Data S1. Pulmonary nodule features extracted by unsupervised learning. The denoising autoencoder lung nodule classification method was used in the present study (Fig. S1). Based on traditional two-layer self-encoding learning, the noise of the specific data distribution was added to the input lung nodule image X and the combined images were added by random mapping. The noise lung nodule image data $\tilde{X}$ were used to extract the features of the nodule image, such as the decreased sensitivity of the trained encoder to the noise and the increased robustness. The training hidden layer and the reconstruction layer were used to restore the original data. The autoencoding operation performed two linear transformations on the lung nodule image. Subsequently, the sigmoid function was selected as the activation function of the autoencoder. Initially, a coding result Y was obtained, and the coding result calculation formula was: $Y = \delta(W\tilde{X} + b)$. Following reconstruction, the reconstructed vector Z was obtained, and the reconstruction result was calculated as $Z = \delta(W'Y + b')$. The overall formula used for the denoising autoencoder was:

$$Z = \delta(W' * \delta\left(W\tilde{X} + b\right) + b') \tag{10}$$

The denoising autoencoder exhibits a powerful non-linear expression capability and therefore, can often fully describe the characteristics unique to individual objects. The denoising autoencoder can decrease the shortcomings of the traditional autoencoder, which was highly sensitive to noise. Following a decrease in noise levels, the autoencoder obtains selected, relatively low-noise data to increase the robustness of the expressed data and improve the generalization ability of encoding. To avoid over-fitting of the input data, the present method included an elastic network regularization term to the recombination cross-entropy loss function of the self-encoder to reduce the error between input and output, and to improve the generalization ability of the model. The term was described as:

$$l_D = -\sum_{i=1}^{N} X_i log Z_i + (1 - X_i) log(1 - \quad Z_i) + \lambda_1 \sum_{i=1}^{N} |\theta_i| + \lambda_2 \sum_{i=1}^{N} {\theta_i}^2 \tag{11},$$

where θ was a network model parameter and λ_1 and λ_2 were the coefficients of constraints l_1 and l_2, respectively.

Data S2. Pulmonary nodule features extracted by supervised learning. In the present study, according to the basic principle of ResNet, an 18-layer ResNet (ResNet-18) was designed to extract the deep semantic features of the lung nodules (Fig. S2). The architecture is depicted in Table SI, which describes the hyper-parameters of ResNet in detail.

Initially, the input of the network was a gray 64x64 pixel^2 image of a lung nodule. Secondly, ResNet-18 adds a shortcut connection to the original 18-layer plain convolution network to form a residual block. This process allows the original input information to be transmitted directly to the layer behind the convolutional layer, which is an acceptable solution to the issue of network degradation. Thirdly, the residual block is composed of batch norm units, rectified linear units and convolutional layers.

In general, use of a rectified linear unit overcomes the issue of gradient disappearance. The Batch Normalization layer (25) is normally distributed following the convolutional layer, and the input distribution of any unit of each layer of the neural network is redistributed to a normal distribution with a mean of 0 and a variance of 1. This process ensures that the activated input value can be placed in the area where the nonlinear function is sensitive to the input, thereby decreasing the training time and the partial gradient disappearance issue. The specific formula used was:

$$\begin{cases} \mu = \frac{1}{n}\sum_{i=1}^{n} x_i , \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 , \\ \hat{x} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}}. \end{cases} \tag{12}$$

In order to avoid a variance of 0, ε was added to the variance.

For ResNet-18, the result X of the previous layer input was the output following construction of the block function $F(X)$ by the residual $H(X) := F(X) + X$. To obtain an identity mapping between the features of the layer and those of the lower layer, the forward process was adjusted as linear for the residual element, and the subsequent input was equal to the result of the input plus the residual element of each time. In the ordinary network, the convolution operation was applied for each layer of convolution. At the end of the network, a fully connected layer of five neurons was placed to classify the lung nodules for malignancy. During re-propagation of the residual network, the gradient from the L-th layer could be stably transmitted to the first layer using the following formula:

$$\frac{dE}{dx_l} = \frac{dE}{dx_L}\left(1 + \frac{d}{dx_l}\sum_{i=l}^{L-1} F(x_i, W_i)\right) \tag{13}.$$

Concomitantly, for mini-batch gradient calculations, the possibility of gradient disappearance was low since for all items, the following condition was taken into account: $\frac{\partial}{\partial x_l} = \sum_{i=l}^{L-1} F(x_i, W_i)$. This value in the sample was not always negative, suggesting that even if the weight was arbitrarily small, the gradient of the layer did not disappear.

Data S3. Pulmonary nodule classification by handcrafted features. The texture feature (TF) is a global feature that describes the surface properties of a nodule in an image or image region, such as the solidity of a nodule or calcification (Fig. S3). Malignant nodules usually have an irregular texture with a low degree of calcification. The TF of the nodule is a powerful indicator for assessing the malignant phenotype. TFs often exhibit rotational invariance and are highly resistant to noise. In the present study, a gray level co-occurrence matrix was selected as the descriptor of the TF. Due to the texture being formed by the gray level distribution, which appeared repeatedly in the spatial position, a certain gray level margin appeared between the two pixels that were separated by a specific distance in the image space. The gray level co-occurrence matrix is a common method used to describe texture by studying the spatial correlation of the gray scale. For fine and regular textures in the image, the two-dimensional pairs of the pixel histograms are evenly distributed, whereas for coarse and regular textures, they are more diagonal in appearance. In case of the pixel value of the i-th row and the j-th column being P(i,j), the gray level co-occurrence matrix feature quantity (26) includes the following equations:

$$Contrast = \sum_i \sum_j (i,j)^2 P(i,j) \quad (14);$$

$$Energy = \sum_i \sum_i P(i,j)^2 \quad (15);$$

$$Entropy = -\sum_i \sum_j P(i,j) \log P(i,j) \quad (16);$$

$$Inverse\ Difference = \sum_i \sum_j \frac{P(i,j)}{1+(i-j)^2} \quad (17);$$

$$Correlation = -\sum_i \sum_j \frac{(i-\mu_x)(j-\mu_y)}{\sqrt{\sigma_x \sigma_y}} \quad (18);$$

$$Homogeneity = \sum_i \sum_j \frac{P(i,j)}{1+|i-j|} \quad (19),$$

where μ_x and μ_y are the row and column means, and σ_x and σ_y are the row and column standard deviations of the image matrix. The latter parameters were calculated as follows:

$$\mu_x = \sum_i i \sum_j P(i,j), \mu_y = \sum_j j \sum_i P(i,j) \quad (20)$$

$$\sigma_x = \sum_i (i-\mu_x)^2 \sum_j P(i,j)$$
$$\sigma_y = \sum_j \left(j-\mu_y\right)^2 \sum_i P(i,j) \quad (21)$$

The shape feature (SF) is one of the simplest classification features. The SF assesses the benign and malignant nodules by extracting the appearance shape of the target nodule, such as the spherical appearance of the nodule, the burring and the feature of the lobulation signs. Benign nodules are usually characterized by a smooth and burr-free appearance, whereas the appearance of malignant nodules is irregular and burr-like. In the present study, the geometric parameter method was selected to extract the shape characteristics of lung nodules (27). The two-dimensional slice CT was used and the main geometric features of the lung nodules included the circumference, area, spherical formation, rectangularity, kurtosis, skewness and eccentricity of the lung nodules. Since the accuracy of the geometric parameters was affected by the target segmentation effect, the extraction of the SFs of the pulmonary nodules was premised by the nodule location. This decision was made by the medical practitioner. The circumference of the pulmonary nodule was calculated by the sum of the number of boundary points marked by the doctor. The area was calculated by the sum of the statistical boundary and its internal pixels. For an image I of size *x* x *y*, the geometrical formula was calculated as follows:

$$Roundness = \frac{4\pi S}{C^2} \quad (22);$$

$$Rectangularity = \frac{S}{S_{MER}} \quad (23);$$

$$Eccentricity = \frac{A_{MER}}{B_{MER}} \quad (24);$$

$$Skewness = \frac{\sum_i \sum_j (I_{i,j}-\mu)^3}{xy\sigma^3} \quad (25);$$

$$Kurtosis = \frac{\sum_i \sum_j (I_{i,j}-\mu)^4}{xy\sigma^4} \quad (26),$$

where S represents the nodule area, C represents the contour perimeter of the nodule, S_{MER} represents the minimum external rectangle representing the nodule, A_{MER} and B_{MER} are the length and width of the rectangle, μ is the mean of the images and σ is the standard deviation.

$$\begin{cases} \mu = \frac{\sum_i \sum_j I_{i,j}}{x*y} \\ \sigma = \sqrt{\frac{1}{x*y-1} \sum_i \sum_j \left(I_{i,j}-\mu\right)^2} \end{cases} \quad (27)$$

Where $I_{i,j}$ represents the pixel value in rows i-th and columns j-th, and x and y represent the length and width of image.

Data S4. Experimental dataset. A total of 4,578 images of pulmonary nodules were extracted, including 469 images of those with a malignancy level of 1. These were used as labels for the model and for the conversion of multiple types of tags into binary data. A total of 772, 1,017, 1,091 and 1,229 images of pulmonary nodules exhibited malignancy levels of 2, 3, 4 and 5, respectively. In order to increase the balance in the dataset, the sub-dataset with a malignancy level of 1 was rotated 180 degrees and expanded two times compared with its original size. A total of 47 images with a malignancy level of 5 were randomly deleted from the sub-dataset (Fig. S4). A total of 5,000 images were used as experimental data. These images and data were used to conduct data augmentation via a horizontal flip, vertical flip and reflection. The experimental dataset included 20,000 images. The experimental platform was equipped with an Intel Xeon E7-4807, NVIDIA GeForce GTX 1080 GPU and 128 GB RAM of memory. The framework used was the TensorFlow framework.

Figure S1. Feature extraction process of the denoising Auto Encoder. $\tilde{X}$ indicates the pulmonary nodule images with noise, Y indicates the latent variables encoded through Denoising Autoencoder and Z indicates the reconstruction images.

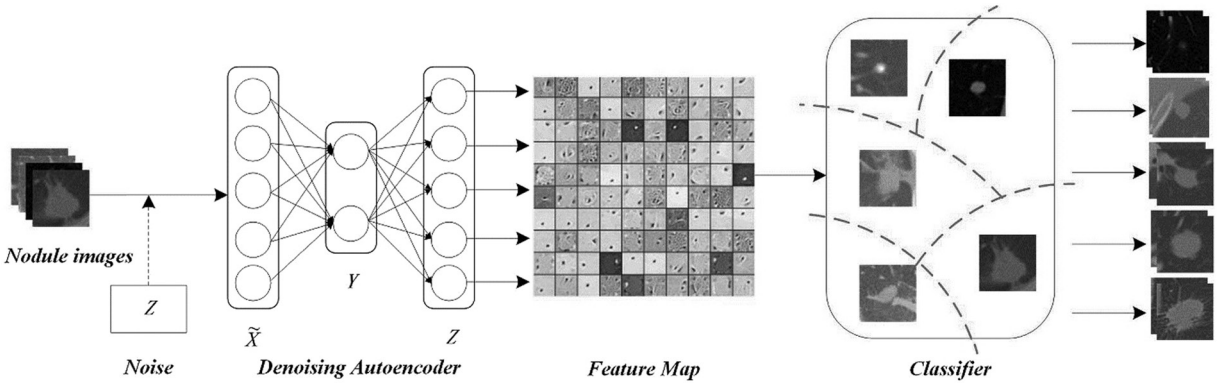

Figure S2. Feature extraction process by ResNet-18.

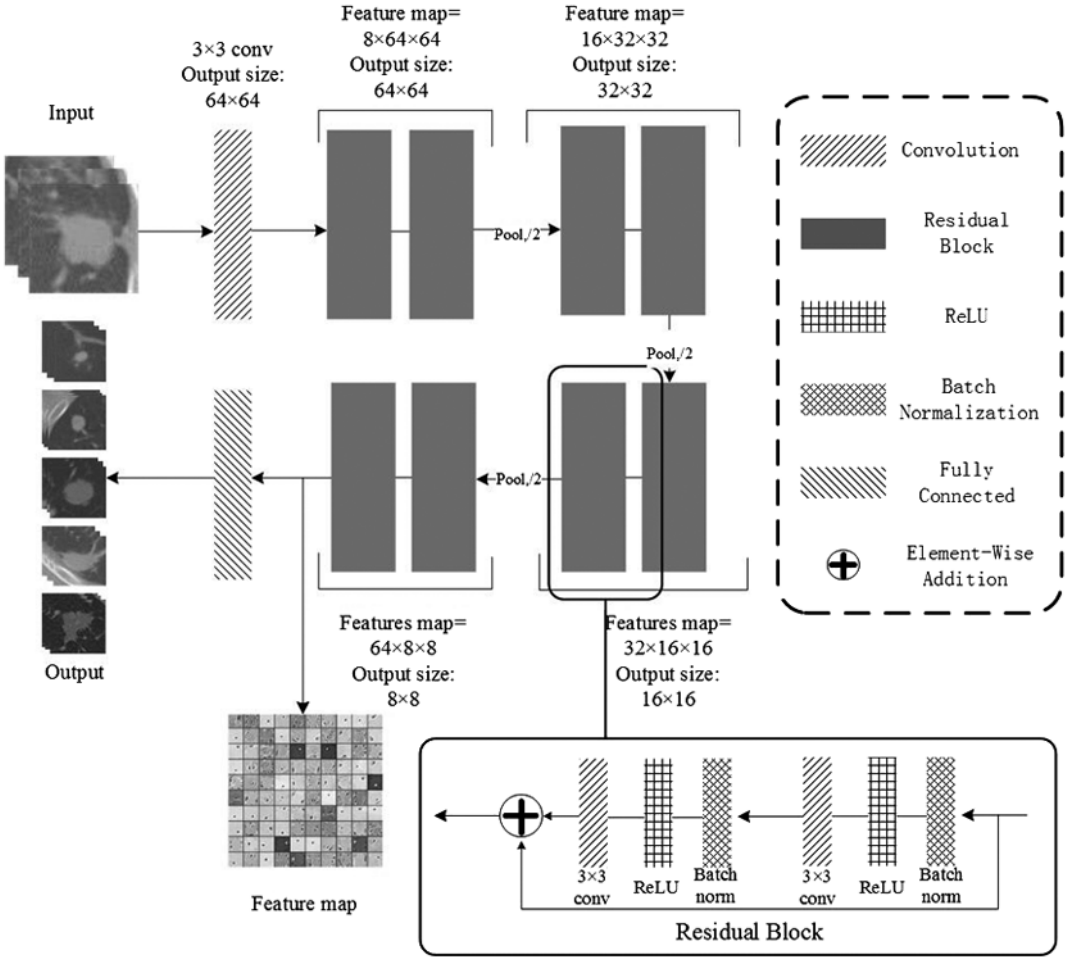

Figure S3. Pulmonary nodules with different texture and shape features.

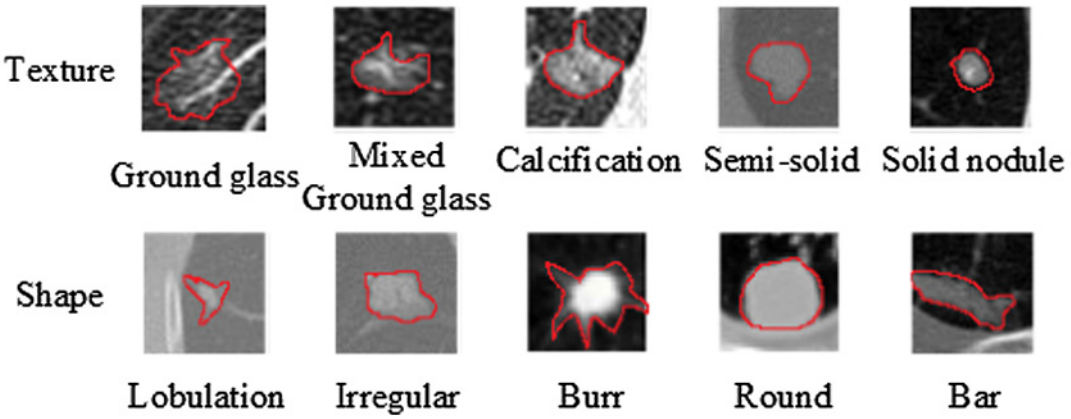

Figure S4. Nodule sample in different malignancy. The different numbers indicate the different malignancy level.

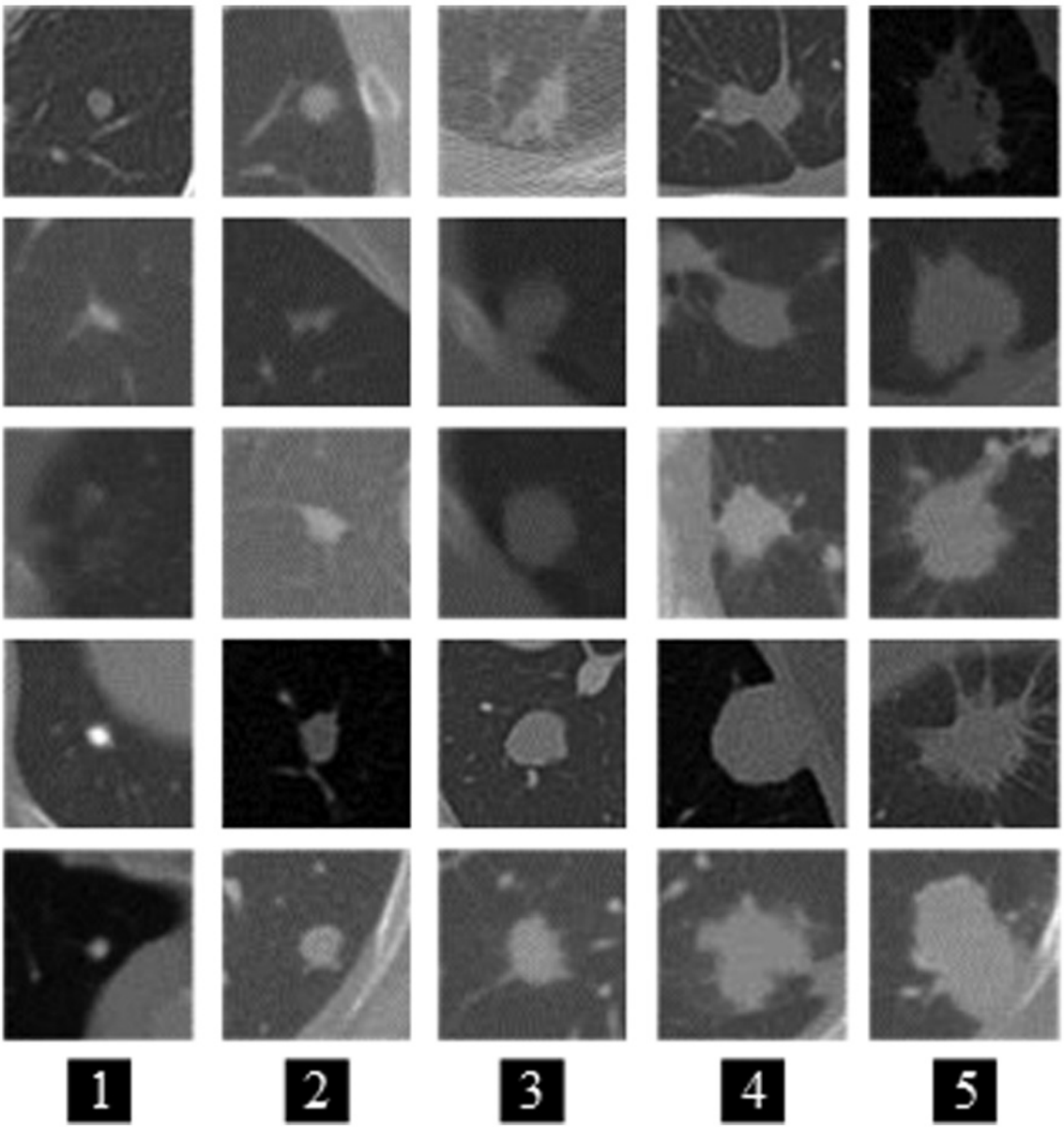

Table SI. Hyper-parameters of the ResNet-18 architecture.

Type	Filter size, pixel^2 and number	Output size, pixel^2	Layers
Convolution	3x3, 64	64x64	1
Max pooling	3x3, 64	32x32	0
Convolution	$\begin{bmatrix}3x3, 64\\3x3, 64\end{bmatrix}$x2	32x32	4
Convolution	$\begin{bmatrix}3x3, 128\\3x3, 128\end{bmatrix}$x2	16x16	4
Convolution	$\begin{bmatrix}3x3, 256\\3x3, 256\end{bmatrix}$x2	8x8	4
Convolution	$\begin{bmatrix}3x3, 512\\3x3, 512\end{bmatrix}$x2	4x4	4
Fully connected	1x1x5	1x1	1