Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/79/20241576

The application progress of Convolutional Neural Networks (CNN) in lung nodule diagnosis

Jingxuan Wu[1,4], **Jiahao Yang**[2,5,*], **Guanlin Peng**[3,6]

1Chengdu No.7 High Shcool, Chengdu, 610041, China
2Xi'an Gaoxin No.1 High School, Xi'an, 710119, China
3Department of Civil engineering, Chongqing Jiaotong University, 400074, Chongqing

[4]2230941139@qq.com
[5]3345180246@qq.com
[6]1731623188@qq.com
*corresponding author

Abstract. With the development of computers, machine learning continues to be widely used in various fields. And there are many application scenarios in the field of medicine. Among these, the broadest one is the field of medical image analysis. Medical image has the characteristics of huge data, excessive noise, and recognition difficulty. And the most difficult one is the analysis of lung medical images. Lung cancer has a higher incidence rate and mortality rate than other cancers. According to the National Cancer Center, about 127,070 people died from lung cancer in 2023, making it the highest death rate in the United States. Therefore, early detection of malignant pulmonary nodules has become crucial in the field of medical imaging. The medical imaging's inadequacies are most noticeable in the pictures of malignant pulmonary nodules, which are difficult for a doctor to identify with their naked eyes. However, pre-processing, segmentation difficulties, and poor fitting impact are the drawbacks of classical machine learning. As a result, we must create fresh approaches to these issues.

Keywords: machine learning, medical image analysis, pulmonary nodules.

1. Introduction

According to the American Thoracic Society (ATS), pulmonary nodule is a relatively prevalent condition, with about half of all adults who get CT scans having at least one lung nodule [1]. However, if we ignore lung nodules, they can develop into lung cancer, which has a high death rate and requires patients to spend a significant amount of time and money to treat. Pulmonary nodules are found in the lungs, which are vital to the human body because they exchange gas between the inside and outside of the body. A pulmonary nodule is an abnormal development in the lung. They are often smaller than 3cm in diameter, formed like a round or irregular model, and have clear or fuzzy borders in lung pictures. Some may manifest as solitary pulmonary nodules, while others may form multiple pulmonary nodules in the lungs simultaneously. Nodules are divided into three types: pure ground-glass nodules, mixed ground-glass nodules, and solid nodules. Pure glass nodules, also known as pure GGO nodules, appear on thin-slice CT scans as frosted glass and are typically benign. As they grow in size and the solid components increase, the density rises and they become mixed ground glass nodules. The mixed ground glass nodule has a solid center and the texture of a pure glass nodule on the outside. Furthermore, the density of solid nodules is greater. It is difficult to catch pulmonary nodules using a chest X-ray, thus it is recommended to use a low-dose spiral CT to identify them.

2. Basic rule of CNN

2.1. Fundamentals of Machine Learning

Machine learning is a core area of artificial intelligence that allows computers to learn from experience and self-improve without explicit programming. The foundation of this discipline lies in pattern recognition and computational learning theory, utilizing algorithms to interpret data, learn underlying patterns, and make decisions or predictions [2]. The rapid development of machine learning technology is due to the interdisciplinary integration of theories and methods from mathematics, statistics, computer science, and information theory.

2.2. Evolution and Impact of CNNs

Among the plethora of machine learning algorithms, the rise of Convolutional Neural Networks (CNNs) has had a revolutionary impact on the field of image processing. CNNs, by mimicking the mechanism of human visual perception, can effectively handle high-dimensional image data. They utilize structures like convolutional layers, pooling layers, and fully connected layers to extract features from images and learn through the backpropagation algorithm.

2.3. The Architecture of Convolutional Neural Networks

The convolutional layer, the heart of CNNs, performs convolution operations on images with filters to extract features at various levels. These features may include edges, corners, or more complex patterns and become more abstract as the network deepens. The pooling layer is responsible for reducing the spatial size of features, enhancing the model's generalization ability, and reducing the consumption of computational resources. Finally, the fully connected layers use the high-level features extracted to perform classification or regression tasks.

2.4. Machine Learning in Practice

The powerful functionality of CNNs owes a lot to the design of "weight sharing" and "local receptive fields," which confer a certain invariance to the network against translations, scaling, and other forms of image transformations. This invariance is why CNNs have demonstrated exceptional performance in fields like image recognition, speech recognition, and natural language processing [3]. Today, machine learning and deep learning have transcended theoretical research and their applications have permeated various industries including autonomous driving, medical diagnostics, financial analysis, and intelligent recommendation systems, becoming a key technology driving societal advancement. With enhanced computational power and the advent of the big data era, the potential for development in machine learning, especially CNNs, remains vast and is expected to continue as a major engine driving the progress of artificial intelligence.

2.5. CNN's Role in Advanced Applications

CNN is often employed in computer vision and natural language processing applications like as image recognition and classification, super-resolution reconstruction, and autonomous automobiles. CNN mimicking animal vision perception tends to directly fetch information in the original picture and its kernel inside the hidden layer connecting with layers enables CNN to lattice features with less computer effort, resolving the traditional neural network's problem of the weighting parameter matrix being too large and reducing the risk of over-fitting.

2.6. Convolutional Processes and Feature Mapping

The primary distinction between CNNs and standard neural networks is that CNN uses the two-dimensional matrix of the original image as input rather than pixel information, whereas classic neural networks compress two-dimensional pictures into one-dimensional image characteristics. CNN contains four layers: an input layer, a convolutional layer, a pooling layer, and a fully-connected layer. Following image input, the convolutional layer divides the picture into various pieces and employs a set of filters to generate feature maps.

2.7. Optimization and Efficiency in CNNs

Convolution uses inner product calculation meaning that each number inside the kernel matrix is multiplied with the value for each pixel it lines up with, and then all the elements of the kernel are summed to output one single value [4]. The sliding step size of kernels, size of kernels, zero padding, and number of kernels are all significant characteristics in the convolutional layer. The bigger the sliding step size of the kernel is, the smaller the feature map that can be obtained, the fewer features that can be obtained, and the cruder the result is [5]. The large sliding size of the kernel, on the other hand, enhances the efficiency of calculating and fitting text jobs. It is the same as the size of kernels; the larger the kernel size, the smaller the feature map that may be obtained. Zero padding is the addition of a circle of zero around the borders of filters, which helps decrease information loss at the boundaries of the input feature map and enhance model performance. The more kernels there are, the more feature maps there are. The kernel parameters are shared, which implies that any section of the image is computed convolutionally with the same kernel, which might lower the calculated amount. To get high-level features, convolution must be performed numerous times. One picture has many color channels, which explains why pictures are three-dimensional.

2.8. The Progression and Refinement of CNNs

To reach the final result, we must put all of the convolution results from different color channels together. The pooling layer's function is to down-sample and selectively leave out essential information. One of the pooling approaches, for example, is max pooling, which involves splitting a single depth slice of a feature map into several parts and picking the largest value of each component to keep. Following pooling, the previously obtained two-dimensional picture must be converted into a one-dimensional vector and then linked with a fully connected layer. The function of the fully linked layer is to obtain the final categorization result. The Fully Connected (FC) layer, which includes weights and biases as well as neurons, is used to connect neurons from different layers. These layers are often placed prior to the output layer and constitute the final few levels of a CNN Architecture [6]. This is how CNN operates, and it has evolved significantly in recent years. From the original AlexNet network to the Vgg network to the GoodNet network, and finally to the very accurate ResNet network that is currently extensively used. The addition of a weight layer to the ResNet network fixes the problem that the error rate grows while convolutional increases indefinitely. We can enhance the right rate by adding a large number of convolutional layers by setting the weight parameter to 0 after the layers with weak convolutional effects.

3. Combine lung images problem with CNN

CNN is commonly utilized as an auxiliary diagnostic tool in medical imaging. CNN offers considerable promise, particularly in the early diagnosis of lung nodule pictures. The conventional early diagnosis of pulmonary nodules is based on doctors' expertise, but it is strongly influenced by many different circumstances, such as patients' physical conditions and ages, while large-scale scene screening is a time-consuming and complex work. As a result, the CNN model is recognized as a tool capable of efficiently replacing manual diagnosis. The researchers proved the potential benefits of CNNS for the identification and categorization of lung nodules by analyzing a large number of lung CT images.

In the study, the standard ED-CNN model evaluated 354 lung CT images and compared them using DICE scores to manual detection. The results reveal that the model's average accuracy is 0.962, with the lowest score being 0.926, the highest score being 0.974, and the standard deviation being 0.008. These findings demonstrate that the CNN model may deliver very accurate results in the identification of lung nodules while also providing clinicians with trustworthy auxiliary information. The accuracy of pulmonary nodule pictures is affected differently by different CNN algorithms and architectures. In the following section, we will examine the CNN model in these two areas.

4. Structure optimization

The structure of the model can have a significant impact on the correctness of the outcome, thus the study of the model structure must be thorough and thorough. In the subject of utilizing CNN on lung nodule pictures, we believe that the model's performance and practicability may be enhanced by taking the following factors into account.

1) Data preprocessing:

- **Image processing:** In order to reduce interference, the lung CT scan images need to be standardized, including grey value normalization, noise removal, and artifact removal.
- **Data enhancement:** Use data enhancement techniques, such as rotation, flipping, scaling, and panning, to expand the training dataset.
- **Network architecture:** Choose the best CNN architecture for the job, such as a combination of the convolutional layer, pooling layer, and fully connected layer.
- **Muti-scale model:** Use muti-scale CNN model to better capture characteristics of different sizes of nodule and improve classification accuracy.
- **Large scale dataset:** To improve the model's effectiveness, train it on a big dataset of lung nodule photos. Make sure the dataset covers a variety of nodule kinds and sizes to strengthen the model's flexibility to varied conditions.
- **Loss function:** Select appropriate loss function according to specific tasks, such as BCE(Binary Cross Entropy)Loss or Dice loss.
- **Intergrated approach:** Use multiple CNN models to do ensemble learning, such as voting, fusion, or stacking, to further improve the classification performance.
- **Interpretability**: Try to design a model with strong interpretability, so that doctors can understand the decision-making process of the model and improve the practical application of the model.

2) Model evaluation:

- **Cross validation:** Use cross-validation to evaluate the performance of the model in order to ensure the stability of the model on different datasets.
- **ROC curve and AUC:** Use the ROC curve and AUC (area under the curve) to assess the model's classification performance, particularly at different thresholds.
- **Real-time consideration:** For clinical applications, optimize the computational speed of the model to ensure that it can be used in real-time or near-real-time environments.
- **Generalization ability:** Verify the generalization ability of the model across different medical centers, different scanning devices, and different patient populations.

5. Optimize algorithm

According on their algorithmic logic, CNNs are classified into several types. Multi-scale CNNsand bilinear CNNs are two of the most often employed in health and life science.

1) **Multi-scale convolutional neural network**

Based on this definition, the multi-scale CNN is designed with convolutional kernels of various sizes to capture features at different scales, thereby obtaining both local and global information simultaneously. For instance, smaller convolutional kernels can capture details and features of small regions, while larger kernels are capable of perceiving a broader range of features. Such a multi-scale design enhances the model's ability to learn hierarchical structural information within images, making it more effective in processing image data of varying resolutions. In the multi-scale CNN, after the first convolutional layer and activation function, the generated feature maps are subjected to a max-pooling layer. The role of the max-pooling layer is to down sample the feature maps, which not only reduces the dimensionality of the data but also increases the translational invariance of the features, enhancing the network's robustness to variations in the input. Ultimately, the fully connected layer transforms the pooled feature maps into a fixed-length feature vector, which contains the high-level features learned by the network and can be used for classification, regression, or other downstream tasks. Furthermore, the multi-scale CNN can also enhance the capability of feature representation by parallelizing the extraction of features at different scales or by gradually merging feature information from various scales. This network structure is particularly effective for complex image processing tasks, such as object detection, semantic segmentation, and image classification, as these tasks often require the integration of visual information across different scales to make accurate decisions. Specifically, the convolution operation between the input feature map f and the convolution kernel h is defined as follows [6].

$$\mathrm{y} = \max(\,0, \sum_{\mathrm{c}} \mathrm{f_c} \times \mathrm{h_c} + \mathrm{b})$$

In the given operation, slices $\mathrm{f_c}$and $\mathscr{h}_\mathrm{c}$ correspond to the $\mathrm{c_{th}}$ segment from the feature map and the convolution kernel respectively, with b serving as the bias term. Throughout the training process, continuous learning is applied to both $\mathscr{h}$ and $\mathscr{b}$ keep learning continuously during training. A rectified linear unit (ReLU) is employed to facilitate a nonlinear transition from input to output dimensions, applying a non-linear effect following each convolution within the equation, represented by y=max(0, x), where x stands for the convolution's resultant output. Subsequent to the convolutional stage, a layer for max-pooling is integrated, which functions to isolate a selected group of features; the formula [7]:

$$\mathrm{y(i.j)} = \max_{0<\mathrm{m,n}<\mathrm{s}} \{\mathrm{x(i \cdot s + m, j \cdot s + n)}\}$$

s is the pooling size and x represents the output of the convolutional layer. The advantage of using a max-pooling layer is its translation invariance, especially when the different nodule images are not well aligned [8]. Additionally, max-pooling can significantly reduce computational requirements because it lowers the spatial dimension of the feature maps by selecting the maximum value within each pooling window. This not only helps prevent overfitting but also reduces the model's sensitivity to noise. This attribute is especially valuable when processing medical imaging data, as it enhances the learning of key features while ignoring irrelevant variations and noise, thereby speeding up the processing without sacrificing diagnostic accuracy. After the max-pooling layer, one or more fully connected layers are usually employed to further synthesize the features and yield predictive outcomes for final decision-making. In this manner, CNNS are capable of extracting and utilizing complex features from raw pixels, ultimately achieving efficient and accurate image classification.

2) Bilinear convolutional neural networks

Bilinear CNNs (BCNN) consist of two BCNNs that work in parallel, each acting as a feature extractor, and their output vectors are bilinearly combined through an outer product function. This structural design allows the BCNN model to generate more rich and detailed information compared to the standard CNN model. Specifically, the bilinear model captures the interactions between different features through the outer product, which is particularly important for fine-grained image classification tasks, such as distinguishing between different species of birds or varieties of plants. In such a model, the two branch networks do not have to be identical; they can be optimized and adjusted for specific tasks. After their output vectors are concatenated via the outer product, the resulting matrix is flattened into a long vector to facilitate classification or other types of prediction tasks. Bilinear features generally have very high dimensions, so dimensionality reduction techniques such as Principal Component Analysis (PCA) or feature selection methods are often used in practical applications to reduce computational load. With such processing, BCNNs maintain their feature representation capabilities while improving computational efficiency, adapting to more complex or detail-oriented visual recognition tasks. It may be formalized as the following quadruple:

$$\mathrm{B_{CNN}} = (\mathrm{f_{1,}f_{2,}p_{f,}c})$$

In the quadruple, $\mathrm{f_1}$ and $\mathrm{f_2}$are the feature extractors for CNN1 and CNN2, respectively, $\mathcal{P}_\mathrm{f}$is the pooling function and $\mathcal{C}$ is the classification function.

The feature extractor generates the mapping function:

$$\mathrm{F}: \mathcal{J} \times \mathcal{L} \to \mathbb{R}^{\mathcal{K}\times\mathcal{D}}$$

It considers images $\mathrm{I} \times \mathcal{J}$ and position $\mathrm{I} \times \mathcal{L}$, and the size of its output feature R is also given. Generally, $\mathcal{L}$ indicates position and scale. To obtain bilinear features, we combine the outputs of $\mathrm{f_1}$ and $\mathrm{f_2}$ at each position by computing the outer matrix product as follows:

$$\mathfrak{b}(e, \mathcal{J}, \mathrm{f_1}, \mathrm{f_2}) = \mathrm{f_1}(1, \mathcal{J})^\mathrm{T}\mathrm{f_2}(1, \mathcal{J})$$

The feature value $\mathcal{K}\mathrm{f_1}$ and $\mathrm{f_2}$ must remain compatible in the same dimension concurrently. The global image descriptor $\varphi(\mathcal{J})$ is obtained when bilinear features at all locations are aggregated of the image via a pooling function $\mathcal{P}_\mathrm{f}$. We can extract the dimensions K×N and K×M of the features by $\mathrm{f_1}$ and $\mathrm{f_2}$, respectively. Finally, the bilinear featurer $\varphi(\mathcal{J})$ is of size M×N of the global image descriptor, which will be used with the classification function $\mathcal{C}$. The bilinear descriptor is reshaped into 1D bilinear vector $\mathcal{V}(\mathcal{J})$, and perform element-wise signed square root operations as follows [9]:

$$\mathcal{Y}(\mathcal{J}) = \mathrm{sign}(\mathcal{V}(\mathcal{J}))\sqrt{\mathcal{V}(\mathcal{J})}$$

The calculation of $\mathcal{Y}(\mathcal{J})$ is followed by two normalizations steps to improve the performance of the model in practice.

6. Clinical application----application of CNN in pulmonary nodules

$$\mathcal{Z}(\mathcal{J}) = \frac{\mathcal{Y}(\mathcal{J})}{\mathcal{J}(\mathcal{J})^2}$$

CNN has been widely used as a tool for auxiliary judgment in medical images. It is difficult to diagnose pulmonary nodules according to the early lung images, and the diagnosis is great relative to the patient's physical condition and age which makes the diagnosis harder. In addition, it is also difficult to perform large-scale scene screening manually. In conclusion, the CNN model can effectively replace manual work for early intervention of pulmonary nodules.

DICESCORE performs manual detection on 354 lung CT images using the standard ED-CNN algorithm. The model's average precision is 0.962, its lowest score is 0.926, its highest score is 0.974, and its standard deviation is 0.008.

Based on the statistics presented above, we may infer that CNN is well suited for medical imaging. With the advancement of CNN, new CNN models with varying methods and topologies are developing. This section focuses on the impact of two optimized algorithms, the SVM and the RF, on their findings. The accuracy of the multi-scale CNN processing is illustrated in Figure 1:

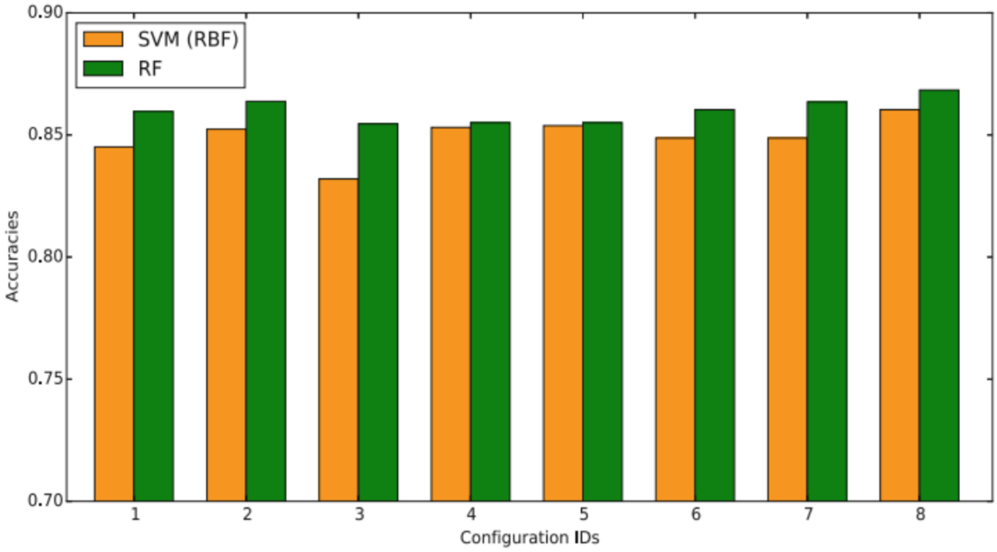

Figure 1. The classification performance of SVM with the RBF kernel and RF is based on features from the MCNN using 8 different configurations. Each configuration is assigned to a unique ID for display convenience [10].

The average accuracy is 84%, with the greatest accuracy being 86.84%. The BCNN has an average accuracy of 91.47% and a maximum accuracy of 91.99%. Meanwhile, when compared to others in the same database, VGG16 and VGG19, the accuracy of BCNN is the greatest, as shown in Table 2:

Table 1. Classification results of BCNN models compared to previous works

Reference	Database	Methods	Accuracy(%)	AUC
MonKam et al.	LIDC-IDRI(2635 nodules)	Three CNNs with various patch sizes	88.28	0.87
Shen et al.	LIDC-IDRI(2618 nodules)	Multi-crop CNN	87.14	0.93
Kaya et al.	LIDC-IDRI(1402 nodules)	AlexNet + cascaded classifier	84.70	--
Shi et al.	LIDC-IDRI(1400 images)	VGG16 with SVM	91.50	--
Zhao et al.	LIDC-IDRI(743 images)	Agile CNN(LeNet + AlexNet)	82.23	0.877
Zhao et al.	LIDC-IDRI(743 images)	Transfer learning CNNs	85.00	0.94
Shen et al.	LIDC-IDRI(4252images)	Hierarchical semantic CNN	84.20	0.856
Our proposed method	LIDC-IDRI(3186images)	$[VGG16]^2$+SVM $[VGG19]^2$+SVM $[VGG16,VGG19]^2$+SVM	91.84 90.58 91.99	0.948 0.94 0.959

The statistics shown above show that different architectures and algorithms have varied picture recognition accuracies. The bilinear convolutional neural network is the most accurate and has clear benefits over other convolutional neural networks. After fitting and analyzing the ConvPath open-source data with the same linear model, we can conclude that the multi-precision model has an average accuracy of 70.1% and a maximum accuracy of 78.62%, while the blinear convolutionary neural network has an average accuracy of 86.38% and a maximum accuracy of 89.12%.

7. Conclusion

After analysis, we find that bilinear convolutional neural network is the most accurate. It has an average accuracy of 86.38%, while multi-precision model only 70.1%. However, bilinear CNNs also face challenges with computational costs during model training and deployment. Due to the high dimensionality of bilinear features, even modern computational hardware may struggle with slow processing speeds and high storage demands. To mitigate these issues, researchers are exploring more efficient network architectures and compression techniques to reduce the complexity of models while maintaining their performance. In real clinical settings, fast model inference times are crucial for improving patient experience and healthcare efficiency. Therefore, new algorithms are being developed to accelerate the computation of bilinear features, along with novel hardware acceleration technologies, such as using Field Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs).Additionally, to address the issue of data scarcity, researchers are developing more advanced data augmentation techniques and semi-supervised learning methods that can make the most of limited data without increasing the cost of annotation. At the same time, privacy protection is an important consideration for medical imaging data. Thus, the development of anonymization and encryption technologies is also key to enabling the practical application of bilinear CNNs in real medical scenarios.

Despite these challenges and limitations, the superior performance of bilinear CNNs in visual recognition tasks makes them a focal point of research. Through continuous technological innovation and interdisciplinary collaboration, it may be possible to resolve these issues, allowing bilinear CNNs to play a significant role in broader fields, especially in the development of precision medicine and personalized treatment strategies. In summary, as a powerful machine learning tool, bilinear CNNs have a promising future in applications, but it requires the joint efforts of scientists, engineers, and medical professionals to overcome current technical and application barriers. With the maturation of related technologies and the emergence of more innovative solutions, we can expect bilinear CNNs to play an increasingly important role in the medical field in the future.

References

[1] Am J Respir Crit Care Med Vol. “What is a Lung Nodule?”. ATS Patient Education Series 193, P11-P12. 2016. https://www.thoracic.org/patients/patient-resources/resources/lung-nodules-online.pdf

[2] Exploration of Teaching Practice for Machine Learning Courses in the Big Data Environment D kejun; -“Internet Weekly”- 2023-05-20)

[3] N yongqi; W dejian; F yanyan; -“Computer Engineering and Applications”- 2022- 09-27 20:19

[4] Moslem Azamfar, Xiang Li, Jay Lee. "Intelligent ball screw fault diagnosis using a deep domain adaptation methodology", Mechanism and Machine Theory, 2020

[5] (https://courses.cs.washington.edu/courses/cse416/22su/lectures/10/lecture_10.pdf)

[6] (R. P. Meenaakshi Sundhari. "Enhanced histogram equalization based nodule enhancement and neural network based detection for chest x-ray radiographs",7 Journal of Ambient Intelligence and Humanized Computing, 2020

[7] (Second-Order Global Attention Networks for Graph Classification and Regression, Beijing, China, August 27-28, 2022)

[8] Jason B ,2019,7 (Gentle Introduction to Pooling Layers for Convolutional Neural Networks)

[9] Mastouri R, Khlifa N, Neji H, Hantous-Zannad S. A bilinear convolutional neural network for lung nodules classification on CT images. Int J Comput Assist Radiol Surg. 2021 Jan;16(1):91-101. doi: 10.1007/s11548-020-02283-z. Epub 2020 Nov 2. PMID: 33140257.

[10] Saberioon M, Císař P, Labbé L, Souček P, Pelissier P, Kerneis T. Comparative Performance Analysis of Support Vector Machine, Random Forest, Logistic Regression and k-Nearest Neighbours in Rainbow Trout (Oncorhynchus Mykiss) Classification Using Image-Based Features. Sensors (Basel). 2018 Mar 29;18(4):1027. doi: 10.3390/s18041027. PMID: 29596375; PMCID: PMC5948703.