scientific reports

OPEN

Constructing a predictive model for acute mastitis in lactating women based on machine learning

Liujing Zhu[1], Zuyan Huang[3], Yan Chen[4], Guangqiu Li[1] & Liwen Liu[2]✉

Acute lactational mastitis is a frequently occurring complication for lactating women, exerting a certain degree of influence on their physical condition, breastfeeding, mental health, and daily life. The etiology of this disease is complex, and the early symptoms lack typicality. Delayed diagnosis often occurs, which further progresses to abscess or more severe infections, adversely affecting the therapeutic effect and eventually leading to a prolonged recovery process and triggering other complications. Nevertheless, currently, the research regarding the risk factors of acute mastitis that occurs during lactation remains unfinished. This study employed a retrospective case-control study approach and collected relevant data from 369 patients with acute mastitis and 447 healthy controls. The involved data covered indicators such as age, parity, history of breast surgery, etc. By using machine learning (ML) algorithms (Logistic Regression (LR), Naive Bayes (NB), XGBoost, Multilayer Perceptron (MLP)) to train and validate the above data, it aimed to construct a predictive model of the risk factors for the occurrence of acute mastitis in lactating women, and simultaneously analyzed the other influences and effects of these factors on acute mastitis. The ML model demonstrated high accuracy in differentiating patients with acute mastitis from non-patients. We evaluated twelve indicators, namely age, primiparity, history of breast surgery, cracked, external breast trauma, puerperium, gestational diabetes, C-reactive protein (CRP), procalcitonin (PCT), neutrophils (NE), white blood cells (WBC), and abnormal nipple discharge, to determine their influence on the occurrence of acute mastitis in lactating women. Prediction models were established using four different ML algorithms. Through analysis, when comparing the four distinct ML models on the test set, the MLP model performed optimally across various evaluation metrics, including the highest area under the receiver operating characteristic (ROC) curve (AUROC) (0.898), sensitivity (0.820), test specificity (0.863), and F1 score (0.849), with an accuracy of 0.840. Decision Curve Analysis (DCA) indicates that within the majority of threshold ranges, the MLP can achieve the highest net benefit. Among these twelve indicators, five are significantly related to the occurrence of acute mastitis, namely age, cracked, CRP, NE, and WBC. We have successfully developed a prediction model for acute lactational mastitis and identified five key indicators closely related to its occurrence. This study effectively predicts the occurrence of acute lactational mastitis and provides a reference for the timely implementation of targeted clinical interventions as well as accurate diagnosis and treatment.

Acute lactational mastitis refers to the acute purulent infection of the breast in lactating women[1]. The pathogenesis is mainly bacterial infection, among which the most common bacteria are Staphylococcus aureus and Streptococcus. These bacteria can enter the breast tissue through nipple fissures, duct blockage, milk stasis and other routes[2,3]. The main symptoms of this disease are breast pain, redness, swelling and fever[3]. These symptoms will not only bring severe discomfort to patients, but also disrupt breastfeeding, resulting in insufficient nutritional intake for infants. Recent epidemiological data show that the incidence of acute mastitis in lactating women ranges from 2.2 to 10% and is on a steady upward trend[4].

Although the pathophysiological mechanism of acute mastitis has been relatively clear, its etiology is still rather complex. Its occurrence involves multiple risk factors[5], breastfeeding techniques[6], abnormal breast structure and immune status[7]. Currently, the clinical prevention of acute mastitis mainly depends on health education. In addition, diagnosis and treatment are usually based on clinical manifestations and laboratory

[1]Department of Galactophore, Liuzhou Hospital, Guangzhou Women and Children's Medical Center, Liuzhou, Guangxi, China. [2]Department of Anesthesia, The Reproductive Hospital of Guangxi Zhuang Autonomous Region, Nanning City, China. [3]Department of Galactophore, Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, Guangxi, China. [4]Obstetrical department, Liuzhou Hospital, Guangzhou Women and Children's Medical Center, Liuzhou, Guangxi, China. ✉email: liwenlmd01@163.com

test results[8]. However, due to the non-specific characteristics of the early symptoms of acute mastitis and the difficulty in predicting the onset of this disease, there are often delays in its diagnosis and treatment. Such delays often make early intervention and treatment of the disease more complicated. These delays may lead to the occurrence of complications, for example, the formation of abscesses. These complications will have a negative impact on the treatment effect[9]. Therefore, it is crucial to identify the key risk indicators for the occurrence of acute mastitis in lactating women. This is not only of great significance for effective prevention but also crucial for timely diagnosis and treatment.

In recent years, big data technology and ML technology have made certain progress. The progress of these technologies enables people to conduct more detailed analysis of medical data, and then realize the prediction of disease risks[10]. Machine learning algorithms have the ability to extract valuable information from complex data sets. They can identify potential risk factors, which have a positive effect on the early detection and intervention of diseases[11]. Previous studies have explored the prediction of acute mastitis using retrospective data. However, many studies have some limitations. For example, some studies have a small sample size, incomplete consideration of risk factors, and insufficient generalization ability of the models[12]. Logistic Regression (LR) and Naive Bayes (NB) are widely used algorithms. However, these two algorithms are relatively simple. When facing complex data sets, they may have difficulty in screening out relevant indicators from them[13]. XGBoost, an enhanced version of gradient boosting, performs better in terms of accuracy, stability, scalability and customizability[14,15]. Multilayer Perceptron (MLP) is another powerful and flexible artificial neural network model. It performs excellently in learning complex patterns through multiplelayer nonlinear transformations. MLP is especially good at capturing complex relationships in high-dimensional data[16,17]. Therefore, it is suitable for tasks such as regression and classification. Different from relatively simple models such as LR or NB, the deep learning ability of MLP enables it to extract complex features and interactions between variables.

In this study, we adopted four ML models, namely LR, NB, XGBoost and MLP. We carried out a retrospective analysis of twelve relevant indicators of 369 patients with acute mastitis and 447 healthy controls. The aim of our study was to develop a more accurate and reliable prediction model for acute mastitis in lactating women, with the MLP model showing the most promise. This model can predict the occurrence of the disease before the clinical manifestations of acute mastitis appear, thus enabling earlier implementation of intervention measures, and ultimately achieving the reduction of the severity of the disease and the improvement of the overall health status of mothers and infants.

Materials and methods
Research object
In recent years, given the rapid advancement of big data and ML technologies, a growing number of studies have employed these technologies to carry out in-depth analyses of medical data for predicting disease risks. Retrospective case-control studies were conducted. A total of 369 patients with acute lactational mastitis were admitted from January 2016 to December 2023 and 447 healthy individuals without breast diseases during the same period (control group) were selected.

Inclusion and exclusion criteria
This study focused on breastfeeding women aged 18 to 50 years, with a clear diagnosis of acute mastitis (whether present or not). Participants were required to have a complete medical history, including information on age, primiparity status, history of breast surgery, nipple trauma, puerperium, gestational diabetes, and relevant biomarkers (e.g., C-reactive protein (CRP), procalcitonin (PCT), neutrophils (NE), white blood cells (WBC)). Symptoms like elevated body temperature, general pain, headache, chills, and abnormal nipple discharge were also considered in the diagnosis.

Exclusion criteria included women outside the 18 to 50 age range, non-lactating women, and cases with incomplete or missing medical data (more than 10%). Other breast diseases that could interfere with the diagnosis of acute mastitis, severe systemic diseases, and medications affecting mastitis occurrence or diagnosis were also excluded. Additionally, women with acute mastitis during pregnancy, unless it occurred during lactation, were excluded from the study. The study received ethical approval from the Ethics Committee of Liuzhou Hospital of Guangzhou Women and Children's Medical Center (approval number: 2024–212).

Determination time of biomarkers
During routine postpartum examinations, blood samples for biomarkers were collected from both the control and case groups, with additional sampling for the case group at the time of acute mastitis diagnosis. These examinations, as standard procedures for lactating women, involve a systematic assessment of maternal health. Measured indicators include CRP, PCT levels, and blood routine tests, which aim to comprehensively evaluate the inflammatory status and exclude potential diseases that may affect breastfeeding and maternal health.

Establishment of machine learning models
When constructing machine learning models, we used algorithms such as LR, NB, XGBoost, and MLP to develop predictive models. The dataset was divided into training and testing subsets to evaluate model effectiveness, and cross-validation was employed to enhance model robustness. Models were assessed using metrics including ROC curves, accuracy, sensitivity, specificity, and F1 score, and confusion matrices were analyzed to understand their strengths and weaknesses in predicting acute mastitis.

During model construction, we performed 5 rounds of 10-fold cross-validation combined with grid search for parameter optimization. This approach aimed to ensure that models would not overfit the training data while maintaining performance and predictive ability on unseen data. Through this method, we confirmed that

the models exhibited good generalization performance. Prior to evaluation, the optimal decision threshold was determined by maximizing the Youden index on the ROC curve of the training data.

Data statistics

During the data preprocessing stage, numerical statistics and proportion computations were carried out on count data to intuitively understand the distribution among various categories. Subsequently, the chi-squared test was adopted to evaluate the association between different categories, a prevailing statistical method suitable for analyzing the association of categorical data.

For continuous data that did not comply with the assumption of normal distribution, the non-parametric Mann-Whitney U test was chosen to compare two independent samples. This test method is independent of the data distribution, making it especially appropriate for handling continuous data that is not normally distributed. Through this approach, the robustness and reliability of the model were ensured, providing considerable data support for subsequent clinical applications and decision-making.

Results

General data comparison

In the absence of acute mastitis and acute breast disease groups, the average age of the non-acute breast disease group was 32.776 years, with a standard deviation of 7.23 years. For the acute mastitis group, the average age was 34.36 years, and the standard deviation was 6.06 years. Regarding CRP, the average value in the non-acute mastitis group was 6.7 mg/L, while that in the acute mastitis group was 21.1 mg/L.

For the acute breast disease group, the average WBC count was 9.239×10^9/L. As for the non-acute mastitis group, the average WBC count was 7.219×10^9/L, with a maximum value of 9.329×10^9/L. Additionally, in the non-acute mastitis group, the average lymphocyte count was 3.72×10^9/L, while in the acute breast disease group, the average cell count was 3.29×10^9/L.

Our research shows that age, cracked, CRP, NE, and WBC all have statistical significance ($p < 0.05$) (Table 1).

Model evaluation and assessment

A machine learning model was constructed based on variables such as age, cracked, CRP, NE, and WBC, encompassing logistic regression analysis, Naive Bayes, XGBoost, and Multilayer Perceptron. The analysis results show that the multilayer perceptron model exhibits a relatively high area under the receiver operating characteristic curve (AUROC) on both the training set and the test set. This highlights its excellent ability to

Variable	Total (*n* = 816)	No acute mastitis (*n* = 447)	Acute mastitis (*n* = 369)	P
Age, Mean ± SD	33.495 ± 6.775	32.776 ± 7.237	34.366 ± 6.064	0.001
Primiparity, n (%)				0.53
No	428 (52.451)	230 (51.454)	198 (53.659)	
Yes	388 (47.549)	217 (48.546)	171 (46.341)	
History of breast surgery, n (%)				0.513
No	772 (94.608)	425 (95.078)	347 (94.038)	
Yes	44 (5.392)	22 (4.922)	22 (5.962)	
Cracked, n (%)				0.042
No	730 (89.461)	391 (87.472)	339 (91.870)	
Yes	86 (10.539)	56 (12.528)	30 (8.130)	
External breast trauma, n (%)				0.182
No	660 (80.882)	369 (82.550)	291 (78.862)	
Yes	156 (19.118)	78 (17.450)	78 (21.138)	
Puerperium, n (%)				0.347
No	604 (74.020)	325 (72.707)	279 (75.610)	
Yes	212 (25.980)	122 (27.293)	90 (24.390)	
Gestational diabetes, n (%)				0.134
No	738 (90.441)	398 (89.038)	340 (92.141)	
Yes	78 (9.559)	49 (10.962)	29 (7.859)	
C-reactive protein, M (Q1, Q3)	9.350 (5.300, 22.725)	6.700 (3.450, 9.650)	21.100 (10.400, 27.900)	< 0.001
Procalcitonin, M (Q1, Q3)	0.042 (0.027, 0.642)	0.039 (0.024, 0.740)	0.044 (0.027, 0.567)	0.134
Neutrophils, M (Q1, Q3)	6.518 (4.277, 9.253)	6.425 (4.315, 8.854)	6.693 (4.234, 9.362)	0.095
White blood cells, M (Q1, Q3)	7.374 (5.634, 10.892)	7.219 (5.299, 9.329)	9.239 (6.392, 26.670)	< 0.001
Abnormal nipple discharge, n (%)				0.645
No	542 (66.422)	300 (67.114)	242 (65.583)	
Yes	274 (33.578)	147 (32.886)	127 (34.417)	
SD, standard deviation				

Table 1. Baseline characteristics in acute mastitis and Non acute mastitis.
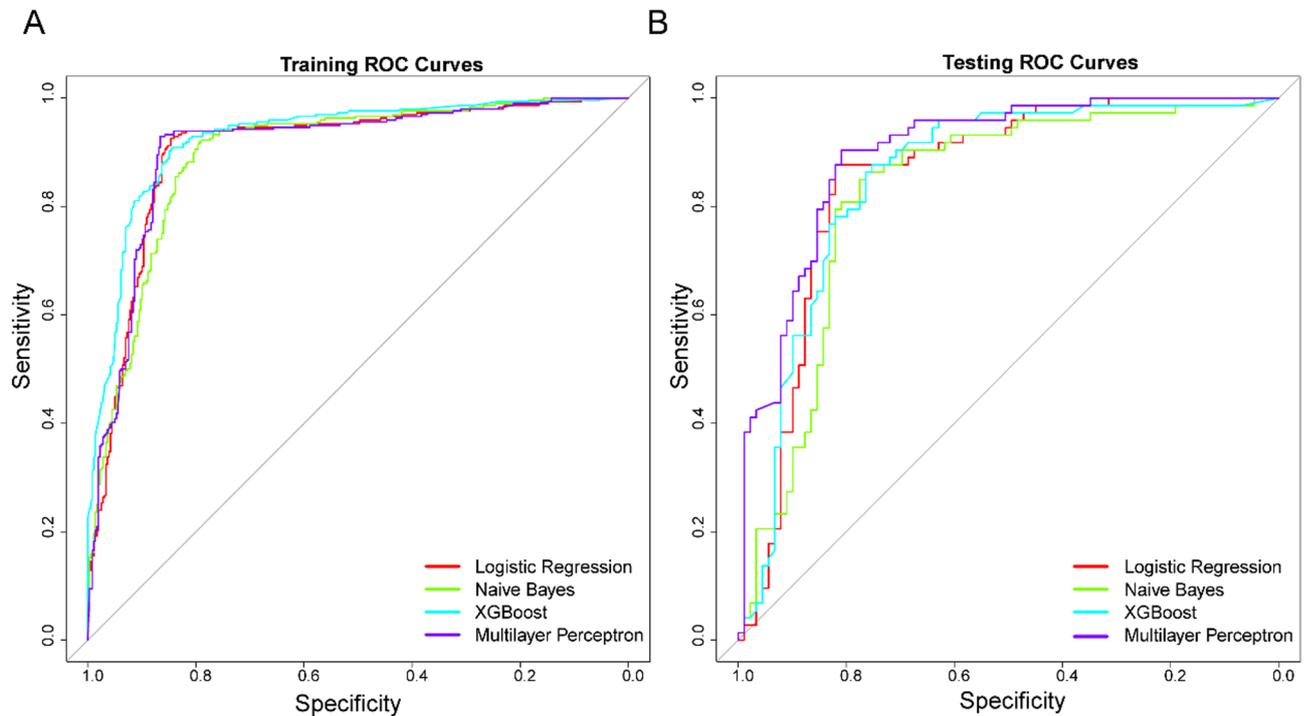
Fig. 1. Evaluation of the machine learning models. (**A**) ROC curve and AUROC for the training set; (**B**) ROC curve and AUROC for the testing set.

Machine learning models	Training accuracy	Training sensitivity	Training specificity	Training F1 score	Training AUROC	Testing accuracy	Testing sensitivity	Testing specificity	Testing F1 score	Testing AUROC
Logistic regression	0.859	0.863	0.855	0.870	0.904	0.809	0.831	0.781	0.827	0.852
Naive bayes	0.787	0.897	0.655	0.822	0.897	0.691	0.843	0.507	0.750	0.826
XGBoost	0.867	0.832	0.909	0.873	0.927	0.796	0.764	0.836	0.805	0.852
Multilayer perceptron	0.890	0.858	0.929	0.895	0.906	0.840	0.820	0.863	0.849	0.898

Table 2. Performance comparison of four algorithms on training and testing datasets.

distinguish between different classes (as shown in Figs. 1A-B). Table 2 shows detailed information on the specific parameters of the models constructed by using different algorithms. Model performance is reported separately for the training set (for model optimization) and testing set (for generalization validation). Testing set metrics represent the predictive efficacy for clinical application.

DCA curve analysis

Decision curve analysis (DCA) is a tool for evaluating the clinical value of a model. It enables us to grasp the practical application value of the model by comparing the expected net benefits at different decision-making thresholds. Through DCA, as shown in Fig. 2, it is found that the MLP model shows the highest net benefit at most thresholds, indicating that it has excellent predictive performance on the test set. This performance advantage not only performs well statistically but also shows considerable clinical value in practical applications, further confirming that the MLP model is one of the most suitable models to use.

Discussion

This study aims to construct a predictive model based on machine learning techniques, identifying the risk factors associated with the occurrence of acute mastitis in lactating women. The data set used in this study comes from a case-control study, involving 369 lactating women diagnosed with acute mastitis and 447 breastfeeding women without this condition. Data covering various potential risk indicators, such as age, primiparity status, history of breast surgery, nipple trauma, external breast trauma, postpartum status, history of gestational diabetes, and abnormal nipple discharge, were collected from these participants.

Machine learning techniques, especially XGBoost and MLP, are used to build prediction models. XGBoost is a powerful variant of gradient boosting and is known for its excellent performance characteristics, including high precision, robustness, scalability, and customizability[18]. On the other hand, MLP is an artificial neural network that is good at capturing complex patterns through multilayer non-linear transformations, making
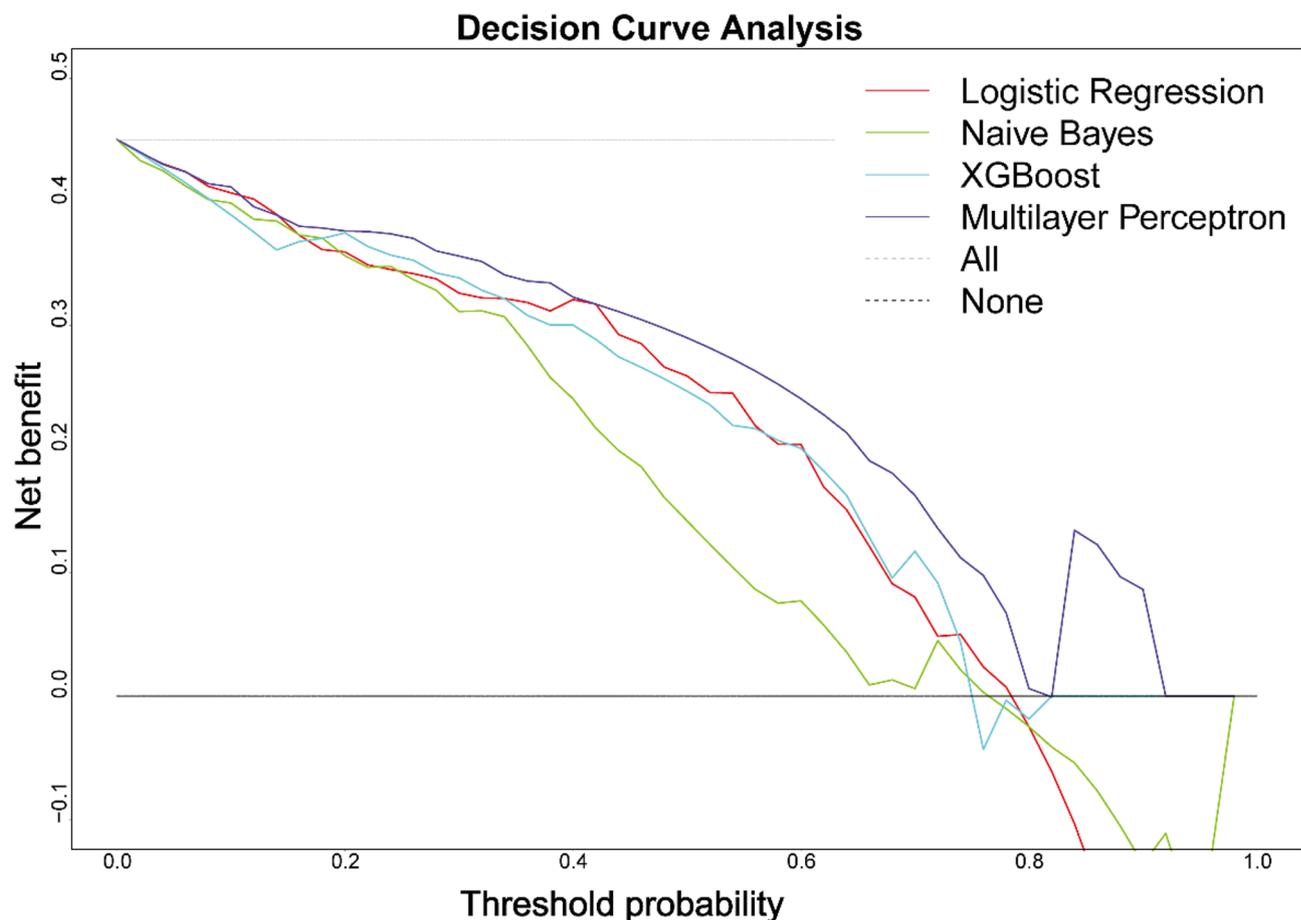
Fig. 2. DCA curves of multiple models in the test set.

it especially good at learning complex relationships in high-dimensional data[16]. This ability makes MLP an efficient tool for regression and classification tasks and it often outperforms traditional algorithms in terms of flexibility and predictive ability[19].

The application of these advanced machine learning methods in the field of medical diagnosis, especially in the prediction and diagnosis of acute mastitis, demonstrates their ability to handle complex high-dimensional data sets and reveal underlying disease patterns[20,21]. For example, XGBoost can assist clinicians by analyzing clinical symptoms, diagnosis results, and imaging data, so as to improve the accuracy and timeliness of mastitis diagnosis[22]. Meanwhile, MLP can reveal non-linear relationships in a wide range of data sets, including patients' medical history and genetic information, providing more personalized treatment options[23]. These machine learning methods not only improve the diagnostic efficiency of acute mastitis but also help promote personalized medicine, making clinical decisions more precise[24].

In this research, the logistic regression model was also applied to fit the data, and the results highlighted several key variables significantly associated with the onset of acute mastitis. These include age, nipple cracking, CRP level, NE count, and WBC count. Using these insights, a comprehensive machine learning based prediction model was developed, which demonstrated its high accuracy and precision. Various machine learning algorithms, including logistic regression, were used to train and validate this model, ultimately resulting in a powerful tool for predicting the probability of acute mastitis in lactating women.

The predictive model discussed in this study exhibited the outstanding ability to differentiate. It could effectively separate patients suffering from acute mastitis from those who were healthy and breastfeeding. It was especially adept at pinpointing individuals at high risk. This skill is of great clinical significance. It offers crucial backing for decisions made in a clinical setting. In comparison with earlier research, the model from this study showed better sensitivity and AUROC[5]. This highlights the promise of machine learning techniques. They can improve assessments of clinical risk for acute mastitis[25].

In this study's findings, age emerged as a significantly important factor related to the development of acute mastitis, with a distinct connection observed both in the disease's onset and progression. Acute mastitis affects a varying percentage (2%–30%) of breastfeeding women globally, and most cases occur in the first three weeks after childbirth[26]. Younger women who have recently given birth, particularly those who start breastfeeding soon after delivery, are at a greater risk of developing acute mastitis. Prior research has noted that the average age of women with lactational mastitis or breast abscesses is 29 years, reinforcing the notion that younger women are more prone to this health issue[5].

Damage to the nipple was found to elevate the risk of acute mastitis. Such damage includes cracks. It also includes abnormalities. These can be inversions or deformities. These issues might lead to blockages in the milk-producing ducts. They can also hinder the normal flow of milk[27,28]. The results of this study confirm earlier findings. They underscore the importance of keeping the nipple in good condition. This is vital for preventing the start of acute mastitis.

CRP levels function as a significant indicator of inflammation. They are recognized for their association with the intensity of acute mastitis. In instances of mastitis, increased CRP levels are typically observed. This elevation occurs in both blood serum and breast milk. It indicates the degree of inflammation affecting the entire body[29,30]. Research conducted on patients with idiopathic granulomatous mastitis (IGM) revealed noteworthy results. In patients experiencing severe IGM, serum CRP and interleukin-6 (IL-6) levels were markedly elevated. This was in contrast to patients with less severe manifestations of the disease[30,31]. These outcomes emphasize the role of CRP. It serves as a marker of inflammation. This is relevant for both the diagnosis and the handling of mastitis cases.

Neutrophils (NE) are pivotal in the inflammatory reaction to mastitis. They operate as the main cells that combat pathogens. One such pathogen is Escherichia coli[32]. In cows with mastitis, changes in the DNA methylation of neutrophils have been observed. These changes are connected to shifts in gene expression and the regulation of microRNA. This may add to the development of mastitis[33]. Furthermore, pro-inflammatory cytokines play a part. An example is tumor necrosis factor-alpha (TNF-α). This cytokine is produced by macrophages in the mammary gland. It aids in bringing neutrophils to the site of infection. In addition, molecules like interleukin-8 (IL-8) increase the arrival of neutrophils[34]. Nitric oxide (NO) also affects neutrophil function. It may control their movement and activity during mastitis[35].

WBC, notably polymorphonuclear neutrophils (PMNs), are crucial in the immune reaction during mastitis. They move from the blood to the mammary gland. Their role is to fight infection[36]. While the inflammatory response is necessary to remove pathogens, too much inflammation can damage tissue. This highlights the importance of properly controlling the immune response. Such control is needed to avoid additional harm to mammary tissue[37]. This process is strictly managed by both the innate and adaptive immune systems. T and B lymphocytes become active when they meet antigen-presenting cells. They also react to other immune signals[38].

The predictive models developed in this study hold significant clinical significance, enabling the identification of women at high risk of acute mastitis. This allows for the timely implementation of interventions and preventive measures, thereby reducing the incidence and severity of mastitis and improving patient outcomes. The models also provide new insights into the etiology and pathophysiological mechanisms of acute mastitis, pointing the way for future research.

However, the clinical application of the models is constrained by the timing of biomarker measurements (e.g., CRP), which were only conducted during routine postpartum examinations and at the time of diagnosis. This may hinder their ability to effectively predict risk before the onset of clinical symptoms. If these biomarkers only exhibit significant changes after symptom occurrence, the models' early predictive capacity will be limited. Future studies should explore the changes in these biomarkers prior to symptom onset to enhance the models' early prediction capability.

Although the findings of this study are encouraging, it is important to acknowledge certain limitations. The first limitation pertains to the source of the data. The data were gathered from one medical institution only. This fact could restrict the ability to apply the results to other areas or groups of people.

In addition, the design of the study did not account for every possible risk factor associated with acute mastitis. Subsequent research endeavors should investigate additional potential risk factors. These could include genetic tendencies, exposure to environmental factors, and lifestyle choices. As a result, the predictive model developed may not encompass all variables that affect the start of the disease. Consequently, it is necessary to validate the model in larger and more varied populations. This step is crucial to evaluate how widely applicable the model is.

Moreover, subsequent research endeavors should investigate additional potential risk factors. These could include genetic tendencies, exposure to environmental factors, and lifestyle choices. Such exploration could refine and improve the model's predictive power.

Conclusion

This study applied machine learning techniques to successfully construct a prediction model for acute mastitis in lactating women, providing important references for clinical practice and scientific research. The research identified age, nipple fissures, CRP, neutrophils, and WBC as the five key predictive indicators. However, the biomarkers used in the current model are only measured during routine postpartum examinations and diagnosis, which may limit the model's predictive ability before the onset of clinical symptoms. Future studies should further optimize the model based on multicenter, large-sample data and explore the temporal dynamic changes of these biomarkers before symptom occurrence, so as to enhance the early predictive value of the model and the accuracy of its practical clinical application.

Data availability

The data set supporting the conclusions of this article is available from the corresponding author on reasonable request.

References

1. Jin, L. et al. Treatment of lactational mastitis with Gualou Xiaoyong soup and painless lactation promoting technique: A case series. *Med. (Baltim).* **102** (48), e36384 (2023).
2. Lisowska-Łysiak K, Dudko P, Kosecka-Strojek M, Walczak J, Wójcik P, Międzobrodzki J. Characteristics of advanced methods used for typing bacterial isolates from mastitis with particular reference to staphylococci. *Pol J Vet Sci* **21**, 229-239 (2018).
3. Kataria, K., Srivastava, A. & Dhar, A. Management of lactational mastitis and breast abscesses: review of current knowledge and practice. *Indian J. Surg.* **75**(6), 430–435 (2013).
4. Ren, L. et al. The efficacy and safety of Chinese massage in the treatment of acute mastitis: A protocol for systematic review and meta-analysis of randomized controlled trials. *Med. (Baltim).* **101** (3), e28504 (2022).
5. Li, D. et al. Risk factors and prognosis of acute lactation mastitis developing into a breast abscess: a retrospective longitudinal study in China. *PLoS One.* **17**, e0273967 (2022).
6. Cullinane, M. et al. Determinants of mastitis in women in the castle study: a cohort study. *BMC Fam Pract.* **16**, 181 (2015).
7. Meretoja, T., Ihalainen, H. & Leidenius, M. Inflammations of the mammary gland. *Duodecim* **133**, 855–861 (2017).
8. Cheng, G. Z., Chen, A., Xin, Y. & Ni, Q. Q. Using the teach-back method to improve postpartum maternal - infant health among women with limited maternal health literacy: a randomized controlled study. *BMC Pregnancy Childbirth.* **23**, 13 (2023).
9. Esmaeil, N. K. et al. Clinical, microbiological, immunological and hormonal profiles of patients with granulomatous mastitis. *Biomed. Rep.* **18**, 41 (2023).
10. Turashvili, G. & Li, X. Inflammatory lesions of the breast. *Arch. Pathol. Lab. Med.* **147**, 1133–1147 (2023).
11. Zhang, Y. et al. The therapy of elimination first for early acute mastitis: a systematic review and meta-analysis. *Evid. Based Complement. Alternat Med.* **2018**, 8059256 (2018).
12. Miller, C., Portlock, T., Nyaga, D. M. & O' Sullivan, J. M. A review of model evaluation metrics for machine learning in genetics and genomics. *Front. Bioinform.* **4**, 1457619 (2024).
13. Li, B. et al. Clinical application of a big data machine learning analysis model for osteoporotic fracture risk assessment built on multicenter clinical data in Qingdao City. *Discov Med.* **37** (192), 55–63 (2025).
14. Chen, H. et al. Unlocking reproducible transcriptomic signatures for acute myeloid leukaemia: integration, classification and drug repurposing. *J. Cell. Mol. Med.* **28** (17), e70085 (2024).
15. Chung, C. C., Su, E. C., Chen, J. H., Chen, Y. T. & Kuo, C. Y. XGBoost - Based simple Three - item model accurately predicts outcomes of acute ischemic stroke. *Diagnostics (Basel).* **13** (5), 842 (2023).
16. Kim, J. & Seo, D. Three - dimensional augmentation for hyperspectral image data of water quality: an integrated approach using machine learning and numerical models. *Water Res.* **251**, 121125 (2024).
17. Dimitsaki, S., Gavriilidis, G. I., Dimitriadis, V. K. & Natsiavas, P. Benchmarking of machine learning classifiers on plasma proteomic for COVID – 19 severity prediction through interpretable artificial intelligence. *Artif. Intell. Med.* **137**, 102490 (2023).
18. Chen, T., Guestrin, C. & XGBoost A scalable tree boosting system. proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. :785–794. (2016).
19. Inoue, T. et al. Xgboost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Rep.* **1**, 8–16 (2020).
20. Li, G., Qi, Y., Zhang, L. & Yu, Y. Development of a machine learning-based diagnostic model using hematological parameters to differentiate periductal mastitis from granulomatous lobular mastitis. *Sci. Prog.* **108**, 352304009 (2025).
21. Li, L., Yang, W. & Jia, H. Deep learning models for predicting the recurrence of idiopathic granulomatous mastitis. *J. Inflamm. Res.* **18**, 2943–2953 (2025).
22. Sun, J. et al. Prediction models for postoperative recurrence of non-lactating mastitis based on machine learning. *BMC Med. Inf. Decis. Mak.* **24**, 106 (2024).
23. Chong PL, Vaigeshwari V, Mohammed Reyasudin BK, Noor Hidayah BRA, Tatchanaamoorti P, Yeow JA, Kong FY. Integrating artificial intelligence in healthcare: applications, challenges, and future directions. *Future Sci OA* **11**, 2527505 (2025).
24. Kayadibi, Y. et al. Differentiation of malignancy and idiopathic granulomatous mastitis presenting as non-mass lesions on mri: radiological, clinical, radiomics, and clinical-radiomics models. *Acad. Radiol.* **31**, 3511–3523 (2024).
25. Nakagami, G. et al. Supervised machine learning - based prediction for in - hospital pressure injury development using electronic health records: a retrospective observational cohort study in a university hospital in Japan. *Int. J. Nurs. Stud.* **119**, 103932 (2021).
26. Suga, Y., Kuehn, T., Knfe, W. A. G. & Teklewold, G. Review of breast abscess cases at a tertiary hospital in addis ababa, Ethiopia. *Int. J. Womens Health.* **15**, 433–442 (2023).
27. Lai, B. Y. et al. Risk factors for lactation mastitis in china: A systematic review and meta - analysis. *PLoS One.* **16** (5), e0251182 (2021).
28. Wilson, E., Woodd, S. L. & Benova, L. Incidence of and risk factors for lactational mastitis: a systematic review. *J. Hum. Lact.* **36**, 673–686 (2020).
29. Amir, L. H., Crawford, S. B., Cullinane, M. & Grzeskowiak, L. E. General practitioners' management of mastitis in breastfeeding women: a mixed method study in Australia. *BMC Prim. Care.* **25**, 161 (2024).
30. Huang, Y. M. et al. Serum C - Reactive protein and Interleukin – 6 levels as biomarkers for disease severity and clinical outcomes in patients with idiopathic granulomatous mastitis. *J. Clin. Med.* **10** (10), 2077 (2021).
31. Zaouak, A. et al. Idiopathic granulomatous mastitis associated with erythema nodosum in a pregnant woman. *Skinmed* **21**, 448–450 (2023).
32. Ju, Z. et al. Genome - wide methylation and transcriptome of blood neutrophils reveal the roles of Dna methylation in affecting transcription of protein - coding genes and Mirnas in e. Coli - infected mastitis cows. *BMC Genom.* **21**, 102 (2020).
33. Elazar, S., Gonen, E., Livneh - Kol, A., Rosenshine, I. & Shpigel, N. Y. Neutrophil recruitment in endotoxin - induced murine mastitis is strictly dependent on mammary alveolar macrophages. *Vet. Res.* **41**, 10 (2010).
34. Jimbo, S., Suleman, M., Maina, T., Prysliak, T. & Mulongo, M. Perez - Casal J. Effect of Mycoplasma Bovis on bovine neutrophils. *Vet. Immunol. Immunopathol.* **188**, 27–33 (2017).
35. Aitken, S. L., Corl, C. M. & Sordillo, L. M. Immunopathology of mastitis: insights into disease recognition and resolution. *J. Mammary Gland Biol. Neoplasia.* **16**, 291–304 (2011).
36. Liu, P., Zhang, P., Yuan, C., Li, J. & Yang, Q. Mechanism of transepithelial migration of lymphocytes into the milk in Porcine mammary glands. *J. Reprod. Immunol.* **149**, 103440 (2022).
37. Shangraw, E. M. & Mcfadden, T. B. Graduate student literature review: systemic mediators of inflammation during mastitis and the search for mechanisms underlying impaired lactation. *J. Dairy. Sci.* **105**, 2718–2727 (2022).
38. Carrasco, P. C. et al. T cell activation and effector function in the human Jurkat t cell model. *Methods Cell. Biol.* **178**, 25–41 (2023).

Acknowledgements

Author contributions

L L is responsible for fund acquisition, project management, software development, supervision, verification,

visualization, and writing initial drafts; L Z participated in data organization, formal analysis, and project management; Z H participated in the survey subjects, methodology, and initial writing draft; Y C participated in the survey subjects and project management; G L is responsible for supervision work.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
This study carried out a retrospective analysis of our hospital's case series by employing the opt-out method. This study has been approved by the Ethics Committee of Liuzhou Hospital of Guangzhou Women and Children's Medical Center (approval number: 2024–212) and was conducted in accordance with the 1964 Helsinki Declaration and its subsequent amendments or similar ethical standards. This study is retrospective in nature, and the requirement for informed consent was waived by Guangzhou Women and Children's Medical Center Liuzhou Hospital Medical Research Ethics Committee.

Additional information
Correspondence and requests for materials should be addressed to L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.