Lung images classification with textural characteristics and hybrid classification schemes

Oleksandr Davydko

[1] *Artificial Intelligence and Cognitive Load Lab, The Applied Intelligence Research Centre, School of Computer Science, Technological University Dublin (TU Dublin), Dublin, D07 EWV4, Ireland*

Abstract

The current study aims to develop lung lesions diagnostics methods on CT images that can be used in novel medical decision support systems. The study considers methods of textural analysis, hybrid classification structures and group method of data handling to improve existing classification solutions and archive higher classification quality. Also, the explainability of the resulting classifier is considered.

1. Context and motivation

The current research was started during the COVID-19 outbreak when the urgent need for efficient medical image diagnostic systems emerged, so the ultimate goal of the study is to propose an efficient pipeline for the classification of lung images in X-Ray and computer tomography modalities. Medical image classification is a well-known task, and there are a lot of known methods for solving it. However, any technique that will increase classification quality is highly appreciated in the medical context because even one misclassified image can have severe consequences. That fact works as a primary motivating factor for the current research. The classifier quality mainly depends on the input features and classification algorithm. Recent work considers methods to improve both of these aspects. The first part of the research lies in the context of texture analysis (TA). Methods of TA are proven to be effective for extracting features for image classification [1, 2, 3]. Many works discuss the usage of global texture descriptors with aggregative second-order statistics like Haralick features [4]. At the same time, other methods for aggregating raw texture information are discovered, much less, making that question actual for consideration. The second part of the research focuses on classification algorithms improvement. Decision tree-based algorithms are one of the oldest but essential in real-world tasks. Still, those algorithms need improvements to compete with modern neural network solutions. The improvements to the classifier could be made in several aspects. Classical Random-Forest features the constant number of trees in the forest defined as an algorithm hyperparameter, but building decision-tree forests with optimal complexity structure can lead to better classifier performance. Principles of the group method of data handling (GMDH) [5] provides tools to archive stated tasks. Paper [6] shows that proposing more sophisticated

The 1st World Conference on eXplainable Artificial Intelligence: Late-breaking work, Demos and Doctoral Consortium, 2023

d22125337@mytudublin.ie (O. Davydko)
0000-0002-6048-4639 (O. Davydko)

methods of tree voting instead of simple majority voting can improve classification quality. This leads to the idea that more complex voting approaches can increase the classifier performance. Finally, integrating different classification structures into a single pipeline allows the usage of each structure's advantages, as shown in [7]. At the same time, when input contains many features, the explainability of such a classifier will drastically degrade as existing model-agnostic methods do not provide sufficient computational speed. So, another question to address is the development of a model-specific explanation method which will provide the researcher with a suitable level of computational speed.

2. Related work

The classification of medical images is a well-known task in machine learning. There are several approaches to building a feature space for the classifier:

- Using pixel intensities as an input
- Using image texture characteristics (local and global)

At the same time, both approaches generate enormous size feature space, which must be reduced. Neural networks are mainly used to address this issue when using pixel intensities. Convolutional and linear layers are intended to extract and transform features from the image matrix input during the forward pass. One of the most successful architectures implementing the described idea is RMT-Net [8] which is built as an extension of vision transformer neural network architecture [9]. Texture characteristics are a big group of different methods, including global and local characteristics. Global characteristics are descriptors of the whole image and form the static feature space. Examples of global characteristics are grey-level co-occurrence matrices (GLCM), grey-level run-length matrices (GLRLM), grey-level size zone matrices (GLSZM), neighbouring grey-tone difference matrices (NGTDM), grey-level dependency matrices (GLDM), grey-level entropy matrices (GLEM) but not limited to that list. Local characteristics are descriptors of the image's local areas, such as pixel neighbourhood zone or image patches. Examples are but not limited to local binary patterns (LBP) [10], scale-invariant texture descriptors (SIFT) [11], speeded-up robust features (SURF) [12]. Works [13, 4, 14] also introduce second-order statistics, which are aggregating GLCM, GLRLM, GLSZM and other matrices into feature vectors and works [1, 2, 3] show the ability of such statistics to act as features for image classification. However, using the texture characteristics matrices without using second-order statistics is not developed very much. Some works discuss such topics but are constrained to using only GLCM and GLRLM matrices, omitting other descriptors. The current research discusses using as many global and local texture descriptors as possible to improve classification quality. The classification quality can also be improved by optimising the classifier structure and improving training algorithms. Works [7, 15, 16] show examples of hybridising neural networks with other classifiers such as support vector machine (SVM) or Random Forest and prove the effectiveness of such hybrid schemes by getting better classification metrics. In this case, the neural network is the feature extractor for the second classification part. The results of these experiments lead us to the idea that improving the results of the second classifier may also improve the overall result of the hybrid structure. In the case of the Random-Forest-like algorithms, improving the classification quality can be archived by optimising tree structures. The principles of GMDH provide us with tools for building each tree and whole forest optimally for particular classification tasks. Also, using the advanced tree voting functions can positively affect classification quality. For instance, work [6] shows that Bayesian Tree Aggregation can deliver better results than standard majority voting schemes. However, this question requires further investigation. To archive better classification results, current work discusses a novel pipeline, which includes the usage of several texture characteristics matrices types in a raw form and a hybrid classifier consisting of a neural network and novel forest-based algorithm with enhanced tree-structure building and logistic voting function with feature blending. Explainability is a crucial part of a designed solution. The SHapley Additive exPlanation (SHAP) [17] and Local Interpretable Model-Agnostic Explanations (LIME) [18] can be used to provide explanations while using the proposed classifier. However, computational speed concerns exist as the number of input features is too large when using textural matrices. A sufficient performance level is archived when using Integrated Gradients, GradCAM, and DeepSHAP methods, but they are limited to neural network usage. Given that, current work also considers the development of the appropriate explainability method to support explanations for the hybrid classification schemes.

3. Research question and hypothesis

The current research considers the following study to be answered: To what extent a hybrid classification structure consisting of a neural network and enhanced decision-tree forest with raw texture characteristics matrices at the input can improve the quality of lung image classification Taking into account research questions, the hypothesis can be formulated in the following way:

Research hypothesis IF texture characteristics matrices(GLCM, GLRLM, GLSZM, GLDM, NGTDM, GLEM) will be used as classifier input along with the original image, and the classifier will be a hybrid structure consisting of neural network for feature extraction and self-organised forest algorithm with logistic voting function for classification, THEN classification F1-score will be statistically significantly higher than for neural network with the original image at the input in the tasks of the lung images in X-Ray and computer-tomography (CT) modalities classification.

4. Research methods

The current section contains details of enhanced classifier implementation. The first step is the feature construction. In the present work, it is proposed to use texture characteristics in their raw matrix forms instead of using aggregate second order-statistics. The resulting feature space will contain approximately n x 256 x 256 features, where n - is the number of the texture matrices used. This can be an issue as classifiers cannot process such feature spaces. To address this issue, it is proposed to use a set of parallel neural networks to compress texture characteristics matrices into compressed feature vectors (CFVs) of significantly lower dimensionality. Next, two other fully-connected networks aggregate parallel network outputs and perform classification. All three components united into a single network on a training stage (figure 1). It allows for

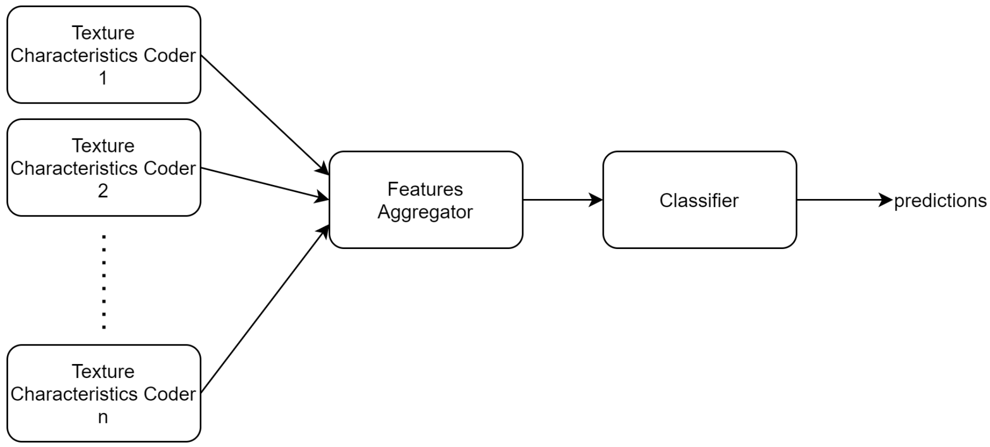

Figure 1: Three component classification scheme. Shows the structure of a three-level classifier with parallel networks to process different texture characteristics and image inputs

forming the classification-optimal features at the output of the aggregation network. Also, such an approach enables the blending of different texture characteristics, initial image intensities etc., as for each input signal, the unique architecture of the encoder network can be used.

The next step is to replace the classification neural network with another classification structure. It is proposed to use an enhanced version of the random forest-like algorithm called a logistic self-organised forest [19]. This algorithm uses a logistic function for tree voting. Also, the outputs of the trees are blended with input features, leading to classification quality improvement. GMDH tools are used for optimal logistic function structure synthesis.

To provide explanations for the developed classifier, a need for a new explanation algorithm emerged. The existing model-agnostic methods' performance is insufficient, while fast model-specific methods are not implemented for hybrid structures. It is proposed to build a new algorithm based on Integrated gradients [20] and SHAP. IG method is used to explain textural matrix elements' contribution to CFV. The SHAP method is used to evaluate CFVs feature contributions for the second part of the hybrid classifier. As CFV does not contain many features, the performance of the SHAP will be sufficient. Finally, IG contributions are multiplied by the SHAP result to scale the importance of the input features by their actual contribution to the final answer. It is proposed to validate the correctness of the new method by generating explanations with it and SHAP values and then comparing generated explanations with the Mann-Whitney U test. The test should indicate no statistically significant differences between the pair of explanations. The dataset with a few features should be used for validation to address the problem with SHAP computation speed.

To test the hypothesis, it is proposed to cross-validate the baseline and the proposed classifiers on 20 sub-sets of 3 datasets. Two groups consisting of 60 samples will be drawn. Each sample in the group will represent the test set F1-score of the classifier, trained with a different dataset split. These two groups will be compared by running the Mann-Whitney

U test. The hypothesis will be accepted if the test results indicate a statistically significant difference between groups, and the group corresponding to the proposed classifier will have a higher mean value.

According to the research questions and hypothesis, the next research objectives were generated:

1. To collect data
   a) To obtain COVID-19 lung X-Ray dataset
   b) To obtain COVID-19 lung computer tomography dataset
   c) To obtain lung cancer X-Ray dataset
2. To prepare and pre-process the data
   a) To apply the window-level operation to threshold Hounsfield units values between -1024 and 300
   b) To align anatomic directions for all CT series
   c) To align all CT series by size and rotation
   d) To resize all X-Rays to the size of 256 by 256 pixels
3. Implement image feature extraction with textural matrices
   a) GLCM
   b) GLRLM
   c) GLSZM
   d) GLDM
   e) NGTDM
   f) GLEM
4. To implement classifier algorithm
   a) To implement neural network architecture for input features compression
   b) To implement classification forest algorithm based on self-organisation principles
   c) To implement logistic function tree voting algorithm
   d) To implement hybrid neural network-classification forest structure
   e) To implement feature selection algorithm
5. To implement an explainability framework for the classifier
   a) To implement hybrid models explanation method based on the Integrated Gradients and SHAP
   b) To apply the existing model-agnostic SHAP explanation method to the hybrid classifier
   c) Compare implemented explanation method with SHAP results.
6. To train and evaluate the classifier on the collected datasets
7. To discuss the results by comparing them with SOTA methods results [8]

Table 1
Comparison of the proposed LSOF-based classifier performance with other classifiers on the testing set

Model	Total accuracy
Neural network generated features + LSOF	0.96
Neural network generated features + Random Forest	0.92
Neural network generated features + neural network classifier	0.92
Second-order stats + Random Forest	0.91
Neural network + Logistic Regression	0.90

5. Current results

At the current stage, the first four objectives are archived. Paper [19] discusses all of the leading research aspects. Firstly, it is shown that using raw texture characteristic matrices can lead to better results than using known second-order statistics. Secondly, it is discovered that using a hybrid classification structure of neural networks and forest algorithms can lead to better classification than using a neural network alone. Finally, it has been shown that improvement in forest training and prediction algorithm according to principles of self-organisation also improves classification properties. Making such improvements increased classification accuracy from 92% to 96% compared to the Random-Forest classifier with aggregative second-order stats in the COVID-19 lung lesions classification task. Detailed results of the proposed classification scheme comparison with the others can be found in table 1

6. Next steps and final contribution

Logistic self-organised forest from [19] and the proposed hybrid classifier must finally be evaluated on several datasets to take or reject the hypothesis. Such evaluation is a part of the next steps. Another question to address is the explainability of the hybrid structure classification results. There are explainable artificial intelligence methods for separate neural networks and forests, but the explainability of the hybrid structures results is an actual research topic. The final contribution is expected to be the novel lung image classification pipeline presentation, which proved more effective than existing solutions while keeping the classifier's results explainable.

References

[1] A. S. Gaudêncio, P. G. Vaz, M. Hilal, G. Mahé, M. Lederlin, A. Humeau-Heurtier, J. M. Cardoso, Evaluation of covid-19 chest computed tomography: A texture analysis based on three-dimensional entropy, Biomedical Signal Processing and Control 68 (2021) 102582. doi:10.1016/j.bspc.2021.102582.

[2] P. K. Mall, P. K. Singh, D. Yadav, Glcm based feature extraction and medical x-ray image classification using machine learning techniques, in: 2019 IEEE Conference on Information and Communication Technology, 2019, pp. 1–6. doi:10.1109/CICT48419.2019.9066263.

[3] R. Usha, K. Perumal, Svm classification of brain images from mri scans using morphological transformation and glcm texture features, International Journal of Computational Systems Engineering 5 (2019) 18–23. doi:10.1504/IJCSYSE.2019.098415.

[4] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (1973) 610–621. doi:10.1109/TSMC.1973.4309314.

[5] H. R. Madala, A. G. Ivakhnenko, Inductive Learning Algorithms for Complex Systems Modeling, CRC Press, 2019. URL: https://doi.org/10.1201/9781351073493. doi:10.1201/9781351073493.

[6] J. Brabec, L. Machlica, Decision-forest voting scheme for classification of rare classes in network intrusion detection, in: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, pp. 3325–3330. doi:10.1109/SMC.2018.00563.

[7] S. Guo, S. Chen, Y. Li, Face recognition based on convolutional neural network and support vector machine, in: 2016 IEEE International Conference on Information and Automation (ICIA), 2016, pp. 1787–1792. doi:10.1109/ICInfA.2016.7832107.

[8] K. Ren, G. Hong, X. Chen, Z. Wang, A COVID-19 medical image classification algorithm based on transformer, Scientific Reports 13 (2023). URL: https://doi.org/10.1038/s41598-023-32462-2. doi:10.1038/s41598-023-32462-2.

[9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations, 2021. URL: https://openreview.net/forum?id=YicbFdNTTy.

[10] L. Liu, P. Fieguth, Y. Guo, X. Wang, M. Pietikäinen, Local binary features for texture classification: Taxonomy and experimental study, Pattern Recognition 62 (2017) 135–160.

[11] S. Hegenbart, A. Uhl, A. Vécsei, G. Wimmer, Scale invariant texture descriptors for classifying celiac disease, Medical Image Analysis 17 (2013) 458–474.

[12] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, Lecture notes in computer science 3951 (2006) 404–417.

[13] M. M. Galloway, Texture analysis using gray level run lengths, Computer Graphics and Image Processing 4 (1975) 172–179. doi:10.1016/S0146-664X(75)80008-6.

[14] G. Thibault, B. FERTIL, C. Navarro, S. Pereira, N. Lévy, J. Sequeira, J.-L. Mari, Texture indexes and gray level size zone matrix application to cell nuclei classification, 2009.

[15] G. Xu, M. Liu, Z. Jiang, D. Söffker, W. Shen, Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning, Sensors 19 (2019). URL: https://www.mdpi.com/1424-8220/19/5/1088. doi:10.3390/s19051088.

[16] M. S. Hammad, V. F. Ghoneim, M. S. Mabrouk, W. I. Al-atabany, A hybrid deep learning approach for COVID-19 detection based on genomic image processing techniques, Scientific Reports 13 (2023). URL: https://doi.org/10.1038/s41598-023-30941-0. doi:10.1038/s41598-023-30941-0.
[17] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017.
[18] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
[19] O. Davydko, Y. Hladkyi, M. Linnik, O. Nosovets, V. Pavlov, I. Nastenko, Hybrid classifiers based on cnn, lsof, gmdh in covid-19 pneumonic lesions types classification task, in: 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), volume 1, 2021, pp. 380–384. doi:10.1109/CSIT52700.2021.9648752.
[20] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, 2017. arXiv:1703.01365.