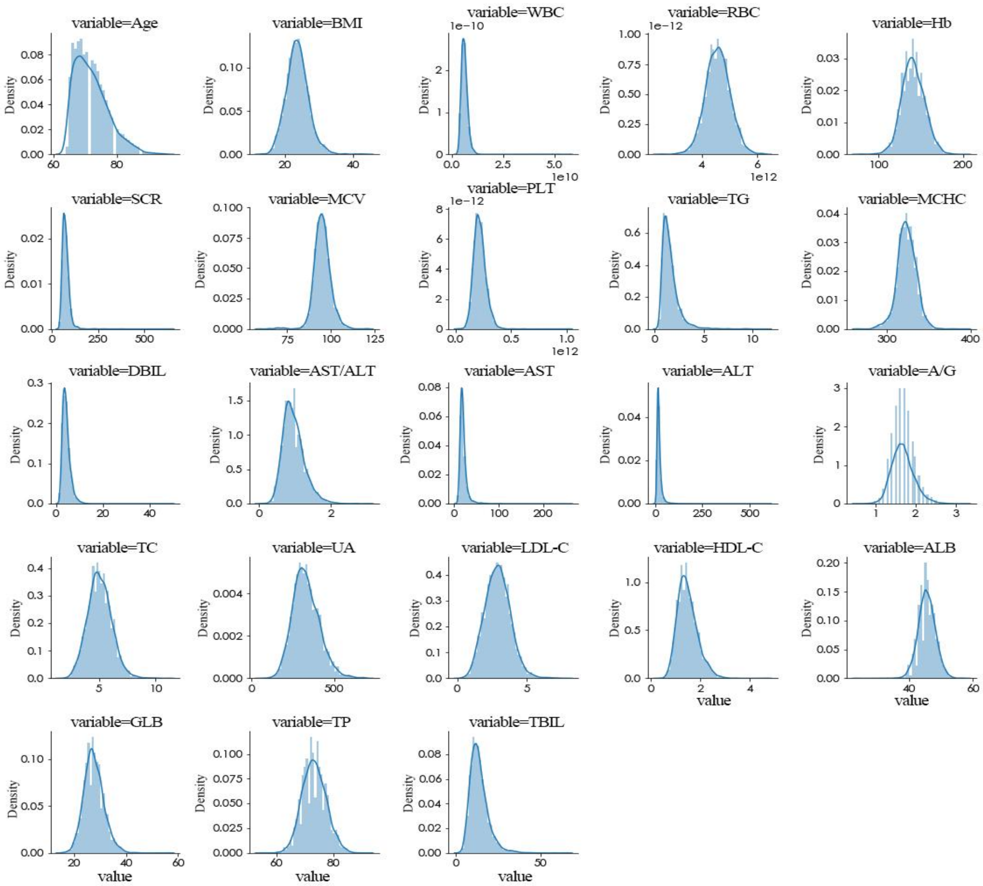


Figure 2. Distribution graphs of various continuous variables. BMI, body mass index; WBC, white blood cell; RBC, red blood cell; Hb, hemoglobin; SCR, serum creatinine; MCV, mean corpuscular volume; PLT, platelets; TG, triglycerides; MCHC, mean corpuscular hemoglobin concentration; DBIL, direct bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; A/G, albumin-globulin ratio; TC, total cholesterol; UA, uric acid; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; ALB, albumin; GLB, globulins; TP, total protein; TBIL, total bilirubin.

six typical feature selection methods: Variance Threshold, Correlation Coefficient, Chi-square test, Mutual Information, Tree-based Feature Selection, and Recursive Feature Elimination.

Variance Threshold selects features based on their variance, by default removing all zero-variance features. The Correlation Coefficient method (Pearson correlation coefficient) calculates the correlation coefficient between features and the target variable, selecting features with a significant correlation with the target. The Chi-Square test evaluates the correlation between features and the target variable based on the chi-square statistic, which can reduce dimensionality and improve model performance. Mutual Information selects features with high mutual information, suitable for both continuous and discrete variables. Tree-based Feature Selection considers interactions among features and can handle nonlinear features and high-dimensional data, appropriate for datasets with complex relationships among features. Recursive Feature Elimination progressively removes features based on model coefficients or feature importance, selecting the most significant features. After comparing six feature subsets, this study chose the eight most commonly used features in clinical practice to construct the prediction model.

Model construction and validation

All processing steps were performed in Python 3.11.4. After feature selection, data normalization was conducted using the StandardScaler from the sklearn.preprocessing module. The selected 8 features were used as independent variables, with the FLD label column as the dependent variable y. The dataset was split into training and testing sets in a 7:3 ratio using train test split, followed by the construction of ML models. This study primarily built six widely used ML models and optimized their performance. For the Categorical Features Gradient Boosting (CatBoost), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) models, the study employed RandomizedSearchCV for random search, which selects a subset of parameter combinations from a preset parameter space and a preset number of trials for training and validation [7-9]. This method returns the best-performing parameter combination from the trials, significantly reducing search time and computational resources [10].

For the Decision Tree (DT), Random Forest (RF), and LR models, the study used GridSearchCV for parameter tuning. GridSearchCV is a hyperparameter optimization tool that automatically adjusts parameters of ML models. It performs grid search by accepting a dictionary of all possible parameter combinations passed to the 'param_grid' parameter. It also requires an integer to be passed to the 'cv' parameter, indicating the number of cross-validation iterations [11].

Performance metrics

The performance of the six ML models used was evaluated using five widely used metrics: accuracy, precision, recall, F1-score, and the area under the curve (AUC). True Positive (TP) indicates that the patient has FLD and the algorithm correctly predicts it as FLD. True Negative (TN) indicates that the patient does not have FLD and the algorithm correctly predicts it as non-FLD. False Negative (FN) indicates that the patient has FLD, but the algorithm predicts it as non-FLD, mistakenly predicting an actual positive as negative. False Positive (FP) indicates that the patient does not have FLD, but the algorithm predicts it as FLD, mistakenly predicting an actual negative as positive.

Accuracy: The proportion of correctly predicted samples out of the total number of samples.

Accuracy =(TP+TN)/(TP+FP+TN+FN)*100% (1)

Precision: The proportion of true positive samples among all samples predicted as positive (TP and FP), reflecting the model's ability to identify positive samples.

Precision =TP/(TP+FP)*100% (2)

Recall: The proportion of true positive samples among all actual positive samples (TP and FN), reflecting the model's ability to detect positive samples.

Recall =TP/(TP+FN)*100% (3)

F1-score: The harmonic mean of precision and recall.

F1-score =2TP/(2TP+FP+FN)*100% (4)

Receiver Operating Characteristic Curve (ROC): the area under the ROC is an important metric for assessing model performance. The ROC curve plots the True Positive Rate on the y-axis and the False Positive Rate on the x-axis, showing the model's performance at different thresholds. The AUC is obtained by calculating the area enclosed by the ROC curve, the x-axis, and the False Positive Rate =1. The closer the

Table 1. Feature selection outcomes of six feature selection methods

	Variance threshold	Correlation coefficient	Chi-square test	Mutual information	Tree-based feature selection	Recursive feature elimination
Age	√	√	√	√		
Sex	√			√		
BMI	√	√	√	√	√	√
WBC	√	√	√	√	√	√
RBC	√	√	√	√	√	√
Hb	√	√	√	√	√	
SCR	√	√	√			√
MCV	√			√		√
PLT	√	√	√	√		√
TG	√	√	√	√	√	√
MCHC	√			√		√
DBIL	√					
AST/ALT		√	√	√	√	√
AST	√	√	√		√	
ALT	√	√	√	√	√	√
A/G						
TC	√					
UA	√	√	√	√	√	√
LDL-C	√					√
HDL-C	√	√	√	√	√	√
ALB	√	√	√	√	√	√
GLB	√			√		√
TP	√	√	√		√	
TBIL	√	√	√			√

Note: BMI, body mass index; WBC, white blood cell; RBC, red blood cell; Hb, hemoglobin; SCR, serum creatinine; MCV, mean corpuscular volume; PLT, platelets; TG, triglycerides; MCHC, mean corpuscular hemoglobin concentration; DBIL, direct bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; A/G, albumin-globulin ratio; TC, total cholesterol; UA, uric acid; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; ALB, albumin; GLB, globulins; TP, total protein; TBIL, total bilirubin.

Table 2. Evaluation metrics for six models

ML model	Accuracy	Precision	Recall	F1-score	AUC
RF	0.73	0.68	0.73	0.70	0.73
DT	0.70	0.68	0.60	0.64	0.69
LR	0.74	0.70	0.71	0.70	0.73
CatBoost	0.74	0.70	0.73	0.71	0.74
XGBoost	0.73	0.67	0.72	0.70	0.72
LightGBM	0.73	0.68	0.74	0.71	0.73

Note: ML, machine learning; AUC, area under the curve; RF, random forest; DT, decision tree; LR, logistic regression; CatBoost, categorical features gradient boosting; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine.

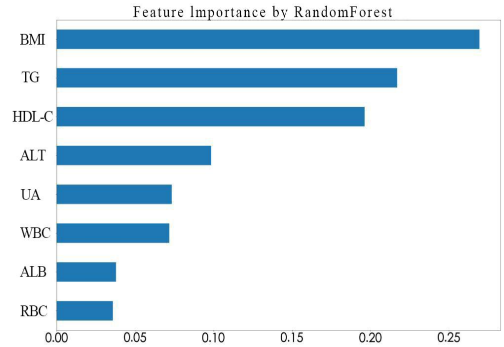


Figure 3. Feature importance ranking based on RF. RF, Random Forest; BMI, body mass index; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; ALT, alanine aminotransferase; UA, uric acid; WBC, white blood cell; ALB, albumin; RBC, red blood cell.

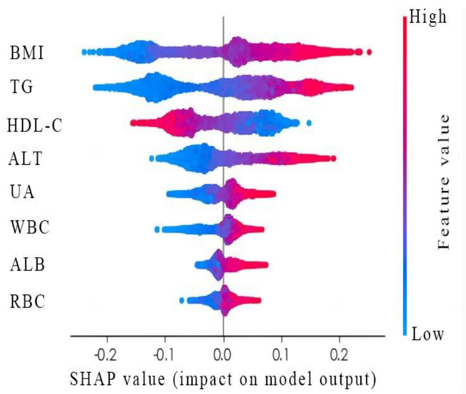


Figure 4. SHAP summary plot. SHAP, shapley additive explanations; BMI, body mass index; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; ALT, alanine aminotransferase; UA, uric acid; WBC, white blood cell; ALB, albumin; RBC, red blood cell.

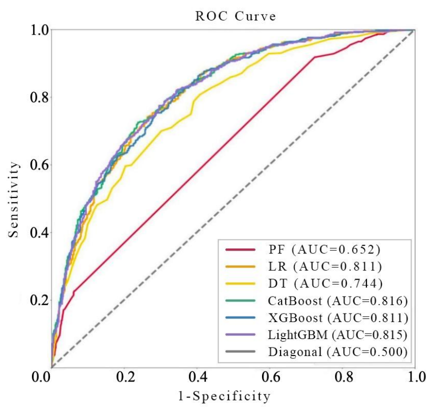


Figure 5. ROC curves of six models. ROC, receiver operating characteristic; RF, random forest; LR, logistic regression; DT, decision tree; CatBoost, categorical features gradient boosting; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine.cell.

AUC is to 1, the better the model's classification performance.

Results

Selected features

Table 1 contains six types of feature selection algorithms, each with its unique selection mechanism and evaluation criteria. It details the specific circumstances under which these feature selection algorithms were used to select features. Ultimately, these selected features were used to construct an FLD risk assessment model, ensuring that the model could perform accurate risk assessments based on the most relevant and predictive features.

By taking the intersection of the six feature selection methods, the eight most meaningful features were identified: HDL-C, BMI, UA, TG, ALB, RBC, WBC, and ALT. Their importance was ranked (**Figure 3**), which illustrates the contribution of each data feature.

Shapley additive explanations (SHAP) is a Python package developed to explain the output of any ML model. All features are considered 'contributors' [12]. The 'summary plot' visualizes the SHAP values for each feature of each sample, providing an intuitive understanding of the overall pattern and helping to identify potential prediction outliers. Each row represents a feature, with the x-axis showing the SHAP values. Each point represents a sample, where the color indicates the feature value (red for high, blue for low), and purple near the mean. The wider the color area, the greater the impact of the feature. As shown in **Figure 4**, higher BMI values increased the predicted risk for FLD, whereas higher HDL-C values decreased the predicted risk for FLD.

Results of performance metrics across different models

Table 2 shows the performance scores for five metrics across six ML models. LR and CatBoost were comparable in terms of accuracy and precision, but CatBoost performed better in the remaining three metrics. Overall, the CatBoost model demonstrated the best performance.

The ROC curves for the six models are shown in **Figure 5**. CatBoost exhibited the highest AUC value at 0.816, followed closely by LightGBM at 0.815. XGBoost and LR were comparable, both with an AUC of 0.811, while RF and DT performed less favorably compared to the other models.

Discussion

The rising incidence of FLD, a prevalent liver disease, has attracted widespread attention from the medical community and society [13, 14]. Early identification and prediction of fatty liver risk is crucial for developing effective prevention and treatment strategies. In recent years, with the rapid development of big data and ML technologies, health risk assessment using these advanced technologies has become a new trend in research [15, 16]. In particular, by mining and analysing key indicators in physical examination data, it can provide a scientific basis for the prediction of fatty liver.

Focusing on the old three-year physical examination data in Tangqiao community, this study aims to predict the risk of fatty liver through ML algorithms, especially CatBoost model.

This study identified eight features through six feature selection methods: BMI, HDL-C, UA, TG, ALB, RBC, WBC, and ALT. Previous studies have highlighted various clinical features associated with FLD. For instance, age, BMI, waist circumference, ALT, fasting glucose, UA, and PLT have been reported as related to FLD [17]. Weng et al. developed an ML model to screen for high-risk populations of FLD, identifying some meaningful clinical features: BMI, ALB, ALT, GB, HDL-C, LDL-C, and TG [18]. Chen et al. used a large dataset and the XGBoost model to investigate the best overall predictive ability for FLD in the diagnosed population, additionally identifying features such as BMI, waist circumference, TG, serum γ-glutamyl transferase, glucose, age, creatinine, gender, and LDL-C as the significant factors [19]. Pei et al. conducted feature importance analysis based on the ranking of each variable, determining that serum UA, BMI, and TC are the three most likely factors, followed by HDL-C and Hb [20]. Su et al. chose waist circumference, BMI, blood pressure, WBC, Hb, PLT, glucose, serum biochemical parameters, and lipids [3]. Guarneros et al., by applying ML techniques, identified the main risk factors for FLD in four clinical datasets of liver disease patients: ALT, alkaline phosphatase, AST, alpha-fetoprotein, and γ-glutamyl transferase [21]. Hence, the features selected in this study hold significant relevance in clinical research on FLD.

CatBoost is a gradient boosting DT-based ML algorithm designed to enhance model performance through ensemble learning and feature handling [22]. Known for its robustness and generalization capabilities, the CatBoost model can automatically adjust learning rates and iteration numbers during training to prevent overfitting. Additionally, CatBoost incorporates randomization strategies to enhance model diversity, thereby improving generalization performance. This study concludes that the CatBoost model has the best overall predictive ability in the preliminary screening of FLD populations.

In the early detection of FLD, the CatBoost algorithm can be used to construct a predictive model to identify potential FLD cases. Through training on a large FLD dataset, the CatBoost model learns to predict the presence of FLD in new samples with high accuracy and sensitivity, effectively detecting early stages of the disease. The algorithm and model obtained in this study offer technical tools for community health management, particularly benefiting residents who may not routinely undergo auxiliary examinations such as ultrasound but receive blood tests during clinical visits. The community health service center can leverage this algorithm and model to intelligently assess the health of these residents, offering early warnings, predictions, and health recommendations for FLD. This supports early diagnosis and treatment by prompting at-risk residents to seek further evaluation and appropriate care.

To improve the model developed in this study, future research should collect more detailed information about patients and further refine the survey factors when conducting statistical analysis of the examined population. Parameters such as waist circumference and γ-glutamyl transferase can also be considered as features to further enhance such ML models by providing more information about the likelihood of disease occurrence and the degree of steatosis and fibrosis.

This study also has certain limitations. First, the patient grouping was based on abdominal ultrasound diagnostic results, which have a lower level of evidence compared to liver biopsy and MRI. This may affect the accuracy of the predictions. Second, the data came from only one street community, and a multicenter dataset and external validation could provide better and more reliable performance. Additionally, this study did not include some clinical parameters, such as waist circumference and γ-glutamyl transferase, which have been proven to be related to FLD. Finally, the dataset lacked patients' clinical information. The presence of diseases such as diabetes, hepatitis B, hepatitis C, and the medications used may affect the predictions of the ML model.

Conclusion

In this study, eight features closely related to fatty liver risk were successfully screened by CatBoost algorithm and six feature selection methods. These features are not only widely recognised in clinical studies, but their validity in fatty liver prediction is further demonstrated by the validation of this study. Therefore, this study is innovative and practical in the field of fatty liver risk prediction, and provides a scientific basis for the prevention and intervention of fatty liver through the combined use of advanced ML algorithms and abundant clinical data.

Supplementary Material: None.

Availability of data and material: None.

Author Contributions: Xiaolei Cai, Qi Sun, Cen Qiu (Lead Author): Led the overall conception and design of the study. Conducted the literature review and synthesized relevant findings to inform the research questions and methodology. Drafted the initial manuscript and incorporated feedback from co-authors to refine the final version.

Qi Sun, Cen Qiu, Jiahao He, Yutong Xie, Siqi Wang, Miao Yu (Data Analyst): Contributed significantly to the data preprocessing and feature selection phases of the study. Developed and implemented the feature selection methods to identify the eight key features for the CatBoost model. Contributed to the writing of the Methods and Results sections, focusing on the technical aspects of the study.

Xiaolei Cai, Sunfang Jiang, Huirong Zhu, Zhenyu Xie, Mengting Tu, Xinran Zhang, Yang Liu (Clinical Expert): Provided critical clinical input throughout the study, ensuring the research questions and methodology were grounded in clinical practice. Reviewed and interpreted the results from a clinical perspective, highlighting their relevance and implications for public health and clinical practice. Contributed to the Discussion section, drawing connections between the findings and existing clinical literature.

Yujing Ren, Chunhong Xue, Xuelin Cheng, Xiaopan Li, Zhaojun Tan, Linrong Yuan (Statistical Advisor): Offered guidance on statistical analysis and modeling techniques. Ensured the rigorous application of machine learning methods, particularly the CatBoost algorithm. Assisted in the interpretation of the results and provided recommendations for enhancing the robustness and reproducibility of the findings. Contributed to the Methods section, emphasizing the statistical rigor of the study.

References

[1] Teng ML, Ng CH, Huang DQ, et al. Global incidence and prevalence of nonalcoholic fatty liver disease. Clin Mol Hepatol 2023;29(Suppl):S32-S42.
[2] Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67(1):328-357.
[3] Su PY, Chen YY, Lin CY, et al. Comparison of Machine Learning Models and the Fatty Liver Index in Predicting Lean Fatty Liver. Diagnostics (Basel) 2023;13(8):1407.
[4] Jiang X, Chen F, Yang X, et al. Effects of personal and health characteristics on the intrinsic capacity of older adults in the community: a cross-sectional study using the healthy aging framework. BMC Geriatr 2023;23(1):643.
[5] Islam MM, Alam MJ, Maniruzzaman M, et al. Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia. PLoS One 2023;18(8):e0289613.
[6] Abnoosian K, Farnoosh R, Behzadi MH. Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. BMC Bioinformatics 2023;24(1):337.
[7] Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. arXiv 2018.
[8] Yang G. Generalized XGBoost Method. arXiv 2022.
[9] Alafate J, Freund Y. Faster Boosting with Smaller Memory. arXiv 2019.
[10] Anghel A, Papandreou N, Parnell T, et al. Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms. arXiv 2019.
[11] Ahmad GN, Fatima H, Ullah S, et al. Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV. IEEE Access 2022;10:80151-80173.
[12] Amoukou SI, Brunel NJB, Salaün T. Accurate Shapley Values for explaining tree-based models. arXiv 2023.
[13] Liu K, Chen Y, Chen J, et al. Genetically determined circulating micronutrients and the risk of nonalcoholic fatty liver disease. Sci Rep 2024;14(1):1105.
[14] Younossi ZM, Golabi P, Paik JM, et al. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. Hepatology 2023;77(4):1335-1347.

[15] Zare A, Shafaei Bajestani N, Khandehroo M. Machine Learning in Public Health. J Res Health 2024;14(3):207-208.
[16] Anuntakarun S, Khamjerm J, Tangkijvanich P, et al. Classification of Long Non-Coding RNAs s Between Early and Late Stage of Liver Cancers From Non-coding RNA Profiles Using Machine-Learning Approach. Bioinform Biol Insights 2024;18:11779322241258586.
[17] Gong XY, Zhang LF, Ge R, et al. Analysis of risk factors of metabolic fatty liver disease based on community physical examination data. J Community Med 2023;21(21):1106-1111.
[18] Weng S, Hu D, Chen J, et al. Prediction of Fatty Liver Disease in a Chinese Population Using Machine-Learning Algorithms. Diagnostics (Basel) 2023;13(6):1168.
[19] Chen YY, Lin CY, Yen HH, et al. Machine-Learning Algorithm for Predicting Fatty Liver Disease in a Taiwanese Population. J Pers Med 2022;12(7):1026.
[20] Pei X, Deng Q, Liu Z, et al. Machine Learning Algorithms for Predicting Fatty Liver Disease. Ann Nutr Metab 2021;77(1):38-45.
[21] Guarneros-Nolasco LR, Alor-Hernández G, Prieto-Avalos G, et al. Early Identification of Risk Factors in Non-Alcoholic Fatty Liver Disease (NAFLD) Using Machine Learning. Mathematics 2023;11(13):3026.
[22] Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data 2020;7(1):94.