

# Evaluation of Neural Network for Automated Classification of Plant Component on Histological Section

Marina A. Nikitina

V. M. Gorbatov Federal Research Centre for Food Systems of RAS

Moscow, Russia

nikitinama@yandex.ru

**Abstract**—Classification of plant component on image histological sections is critical for determining non-compliance type of undeclared additive and further action for technologist, or other responsible person. However, this task is often challenging due to the absence of professional histologists or non-compliance with the conditions of microstructural analysis and also the subjective criteria for evaluation. In this study, we propose a machine learning model that automatically classifies the plant component on images histological sections. Our model uses a convolutional neural network to identify regions of plant components, then aggregates those classifications to infer predominant and minor plant components on histological sections image. We evaluated our model on an independent set of 95 images histological sections. It achieved a kappa score of 0,525 and an agreement of 66,6% with three histologists for classifying the predominant plant component, slightly higher than the inter-histologists kappa score of 0,485 and agreement of 62,7% on this test set. All evaluation metrics for our model and the three histologists were within 95% confidence intervals of agreement.

**Keywords**— convolutional neural network, activation functions, intelligent system, control methods, database and knowledge, decision making

## 1. INTRODUCTION

The need for automation of visualization processes is out of question. But the question is how to implement ANN as effectively as possible in the workflow of histologists. We do not want to repeat the mistakes of the Digital Pathology Association, which were made several years ago when using automatic diagnosing [1-3]. The statistical analysis of the application of this method in mammography featured many errors of the first kind (erroneous deviation from the null hypothesis). There are many methods and algorithms for solving the problem of recognizing objects in an image [4-8], but all these ideas are not inferior in accuracy of the result, simplicity and speed of artificial neural networks [9]. The modern deep neural networks are usually based on convolutional network-based architectures such as cognitron and neocognitron [10]. Their effectiveness and rapid development is due to a hybrid approach to architectural solutions, the development of learning methods, additional error protection methods. In this regard, the possibility of integrating a CNN-based analysis tool to solve the problems of evaluating histological sections of food raw materials and products was of great interest. The study aimed at determining the main microstructural characteristics (classification parameters) for the identification of plant components in meat raw materials and finished products with CNN.

Step-by-step solution to the problem of identifying an image of a histological cut is shown in Figure 1.

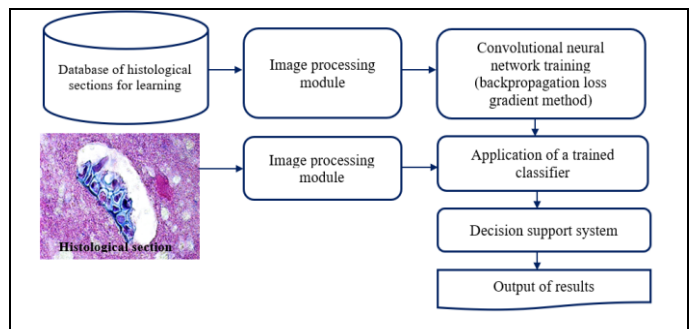


Fig. 1. Step-by-step solution algorithm

To solve this problem, it is required to collect a representative sample - DataSet with marked areas on histological sections. Semantic segmentation is carried out in order to divide the image into separate groups of pixels that correspond to any one object (for example, carrageenan, vegetable gum, etc.). For this purpose, a database structure has been developed, in which identification features and information about marks for testing and training are stored for each image. Before placing an image in the database, it undergoes pre-processing, segmentation and classification (assignment to a particular class of undeclared components). At the training stage of the neural network, augmentation was used (changing the image, for example, flip the image horizontally, vertically, at an angle of 45°) to increase the sample size. At least 10 variants of histological sections were presented for each type of “inclusion”. In our work, DataSet for one inclusion: the size of the training sample is 672, the test sample is 168. The number of epochwise trainings is 50.

## 2. DATABASE OF HISTOLOGICAL SECTIONS

First it is necessary to collect a database of histological sections. The data must contain the same set of object classes for their segmentation. After collecting the data, it is necessary to mark up the images. The marking means recording the coordinates of the polygons of each class object in a separate file in the \*.txt or \*.xml format. To find the main microstructural characteristics (classification parameters/situational classifier) of identification, on the example of plant components in the composition of meat raw materials and finished products, the structure of a unified information database of histological indicators was developed in the course of the work. The first stage included the selection of identification features that must necessarily be included in the database structure – particle shape, size, tinctorial properties of particles (the ability to be stained with

histological dyes), and fragments of the soybean shell for protein components. In this subject area, an Entity is a plant component, while an attribute is data describing the properties/attributes of the Entity. Table 1 shows the database structure. Each attribute is assigned a letter designation «Table 1, Column 3».

**TABLE I. RELATIONAL DATABASE STRUCTURE**

Entity	Attribute value	Designation
Name of components	- Particle shape	A <sub>1</sub>
	- Size	A <sub>2</sub>
	- Tinctorial properties (at hematoxylin and eosin staining)	A <sub>3</sub>
	- Tinctorial properties (at Lugol's solution staining)	A <sub>4</sub>
	- Characteristic microstructure (photo)	A <sub>5</sub>

The signs specified in Column 3 corresponding to the key concepts can take certain syntactic values, which are considered as constants. If the attribute takes this value, it is assumed that the corresponding variable is equal to this value. Based on the production rules, the process will be carried out as follows. An image of the histological section *A* is fed to the input, and then it is processed.

### 3. ALGORITHM TO AUTOMATE THE DETECTION OF INCONSISTENCIES

To implement the stages of processing images of histological sections of meat samples, the following algorithm is proposed for their implementation. Processing of a histological section image *A*, in general, includes the following steps: 1) preprocessing of section images (noise removal, palette optimization, etc.); 2) color segmentation based on palette minimization; 3) approximation of boundaries of the areas highlighted in the image; 4) area size determination; 5) particle shape determination; 6) particle color determination; 7) identification of the presence of counterfeits; 8) results' output regarding the determination of the presence of counterfeits. Consistent implementation of the above steps will allow obtaining an informed decision on the presence or absence of certain types of counterfeits of meat products. In neural networks, both the activation function and the error handling methods on each layer of the neural network play an important role. The accuracy of the classification made by the artificial neural network relative to the object on the histological section depends on these two indicators. The network is considered trained when the comparison error is equal to zero or corresponds to the maximum allowable error value.

### 4. ACTIVATION FUNCTION

At this stage of the study, the *TensorFlow* library was used to study and compare the effect of several activation functions on classification results. The study of activation functions resulted in the obtainment of the following accuracy indicators. We got the best result when using the ReLu activation function (0.9843). The next result was shown by the SoftPlus (0.9765), eLu (0.9687), Sigmoid (0.7851), SoftSign (0.7773) and TanH (0.7591) activation function. Thus, the most accurate classification is achieved by using the ReLu activation function to train a convolutional neural network (CNN) in this study.

### 5. NEURAL NETWORK ARCHITECTURE

At this stage of the study, a kind of convolutional neural network architecture with two layers of convolution (Convolutional, C-Layer) and pooling (Subsampling, S-Layer), which alternate one after another, is proposed. The CNN architecture receives a normalized image with a standardized size on the input layer, if required. Inner layers are consecutive convolution layers, with normalization and pooling layers. The main thing to understand is that a convolution layer means a layer that transforms a part of the input image, a 3 x 3 matrix, into a 1 x 1 pixel using matrix transformations. On the output layer of this type of architecture, we get a set of features of this image. We evaluated our model on an independent set of 95 images histological sections. It achieved a kappa score of 0,525 and an agreement of 66,6% with three histologists for classifying the predominant plant component, slightly higher than the inter-histologists kappa score of 0,485 and agreement of 62,7% on this test set. All evaluation metrics for our model and the three histologists were within 95% confidence intervals of agreement.

### ACKNOWLEDGMENT

This article is published as part of scientific research project No. 0585-2019-0007 under the state assignment of the Federal state budgetary scientific institution 'V.M. Gorbatov's Federal Research Centre for Food Systems' of RAS.

### REFERENCES

- [1] Computer-aided Diagnosis: The Tipping Point for Digital Pathology. Digital Pathology [Electronic resource]. — Access mode: <https://digitalpathologyassociation.org/blog/computer-aided-diagnosis-the-tipping-point-for-digital-pathology/> (01.11.2021).
- [2] Nikitina, M.A. Intelligent analysis of the microstructure of the meat system based on agent and neural network technologies / M.A. Nikitina // In: V All-Russian Pospelov conference with international participation "Hybrid and synergistic intelligent systems" - HASIS-2020. – 2020. – P. 318–326.
- [3] Nikitina, M.A. Identifying States of Complex Technological and Biotechnological Systems under Uncertainty Conditions / M.A. Nikitina // In: 7th International Conference on Control, Decision and Information Technologies – Codit'2020. – 2020. – P. 825–830.
- [4] Gonzalez, R.C. Digital image processing / R.C. Gonzales, R.E. Woods - Prentice-Hall of India Pvt. Ltd., 2005. – 976 p.
- [5] Callan, R. The Essence of neural networks / R. Callan – Prentice Hall PTR, 1999. – 232 p.
- [6] Huang, G. Densely connected convolutional networks / G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger // In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Proceedings. – 2017. – P. 2261–2269.
- [7] Hardt, M. Identity matters in deep learning / M. Hardt, T. Ma // In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, Proceedings. – 2017. – P. 149804
- [8] Abadi, M. TensorFlow: A system for large-scale machine learning / M. Abadi, P. Barham et al. // In: 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI, Savannah, Georgia, US, Proceedings. – 2016. – P. 265–283
- [9] Deepthi, D.R. A new hierarchical pattern recognition method using mirroring neural networks / D.R. Deepthi, K. Eswaran // Int. J. Com. App. – 2010. – Vol. 1(12). – P. 70–78.
- [10] Yang, L. Real-time object recognition algorithm based on deep convolutional neural network / L. Yang, L. Wang, S. Wu // In: IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, Proceeding. – 2018. – P. 331–335.