



Research Article

A Novel N -Gram-Based Image Classification Model and Its Applications in Diagnosing Thyroid Nodule and Retinal OCT Images

Guanfang Wang,¹ Xianshan Chen,² Geng Tian ¹ and Jiasheng Yang ³

¹Genies Beijing Co., Ltd., Beijing 100102, China

²Department of Thoracic Surgery, Hainan General Hospital, Haikou, Hainan 570311, China

³School of Electrical and Information Engineering, Anhui University of Technology, Maanshan, Anhui 243002, China

Correspondence should be addressed to Geng Tian; tiang@geneis.cn and Jiasheng Yang; jsyang.mcc@gmail.com

Received 6 December 2021; Revised 14 April 2022; Accepted 16 April 2022; Published 2 May 2022

Academic Editor: Huiling Chen

Copyright © 2022 Guanfang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Imbalanced classes and dimensional disasters are critical challenges in medical image classification. As a classical machine learning model, the n -gram model has shown excellent performance in addressing this issue in text classification. In this study, we proposed an algorithm to classify medical images by extracting their n -gram semantic features. This algorithm first converts an image classification problem to a text classification problem by building an n -gram corpus for an image. After that, the algorithm was based on the n -gram model to classify images. The algorithm was evaluated by two independent public datasets. The first experiment is to diagnose benign and malignant thyroid nodules. The best area under the curve (AUC) is 0.989. The second experiment is to diagnose the type of fundus lesion. The best result is that it correctly identified 86.667% of patients with dry age-related macular degeneration (AMD), 93.333% of patients with diabetic macular edema (DME), and 93.333% of normal individuals.

1. Introduction

Medical imaging is essential in diagnosis. However, the huge amount of image data generated in exponential manner has made the image-processing task time-consuming [1]. Furthermore, factors such as irregularity, inexperience, or fatigue may lead to misdiagnosis [2].

With the rapid development of medical image acquisition and storage technology, computer-aided diagnosis (CAD) [3] systems have become an attractive option. For example, a CAD system developed by Raghavendra et al. [4] uses an optimized multilevel elongated binary model to characterize thyroid nodules. Yu et al. [5] proposed a weakly supervised segmentation neural network approach to alleviate the differences in nodule size in thyroid ultrasound images which may cause under- and/or over-segmentation problems. Koundal et al. [6] proposed an automated Intuitionistic Fuzzy Active Contour Method (IFACM) that inte-

grates intuitionistic fuzzy clustering with the active contour for the segmentation of thyroid nodules' ultrasound images. Abdolali et al. [7] presented a novel deep neural network architecture that can detect various types of thyroid nodules. Pekala et al. [8] described a new method combining fully convolutional networks (FCN) with Gaussian processes for post processing which could be used for the automated fine-grained segmentation of spectral domain optical coherence tomography (OCT) images of the retina. Kauer-Bonin et al. [9] proposed a real-time feedback system for automatic quality assessment of retinal OCT images.

One of the most important goals of CAD is to classify medical images into different categories to help doctors in disease diagnosis or further research. It is general practice to extract features from a preprocessed image before passing them to a classifier for classification. Feature extraction and selection are key players in classifier performance. The widely used medical image features are mostly based on

the underlying information of the image, such as color, texture, and shape [10]. These are content-based image classification technologies. However, the computer struggles to understand the meaning of the image from the underlying information, resulting in a “semantic gap” [11] between the “semantic similarity” of human understanding and the “visual similarity” of computer understanding.

Semantic-based classification methods, mainly deep learning-based convolutional neural network (CNN) models [12], part-based models [13], and bag-of-words (BOW) models [14], have received eager attention and application in recent years.

Among them, the CNN model avoids the feature extraction and data reconstruction process, but the model has some drawbacks in clinical applications, such as reliance on high-quality big data [15–17], weak generalization ability [18], poor interpretability [19], and high operational costs [20]. Part-based models can accurately describe the relationships between local features, but training the models is complex [21]. In 2003, Sivic and Zisserman [22] proposed the visual bag-of-words model, which introduced the bag-of-words model into the field of computer vision with great success. However, this model treated local features as independent individuals without accounting for the connection between features [23].

In general, features are classified as suitable, unnecessary, or repeated [24]. For the extracted image features, it may still have a large dimensionality and may contain certain irrelevant or redundant features. Therefore, in recent years, researchers have proposed and developed many methods and techniques to reduce the high dimensionality of data and improve model accuracy. Alharan et al. [25] proposed to evaluate the extracted features by five techniques (information gain, gain ratio, oneR, ReliefF, and symmetric). Then, based on the assessment results, feature selection was accomplished by utilizing the K -mean clustering algorithm. Li et al. [26] developed a radiological model based on multivariate logistic regression with the optimal 12 radiomic features after feature dimensionality reduction. The model performed well for the classification accuracy of the PTC and NG thyroid nodules in the training group and validation group. Lai and Deng [27] developed a deep learning model called Coding Network with Multilayer Perceptron (CNMP), which combines high-level features extracted from deep convolutional neural networks with some selected traditional features. In order to decipher the intrinsic structure of the benign and malignant thyroid nodule, Raghavendra et al. [28] used spatial gray level dependence features and fractal textures while Yu et al. [29] extracted 2 morphological features and 65 textural features of the region of interest to describe the thyroid nodules. Ardakani et al. [30] proposed 49 valid features (9 morphological features and 40 textural features) to classify thyroid nodules.

To overcome these concerns about feature extraction and feature selection methods, this article investigated whether the image content can be explained by the n -gram model. The n -gram model is a classical language model which is based on the probability of the $(n - 1)$ -order Markov chain [31]. This model overcomes the language depen-

dence and skips the processing text content by using linguistics [32]. It can extract potential features and take into account the context information of local features; therefore, the n -gram model has been widely used in text retrieval and classification, information retrieval [33], probability theory [34], computer language [35], communication theory [36], computational biology [37], etc. Nevertheless, it has not been widely used in the field of medical image classification.

Applying the n -gram model to medical image classification is challenging because (1) the text is a one-dimensional vector, whereas the image is a two-dimensional matrix; (2) in the n -gram model, the number of all possible word combinations grows exponentially as n increases, and these high-dimensional data can lead to dimensional disasters. To solve these problems, we first proposed a method to extract features from grayscale images using the n -gram model. This procedure maps the image into a character matrix and then extracts all the n -gram character combinations from this matrix sequentially. We then designed a novel feature selection method to reduce the dimensionality of the feature vectors, using the rank sum test and the expected cross-entropy. We validated the feasibility of the proposed approach using two medical datasets, a clinical dataset of ultrasound images of thyroid nodules in China and a human visual optical coherence tomography (OCT) dataset.

The remainder of this paper is structured as follows. Section 2 gives the definitions and application principles of the basic model. Section 3 details the proposed method. Section 4 presents the classification results of two datasets, and Section 5 concludes our work and discusses of the future.

2. Method Description

2.1. Datasets. In this study, we used two different datasets, hereafter referred to as dataset 1 [38] and dataset 2 [39]. Dataset 1 contained 93 malignant and 415 benign ultrasound images of thyroid nodules, which were provided by the Third Hospital of Hebei Medical University in China. Each image in this dataset represented one patient and contained only one nodule. Dataset 2 was an open-source retinal OCT image dataset, and each image was manually labeled by ophthalmologists. This dataset included volume scans of 45 patients, 15 normal patients, 15 patients with dry AMD, and 15 patients with DME.

2.2. N -Gram Method. The n -gram model is a bag-of-phrases model with n -consecutive words as elements. For example, if there is a hexadecimal string “09 EC 84 35 A6 78,” the 2-gram corpus of the string will be “09 EC,” “EC 84,” “84 35,” “35 A6,” and “A6 78.” The feature division process is shown in Figure 1.

In the n -gram model [40], the storage space, computation, and computational complexity increase with n . Therefore, the default value of n is generally 1, 2, and 3. At this point, there are at most $W + W^2 + W^3$ possible word combinations in a text containing W words.

2.3. Application of N -Gram Model in Image Classification. The n -gram model is based on three key principles. First, it

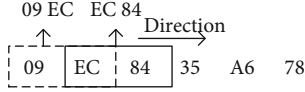


FIGURE 1: Hex string 2-gram partition example.

chooses a part of the dataset as a template. Second, it searches all the segments to get the occurrence frequency of each segment. Finally, it identifies the representative segments of the dataset. To combine the n -gram model with image processing tightly and effectively, a new model was proposed in this paper based on the above principles. This model could describe the deeper semantic information of an image with a limited but representative combination of words. The details of the proposed technique are given in the next chapter.

3. Method Improvement

3.1. Image Preprocessing. The adjacent tissues of thyroid nodules in ultrasound images could interfere with the classification process. The accurate segmentation of the nodule area is conducive to extracting the essential characteristics of pathology. In this paper, the thyroid nodule images were segmented under the guidance of pathologists. On this basis, in order to remove the influence of the nodule shape angle, the longest diameter of the nodule was placed in a horizontal position. The preprocessing procedures are shown in Figure 2.

Speckle noise is the main reason of poor OCT image quality [41]. In this paper, the anisotropic diffusion model is used to denoise retinal OCT images [42, 43]. Another problem is the irregular white edges of the pictures. The solution is to identify the edges and then fill the deselected area with black. The process is shown in Figure 3.

3.2. Medical Image Feature Extraction Based on N -Gram Model

3.2.1. Transformation from Medical Image to Character Matrix. All images in this paper are represented in grayscale, with values between 0 and 255. This is equivalent to giving a dictionary with 256 words. Then, the number of all possible n -grams is 256^n . When " $n = 1, 2, \text{ and } 3$," the number is $\sum_{n=1}^3 256^n$. Such high-dimensional data can cause information redundancy or noise and dimensional disasters. Therefore, we converted the gray value matrix into a character matrix consisting of 16 letters point by point by using the partition function as shown in the following equation.

$$y_{ij} = \begin{cases} A, x_{ij} \in [0, 15] \\ B, x_{ij} \in [16, 31] \\ \vdots \\ P, x_{ij} \in [240, 255] \end{cases}, x_{ij} \in X, y_{ij} \in Y. \quad (1)$$

In the above formula, X is the numerical matrix of the grayscale image, x_{ij} is the grayscale value at (i, j) th of X , Y

is the character matrix, and y_{ij} is the characters located at the (i, j) th element of Y .

The primary purpose of this step is to convert the image to text to reduce the dimensionality of the image. As shown in equation (3), the corresponding character matrix Y can be calculated by equation (1) for a given numerical matrix X .

$$X = \begin{bmatrix} 255 & 255 & 255 & \cdots & 255 & 255 & 255 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 204 & 174 & 165 & \cdots & 84 & 147 & 223 \\ 211 & 196 & 192 & \cdots & 135 & 200 & 249 \\ 213 & 200 & 189 & \cdots & 174 & 239 & 255 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 255 & 255 & 255 & \cdots & 255 & 255 & 255 \end{bmatrix}, \quad (2)$$

$$\Rightarrow Y = \begin{bmatrix} P & P & P & \cdots & P & P & P \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ M & K & K & \cdots & F & G & N \\ N & M & M & \cdots & I & M & P \\ N & M & L & \cdots & K & O & P \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ P & P & P & \cdots & P & P & P \end{bmatrix}. \quad (3)$$

3.2.2. Feature Extractor Based on N -Gram Model. Text retrieval uses a contiguous sequence of n words to represent an n -gram. Similarly, we collected all n -consecutive letters from the character matrix Y and defined them as image features. A feature can be interpreted as a combination of grayscale intensities of n -consecutive pixels in an image. The detailed process was as follows.

First, a picture in the dataset was converted into a character matrix. Then, all combinations of characters were extracted as a subset of features by letting a window of length n characters slide from left to right in each row of the matrix. The above procedure was repeated for each image in the dataset. All the feature subsets were combined to be the set of features of the whole dataset.

The n -gram ($n = 1, 2, \text{ and } 3$) features used to describe the entire dataset in this paper were generated from the training set.

3.2.3. Calculation of Feature Values. The method of calculating the eigenvalues was crucial to the classification ability of the features in the model. So this paper used both Term Frequency (TF) and Term Frequency-Relevance Frequency (TF-RF) [44] methods.

TF represents the frequency of a given word t_i in a document d_j . The frequency value was positively correlated with the importance of the word to the text. The word frequency of t_i can be expressed as

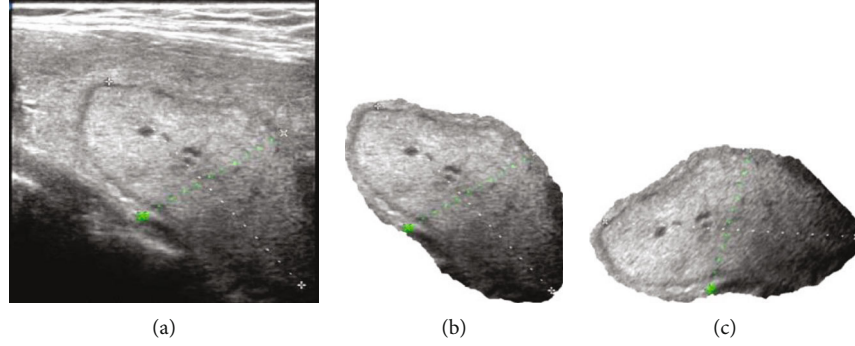


FIGURE 2: Ultrasonic image preprocessing of thyroid nodules. (a) The original image. (b) Image after removing adjacent tissues of nodules. (c) Image after adjusting nodule angle.

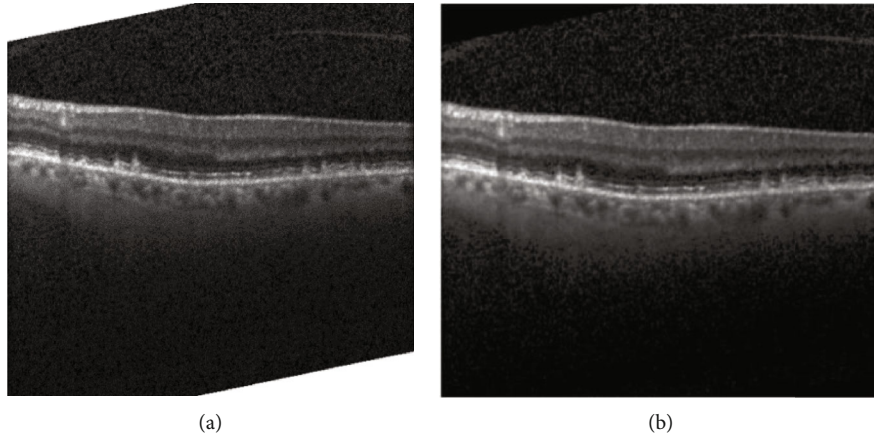


FIGURE 3: Retinal OCT image preprocessing. (a) The original image. (b) The preprocessed image.

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}. \quad (4)$$

In the above formula, n_{ij} is the number of occurrences of word t_i in document d_j and $\sum_k n_{kj}$ is the sum of the number of occurrences of all words in document d_j .

The basic idea of RF is that the higher the concentration of a high-frequency term in the positive category, the greater its contribution to selecting positive samples from the negative category. For the word t_i in the corpus, its formula is expressed as

$$rf_i = \log_2 \left(2 + \frac{a}{\max(1, c)} \right). \quad (5)$$

In the above formula, a is the number of documents containing the word t_i and belonging to the positive category and c is the number of documents containing the word t_i and belonging to the negative category. Considering the imbalance of medical image samples, the final TF-RF calculation formula is as follows:

$$tf_rf_i = tf_i * \log_2 \left(2 + \frac{N}{P} * \frac{a}{\max(1, c)} \right). \quad (6)$$

TABLE 1: Parameter settings for both classifiers.

libsvm		newff			
c	g	Si	Epochs	Goal	lr
10	0.01	3, 6, 3	100	$1e-5$	0.01

In the above formula, P is the total number of positive categories in the training set and N is the total number of negative categories in the training set.

3.3. Feature Selection. A new feature selection method with two steps was applied to remove redundant and irrelevant features. The detailed steps were as follows. In step one, a rank sum test [45] was performed for each feature using the training set data, and only the features with significant differences were retained, with the following assumptions.

$$\begin{aligned} H_0 &: \text{The two populations have equal variance or spread, } \alpha = 0.05. \\ H_1 &: \text{The two populations are independent of one another,} \end{aligned} \quad (7)$$

If the calculated p value for a feature was less than α , H_0 was rejected, meaning that the difference between the features

Input: A gray image f in the test set.
Output: Category labels.

1. Make all the extracted salient features form a set T' .
 - a. Convert an image x to a character matrix d_x .
 - b. Extract n -gram ($n = 1, 2, \text{ and } 3$) corpus of d_x , and make the corpus of the whole training set form a set T .
 - c. Select the feature subset T' by testing the feature vector T_x of T .
2. Train a classifier.
 - a. Select the subset T_x' from the feature vector T_x of the training set according to T' .
 - b. Train a classifier with T_x' .
3. Categorize test sets.
 - a. Convert the image f to a character matrix d_f .
 - b. Calculate the eigenvectors T_f' of set T' using the matrix d_f .
 - c. Use T_f' as the input of the classifier.
 - d. Output the label of f .

ALGORITHM 1: Proposed algorithm.

on positive and negative samples was statistically significant. Only such features were chosen for subsequent analysis.

In step two, based on the previous step, the importance of each feature was measured using the expected cross-entropy (ECE) [46], and the features were sorted in descending order by entropy value. The formula was shown as follows.

$$\text{ECE} = p(t_i) * \sum_j p(C_j|t_i) * \log_2 \left(\frac{p(C_j|t_i)}{p(C_j)} \right), \quad (8)$$

where C_j is the j th category, $p(t_i)$ is the sample frequency of feature t_i , and $p(C_j|t_i)$ is the sample frequency of category C_j under the condition of feature t_i .

3.4. Classification. On each dataset, we trained two different classifiers, the Support Vector Machine (SVM) [47, 48] and Back Propagation (BP) neural network [49], and ran the classification experiments using 10-fold cross-validation. All experiments were performed using MATLAB R2019a.

The first experiment is to diagnose whether a nodule is benign or malignant based on dataset 1. We used the receiver operating characteristic (ROC) curve as well as the area under the ROC curve (AUC) to evaluate the performance of the method in this experiment. The closer the ROC curve is to the upper left corner, the closer the AUC value is to 1, indicating better performance [50]. The second experiment is to diagnose whether the type of fundus lesion is AMD, DME, or normal based on dataset 2. Its classifier is a ternary-classifier consisting of three small two-classifiers. The classification result is determined by the vote of each subclassifier. The parameter settings of all classifiers are shown in Table 1, and other parameters not listed in the table were default values.

4. Discussion

According to the characteristics of medical images, an improved classification algorithm based on the n -gram model was proposed in this paper. The algorithm has been

TABLE 2: Characteristic dimension.

Mode	Essential feature	Distinguishing feature			
		100%	75%	50%	25%
TF	3590	1463	1098	732	366
TF-RF	3590	1338	1004	669	335

described in detail previously. The core steps of the algorithm were described here in conjunction with specific experiments, as shown in Algorithm 1.

Take the experiment with dataset 1 as an example. First, the basic feature set T and the feature matrix T_x are obtained from the training set as described in Section 3.2. Second, a subset T' of T and a submatrix T_x' of T_x are selected, according to the method in Section 3.4. Next, we train the classifier using T_x' . Finally, we predict the labels of the test set samples, as shown in Step 3 in Algorithm 1.

4.1. Experimental Results of Dataset 1. The impact of two eigenvalue calculation methods and four feature dimensions on classification was explored using dataset 1. The specific work is as follows. First, the eigenvalues of the training set were calculated using the two equations TF and TF-RF introduced in Section 3.2.3 to obtain two feature matrices. Second, the same two-step operation to select the submatrices to control the dimensionality was performed for these two matrices as follows. The first step was to calculate the ECE values of the features after selecting the features with the rank sum test according to Section 3.3. The second step was that features were ranked according to ECE values, and 100%, 75%, 50%, and 25% features were selected from the highest to the lowest. After the above work, eight feature submatrices were obtained as the input matrix for training the classifier.

Based on the average of the 10-fold cross-validation results, we observed that the number of valid features was less than the number of all features, and the method of calculating the feature values had an impact on the validity of the features. The variation of feature dimensions is shown in Table 2.

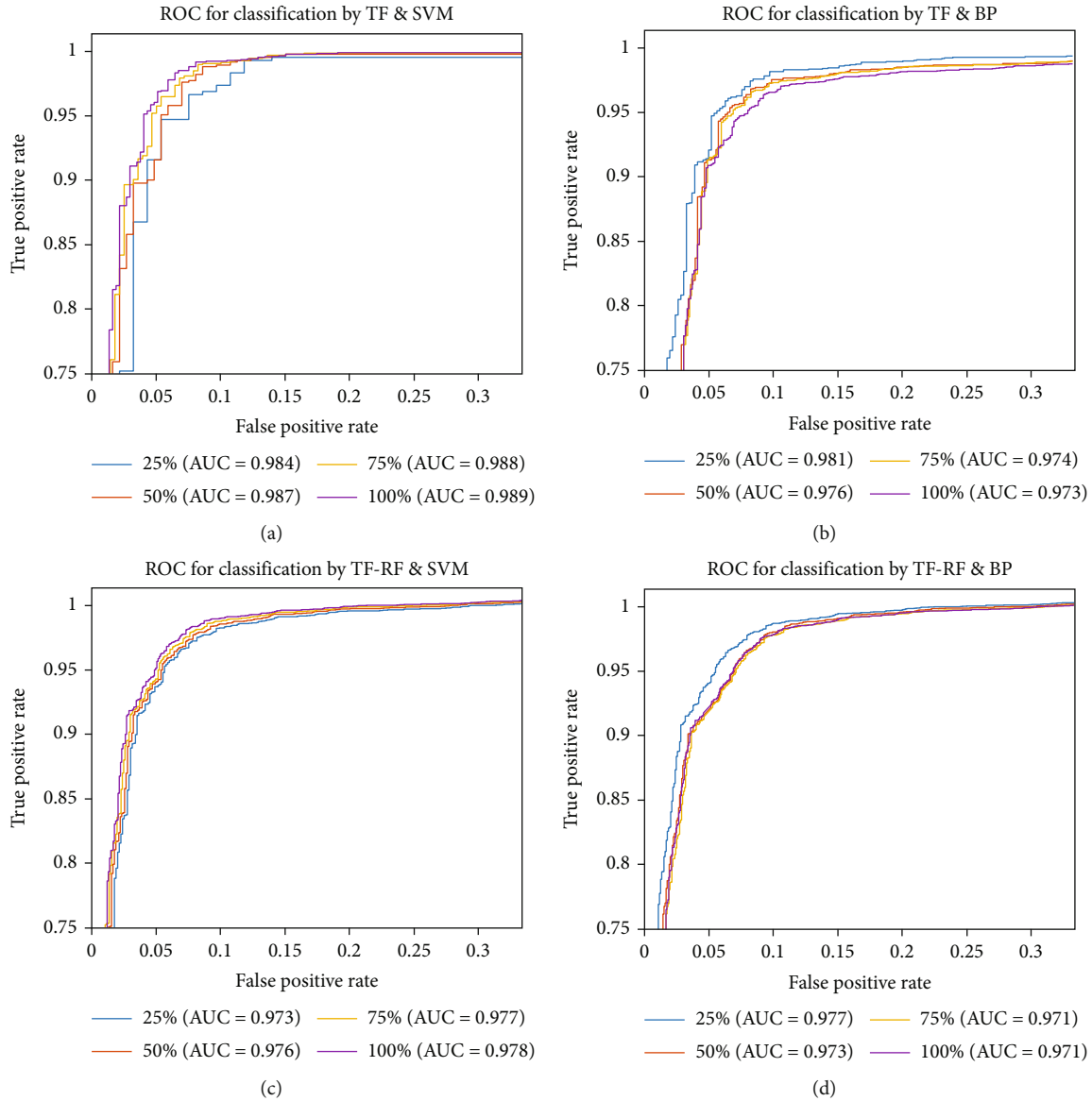


FIGURE 4: The results of 10-fold cross-validation. (a) The classifier with SVM and TF achieves the best AUC of 0.989 using all features as inputs. (b) The classifier with BP and TF achieves the best AUC of 0.981 using 25% features as inputs. (c) The classifier with SVM and TF-RF achieves the best AUC of 0.978 using all features as inputs. (d) The classifier with BP and TF-RF achieves the best AUC of 0.977 using 25% features as inputs.

TABLE 3: Fraction of correctly classified during cross-validation.

Class	TF-SVM	TF-BP	TF-RF-SVM	TF-RF-BP
AMD	13/15 = 86.667%	13/15 = 86.667%	13/15 = 86.667%	12/15 = 80.000%
DME	13/15 = 86.667%	11/15 = 73.333%	14/15 = 93.333%	12/15 = 80.000%
Normal	14/15 = 93.333%	14/15 = 93.333%	14/15 = 93.333%	14/15 = 93.333%

In summary, we conducted a total of 160 experiments on dataset 1, and the cross-validation results are shown in Figure 4. The best diagnostic result was achieved in the experiment with an AUC of 0.989 under the condition that the input was all valid features, the classifier was an SVM, and TF was used to calculate the eigenvalues.

Overall, this experiment showed that SVM outperformed the BP neural network for classification and TF outperformed TF-RF for discriminating samples. It also showed that when 25% of the valid features were selected, the AUC fluctuated only 0.5% from the optimal level, which could balance feature dimensionality and accuracy. The average

compression rate of the algorithm for this dataset is 4.7% based on the 126M original size of the dataset and the 5.91M average size of the feature matrix in the 10-fold cross-validation. The average computation time for this experiment is 267 ms per image.

4.2. Experimental Results of Dataset 2. The triple classification experiment for dataset 2 selected 25% of the valid features as input to the SVM and BP neural network classifiers, and its 10-fold cross-validation results are shown in Table 3. When the formula for calculating feature values was TF-RF and the classifier was SVM, this experiment achieved the best diagnostic result of correctly identifying 86.667% of AMD patients, 93.333% of DME patients, and 93.333% of normal patients.

Overall, this experiment shows that SVM outperforms the BP neural network for classification and TF-RF discriminates the samples better than TF. The average compression rate of the algorithm for this dataset was calculated to be 6.7% based on the original size of the dataset of 946M and the average size of the feature matrix of 10-fold cross-validation of 63M. The average computation time for this experiment is 760 ms per image.

5. Conclusions and Future Works

In this paper, a new medical image classification method is proposed and experimentally validated using two datasets. Dataset 1 is clinically unbalanced ultrasound images of thyroid nodules, and dataset 2 is OCT images of the fundus. The algorithm showed good classification performance on both datasets, which reflected its generalization ability.

The algorithm performs the rank sum test for each feature to ensure its robustness. Therefore, the performance of the algorithm remains stable under the condition that the preprocessing is coarse, and only denoising is performed without steps such as histogram equalization, brightness correction, and geometric transformation. One of the key works is to propose an algorithm for extracting image features based on the n -gram model, which maps the image into a character matrix and then applies a text-processing approach to extract image features instead of relying on traditional morphological and texture features. Another key work is to reduce the number of features by selecting features in two steps. The intent of these two works is to express the deep semantic information of the image with the appropriate number of features. In conclusion, the method is simple and efficient compared to more complex models and has important theoretical implications and potential for clinical applications.

In the future, we will try to optimize the model in several aspects such as spatial information [51, 52], distribution of features [53, 54], clustering of features [55, 56], selection of features [57], and evaluation of features [58]. We explore the use of some representative computational intelligence algorithms [59, 60] to solve the problem, such as monarch butterfly optimization (MBO), earthworm optimization algorithm (EWA), elephant herding optimization (EHO), moth search (MS) algorithm, slime mould algorithm

(SMA), hunger games search (HGS), Runge-Kutta optimizer (RUN), colony predation algorithm (CPA), and Harris hawks optimization (HHO).

Data Availability

We have uploaded the source code to the GitHub (<https://github.com/Wanggf618/Medical-Image-Classification-Algorithm-Based-on-N-Gram-Model.git>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Guanfang Wang and Xianshan Chen contributed equally to this work.

References

- [1] G. J. Tearney, M. E. Brezinski, B. E. Bouma et al., "In vivo endoscopic optical biopsy with optical coherence tomography," *Science*, vol. 276, no. 5321, pp. 2037–2039, 1997.
- [2] L. G. Quekel, A. G. Kessels, R. Goei, and J. M. van Engelshoven, "Miss rate of lung cancer on the chest radiograph in clinical practice," *Chest*, vol. 115, no. 3, pp. 720–724, 1999.
- [3] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang, "Computer-aided diagnosis in radiology: potential and pitfalls," *European Journal of Radiology*, vol. 31, no. 2, pp. 97–109, 1999.
- [4] U. Raghavendra, A. Gudigar, M. Maithri et al., "Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images," *Computers in Biology and Medicine*, vol. 95, pp. 55–62, 2018.
- [5] M. Yu, M. Han, X. Li et al., "Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: thyroid ultrasound image case study," *Computers in Biology and Medicine*, vol. 144, p. 105347, 2022.
- [6] D. Koundal, B. Sharma, and Y. Guo, "Intuitionistic based segmentation of thyroid nodules in ultrasound images," *Computers in Biology and Medicine*, vol. 121, p. 103776, 2020.
- [7] F. Abdolali, J. Kapur, J. L. Jaremko, M. Noga, A. R. Hareendranathan, and K. Punithakumar, "Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 122, p. 103871, 2020.
- [8] M. Pekala, N. Joshi, T. Y. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Computers in Biology and Medicine*, vol. 114, p. 103445, 2019.
- [9] J. Kauer-Bonin, S. K. Yadav, I. Beckers et al., "Modular deep neural networks for automatic quality control of retinal optical coherence tomography scans," *Computers in Biology and Medicine*, vol. 141, p. 104822, 2022.
- [10] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Studies in Media and Communication*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [11] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

- [12] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [13] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004.
- [14] C. Wang and K. Huang, "How to use bag-of-words model better for image classification," *Image and Vision Computing*, vol. 38, 2015.
- [15] G. Algan and I. Ulusoy, "Label noise types and their effects on deep learning," 2020, <https://arxiv.org/abs/2003.10471>.
- [16] H. Long, B. Liao, X. Xu, and J. Yang, "A hybrid deep learning model for predicting protein hydroxylation sites," *International Journal of Molecular Sciences*, vol. 19, no. 9, p. 2817, 2018.
- [17] H. Liu, C. Qiu, B. Wang et al., "Evaluating DNA methylation, gene expression, somatic mutation, and their combinations in inferring tumor tissue-of-origin," *Frontiers in Cell and Developmental Biology*, vol. 9, p. 619330, 2021.
- [18] H. Lee, S. Tajmir, J. Lee et al., "Fully automated deep learning system for bone age assessment," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 427–441, 2017.
- [19] X. Gong, X. Xia, W. Zhu, B. Zhang, D. Doermann, and L. A. Zhuo, "Deformable Gabor feature networks for biomedical image classification," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4003–4011, Waikoloa, HI, USA, 2021.
- [20] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.
- [21] D. Forsyth, "Object detection with discriminatively trained part-based models," *Computer*, vol. 47, no. 2, pp. 6–7, 2014.
- [22] J. Sivic and A. Zisserman, "Video Google: a text retrieval approach to object matching in videos," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Nice, France, 2003.
- [23] H. Sun, X. Zhang, X. Han, X. Jin, and Z. Zhao, "Commodity image classification based on improved bag-of-visual-words model," *Complexity*, vol. 2021, Article ID 5556899, 10 pages, 2021.
- [24] R. Zebari, A. Mohsin Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.
- [25] A. Alharan, H. Fatlawi, and N. Ali, "A cluster-based feature selection method for image texture classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1433–1442, 2019.
- [26] Z. Li, H. Zhang, W. Chen, and H. Li, "Contrast-enhanced CT-based radiomics for the differentiation of nodular goiter from papillary thyroid carcinoma in thyroid nodules," *Cancer Management and Research*, vol. 14, pp. 1131–1140, 2022.
- [27] Z. Lai and H. Deng, "Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 2061516, 13 pages, 2018.
- [28] U. Raghavendra, U. Rajendra Acharya, A. Gudigar et al., "Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions," *Ultrasonics*, vol. 77, pp. 110–120, 2017.
- [29] Q. Yu, T. Jiang, A. Zhou, L. Zhang, C. Zhang, and P. Xu, "Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images," *European Archives of Oto-Rhino-Laryngology*, vol. 274, no. 7, pp. 2891–2897, 2017.
- [30] A. A. Ardakani, A. Mohammadzadeh, N. Yaghoubi et al., "Predictive quantitative sonographic features on classification of hot and cold thyroid nodules," *European Journal of Radiology*, vol. 101, pp. 170–177, 2018.
- [31] P. Brown, V. Dellapetra, P. Souza, J. Lai, and R. Mercer, "Class-based n-gram models of natural language," *Computational Linguistics*, vol. 18, pp. 467–479, 1992.
- [32] A. Broder, S. Glassman, M. Manasse, and G. Zweig, "Syntactic clustering of the web," *Computer Networks and ISDN Systems*, vol. 29, no. 8–13, pp. 1157–1166, 1997.
- [33] E. Miller, D. Shen, J. Liu, C. Nicholas, and T. Chen, "Techniques for gigabyte-scale N-gram based information retrieval on personal computers," in *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications*, pp. 1410–1416, Las Vegas, Nevada, USA, June 28 - July 1, 1999.
- [34] C. Shaoul, H. Baayen, and C. Westbury, "N-gram probability effects in a cloze task," *The Mental Lexicon*, vol. 9, no. 3, pp. 437–472, 2014.
- [35] U. Nandagopal and S. Thirumalaivelu, "Classification of malware with MIST and N-gram features using machine learning," *International Journal of Intelligent Engineering and Systems*, vol. 14, 2021.
- [36] J. Milička, V. Cvrček, and L. Lukešová, "Modelling crosslinguistic n-gram correspondence in typologically different languages," *Languages in Contrast*, vol. 21, no. 2, pp. 217–249, 2021.
- [37] S. Sabra and V. Sabeeh, "A comparative study on N-gram and skip-gram for clinical concepts extraction," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2020.
- [38] X. Ma, B. Xi, Y. Zhang et al., "A machine learning-based diagnosis of thyroid cancer using thyroid nodules ultrasound images," *Current Bioinformatics*, vol. 15, no. 4, pp. 349–358, 2020.
- [39] P. P. Srinivasan, L. A. Kim, P. S. Mettu et al., "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical Optics Express*, vol. 5, no. 10, pp. 3568–3577, 2014.
- [40] M. Farhoodi, A. Yari, and A. Sayah, "N-gram based text classification for Persian newspaper corpus," in *The 7th International Conference on Digital Content, Multimedia Technology and its Applications*, Busan, South Korea, 2011.
- [41] X. Fu, J. Wang, Z. Hu, Y. Guo, and R. Wang, "Reference variance and adjustment factor based wave atoms image speckle reduction method for optical coherence tomography images," *Optik*, vol. 224, p. 165651, 2020.
- [42] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [43] Y. Yu and S. Acton, "Speckle reducing anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260–1270, 2002.

- [44] M. Lan, C. L. Tan, J. Su, and Y. Lu, "Supervised and traditional term weighting methods for automatic text categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 721–735, 2009.
- [45] H. Mann and D. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, vol. 18, 1947.
- [46] D. Mladenic and M. Grobelnik, "Feature selection for unbalanced class distribution and naive Bayes," in *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 258–267, Bled, Slovenia, 1999.
- [47] C. Cortes and V. Vapnik, "Support-vector networks," *Chemical Biology and Drug Design*, vol. 297, pp. 273–297, 2009.
- [48] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [49] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [50] F. Luo, M. Wang, Y. Liu, X. M. Zhao, and A. Li, "DeepPhos: prediction of protein phosphorylation sites with deep learning," *Bioinformatics*, vol. 35, no. 16, pp. 2766–2773, 2019.
- [51] R. Wang, Z. Wu, J. Lou, and Y. Jiang, "Attention-based dynamic user modeling and deep collaborative filtering recommendation," *Expert Systems with Applications*, vol. 188, p. 116036, 2022.
- [52] Z. Wu, R. Wang, Q. Li et al., "A location privacy-preserving system based on query range cover-up or location-based services," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5244–5254, 2020.
- [53] Z. Wu, S. Shen, X. Lian, X. Su, and E. Chen, "A dummy-based user privacy protection approach for text information retrieval," *Knowledge-Based Systems*, vol. 195, p. 105679, 2020.
- [54] Z. Wu, R. Li, Z. Zhou, J. Guo, J. Jiang, and X. Su, "A user sensitive subject protection approach for book search service," *Journal of the Association for Information Science and Technology*, vol. 71, no. 2, pp. 183–195, 2020.
- [55] B. Bai, G. Li, S. Wang, Z. Wu, and W. Yan, "Time series classification based on multi-feature dictionary representation and ensemble learning," *Expert Systems with Applications*, vol. 169, p. 114162, 2021.
- [56] W. Yan, G. Li, Z. Wu, S. Wang, and P. S. Yu, "Extracting diverse-shapelets for early classification on time series," *World Wide Web*, vol. 23, no. 6, pp. 3055–3081, 2020.
- [57] Z. Wu, H. Zhu, G. Li et al., "An efficient Wikipedia semantic matching approach to text document classification," *Information Sciences*, vol. 393, pp. 15–28, 2017.
- [58] Z. Wu, L. Lei, G. Li et al., "A topic modeling based approach to novel document automatic summarization," *Expert Systems with Applications*, vol. 84, pp. 12–23, 2017.
- [59] M. Li and G. G. Wang, "A review of green shop scheduling problem," *Information Sciences*, vol. 589, pp. 478–496, 2022.
- [60] W. Li, G. G. Wang, and A. H. Gandomi, "A survey of learning-based intelligent optimization algorithms," *Archives of Computational Methods in Engineering*, vol. 28, no. 5, pp. 3781–3799, 2021.