Research Article

A Machine Learning Applied Diagnosis Method for Subcutaneous Cyst by Ultrasonography

Hao Feng,[1] Qian Tang,[1] Zhengyu Yu [iD],[2] Hua Tang,[1] Ming Yin,[1] and An Wei [iD][3]

[1]*Department of Dermatology, Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha 410005, China*
[2]*Faculty of Engineering and IT, University of Technology, Sydney, Sydney, NSW 2007, Australia*
[3]*Department of Ultrasound, Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha 410005, China*

Correspondence should be addressed to An Wei; weian1976@163.com

For decades, ultrasound images have been widely used in the detection of various diseases due to their high security and efficiency. However, reading ultrasound images requires years of experience and training. In order to support the diagnosis of clinicians and reduce the workload of doctors, many ultrasonic computer aided diagnostic systems have been proposed. In recent years, the success of deep learning in image classification and segmentation has made more and more scholars realize the potential performance improvement brought by the application of deep learning in ultrasonic computer-aided diagnosis systems. This study is aimed at applying several machine learning algorithms and develop a machine learning method to diagnose subcutaneous cyst. Clinical features are extracted from datasets and images of ultrasonography of 132 patients from Hunan Provincial People's Hospital in China. All datasets are separated into 70% training and 30% testing. Four kinds of machine learning algorithms including decision tree (DT), support vector machine (SVM), *K*-nearest neighbors (KNN), and neural networks (NN) had been approached to determine the best performance. Compared with all the results from each feature, SVM achieved the best performance from 91.7% to 100%. Results show that SVM performed the highest accuracy in the diagnosis of subcutaneous cyst by ultrasonography, which provide a good reference in further application to clinical practice of ultrasonography of subcutaneous cyst.

1. Introduction

Subcutaneous cyst occurs especially at younger age, especially in the head, arms, and back in youth. It is a soft or a plurality of soft or firm balls with diameters ranging from 1 to approximately 3 cm. Subcutaneous cyst is buried in the skin or subcutaneous tissue and adheres to the skin and the base can move. There are small openings on the skin. When the cyst is pushed, it adheres tightly to the skin. Small pit appears when it is slightly depressed, which is the opening of the duct where the gland directly reaches the skin surface. Some openings are stuffed with a black pimple-like plug to squeeze out white wax-like substances [1–3]. The ultrasonic images of superficial epidermoid cyst have certain specificity. The advantages of ultrasound include high spatial resolution, portability, convenience, and low cost. It is important to be able to combine the physical examination results with the patient's medical history in ultrasonic examination. In addition, real-time imaging allows manual compression, limb movement, muscle contraction, and direct interaction with patients during scanning [4]. By summarizing and analyzing its acoustic features, it can effectively improve the correct rate of clinical diagnosis and reduce misdiagnosis and missed diagnosis [5, 6]. However, one disadvantage of ultrasound is when the disease occurs in deeper soft tissues. In these cases, the image resolution is reduced, and auxiliary information about the mass, such as physical examination results and medical history, may be blurred [7].
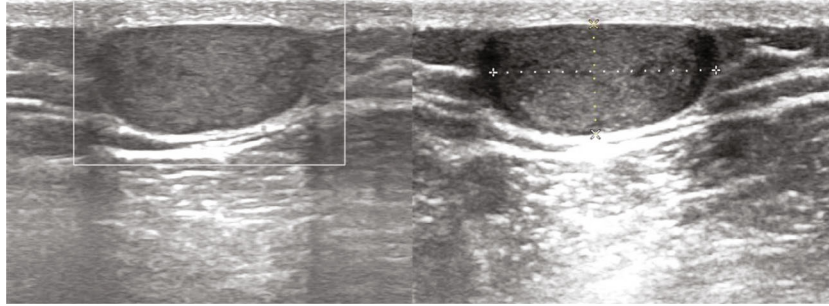
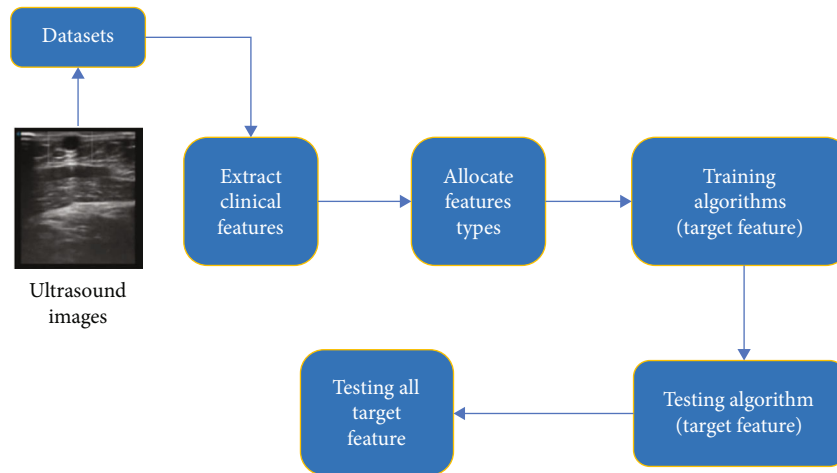Figure 1: Sample ultrasound images of subcutaneous cyst.

Figure 2: Workflow diagram for machine learning applied diagnosis method.

Machine learning is one of the branch in artificial intelligence and has been widely used in multidisciplinary research fields [8, 9], which is related to computer science, statistics, and information theory [10, 11]. Algorithms are used to analyze data and try to explore potential patterns hidden in data to predict new information in machine learning [8, 12], with high precision, high speed, and convenient expansion [13, 14]. The main application fields of deep learning in ultrasound computer-aided diagnosis system include breast disease diagnosis [15], liver disease diagnosis [16], fetal ultrasound standard plane detection [17], thyroid nodule diagnosis [18], and carotid artery ultrasound image classification [19]. In recent years, machine learning algorithms, such as decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), and neural networks (NN), have been more and more frequently applied in medical field [20–23].

However, limited relative studies on machine learning in the diagnosis of subcutaneous cyst based on ultrasonography have been reported. Thus, the motivation and novelty of this study were to apply several machine learning algorithms to diagnose subcutaneous cyst from clinical features extracted from datasets and images of ultrasonography and find a powerful alternative method for ultrasonic diagnosis of subcutaneous cysts.

2. Materials and Methods

2.1. Data Acquisition.
All datasets in this article are from Hunan Provincial People's Hospital in China. There are 133 patients that participated. Each patient has five ultrasound images record with 19 clinical features extracted including gender, age, none blood flow, dotted blood flow, dot-bar blood flow, rich blood flow, none echo, mixed echo, low echo, medium echo, high echo, uniform internal echo, nonuniform internal echo, clear boundary, unclear boundary, regular form, irregular form, strong spot, and changes of parenchyma echo. Figure 1 illustrates sample ultrasound images of subcutaneous cyst.

2.2. Data Preparation.
A machine learning applied method is proposed to diagnose subcutaneous cyst in this article. Clinical features are extracted from datasets and images of ultrasonography. Figure 2 displays the workflow processes for the machine learning applied diagnosis method in several steps, which are presented in the following:

(1) Extract clinical features from datasets and ultrasound images

(2) Allocate feature types: there are two different feature types in our method, categorical and numerical. To

improve the accuracy of the diagnosis method, we allocated these features into two types, Table 1 presents all the features and types

(3) Training algorithms: machine learning algorithms, such as decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), and neural networks (NN), had been approached to determine the best performance. The result indicated that the best accurate algorithm is SVM. Feature 1 (gender) and feature 2 (age) are excluded from target features, as they are basic information of the patient, they do not change or are not affected by cysts lesions. Therefore, the target features in this research are 17 features, which are from feature 3 to feature 19. In the training stage, we start from target feature 3, the results indicated that SVM provided the highest accuracy

(4) Testing algorithm: in this stage, the rest 30% of the datasets is processed by SVM. We use the same target feature as the previous training. SVM achieves 100% accuracy

(5) Testing target features: each target feature is trained and tested to validate the reliability of SVM. The rest target features are from feature 4 to feature 19. We compared with all the results from each feature, SVM still achieved the best performance from 91.7% to 100%. This is a significant accuracy in diagnosing subcutaneous cyst

Statistical Processing

3. Results

3.1. Training Process. Four different types of machine algorithms are applied. We separated datasets into two parts, 70% for training and 30% for testing. 19 features from 132 patients are computed. Since feature 1 and feature 2 are gender and age, they are general information of patients and will not be affected by any cyst lesions, and the target features in this research are from feature 3 to feature 19. Initially, we set feature 3, none blood flow, as the target feature, and the rest of the features are attributes. To avoid datasets overfitting, we use 5 cross-validation folds. After processing the training by using DT, SVM, KNN, and ANN, the accuracy results are 99.2%, 100%, 92.4%, and 98.5%, respectively. SVM provides best result for the diagnosis method. Table 2 indicates the accuracy result for the training process.

3.2. Testing Process. To validate the reliability of the method, SVM is tested by using the rest 30% of the datasets. The target features are the same as the previous training, which are from feature 3 to feature 19. We start setting the target feature from feature 3. We also use 5 cross validation folds to avoid datasets overfitting. As shown in Table 3, the testing results slightly changed to DT (96.2%), SVM (100%), KNN (95.5%), and ANN (98.5). Comparing with the results, SVM still achieves 100% accuracy. It is the best algorithm for the diagnosis method when we set the target feature as feature 3.

TABLE 1: 19 features extracted from datasets.

ID	Feature	Type
1	Gender	Categorical
2	Age	Numerical
3	None blood flow	Categorical
4	Dotted blood flow	Categorical
5	Dot-bar blood flow	Categorical
6	Rich blood flow	Categorical
7	None Echo	Categorical
8	Mixed Echo	Categorical
9	Low Echo	Categorical
10	Medium Echo	Categorical
11	High Echo	Categorical
12	Uniform internal Echo	Categorical
13	Nonuniform internal Echo	Categorical
14	Clear boundary	Categorical
15	Unclear boundary	Categorical
16	Regular form	Categorical
17	Irregular form	Categorical
18	Strong spot	Categorical
19	Changes of parenchyma Echo	Categorical

TABLE 2: Machine learning algorithms for training.

Algorithm	Accuracy (%)
Decision tree (DT)	99.2
Support vector machine (SVM)	100
K-nearest neighbors (KNN)	92.4
Artificial neural network (ANN)	98.5

TABLE 3: Machine learning algorithms for testing.

Algorithm	Accuracy (%)
Decision tree (DT)	96.2
Support vector machine (SVM)	100
K-nearest neighbors (KNN)	95.5
Artificial neural network (ANN)	98.8

3.3. Training and Testing for Each Target Feature. In the previous stage, we trained and tested feature 3 as a target feature, and the result indicated that SVM achieved the best accuracy. Further training and testing are approached for each target feature, from feature 4 to feature 19. Table 4 indicates the testing results for 17 target features. In Table 4, the target features are trained and tested by using DT, SVM, KNN, and ANN. Comparing with the results in each column, SVM achieved the best performance (from 91.7% to 100%) for all the target features. The lowest accuracy from SVM is 91.7% at target feature 19. The reason could be that the datasets for feature 19 contains noise and uncertainties that reduces the accuracy from all the algorithms. In that column, we can find that DT is 87.1%, KNN is 90.9, and ANN is 86.5. We compare the results, SVM (91.7) is still

TABLE 4: Testing results for target features.

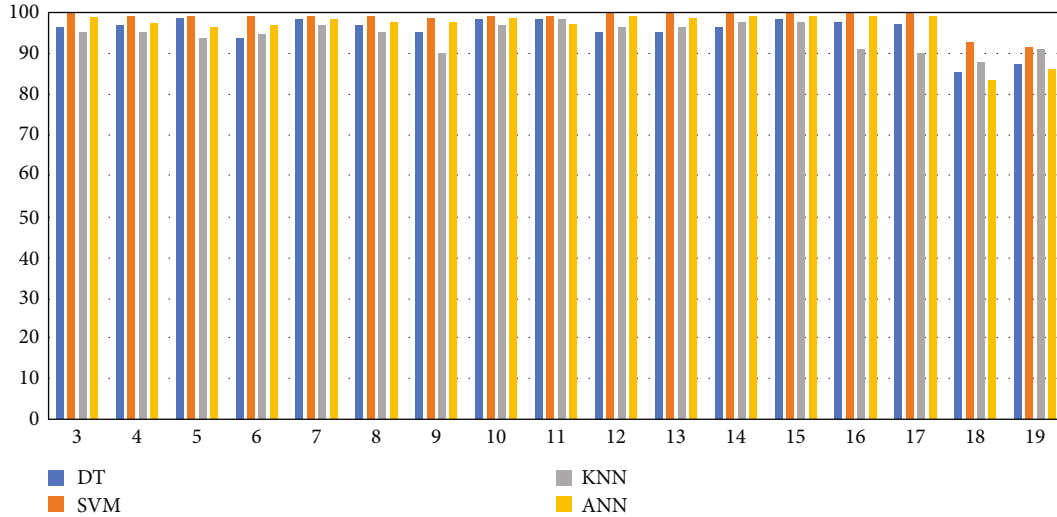| Accuracy (%)	3	4	5	6	7	8	9	10
DT	96.2	97.0	98.5	93.9	98.5	97.0	95.6	98.5
SVM	100	99.2	99.2	99.2	99.2	99.2	98.8	99.2
KNN	95.5	95.5	93.9	94.7	97.0	95.5	90.2	97.0
ANN	98.8	97.7	96.2	97.0	98.5	97.7	97.7	98.5

FIGURE 3: Column chart of testing results.

TABLE 5: Testing results for target features.

Accuracy (%)	11	12	13	14	15	16	17	18	19
DT	98.5	95.5	95.5	96.2	98.5	97.7	97.0	85.6	87.1
SVM	99.2	100	100	100	100	100	100	92.9	91.7
KNN	98.5	96.2	96.2	97.7	97.7	90.9	90.2	87.9	90.9
ANN	97.0	99.2	98.5	99.2	99.2	99.2	99.2	83.3	86.4

the best. Figure 3 is the column chart associated with Table 5. SVM is in orange and is the highest in each column. In this article, we found a machine learning method with SVM achieved the significant accuracy (from 91.7 to 100%) to diagnose subcutaneous cyst in addressing blood flow, echo, internal echo, boundary, strong spot, and changes of parenchyma echo.

4. Discussion and Conclusions

In this study, a total of 132 patients from Hunan Provincial People's Hospital in China participated. Four different types of machine algorithms including decision tree (DT), support vector machine (SVM), K-nearest neighbors (KNN), and neural networks (NN) in machine learning method are applied for the diagnose of subcutaneous cyst by ultrasonography. Decision tree is an effective ultrasonic image classification algorithm. It can learn classification rules from unordered data [24, 25]. The decision tree algorithm uses the divide and conquer strategy to divide the search space

of the problem into several subsets. From top to bottom, each node determines the next node by calculating the eigenvalue of the input sample. In the leaf node, the final classification result is given [26]. In the case of small amount of data and nondiversified eigenvalues, the construction of decision tree is simple and fast. However, if the data volume is large and the eigenvalues are large, the complexity of the decision tree algorithm will be large. Support vector machine has good performance on both small datasets and large datasets. However, with the increase of dataset size, the complexity of support vector machine also increases. At the same time, the choice of kernel function will also affect the performance of support vector machine [27–29]. K-nearest neighbors is a lazy and nonparametric algorithm that has the good characteristics of being simple and easy to use and has a reasonable accuracy [30].

Our research has limitations. First of all, the training set data came from the same hospital, and we did not summarize the basic information of patients and cysts. Second, no matter training or testing datasets, the sample size is

relatively small. Therefore, these results need to be validated in a larger cohort to determine the value of our model in clinical practice.

All datasets are separated into two parts, which is 70% for training and 30% for testing. Results show that SVM achieved the best performance (from 91.7% to 100%) for all the target features. The machine learning method developed in this study can help doctors diagnose the ultrasonic images of patients with subcutaneous cysts more accurately. However, other prospective cohort studies should be conducted externally.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Ethical Approval

This retrospective study was approved by the Ethics Committee of Hunan Provincial People's Hospital as an exempt study without a registration number.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

References

[1] J. J. Krant and J. A. Carucci, "Benign subcutaneous lesions: cysts and lipomas," in *Surgery of the Skin*, pp. 767–776, Elsevier, 2005.

[2] R. D. Tyler, J. H. Meinkoth, R. L. Cowell, C. G. MacAllister, and K. J. Caruso, *Cutaneous and subcutaneous lesions: masses, cysts, and fistulous tracts*, Elsevier, 2002.

[3] S. R. Thaller and B. S. Bauer, "Cysts and cyst-like lesions of the skin and subcutaneous tissue," *Clinics in Plastic Surgery*, vol. 14, no. 2, pp. 327–340, 1987.

[4] J. A. Jacobson, W. D. Middleton, S. J. Allison et al., "Ultrasonography of superficial soft-tissue masses: Society of Radiologists in Ultrasound consensus conference statement," *Radiology*, vol. 304, no. 1, pp. 18–30, 2022.

[5] V. Quaresma, L. Marconi, M. Lopes et al., "Diagnostic performance of contrast-enhanced ultrasonography for the evaluation of malignancy in complex cystic masses," *European Urology Open Science*, vol. 33, pp. S172–S173, 2021.

[6] H. Athar, A. Q. Mir, M. B. Gugjoo, R. A. Ahmad, and H. M. Khan, "Ultrasonography: an affordable diagnostic tool for precisely locating coenurosis cyst in sheep and goats," *Small Ruminant Research*, vol. 169, pp. 19–23, 2018.

[7] L. H. Goldman, L. Perronne, E. F. Alaia et al., "Does magnetic resonance imaging after diagnostic ultrasound for soft tissue masses change clinical management?," *Journal of Ultrasound in Medicine*, vol. 40, no. 8, pp. 1515–1522, 2021.

[8] K.-Q. Li, Y. Liu, and Q. Kang, "Estimating the thermal conductivity of soils using six machine learning algorithms," *International Communications in Heat and Mass Transfer*, vol. 136, article 106139, 2022.

[9] M. Zhu, J. Wang, X. Yang et al., "A review of the application of machine learning in water quality evaluation," *Eco-Environment & Health*, vol. 1, no. 2, pp. 107–116, 2022.

[10] B. Ratner, *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data*, CRC Press, 2017.

[11] C. Ma, H. H. Zhang, and X. Wang, "Machine learning for big data analytics in plants," *Trends in Plant Science*, vol. 19, no. 12, pp. 798–808, 2014.

[12] Y. Liu, T. Zhao, W. Ju, and S. Shi, "Materials discovery and design using machine learning," *Journal of Materiomics*, vol. 3, no. 3, pp. 159–177, 2017.

[13] N. Geetha and P. Bridjesh, "Overview of machine learning and its adaptability in mechanical engineering," Materials Today: Proceedings, 2020.

[14] Z. Zhang, Y. Wang, and K. Wang, "Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks," *The International Journal of Advanced Manufacturing Technology*, vol. 68, no. 1-4, pp. 763–773, 2013.

[15] S. Han, H.-K. Kang, J.-Y. Jeong et al., "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Physics in Medicine & Biology*, vol. 62, no. 19, pp. 7714–7728, 2017.

[16] T. M. Hassan, M. Elmogy, and E.-S. Sallam, "Diagnosis of focal liver diseases based on deep learning technique for ultrasound images," *Arabian Journal for Science and Engineering*, vol. 42, no. 8, pp. 3127–3140, 2017.

[17] Z. Yu, E.-L. Tan, D. Ni et al., "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 874–885, 2017.

[18] J. Ma, F. Wu, T. A. Jiang, J. Zhu, and D. Kong, "Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images," *Medical Physics*, vol. 44, no. 5, pp. 1678–1691, 2017.

[19] R.-M. Menchón-Lara, J.-L. Sancho-Gómez, and A. Bueno-Crespo, "Early-stage atherosclerosis detection using deep learning over carotid ultrasound images," *Applied Soft Computing*, vol. 49, pp. 616–628, 2016.

[20] B. Ambale-Venkatesh, X. Yang, C. O. Wu et al., "Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis," *Circulation Research*, vol. 121, no. 9, pp. 1092–1101, 2017.

[21] N. Ludwig, T. Fehlmann, F. Kern et al., "Machine learning to detect Alzheimer's disease from circulating non-coding RNAs," *Genomics, Proteomics & Bioinformatics*, vol. 17, no. 4, pp. 430–440, 2019.

[22] P. Ferroni, F. M. Zanzotto, S. Riondino, N. Scarpato, F. Guadagni, and M. Roselli, "Breast cancer prognosis using a machine learning approach," *Cancers*, vol. 11, no. 3, p. 328, 2019.

[23] A. D. Hardie, J. H. Chamberlin, J. H. Boyum et al., "Multi-center follow-up study to develop a classification system which differentiates mucinous cystic neoplasm of the liver and

benign hepatic cyst using machine learning," *Academic Radiology*, vol. 29, no. 8, pp. 1149–1156, 2022.

[24] U. R. Acharya, P. Chowriappa, H. Fujita et al., "Thyroid lesion classification in 242 patient population using gabor transform features from high resolution ultrasound images," *Knowledge-Based Systems*, vol. 107, pp. 235–245, 2016.

[25] U. R. Acharya, S. V. Sree, M. M. R. Krishnan et al., "Computer-aided diagnostic system for detection of Hashimoto thyroiditis on ultrasound images from a polish population," *Journal of Ultrasound in Medicine*, vol. 33, no. 2, pp. 245–253, 2014.

[26] T. S. Rao, "Classification, parameter estimation and state estimation-an engineering approach using MATLAB," *Journal of Time Series Analysis*, vol. 32, pp. 194–194, 2011.

[27] L. Cai, X. Wang, Y. Wang, Y. Guo, J. Yu, and Y. Wang, "Robust phase-based texture descriptor for classification of breast ultrasound images," *Biomedical Engineering Online*, vol. 14, no. 1, pp. 1–21, 2015.

[28] P.-L. Yen, H.-K. Wu, H.-S. Tseng et al., "Vascular morphologic information of three-dimensional power Doppler ultrasound is valuable in the classification of breast lesions," *Clinical Imaging*, vol. 36, no. 4, pp. 267–271, 2012.

[29] Q. Zhang, J. Suo, W. Chang, J. Shi, and M. Chen, "Dual-modal computer-assisted evaluation of axillary lymph node metastasis in breast cancer patients on both real-time elastography and b-mode ultrasound," *European Journal of Radiology*, vol. 95, pp. 66–74, 2017.

[30] K. Li, Y. Xu, Z. Zhao, and M. Q.-H. Meng, "Automatic recognition of abdominal organs in ultrasound images based on deep neural networks and k-nearest-neighbor classification," in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1980–1985, Sanya, China, 2021.