

# Adaptation of Deep Convolutional Neural Networks for Cancer Grading from Histopathological Images

Stefan Postavaru<sup>1,2</sup>, Ruxandra Stoean<sup>3(✉)</sup>, Catalin Stoean<sup>3</sup>,  
and Gonzalo Joya Caparros<sup>4</sup>

<sup>1</sup> Faculty of Mathematics and Computer Science,  
University of Bucharest, Bucharest, Romania

<sup>2</sup> Bitdefender, Bucharest, Romania  
spostavaru@bitdefender.com

<sup>3</sup> Faculty of Sciences, University of Craiova, Craiova, Romania  
{ruxandra.stoean,catalin.stoean}@inf.ucv.ro

<sup>4</sup> School of Telecommunication Engineering, University of Malaga, Málaga, Spain  
gjoya@uma.es

**Abstract.** The paper addresses the medical challenge of interpreting histopathological slides through expert-independent automated learning with implicit feature determination and direct grading establishment. Deep convolutional neural networks model the image collection and are able to give a timely and accurate support for pathologists, who are more than often burdened by large amounts of data to be processed. The paradigm is however known to be problem-dependent in variable setting, therefore automatic parametrization is also considered. Due to the large necessary runtime, this is restricted to kernel size optimization in each convolutional layer. As processing time still remains considerable for five variables, a surrogate model is further constructed. Results support the use of the deep learning methodology for computational assistance in cancer grading from histopathological images.

**Keywords:** Image processing · Histopathological slides · Classification · Deep convolutional neural networks · Parametrization

## 1 Introduction

As with all types of cancer, early detection and correct grading (made possible by modern medical imaging technologies) increase the chances of cure by allowing timely counteractive treatment measures. Hence, screening interventions are more and more regular nowadays and the consequent amount of related imaging data is growing vertiginously. With the additional known considerable inter- and intra-observer variability in interpreting the resulting images, computational image evaluation for the automated detection of landmarks and determination of cancer stage is getting a high interest as it can efficiently assist the medical

decision-making in terms of precision, invariability and time economy. Among such computational approaches, those that neither need further supervision from the medical experts, in terms of manual delineation or feature specification, nor manual parametrization from the computer scientists are obviously preferred in general, for any diagnosis task in medicine [9–12, 21, 37].

The current study thus attempts to tackle the medical problem of grading colorectal cancer from available histopathological slides by convolutional neural networks (CNN). The technique is recognized for its high-performing and autonomous learning capability as a result of its special layer types and hierarchical arrangement. It would thus be indeed supportive in diminishing the routine work for the medical experts, since it can configure the important features by itself [36] for the common images and need expert reconfirmation only for the problematic ones. The CNN is nevertheless problem-dependent with respect to parametrization and the paper will consequently furthermore investigate an inexpensive option in design to allow automatic tuning of CNN sensitive variables. A surrogate model is thus constructed to streamline the framework as regards runtime but still maintain accuracy, and an heuristic approach [6, 25] is appointed to search for the optimal parameters. The work is not only important from the medical perspective, but also from the CNN side, as the trust in its potential must be also shown on real-world data sets and not only on toy problems [36].

The paper is organized as follows. Section 2 describes the real-world data set containing histopathological colorectal cancer slides with the earlier models for this collection and outlines the state of the art in CNN application for medical image interpretation. The CNN architecture is presented in Sect. 3, together with its augmentation in terms of automatic tuning with an interest in runtime efficiency. Experimental findings are given in Sect. 4, while the conclusions are drawn in Sect. 5.

## 2 Problem and Prior Medical Computational Diagnosis

The medical data set contains histopathological digitalized slides for colorectal tissues from the Emergency Hospital of Craiova, Romania, with representatives of both benign (denoted by G0) and malignant (G1–G3) classes. The collection<sup>1</sup> contains 357 images (at resolution  $800 \times 600$ ), where each is assigned to only one cancer class. There are 62 G0 entries, 96 G1 records, 99 G2 and 100 G3 slides.

### 2.1 Previous Attempts to Classify the Data

Within the examination of the histopathological slides for the detection of colorectal cancer, the medical expert looks for a uniform pattern of glands and nuclei as a sign of the absence of cancer, while, conversely for a malignant tissue, for a high variation in size, shape and texture. The correct segmentation of glands and nuclei and the appearance of their features is thus important for

---

<sup>1</sup> IMEDIATREAT Project, <https://sites.google.com/site/imediatreat/>.

the recognition of cancer. Earlier attempts on the current problem consequently started from the segmentation of glands through a watershed algorithm parameterized by evolutionary algorithms (EA) [30,34]. A distance transform algorithm was further employed to additionally find the nuclei. Departing from the discovered landmarks, measurements quantifying the number, area, perimeter, radius for the glands and nuclei, as well as for corresponding Delaunay triangles and Voronoi polygons, resulted into 76 numerical extracted features. Support vector machines (SVM) achieved 79.89% classification accuracy on this numerical collection [33]. Feature selection was further employed to help the SVM and considered a consistency-based filter, a correlation mechanism, principal component analysis and a genetic algorithm (GA). The last led to an accuracy increase of 4 percents and the determination of the more important features [31].

The current work however uses the CNN to allow for direct classification into G0–G3 with its own implicit inner feature detection. This will therefore be an alternative “all-in-one”, expert-independent approach to slide interpretation.

## 2.2 Convolutional Neural Networks for Medical Image Diagnosis

First introduced as “self-organizing” neural networks unaffected by position shifts in [8], CNN have developed over years into powerful learning models, trainable with the efficient method of back-propagation. As presented in [20], convolutional layers present a degree of robustness to translation and distortion, thus being well-suited for problems where the input presents spatial relations. The advancement of computing power in recent years made a range of solutions based on deep-learning with CNN possible, with better results than previous state-of-the-art methods: CIFAR image classification [18], video classification [16], action-recognition [14] and others.

Given their high performance on tasks of artificial vision and even prophesied to be “the most disruptive technology [...] since the advent of digital imaging” [7], CNN were naturally recently applied for medical image classification and diagnosis. [26] implemented a CNN on images obtained with computed tomography to classify the given input into one of five classes. They used an augmented data set with 4298 entries to train and test the CNN model, obtaining an accuracy of 94.1% on the test set. A solution for detecting mitosis in breast cancer histology images with CNN was proposed by [5]. A sliding-window was moved over the full image in order to capture a context for each pixel. These patches were fed into the learned model and labeled accordingly. The IEEE Transactions on Medical Imaging journal even had a dedicated special issue in 2016, i.e. *Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique*, where CNN was employed for learning from crowds in detecting mitosis in breast cancer images [2], for lung cancer pattern classification from high resolution computed tomography [3] and for brain tumor segmentation from MRI images [24]. Among them there is also the study of [28], which is specifically for the colorectal cancer problem at hand and concerns the detection of nuclei (from images with 20000 annotated nuclei categorized into four classes) by CNN and subsequent, separate classification by a neighboring assemble predictor.

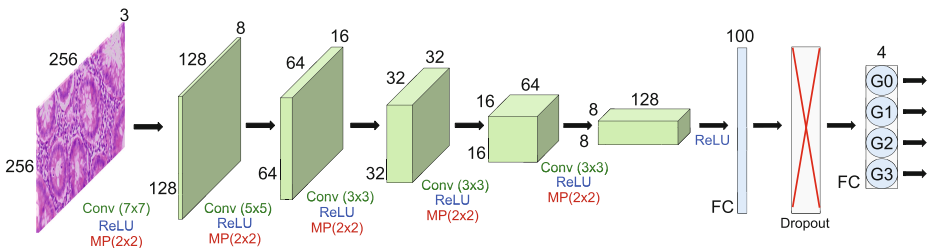
In opposition, the present aim is to use the CNN for direct classification of the histopathological slides into four classes, i.e. normal (G0) and cancer grades G1–G3, with neither intermediate segmentation nor in tandem with other classifiers.

### 3 Proposed Convolutional Neural Network Methodology

The task of diagnosing histopathological images can be formulated as a supervised learning problem: having a data set of pairs  $(X_i, c_i)$ , where  $c_i$  represents the ground-truth class of the image  $X_i$ , the aim is to train a model capable of discriminating between the four possible classes.

#### 3.1 Convolutional Neural Network Architecture

A CNN learning model is proposed to solve this task. It is carefully designed to avoid overfitting, given the small data set size, while being capable to generalize well enough in order to provide a competitive accuracy on the final test suite. The CNN (Fig. 1) receives a volume of  $256 \times 256 \times 3$  (image width  $\times$  image height  $\times$  three color channels) representing a scaled version of the original images and will output a 4-dimensional vector, interpreted as unnormalized scores for each class. The volume is transformed through 5 convolutional layers, with weight-sharing kernels of sizes (lengths) -  $KS$  (7, 5, 3, 3, 3) (i.e. filter sizes of  $KS \times KS$ ), padding (of the volume with zero around the border) chosen such that the first two dimensions of the input are preserved. The first kernel depth -  $KD$  (number of filters) is 8, and each consecutive depth is multiplied by 2. Every convolution layer slides each of the  $KS \times KS \times KD$  filters over all spatial locations of the image by computing dot products between the shared weight vector within the slice and the image piece. The result of convolving each filter with the image is an activation map of  $N \times N$ , where there are  $N$  unique positions to place the  $KS \times KS$  filter. Large values on the activation map are assigned to patters that better stimulate the neurons on that filter.



**Fig. 1.** CNN Architecture: each convolution function (written in green) is followed by ReLU non-linearity (blue) and a Max-pooling ( $2 \times 2$ ) operation (red). Thus, each consecutive volume doubles in depth and halves its width and height. A dropout layer cuts a subset of activations from the preceding fully-connected one. The last fully connected layer outputs the unnormalized scores for each class. (Color figure online)

The obtained activation maps (whose number is equal to the number of filters) are stacked in a volume  $N \times N \times KD$  to enter the next layer [15]. Convolution is followed by a ReLU transfer layer (a Rectified Linear Unit to add nonlinearity) and a Max-Pooling layer (for a downsampling operation of the image following the maximum value in the window) with window size  $2 \times 2$ , effectively reducing the input volume and making the network resistant to spatial shifts.

The learning speed was greatly increased by applying batch-normalization [13] to the last 4 convolutional layers, achieving the same accuracy within 10% of the original number of iterations. The convolutional layers are followed by two fully connected ones. Dropout is also used between the last two fully-connected layers (where neuron activations are randomly blocked within training) [29]. During hyper-parameter tuning, it was found that a dropout rate of 90% manages to greatly reduce overfitting, while maintaining the ability of the network to learn. During training, the KL-Divergence [19] between the ground-truth class distribution and the network outputs interpreted as probabilities is minimized through a softmax function.

### 3.2 Kernel Size Parametrization

Although acknowledged for autonomous classification and implicit feature learning, parametrization of the network is known to be problem-dependent. To the best of our knowledge, it is only the recent study of [36] that accomplishes the tuning of six hyper-parameters through EA with message passing interface to distribute fitness computation and hence accelerate runtime. The CNN architecture has three convolutional layers and, for each of those, two parameters are heuristically determined: the number of filters and the kernel size. Although known to be flexible, EA are too expensive in runtime and hence the current study alternatively employs Latin hypercube sampling (LHS, [22]) for kernel size automatic generation from the five CNN convolutional layers.

### 3.3 Surrogate Model Design

In order to reduce runtime, the present paper proposes an accompanying surrogate model to the automatic setting of kernel sizes. Such regression options proved to be efficient before for the same purpose of parametrization [32]. A data set of input parameter values and the obtained CNN accuracy result on validation samples is learnt by the regression models that represent the surrogates. Each such model is next used to estimate the accuracy for new input parameters. A GA [17] is used to search the parameter space for the optimal settings, with the output estimated by the regression models.

## 4 Experimental Results

The experimental section is split into two parts. While the aim of the first experiment is to reach significantly better classification accuracy results than in the

previous attempts on the same data set, the goal of the latter is to better understand the choice for the kernel size parameter. The organization of the two experiments follows the guidelines in [4]. In order to deploy the central architecture, a Python with Tensorflow [1] framework was used.

#### 4.1 Experiment 1: Manual Tuning

**Pre-experimental Planning.** From the initial experiments it was noticed that the runtime is relatively high, which would have restricted the number of setups to be tried. In search of a balance between accelerating runtime but not decreasing quality, the pictures were resized to a resolution of  $256 \times 256$  pixels, from the original  $800 \times 600$ . This ensures a smoother training and also allows the usage of a low-volume model, as one of the measures implied in preventing overfitting.

**Task.** The goal is to test the CNN architecture, with its own implicit inner feature detection, on the colorectal cancer images and achieve a classification accuracy superior to that of the previous multi-stage approaches [31, 33].

**Experimental Setup.** The samples are split into training/validation/test sets with the ratios 0.5, 0.25 and 0.25. This partitioning was chosen for a better control and measurement of the generalization power for the trained model. Given the small data set size, an extra step in ensuring that every class has the same percent of individuals in each set was taken, in order to avoid situations where a class is poorly represented by the training suite.

For fine-tuning the network in search of a successful set of hyper-parameters, it is aimed to maximize the accuracy on the validation set. In order to assign a score to a vector of hyper-parameters, the average accuracy obtained on validation sets of random partitions of the data collection by models trained with the respective hyper-parameters was used, as given by Eq. (1), where  $M_{h,X}$  is the model trained with hyper-parameters  $h$  and training set  $X$ ,  $Acc_V(M)$  is the accuracy of model  $M$  computed on the validation set  $V$  as  $\frac{\#correct}{|V|}$  and  $\mathbf{E}_X$  is the expected value with respect to the distribution of  $X$ .

$$score(h) = \mathbf{E}_X[Acc_V(M_{h,X})] \quad (1)$$

After training the model 30 times with random initialization, a mean accuracy of 92.5% is obtained on the validation sets, with learning rate  $10^{-3}$ , batch size 20, kernel sizes (7, 5, 3, 3, 3) and first kernel depth of 16.

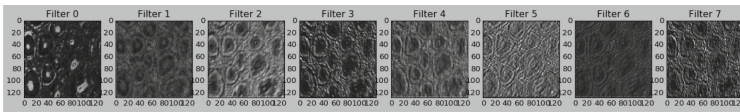
**Results and Visualization.** On the test sets, the model was able to correctly classify, on average, 90.15% of examples. An example confusion matrix is given in Table 1. Figure 2a and b depict the activation intensities of the first and second convolutional layers, given a G0 image.

**Table 1.** Example of a confusion matrix on one of the test sets.

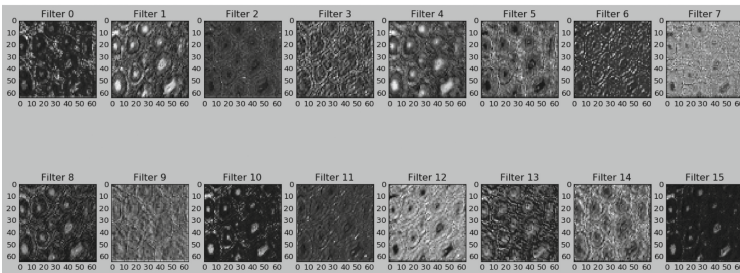
	G0	G1	G2	G3
G0	0.91	0	0.09	0
G1	0	1	0	0
G2	0	0.2	0.73	0.07
G3	0	0	0	1

**Discussion.** Overall, the classification results are significantly better than those obtained in [31,33], as the percentage increase is of over 10%. The confusion matrix in Table 1 is obtained from a single run on the test set and it reveals that samples from G1 and G3 categories are classified with no error and 9% of the normal tissues are classified as grade 2 cancer. This type of errors (from G0 to G2) also occurred in [31], they are probably due to the manner of obtaining the G0 slides, as they are cut from larger images that contained borders between healthy and cancerous tissues and G2 had many instances [33]. The most problematic category is G2, with an accuracy of only 73%, but the vast majority of misclassified examples are labelled as G1, and none are considered normal tissues. Still, the confusion matrix is only computed to have a general overview, as in other runs the distribution might differ.

As observed in Fig. 2a and b, the network learned filters to recognize gland interior, shape and nuclei. They represent sections of the volume that passes through the network. The number of slices on each layer is determined by the depth of the convolution kernels, i.e. 8 and 16. Highlighted areas correspond to image regions producing larger values on that layer, in contrast to darker areas.



(a) Activations of the first layer



(b) Activations of the second layer

**Fig. 2.** Activation maps for a histological slice with normal tissue.

## 4.2 Experiment 2: Inspection of the Kernel Size Parameters Space

**Pre-experimental Planning.** Despite the image resizing, the runtime of the CNN remains relatively high in order to allow for the investigation of a large variety of parameter settings. The first attempt was to employ a random walk procedure to search for a better set of hyper-parameters. However, the search was still slow, as, in order to reach a relevant score, 10 different models were trained for every solution and the average on the validation sets was taken into account. Next, LHS was considered in order to have a better covering of the search space. All the tried combinations were therefore gathered, as they provided valuable information regarding the proper choice of the values for the kernel sizes. However, an optimization procedure (e.g., a GA) that would automatically search for the best combination of hyper-parameters was still not possible due to the time-consuming evaluations. Each fitness evaluation would assume the run of the CNN using the values provided by the GA candidate solutions. Consequently, the complement solution that was imagined during pre-experimentation was to build surrogate models. These would learn the correspondence between the input parameters and the accuracies obtained on the validation sets and would provide new outputs for novel configurations.

**Task.** The aim is to investigate the choice of the CNN kernel size hyper-parameters to improve the classification accuracy on the validation and test sets.

**Experimental Setup.** A random walk algorithm that had a budget of 10 evaluations per solution was firstly considered. Each such evaluation attempts to produce a score for a vector of hyper-parameters, according to Eq. (1). Given the computational constraints, the average validation score of only 10 models is taken. The solution with the highest score is taken to the test phase, where the performance of 30 models with the same hyper-parameters is evaluated.

The total number of tried LHS configurations is 75. The 5 kernel size parameters are chosen from the following intervals: [6, 16], [4, 12], [2, 6], [2, 6] and [2, 6]. As in the case of the random walk procedure, 10 models are considered and the average on the validation set represents the final outcome.

The random walk and the LHS configurations are next gathered and used by 4 different regression models. These are subsequently used to simulate further results in points that have not been previously explored with CNN. In order to search for better parameter settings, a canonical GA is used and the result provided by the surrogate models for each explored configuration is the value of the fitness evaluation. The regression models that are tried in the current experiment are: a linear model, a SVM with radial and linear kernels and regression decision trees. This conducts to 4 versions of the GA, as each one has a different fitness evaluation function. The population size of the GA is of 50, 150 iterations are considered as stop condition, while the crossover and mutation probabilities are of 0.8 and 0.1, respectively. The regression models and the GA are encoded into R implementations [23, 27, 35].



**Table 2.** The hyper-parameter values as discovered by the GA with the surrogate models. The associated percentages show the regression model estimations beside the actual evaluations of the CNN on the validation and test sets.

Regression model	Kernel size values	Surrogate output	CNN validation	CNN test
Linear model	(6, 11, 2, 5, 5)	93.93%	93.26%	90.91%
SVM radial	(12, 11, 4, 5, 5)	93.87%	93.41%	89.05%
SVM linear	(6, 11, 2, 5, 5)	93.91%	93.26%	90.91%
Regression trees	(10, 6, 3, 2, 5)	94.35%	94.17%	91.44%

**Results and Visualization.** The best result on the test set as obtained from the random walk is of 91.06% and is achieved from the parameter values (7, 4, 4, 4, 4). Table 2 shows the combinations of kernel size parameters given by the GA with the considered surrogate models and the results obtained for those settings.

**Discussion.** Generally, Table 2 indicates that there is a good agreement between the surrogate estimated outputs and the CNN actual results on the validation set. As concerns the prediction accuracy on the test set, the only weak result is obtained for the parameters discovered when the GA used the SVM with a radial kernel as the fitness function. Table 2 indicates that there is a wide variety of input values for the kernel sizes that achieve good results. As each surrogate approach models the solution landscape differently, the GA naturally found various sets of parameters, which generally proved to be successful (except for SVM radial) when the actual CNN was tested on them. The most appropriate found solution is the one discovered through the regression decision trees.

## 5 Conclusions and Future Work

A CNN approach is considered for the automated diagnosis of a set of colorectal cancer histopathological slides. The method provides significantly better results as compared to the previous approaches on the same data set. A drawback of the method is the large training time which does not permit the user to try a wide-range of CNN hyper-parameters. In this sight, by departing from a set of parameters and their results on the validation set, several regression models were employed to explore the regions of a surrogate parameter search landscape. A GA is used to intensify the search and several sets of parameter values are suggested. The approximations on the validation set coincide to a high extent with the actual values obtained by the CNN.

As concerns future work, transfer-learning seems a viable approach to bypass the generalization limits imposed by the relatively small dataset size. By training a set of convolutional layers on a similar dataset and using the resulting weights as a starting point of the main model afterwards, one can possibly build a more robust solution. Also, new images are collected in the IMEDIATREAT project, so the data collection will expand, profiting even more from the CNN potential.

**Acknowledgments.** The second and third authors acknowledge the support of the research grant no. 26/2014, IMEDIATREAT - Intelligent Medical Information System for the Diagnosis and Monitoring of the Treatment of Patients with Colorectal Neoplasm -of the Romanian Ministry of National Education - Research and the Executive Agency for Higher Education Research Development and Innovation Funding.

## References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015). <http://tensorflow.org/>. Software available from tensorflow.org
2. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., Navab, N.: Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Trans. Med. Imaging* **35**(5), 1313–1321 (2016)
3. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Med. Imaging* **35**(5), 1207–1216 (2016)
4. Bartz-Beielstein, T., Preuss, M.: The future of experimental research. In: Bartz-Beielstein, T., Chiarandini, M., Paquete, P., Preuss, M. (eds.) *Experimental Methods for the Analysis of Optimization Algorithms*, pp. 17–49. Springer, Heidelberg (2010)
5. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013. LNCS*, vol. 8150, pp. 411–418. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40763-5\\_51](https://doi.org/10.1007/978-3-642-40763-5_51)
6. Czibula, G., Crişan, G.C., Pinteă, C.M., Czibula, I.G.: Soft computing approaches on the bandwidth problem. *Informatica* **24**(2), 169–180 (2013). <http://dl.acm.org/citation.cfm?id=2773202.2773203>
7. Daher, N.M.: Deep learning in medical imaging: the not-so-near future. *Diagnostic Imaging*. <http://www.diagnosticimaging.com/pacs-and-informatics/deep-learning-medical-imaging-not-so-near-future>
8. Fukushima, K.: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**(4), 193–202 (1980)
9. Gorunescu, F., Belciug, S.: Boosting backpropagation algorithm by stimulus-sampling: application in computer-aided medical diagnosis. *J. Biomed. Inform.* **63**, 74–81 (2016)
10. Gorunescu, F., Belciug, S., Gorunescu, M., Badea, R.: Intelligent decision-making for liver fibrosis stadialization based on tandem feature selection and evolutionary-driven neural network. *Expert Syst. Appl.* **39**(17), 12824–12832 (2012). <http://dx.doi.org/10.1016/j.eswa.2012.05.011>
11. Iantovics, B.L.: Agent-based medical diagnosis systems. *Comput. Inform.* **27**(4), 593–625 (2012). <http://www.cai.sk/ojs/index.php/cai/article/view/234>

12. Iliescu, D.G., Dragusin, R.C., Cernea, D., Patru, C.L., Florea, M., Tudorache, S.: Intrapartum ultrasound - an integrated approach for best prognosis. *Med. Ultrasonography* **19**(1), 932 (2017)
13. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*, pp. 448–456 (2015)
14. Ji, S., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(1), 221–231 (2013)
15. Karpathy, A.: Stanford university cs231n: convolutional neural networks for visual recognition. <http://cs231n.github.io/convolutional-networks/>
16. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732 (2014)
17. Kramer, O.: *Genetic Algorithm Essentials*. SCI, vol. 679. Springer, Cham (2017). doi:[10.1007/978-3-319-52156-5](https://doi.org/10.1007/978-3-319-52156-5)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1106–1114 (2012)
19. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79–86 (1951)
20. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
21. Li, R., Emmerich, M.T.M., Eggermont, J., Bovenkamp, E.G.P., Bäck, T., Dijkstra, J., Reiber, J.H.C.: Optimizing a medical image analysis system using mixed-integer evolution strategies. In: Cagnoni, S. (eds.) *Evolutionary Image Analysis and Signal Processing*. SCI, vol. 213, pp. 91–112. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-01636-3\\_6](https://doi.org/10.1007/978-3-642-01636-3_6)
22. McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**(2), 239–245 (1979). <http://www.jstor.org/stable/1268522>
23. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.: e1071: Misc Functions of the Department of Statistics, TU Wien (2015)
24. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Trans. Med. Imaging* **35**(5), 1240–1251 (2016)
25. Riera-Ledesma, J., Salazar-Gonzalez, J.J.: A heuristic approach for the traveling purchaser problem. *Eur. J. Oper. Res.* **162**(1), 142–152 (2005). <http://www.sciencedirect.com/science/article/pii/S037722170300821X>. *Logistics: From Theory to Application*
26. Roth, H.R., Lee, C.T., Shin, H.C., Seff, A., Kim, L., Yao, J., Lu, L., Summers, R.M.: Anatomy-specific classification of medical images using deep convolutional nets. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 101–104 (2015)
27. Scrucca, L.: GA: a package for genetic algorithms in R. *J. Stat. Softw.* **53**(4), 1–37 (2013)
28. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R.J., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans. Med. Imaging* **35**(5), 1196–1206 (2016)

29. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
30. Stoean, C., Stoean, R., Sandita, A., Mesina, C., Ciobanu, D., Gruia, C.L.: Investigation on parameter effect for semi-automatic contour detection in histopathological image processing. In: 2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 445–451 (2015)
31. Stoean, C.: In search of the optimal set of indicators when classifying histopathological images. In: 2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 449–455 (2016)
32. Stoean, C., Preuss, M., Stoean, R.: EA-based parameter tuning of multimodal optimization performance by means of different surrogate models. In: Genetic and Evolutionary Computation Conference, GECCO 2013, pp. 1063–1070. ACM (2013)
33. Stoean, C., Stoean, R., Sandita, A., Ciobanu, D., Mesina, C., Gruia, C.L.: SVM-based cancer grading from histopathological images using morphological and topological features of glands and nuclei. In: Pietro, G., Gallo, L., Howlett, R.J., Jain, L.C. (eds.) *Intelligent Interactive Multimedia Systems and Services 2016*. SIST, vol. 55, pp. 145–155. Springer, Cham (2016). doi:[10.1007/978-3-319-39345-2\\_13](https://doi.org/10.1007/978-3-319-39345-2_13)
34. Stoean, C., Stoean, R., Sandita, A., Mesina, C., Gruia, C.L., Ciobanu, D.: Evolutionary search for an accurate contour segmentation in histopathological images. In: *The ACM Genetic and Evolutionary Computation Conference Companion (GECCO 2015)*, pp. 1491–1492 (2015)
35. Therneau, T., Atkinson, B., Ripley, B.: *rpart: Recursive Partitioning and Regression Trees* (2015)
36. Young, S.R., Rose, D.C., Karnowski, T.P., Lim, S.H., Patton, R.M.: Optimizing deep learning hyper-parameters through an evolutionary algorithm. In: *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, pp. 4:1–4:5. ACM (2015)
37. Zaharie, D., Lungeanu, D., Zamfirache, F.: Interactive search of rules in medical data using multiobjective evolutionary algorithms. In: *Proceedings of the 10th Annual Conference Companion on Genetic and Evolutionary Computation, GECCO 2008*, pp. 2065–2072. ACM, New York (2008). <http://doi.acm.org/10.1145/1388969.1389023>