Research Article

Breast Cancer Recognition Using Saliency-Based Spiking Neural Network

Qiang Fu and Hongbin Dong

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Hongbin Dong; donghongbin@hrbeu.edu.cn

Received 19 January 2022; Accepted 14 March 2022; Published 24 March 2022

Academic Editor: M. Hassaballah

Copyright © 2022 Qiang Fu and Hongbin Dong. 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.

The spiking neural networks (SNNs) use event-driven signals to encode physical information for neural computation. SNN takes the spiking neuron as the basic unit. It modulates the process of nerve cells from receiving stimuli to firing spikes. Therefore, SNN is more biologically plausible. Although the SNN has more characteristics of biological neurons, SNN is rarely used for medical image recognition due to its poor performance. In this paper, a reservoir spiking neural network is used for breast cancer image recognition. Due to the difficulties of extracting the lesion features in medical images, a salient feature extraction method is used in image recognition. The salient feature extraction network is composed of spiking convolution layers, which can effectively extract the features of lesions. Two temporal encoding manners, namely, linear time encoding and entropy-based time encoding methods, are used to encode the input patterns. Readout neurons use the ReSuMe algorithm for training, and the Fruit Fly Optimization Algorithm (FOA) is employed to optimize the network architecture to further improve the reservoir SNN performance. Three modality datasets are used to verify the effectiveness of the proposed method. The results show an accuracy of 97.44% for the BreastMNIST database. The classification accuracy is 98.27% on the mini-MIAS database. And the overall accuracy is 95.83% for the BreaKHis database by using the saliency feature extraction, entropy-based time encoding, and network optimization.

1. Introduction

The incidence rate of breast cancer ranks first among female malignancies [1]. How to early detect, standardize diagnosis, and reduce mortality is a major challenge facing the medical community all over the world. Medical image analysis is an important means of early detection of breast cancer lesions risk. It is also recognized as the most effective and reliable medical examination method, and it can find the related clues of early breast cancer lesions. The combination of early prevention and control measures can significantly reduce the cancer rate and mortality of the screening population. The current medical technology for breast cancer has no effective means of prevention. How to detect early is an important way to combat breast tumors. With the rapid development of computer technology, the use of the computer to achieve digital diagnosis systems has become the development trend of medical technology. Computer-aided diagnosis of breast tumors has become an inevitable trend.

In recent years, many deep learning detection methods based on breast images have been proposed [2–4]. However, training deep neural networks (DNNs) requires too many samples and annotations. Moreover, since the DNNs contain multiple hidden layers, training a DNN needs a lot of computing resources and time costs. The acquisition of medical images is not easy, and senior doctors need to label the lesions. Compared with DNN, SNN is a more available tool. SNN simulates the information processing mechanism of the brain and takes the spike neurons as the basic units. It simulates the information processing mechanism of nerve cells, to realize the pattern recognition task [5]. SNN represents information in spike patterns, and each spike neuron experiences rich dynamic behaviors. In addition to the spread of information in the spatial domain, the
current state is also affected by the past state in the time domain specifically. Therefore, SNN usually has more time commonality. Due to the biological characteristics of SNN, it has a lower computational cost in theory. However, the accuracy of SNN is lower than that of ANN realized by spatial propagation and continuous activation function. On the other hand, SNN is rarely used in the field of medical image recognition. Therefore, this paper focuses on the SNN for breast cancer image recognition.

In this paper, SNN is employed to recognize breast cancer on three modalities of datasets. The network consists of three parts (i.e., input layer, reservoir layer, and readout layer). In the input layer, the image is encoded to spike time, and the saliency features of the input image are extracted to better identify the lesion information. To obtain the optimal architecture, an evolutionary algorithm, i.e., FOA, is employed to optimize the architecture of the SNN.

The main contributions of this work are as follows.

1. Two-time encoding schemes (i.e., linear time encoding and entropy-based time encoding) are proposed to encode the input information. The encoding methods are compared and analyzed in the

Figure 1: Schematic diagram of the proposed methods.

Figure 2: Schematic diagram of the three-layer structure in the cerebellum with that in LSM.

Figure 3: Illustration of linear spike time encoding of the image.
experimental part. The manner of linear time encoding maps pixels of the input image into spike time linearly. The entropy-based encoding scheme calculates the statistical form of image features. This method includes not only the aggregation feature of image-level but also the spatial feature of image distribution.

(2) A method of saliency feature extraction is proposed to detect the lesion features of breast images so that the network can better detect the images containing lesions. This method uses a spike convolution network to extract input features. According to the output feature maps, the heat map of the image is reconstructed, and then, the mask is calculated to obtain the salient region of the image.

(3) To improve the SNN performance, an evolutionary computing method, namely, Fruit Fly Optimization Algorithm (FOA) [6], is employed to optimize the SNN architecture. The optimized SNN via FOA can improve the classification accuracy and reduce the number of synaptic connections in the reservoir. The structure of the network in our work is determined by the neurons and the connections of neurons in the reservoir layer. The network structure determines future performance in a great measure. The super parameters that determine the structure

```python
# Random initial fruit fly swarm location, radius (R) and number of optimization variables (D):
Init X_axis, Y_axis, Z_axis, R, D
# Enter iterative optimization to repeat the implementation, then judge if the smell concentration is superior to the previous iterative smell concentration.
for i in range (popsize):
    X(i) = X_axis + R*(random (1,D)-1)
    Y(i) = Y_axis + R*(random (1,D)-1)
    Z(i) = Z_axis + R*(random (1,D)-1)
    # The distance to the origin is estimated.
    Dist(i) = sqrt(X(i)**2+Y(i)**2+Z(i)**2)
    # the smell concentration judgment value is calculated.
    S(i) = 1/Dist(i)
    # Substitute smell concentration judgment value (S) into smell concentration judgment function.
    Smell(i) = function(S(i))
    # Finding out the best value and the location index
    Smellbest, index = min(Smell), argmin(Smell)
    bestSmell = Smellbest
    X_axis = X(index)
    Y_axis = Y(index)
    Z_axis = Z(index)
```

Algorithm 1

Figure 4: Saliency-based feature extraction using spiking CNN. The spike convolution is used to obtain the 2D feature maps; then, the mask map is calculated by 2D feature maps (i.e., the process indicated by the red arrow). The original image and mask image are used to calculate the dot product to generate the saliency feature map, and the symbol ⊙ indicates the dot product.
are usually empirical values. It is difficult to select the appropriate value manually. Therefore, the evolutionary algorithm is used to obtain the optimal structure in this work.

The rest of the paper is organized as follows. The related works are provided in Section 2. Section 3 proposes two methods of spike time encoding, the saliency feature extraction method, and the spiking reservoir neural network. Section 4 presents a manner of SNN architecture optimization using FOA. The experimental results are provided in Section 5, which demonstrates the performance of the proposed methods under three breast cancer recognition tasks. Section 6 concludes this paper.

### 2. Related Works

Interventional therapy of medical imaging has significantly improved the level of early diagnosis of breast cancer. With the application of artificial intelligence in the field of healthcare, researchers use image processing and computer vision technology to design effective intelligent computer-aided detection and diagnosis systems.

Deep learning has made great progress in the field of pattern recognition. Various DNNs based feature extraction architectures are proposed for breast cancer detection and classification [7]. A deep convolution neural network (CNN) framework is proposed by [8]. Features of input images are extracted by using three pretrained CNN architectures. Features are fed into a fully connected layer for the recognition of malignant and benign tumors using average pooling. Whole slide images of breast biopsies are classified into five categories, and a saliency detector is proposed in [9], which uses a pipeline of four fully CNNs. The network fused saliency and classification maps for final categorization. An automated multiscale end-to-end DNN framework is proposed for mammogram classification [10]. It only requires mammogram images with annotations. The model generates three scales of feature maps that make the classifier combine global information with the local lesions for classification. To detect breast cancer automatically, the software is developed by [11]. It also proposes an algorithm to extract the features based on biodata, image analysis, and image statistics. These features are used to distinguish the images as normal or suspected by using CNNs and optimized by the Bayes algorithm. [12] proposes a hybrid deep CNN and RNN to classify breast cancer histopathological image.

Unlike deep CNNs, there is limited work done for SNNs in the field of breast cancer detection and diagnosis. [13] proposes that the discontinuous spike time is regarded as noise and the membrane potential is regarded as continuous signal, so that SNN can be trained by back propagation algorithm based on gradient descent. [14] proposes to train deep...
3. Proposed Methods

A spiking reservoir neural network architecture, two manners of spike encoding (i.e., linear time encoding and entropy-based time encoding), and a saliency-based feature detection network are presented in this section. A diagram of the proposed method is shown in Figure 1. Compared with the existing spike encoding schemes, the linear spike time encoding method proposed in this paper is more concise in the calculation. The entropy-based spike time encoding calculates the statistical form of features and includes the spatial characteristics of the gray distribution.

3.1. Spiking Reservoir Neural Network Architecture. SNN is the third generation artificial neural network (ANN), which is most biologically inspired and event-driven. A recursive architecture of the SNN, namely, the spike reservoir neural network, to perform breast image recognition is employed in this work. Spike reservoir neural network, also called liquid state machine (LSM) [16], contains an input layer, reservoir computing layer, and readout layer. LSM is a real-time neural network model, which transforms time-varying input flow into high-dimensional space. The study in [17] linked the three-layer structure in the cerebellum (i.e., granule layers, mossy fibers, and Purkinje cells) with that in LSM. In the cerebellar cortex, Purkinje cells learn connections between learning processes from the high-dimensional representation of information. The input of granule cells is random and will not be modified in the learning process; only Purkinje cells can learn through unsupervised plasticity [18]. This form is similar to the learning process of LSM, as shown in Figure 2. Besides, the information processing mechanism in the prefrontal cortex (PFC) region of the brain is also assumed to be similar to LSM [19]. The mixing selectivity was studied in [19]. The PFC was simulated by the hidden layer, and the striatum was simulated by the readout layer. And [19] discussed the characteristics of reservoir computing networks observed in the cortical regions of cats, rats, and monkeys [18].

As shown in Figure 2, LSM includes three parts, i.e., input layer, reservoir or liquid layer, and a memoryless readout layer. The reservoir consists of Leaky Integrate and Fire (LIF) neurons. The LIF neurons are connected recursively through dynamic synaptic connections in the reservoir part. The red solid circle in the reservoir represents the excitatory neurons, and the black represents the inhibitory neurons. The number of excitatory neurons is selected as four times the number of input neurons (i.e., linear time encoding and entropy-based time encoding) of the proposed method is shown in Figure 1. Compared with the existing spike encoding schemes, the linear spike time encoding method proposed in this paper is more concise in the calculation. The entropy-based spike time encoding calculates the statistical form of features and includes the spatial characteristics of the gray distribution.

Table 2: The number of images in each category.

<table>
<thead>
<tr>
<th>Classes of images</th>
<th>Number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>133</td>
</tr>
<tr>
<td>Benign</td>
<td>437</td>
</tr>
<tr>
<td>Malignant</td>
<td>210</td>
</tr>
<tr>
<td>Total</td>
<td>780</td>
</tr>
</tbody>
</table>

Table 3: The network performance comparison of different methods on BreastMNIST.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Training accuracy</th>
<th>Test accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original SNN</td>
<td>74.36%</td>
<td>81.19%</td>
</tr>
<tr>
<td>SNN with linear time encoding</td>
<td>77.67%</td>
<td>89.32%</td>
</tr>
<tr>
<td>SNN with entropy-based time encoding</td>
<td>86.19%</td>
<td>92.74%</td>
</tr>
<tr>
<td>Silence-based SNN</td>
<td>85.20%</td>
<td>92.74%</td>
</tr>
<tr>
<td>Improved SNN</td>
<td>88.57%</td>
<td>97.44%</td>
</tr>
</tbody>
</table>
neuron, excitatory neuron, and inhibitory neuron is $N_p$, $N_E$, and $N_I$, respectively. Instantaneous current consists of three parts, i.e., the current of input, inhibitory, and excitatory neurons. $\delta(t - t_p)$ means that the current of the input neuron integrates the sum of presynaptic spikes, $t_p$ is the firing time, and $W_{pi}$ is the corresponding synaptic weight. Similarly, $\delta(t - t_i)$ represents the sum of presynaptic spikes of excitatory neurons, and $\delta(t - t_k)$ represents the sum of presynaptic spikes of inhibitory neurons. $W_{ji}$ is corresponding synaptic weights of excitatory neurons, and $W_{ki}$ is corresponding synaptic weights of inhibitory neurons.

The neurons in the reservoir act as liquid filters $L^M$. It maps the vector of the input spike to the vector of continuous function. When the input function is $u(t)$, $L^M$ maps $u(t)$ to internal states $x^M(t)$. It can be calculated by

$$x^M(t) = (L^Mu)(t).$$  \(3\)

Figure 7: The classification performance curve of SNN on BreastMNIST database.
The output state of the reservoir is taken as the input of the readout layer, and the state is converted to the output \( y(t) \) of the readout layer at each time \( t \). It can be calculated by
\[
y(t) = f^M(x^M(t)). \tag{4}
\]

The readout layer of LSM usually uses the exponential filter to convert the spike output of the reservoir into the output of a continuous signal for linear regression training. The LSM model in [21] is trained by the recursive least square method using liquid filtered output. It results in the loss of the accurate spike time information produced by the liquid neuron [22]. In this work, readout neurons are trained by the ReSuMe algorithm [23, 24]. The ReSuMe is a biologically logically plausible algorithm of supervised learning. It allows neurons to adjust their synaptic weights so that it can learn any spike pattern corresponding to a given synaptic stimulus. The ReSuMe can guarantee convergence for a single input spike on each synapse. In the case of multiple spikes, learning rate with a low value can be used to ensure algorithm performance. The weight adjustment can be calculated by
\[
\frac{d}{dt} w_{oi}(t) = |S_d(t) - S_o(t)| \left[ a_d + \int_0^\infty a_d(s) S_i(t-s) ds \right], \tag{5}
\]
where \( w_{oi} \) is the \( i \)-th output weights of neuron \( o \), \( S_d(t) \) is desired spike trains, and \( S_o(t) \) is actual spike trains. The learning rate is \( a_d \), and \( S_i(t) \) is an input spike train of the \( i \)-th neuron. \( a_d \) is a kernel function; it can be calculated by
\[
a_d(s) = A_d e^{-|s|/\tau_d}, \tag{6}
\]
where \( A_d \) is the constant of the learning rate and \( \tau \) is a constant.

3.2. Spike Encoding Scheme. For the human visual system (HVS), neural coding is helpful to understand how neural spike activity represents a visual scene and then decodes neural spike activity to represent given visual information. Therefore, in SNN, how to encode the numerical input data from the input image into the input neuron spikes is the primary problem to be solved. ANN uses matrix-vector operation in the transformation of input to output, which can directly calculate the value. In SNN, numerical data cannot be operated directly. On the contrary, these values must be converted into spike or spike trains. The conversion from numerical data to spike or spike trains is very important. The choice of encoding scheme not only affects the accuracy of practical application but also affects the speed of data processing and the energy efficiency of the system. This section proposes two different input encoding schemes, i.e., linear spike time encoding and entropy-based spike time encoding are proposed, and the effectiveness in medical image recognition tasks is evaluated in Section 5.

---

Table 4: Performance comparison of several networks on the BreastMNIST database (\( z = 1.96 \)).

<table>
<thead>
<tr>
<th>Methods</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-18 (28) [32]</td>
<td>0.821</td>
<td>85.9%</td>
<td>(0.8328, 0.8817)</td>
</tr>
<tr>
<td>ResNet-18 (224) [32]</td>
<td>0.857</td>
<td>87.8%</td>
<td>(0.8534, 0.8993)</td>
</tr>
<tr>
<td>ResNet-50 (28) [32]</td>
<td>0.839</td>
<td>85.3%</td>
<td>(0.826, 0.8757)</td>
</tr>
<tr>
<td>ResNet-50 (224) [32]</td>
<td>0.818</td>
<td>83.3%</td>
<td>(0.8055, 0.8578)</td>
</tr>
<tr>
<td>Auto-sklearn [33]</td>
<td>0.673</td>
<td>80.8%</td>
<td>(0.7786, 0.8338)</td>
</tr>
<tr>
<td>AutoKeras [34]</td>
<td>0.646</td>
<td>80.1%</td>
<td>(0.7719, 0.8278)</td>
</tr>
<tr>
<td>Google AutoML Vision [31]</td>
<td>0.932</td>
<td>86.5%</td>
<td>(0.8396, 0.8876)</td>
</tr>
<tr>
<td>This work</td>
<td>0.997</td>
<td>97.4%</td>
<td>(0.9608, 0.9834)</td>
</tr>
</tbody>
</table>

---

Figure 8: Illustration of the optimization process of network architecture on the BreastMNIST database.
3.2.1. Linear Spike Time Encoding. The purpose of the time encoding is to generate the corresponding spike pattern representing the input. To supply the input layer of SNN, each input representation is converted into a corresponding time pattern. The spike time encoding method uses the time of the spike to express the stimulation of the input signal. This form of time encoding is the basis for the back propagation algorithm in SNNs, such as SpikeProp [25]. Time encoding is also used with other learning algorithms, such as spike time-dependent plasticity (STDP) [26].

In this section, a linear spike time encoding scheme is presented. Using this method, the pixel information of the image can be linearly mapped to the firing time of the spike. It can be calculated by

\[ t_f^{(i)} = \frac{P_i - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}} \times T_{\text{max}} \text{ (ms)}, \]  

where \( t_f^{(i)} \) is the spike time encoded by linear method, \( P_i \) is the pixel corresponding to the input image, the minimum value of input pixels is set to \( P_{\text{min}} \), and the maximum value of input pixels is set to \( P_{\text{max}} \). \( T_{\text{max}} \) is the maximum firing time. For example, as shown in Figure 3, suppose that \( T_{\text{max}} = 10 \), the pixel range of the image is 0-255 gray value. The pixel values 0 and 110 are mapped to 0 ms and 4.314 ms, respectively.

This encoding method is a form of time-to-first-spike. According to neuroscience theory, the first spike fired by neurons carries the most information, so it is theoretically reliable.

3.2.2. Entropy-Based Spike Time Encoding. Spike time encoding method based on information entropy calculates the statistical form of image features. The average information in the image is taken as the time of firing spike. This method includes not only the aggregation feature of gray level but also the spatial feature of gray distribution. The adjacent gray value is selected as the spatial feature of the gray distribution. The adjacent gray value and the pixel of the image form a feature tuple, i.e., \((i, j)\), where \( i \) is the pixel gray value \((0 \leq i \leq 255)\) and \( j \) is the adjacent gray value \((0 \leq j \leq 255)\). It can be calculated by

\[ P_{ij} = \frac{f(i, j)}{N^2}. \]  

The above formula reflects the gray value of a pixel position and the comprehensive characteristics of the distribution, where \( f(i, j) \) is the frequency of feature tuple \((i, j)\) and \( N \) represents the image scale. The information entropy of an input image can be calculated by

\[ H = \sum_{i=0}^{255} P_{ij} \log P_{ij}. \]

According to the information entropy, the time of spike firing can be calculated by

\[ t_f^{(i)} = H \times T_{\text{max}} \text{ (ms)}, \]

where \( t_f^{(i)} \) is the time of firing spike.
The two-dimensional feature maps mainly depend on convolution calculation. It can be calculated by
\[
c_{p,q} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m,n) \cdot R(p-m, q-n),
\]
where \( c \) represents the 2D feature map generated by spike convolution, \( R \) is the receptive field, and \( M \) and \( N \) are the numbers of rows and columns of the image. The sum of multiple 2D feature maps can be calculated by
\[
C = \sum_{i=1}^{N} c_{i},
\]
where \( C \) is the sum of \( N \) 2D feature maps, \( n \) represents the number of feature maps, and \( c_{i} \) is the \( i \)-th feature map. The mean value \( \bar{C} \) of \( C \) is calculated, and then, the value of each element \( C_{ij} \) in \( C \) with \( \bar{C} \) is compared, to calculate the mask map. If the \( C_{ij} \) is greater than the \( \bar{C} \), the element value \( m \) in the mask map is 1; otherwise, the element value \( m \) in the mask map is 0. It can be calculated by
\[
\text{Mask} = \begin{cases} 
  m = 1, & \text{if } C_{ij} > \bar{C} \\
  m = 0, & \text{if } C_{ij} \leq \bar{C},
\end{cases}
\]
where \( \text{Mask} \) is the obtained mask map and \( m \) is the element value in Mask. In this manner, the original image and mask image can be employed to calculate the dot product to get the saliency feature map. It can be calculated by
\[
S = I \odot \text{Mask}.
\]

### 3.3. Visual Saliency Detection

The visual data entering human eyes is about 108 to 109 bits per second [27]. For the human visual system (HVS), real-time processing of these data streams is an extremely heavy task. The HVS only understands and processes a part of the information. This selection mechanism is named visual attention. This kind of attention behavior is considered to be led by two mechanisms, i.e., stimulus-driven bottom-up and expectation-driven top-down mechanisms [28]. Bottom-up attention is mainly driven by the orientation, contrast, color, action, and other attributes of the visual scene. In the field of computer vision, top-down attention is related to cognitive aspects such as memory, experience, and cultural background. Because of the simplicity of visual attention, visual attention is related to the former attention mechanism mainly, which is often called visual salience [29].

In this section, a saliency-based calculation model is proposed, as shown in Figure 4. A malignant tumor image is taken as an example. The spiking CNN is employed to extract the features. Then, the two-dimensional feature maps generated by the spike convolution layer are summed, and the mask is calculated to obtain the saliency feature map. However, for medical imaging devices, the relationship between the input energy and the brightness of the color recorded in the image file is linear. As a result, the image displayed on the device is inconsistent with the actual image captured by the camera. To correct this difference, gamma nonlinearity is performed before the saliency calculation. The gamma nonlinearity can be calculated by
\[
f(I) = I^{\gamma}.
\]

When \( \gamma < 1 \), the dynamic range gets larger, and the image contrast is enhanced in the low gray value region. The dynamic range gets smaller, the image contrast decreases, and the overall gray value of the image gets larger in the high gray value region. The dynamic range of the low gray value region gets smaller, when \( \gamma > 1 \), and the dynamic range of the high gray value region gets larger, which reduces the contrast of the low gray value region image and improves the contrast of the high gray value region image. At the same time, the gray value of the whole image gets smaller.

### 4. SNN Architecture Optimization

The connection of neurons in the LSM neural network includes the connection between input neurons and excitatory (or inhibitory) neurons in the reservoir, interconnection of neurons in the reservoir, and connection between neurons in the reservoir and neurons in the readout layer. There is no interconnection between neurons in the input layer (or readout layer). There are four types of recursive connections between the synapses in the reservoir, i.e., the connection between excitatory and excitatory synapses \((E \rightarrow E, W_{EE})\), the connection between excitatory and the inhibitory synapses \((E \rightarrow I, W_{EI})\), the connection between inhibitory and the excitatory synapses \((I \rightarrow E, W_{IE})\), and the connection between inhibitory and the inhibitory synapses \((I \rightarrow I, W_{II})\).

The performance of the LSM model in pattern recognition tasks depends not only on the strength of the connections between neurons but also on the number of neurons.
and the probability of synaptic connection. To design an efficient reservoir layer that performs desired kernel functions, these parameters need to be optimized. Therefore, the FOA [6] is employed to search for the best network architecture, that is, the parameters of connection probability $p$ and the number of neurons.

$N$ in the reservoir, which determines the network architecture. The steps of FOA are as follows.

FOA is an algorithm of global optimization. It is based on the foraging behavior of fruit flies, as shown in Figure 5. In sensory perception, the fruit fly is superior to other species; especially in the sense of smell and vision, the olfactory organs of the fruit fly can collect all kinds of odors floating in the air and even smell food sources 40 km away. After flying close to the food location, the fruit fly can also use their eyes to find the location where the food and its companions gather and fly in this direction. The random direction and distance of individual fruit fly searching for food can be calculated by

$$x_i = x_{\text{init}} + \text{random value},$$

\[ (17) \]
yi = y_{init} + \text{random value}, \quad \text{(18)}
\quad
z_i = z_{init} + \text{random value}, \quad \text{(19)}

where \( x_{init} \) is the initial position of the individual in the \( x \)-axis direction, \( y_{init} \) is the initial position of the individual in the \( y \)-axis direction, and \( z_{init} \) is the initial position of the individual in the \( z \)-axis direction. According to an individual position, the distance and smell concentration can be calculated by

\[
\text{dist} = \sqrt{x_i^2 + y_i^2 + z_i^2},
\]

\[
S_i = \frac{1}{\text{dist}},
\]

where \( \text{dist} \) is the distance between the individual and the target and \( S_i \) is used to judge the value of smell concentration. The goal of SNN is to minimize the gap between the desired output and the actual output on the training sample. It can be calculated by

\[
E = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{a} - y_{d})^2},
\]

where \( N \) is the number of samples, \( y_{a} \) is the actual output of the network, and \( y_{d} \) is the desired output of the network. In addition to minimizing the error on the training set, the number of neurons in the reservoir (\( n_r \)) to be as small as possible and the connections between neurons as few as possible are also expected, to simplify the network structure. Therefore, the objective function is set to

\[
\text{fitness} = (w_1 E + w_2 n_r + w_3 p)_{\text{min}},
\]

where \( w_1 + w_2 + w_3 = 1 \). \( w_1 \) is the weight of \( E \) in the objective function, \( w_2 \) is the weight of \( n_r \) in the objective function, and \( w_3 \) is the weight of \( p \) in the objective function. The value of parameter \( p \) is the connection probability between neurons in the reservoir. The larger the value of \( w_1 \), the more emphasis on minimizing the training error. The smaller the value of \( w_1 \), the more emphasis on simplifying the network structure, thus reducing the training time. If \( w_1 \) is too large, the algorithm will overemphasize parameter optimization and ignore structural optimization. If \( w_2 \) and \( w_3 \) are too large, it will lead to the poor performance of the network. In this work, \( w_1 \) is set to 0.6, \( w_2 \) is set to 0.2, and \( w_3 \) is set to 0.2.

5. Experimental Results

To effectively verify the performance of the network, three different modalities of breast image datasets are selected to test the proposed network, namely, breast ultrasound images, breast X-ray images, and breast histopathological images.
5.1. Experiment Settings. The number of input neurons is set according to the image size. The image size of the three datasets is different, so it is necessary to set a different number of input neurons. Other parameters in the network are set according to experience, as shown in Table 1. $V_{\text{rest}}$ is the membrane potential of neurons in a resting state. In this paper, it is set to 0 mV. $V_{\text{threshold}}$ is the threshold that determines the spike firing or not. $w_{\text{max}}$, $w_{\text{min}}$, $A_{\text{pre}}$, and $A_{\text{post}}$ are used to perform STDP training. $n_r$ is the total neurons in the initial reservoir, which determines the size of the network. In the initial parameter set up of the FOA, the random initialization fruit fly swarm location range is [0, 300], the random fly direction and distance zone of iterative fruit fly food searching is [-100, 100], fruit fly population size is 1000, and iterative number is 120. The Gaussian distribution hypothesis of classification accuracy is used to calculate the confidence intervals. It can be calculated by

$$l = \frac{2 \ast n \ast p + z - z \ast \left(\sqrt{z + 4 \ast n \ast p \ast (1 - p)}\right)}{2(n + z)}, \quad (24)$$

$$u = \frac{2 \ast n \ast p + z + z \ast \left(\sqrt{z + 4 \ast n \ast p \ast (1 - p)}\right)}{2(n + z)}, \quad (25)$$

where $l$ and $u$ are the upper and lower bounds of the confidence interval, respectively, $p$ is the classification accuracy, $n$ is the sample size, and $z$ is the critical value of Gaussian distribution.

5.2. BreastMNIST Database. The dataset of breast ultrasound images [30] consists of 780 images. It can be categorized into 3 classes, i.e., malignant, benign, and normal tumors, as shown in Figure 5. In our work, the BreastMNIST database [31] is used for testing the proposed SNN. It is based on the breast ultrasound images. The dataset uses low-resolution images to simplify the task to binary classification. This work combines normal and benign into negative and classifies malignant as positive. And the source images with an average image size of 500 × 500 pixels are resized into 28 × 28.

The three types in the dataset are shown in Figure 6. Figure 6(a) is a normal image, Figure 6(b) is a benign tumor image, and Figure 6(c) is a malignant tumor image. The images in each category are shown in Table 2. As shown in
on the BreastMNIST database is shown in Figure 8. The optimization process of SNN architecture using the ResNet-18. Using the ResNet-50, the accuracy of 85.9% accuracy and 87.8% accuracy can be obtained by the spike time encoding scheme based on information entropy. After employing the saliency-based feature module, the training accuracy is 89.75%. The training accuracy of 94.72% and the test accuracy is 92.74%. Further optimization of the network architecture parameters can obtain the optimal training accuracy and test accuracy, which are 85.3% and 83.3%. The accuracy of 80.8% and the accuracy of 80.1% are obtained by the Auto-sklearn and AutoKeras, respectively. The Google AutoML Vision obtains 86.5% accuracies. Our work can get the best result, i.e., 97.4%, on the BreastMNIST database.

5.3. Mini-MIAS Database. Breast X-ray image has good contrast, and resolution can distinguish the difference of microstructure density between tissues, and it is easy to operate, relatively cheap, easy to accept, and high diagnostic accuracy. It is internationally recognized as an effective measure for early opportunistic screening and early detection of breast cancer. 322 images from 161 patients are included in the Mammographic Image Analysis Society (MIAS) dataset. This work uses the mini-MIAS database, which contains images with a size of $1024 \times 1024$ pixels [35].

The three types of images in the dataset are shown in Figure 9. Figure 9(a) is the normal image, Figure 9(b) is the benign tumor image, and Figure 9(c) is the malignant tumor image. To get an effective model and balance all kinds of data, the original image is rotated at different angles to expand the data. The experimental results are shown in Table 5.

The SNN performance comparison of different methods on the BreastMNIST dataset is shown in Table 5. It shows that 81.99% training accuracy and 83.54% test accuracy can be obtained by setting the LSM with empirical parameters only. Using the linear time coding algorithm, the training accuracy of the network is 92.85%, and the test accuracy is 95.03%. The training accuracy of 95.03% are obtained by the spike time encoding scheme based on information entropy. After employing the saliency-based feature module, the training accuracy is 85.20%, and the test accuracy is 92.74%. Further optimization of the network architecture parameters can obtain the optimal training accuracy and test accuracy, which are 88.57% and 97.44%, respectively.

The performance curve is shown in Figure 7. Figure 7(a) represents the performance curve and area under the curve on training data, and Figure 7(b) represents the performance curve and area under the curve on test data. As shown in the figure, the area on training data and test data are 0.90 and 0.99, respectively.

In the optimized SNN, the total number of neurons in the reservoir $n_r$ is 300; the connection probability $p$ of neurons is 0.1. The optimization process of SNN architecture on the BreastMNIST database is shown in Figure 8.

Performance comparison of different algorithms on the BreastMNIST database is shown in Table 4. It shows that 85.9% accuracy and 87.8% accuracy can be obtained by using the ResNet-18. Using the ResNet-50, the accuracy of the network is 85.3% and 83.3%. The accuracy of 80.8% and the accuracy of 80.1% are obtained by the Auto-sklearn and AutoKeras, respectively. The Google AutoML Vision obtains 86.5% accuracies. Our work can get the best result, i.e., 97.4%, on the BreastMNIST database.

Table 2, the dataset contains 133 normal images, 437 benign images, and 210 malignant images.

The network performance comparison of several models on the BreastMNIST dataset is shown in Table 3. It shows that 74.36% training accuracy and 81.19% test accuracy can be obtained by setting the LSM with empirical parameters only. Using the linear time coding algorithm, the training accuracy of the network is 77.67%, and the test accuracy is 89.79%. The training accuracy of 86.19% and the test accuracy of 92.74% are obtained by the spike time encoding scheme based on information entropy. After employing the saliency-based feature module, the training accuracy is 85.20%, and the test accuracy is 92.74%. Further optimization of the network architecture parameters can obtain the optimal training accuracy and test accuracy, which are 88.57% and 97.44%, respectively.

Table 3: Performance comparison of different methods on the BreastMNIST dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Training accuracy (%)</th>
<th>Test accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved SNN</td>
<td>91.10</td>
<td>95.54</td>
</tr>
<tr>
<td>Saliency-based SNN</td>
<td>89.92</td>
<td>93.25</td>
</tr>
<tr>
<td>SNN with entropy-based time encoding</td>
<td>86.56</td>
<td>85.33</td>
</tr>
<tr>
<td>SNN with linear time encoding</td>
<td>80.20</td>
<td>82.90</td>
</tr>
<tr>
<td>Original SNN</td>
<td>76.33</td>
<td>76.31</td>
</tr>
</tbody>
</table>

Table 4: Accuracy comparison of the proposed method and several other methods over the BreaKHis database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Training accuracy (%)</th>
<th>Test accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet [43]</td>
<td>90.0%</td>
<td>84.6%</td>
</tr>
<tr>
<td>PFTAS+RF [42]</td>
<td>81.4%</td>
<td>85.5%</td>
</tr>
<tr>
<td>PFTAS+SVM [42]</td>
<td>81.7%</td>
<td>90.1%</td>
</tr>
<tr>
<td>CSDCNN [44]</td>
<td>97.1%</td>
<td>95.7%</td>
</tr>
<tr>
<td>Inception-v1 [45]</td>
<td>89%</td>
<td>94%</td>
</tr>
<tr>
<td>MobiHisNet [46]</td>
<td>91.42%</td>
<td>85.84%</td>
</tr>
<tr>
<td>The ensemble SNN [5]</td>
<td>98.7%</td>
<td>97.5%</td>
</tr>
<tr>
<td>This work</td>
<td>96.3%</td>
<td>98.4%</td>
</tr>
</tbody>
</table>

Table 5: Accuracy comparison of different methods on the BreaKHis database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Magnification factors</th>
<th>Training accuracy (%)</th>
<th>Test accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet [43]</td>
<td>40x</td>
<td>88.4%</td>
<td>91.6%</td>
</tr>
<tr>
<td>PFTAS+RF [42]</td>
<td>100x</td>
<td>84.6%</td>
<td>85.5%</td>
</tr>
<tr>
<td>PFTAS+SVM [42]</td>
<td>200x</td>
<td>85.5%</td>
<td>85.5%</td>
</tr>
<tr>
<td>CSDCNN [44]</td>
<td>400x</td>
<td>95.7%</td>
<td>95.7%</td>
</tr>
<tr>
<td>Inception-v1 [45]</td>
<td>40x</td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>MobiHisNet [46]</td>
<td>100x</td>
<td>93.8%</td>
<td>93.8%</td>
</tr>
<tr>
<td>The ensemble SNN [5]</td>
<td>200x</td>
<td>93.8%</td>
<td>93.8%</td>
</tr>
<tr>
<td>This work</td>
<td>400x</td>
<td>93.8%</td>
<td>93.8%</td>
</tr>
</tbody>
</table>
Performance comparison of different algorithms on the mini-MIAS database is shown in Table 6. The method of adaptive thresholding provides 93% accuracy in [36]. An accuracy of 94.57% is achieved in [37] using Fisher linear discriminant analysis features of neighborhood structural similarity. An accuracy of 96.7% is obtained in [38] using deep distance metric learning. Using texture features with a neural network classifier provides 95.2% accuracy [39].

Superresolution reconstruction module with texture features provides 96.7% accuracy in [40]. The Hanman transform classifier and the hesitancy-based Hanman transform classifier that are used in [41] can achieve 100% accuracy. Although 98.27% accuracy is provided, only one LSM classifier is used for classification in this work.

5.4. BreaKHis Database. To further test the performance of the reservoir SNN model in breast cancer image recognition, the proposed SNN is applied to the breast histopathological images, i.e., the BreaKHis database [42]. The dataset contains four types of histologically benign tumors (phyllodes tumor, fibroadenoma, tubular adenoma, and adenosis) and four types of malignant tumors (mucinous, lobular, papillary, ductal). Each type of image is shown in Figure 12.

This database contains 7909 images and four magnification factors, i.e., 40x, 100x, 200x, and 400x are acquired. The database includes 2480 images of benign and 5460 malignant tumors, respectively. The size of each image is 700 × 460 pixels. The detailed information of the database is shown in Table 7.

The SNN performance comparison of several methods on the BreaKHis dataset is shown in Table 8. It shows that on the data of 40x image, 76.33% training accuracy, and 78.20% test accuracy can be obtained by setting the LSM with empirical parameters only. Using the linear time encoding algorithm, the training accuracy of the network is 80.20%, and the test accuracy is 84.54%. The training accuracy of 86.56% and the test accuracy of 89.90% are obtained by using the spike time encoding scheme based on information entropy. After adding the saliency analysis module, the training accuracy is 89.92%, and the test accuracy is 94.46%.

<table>
<thead>
<tr>
<th>Methods</th>
<th>40x</th>
<th>100x</th>
<th>200x</th>
<th>400x</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet [43]</td>
<td>(0.8857, 0.9121)</td>
<td>(0.8697, 0.8972)</td>
<td>(0.8296, 0.8611)</td>
<td>(0.8443, 0.8761)</td>
</tr>
<tr>
<td>PFTAS+SVM [42]</td>
<td>(0.7984, 0.8324)</td>
<td>(0.7813, 0.8158)</td>
<td>(0.8348, 0.8659)</td>
<td>(0.8049, 0.8399)</td>
</tr>
<tr>
<td>PFTAS+RF [42]</td>
<td>(0.8005, 0.8343)</td>
<td>(0.7958, 0.8293)</td>
<td>(0.8183, 0.8507)</td>
<td>(0.7912, 0.8273)</td>
</tr>
<tr>
<td>PFTAS+QDA [42]</td>
<td>(0.8213, 0.8536)</td>
<td>(0.8041, 0.8371)</td>
<td>(0.8254, 0.8573)</td>
<td>(0.8015, 0.8368)</td>
</tr>
<tr>
<td>CSDCNN [44]</td>
<td>(0.9626, 0.9774)</td>
<td>(0.9476, 0.9651)</td>
<td>(0.9585, 0.9741)</td>
<td>(0.9468, 0.9655)</td>
</tr>
<tr>
<td>Inception-v1 [45]</td>
<td>(0.8752, 0.9027)</td>
<td>(0.9073, 0.9307)</td>
<td>(0.9287, 0.9495)</td>
<td>(0.8854, 0.913)</td>
</tr>
<tr>
<td>MobiHisNet [46]</td>
<td>(0.9012, 0.9258)</td>
<td>(0.8854, 0.9113)</td>
<td>(0.9148, 0.9376)</td>
<td>(0.8414, 0.8735)</td>
</tr>
<tr>
<td>The ensemble SNN [5]</td>
<td>(0.981, 0.9911)</td>
<td>(0.9409, 0.9595)</td>
<td>(0.9585, 0.9741)</td>
<td>(0.9671, 0.9815)</td>
</tr>
<tr>
<td>This work</td>
<td>(0.9537, 0.9703)</td>
<td>(0.9227, 0.9441)</td>
<td>(0.9416, 0.9603)</td>
<td>(0.9772, 0.9889)</td>
</tr>
</tbody>
</table>

**Table 10: Confidence interval comparison of the proposed method and several other methods over the BreaKHis database (z = 1.96).**

![Figure 13: Illustration of the optimization process of network structure on the BreaKHis database.](image-url)
Further optimization of network architecture can obtain optimal training accuracy and test accuracy, which are 91.10% and 96.27%, respectively. On the 100x images, the optimal training accuracy is 89.60%, and the test accuracy is 93.35%. On the 200x magnified image, the optimal training accuracy is 90.33%, and the test accuracy is 95.24%. On the 400x images, the optimal training accuracy is 93.06%, and the test accuracy is 98.44%.

The results of several other studies on the BreaKHis dataset are used for performance comparison. Table 9 shows accuracies of these approaches, and Table 10 shows confidence interval comparison of the proposed method and several other methods over the BreaKHis database. The results of [43] report accuracies (86%-90%) by using the AlexNet. The experimental results of [42] demonstrated that the QDA classifier can get higher accuracies than RF and SVM classifiers on BreaKHis. A deep convolutional neural network is used to achieve 95.7%-97.1% classification accuracies in [44]. [45] investigated the performance of five CNN architectures (i.e., LeNet-5, AlexNet, VGG-16, ResNet-50, and Inception-v1) on the basis of test accuracy. The Inception-v1 can achieve the test accuracy of 89%, 92%, 94%, and 90%, respectively, at 40x, 100x, 200x, and 400x magnification factor classification. [46] proposes an efficient and lightweight CNN model for histopathological image classification based on MobileNet. It achieves the test accuracy of 91.42%, 89.93%, 92.70%, and 85.84%, respectively, at 40x, 100x, 200x, and 400x magnification factor classification. The ensemble SNN [5] can achieve 98.7%, 95.1%, 96.7%, and 97.5% accuracies, respectively, which are higher than other approaches. The result of our work is not the best on the magnification of 40x, 100x, and 200x. However, it is better than others on the 400x images.

In the optimized SNN, the total neurons in the reservoir \( n_r \) are selected as 400, and the connection probability \( p \) of neurons is 0.2. The optimization process of SNN architecture on the BreaKHis dataset is shown in Figure 13. In the BreaKHis dataset, the number of images of ductal carcinoma accounts for about two-fifths of the total. The number of images of ductal carcinoma is the most in the images with magnifications of 100 and 200. It leads to data imbalance between categories. Therefore, the recognition results of images with magnifications of 100 and 200 are significantly lower than those of images with magnifications of 40 and 400. In the experiment, some ductal carcinoma images are mis-classified as lobular carcinoma. This is also the reason why the recognition results of images with magnifications of 100 and 200 are significantly lower than that of images with magnifications of 40 and 400.

### 6. Conclusions and Future Work

A saliency-based SNN with breast cancer recognition capability, underpinned by the ReSuMe learning algorithm and spike time encoding scheme, has been presented in this paper. To improve the performance, the FOA was employed to optimize the architecture of the SNN. The performance of the proposed methods demonstrates that the network is effective for breast images. Experimental results show that the SNN with entropy-based time encoding can get better performance than the SNN with a linear time encoding scheme. The saliency model and optimized SNN can further improve the classification accuracy. However, on the multi-classification task of the BreaKHis database, the performance of the proposed SNN is still insufficient. Future work will focus on multiclassification learning using the SNN and further research on SNN architecture, such as multireservoir cascade structure and parallel reservoir architecture.

### Data Availability


### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work is supported by the Natural Science Foundation of Heilongjiang Province under Grant LH2020F023.

### References


