Computing, Telecommunication and Control, 2024, Vol. 17, No. 2, Pp. 62–70.
Информатика, телекоммуникации и управление. 2024. Том 17, № 2. С. 62–70.

Intelligent Systems and Technologies
Интеллектуальные системы и технологии

Research article
DOI: https://doi.org/10.18721/JCSTCS.17206
UDC 004.932.72

COMPARISON AND SELECTION OF RADIOMIC AND DEEP CONVOLUTIONAL FEATURES FOR IMPROVING THE ACCURACY OF CT-IMAGE TEXTURE CLASSIFICATION

F. Shariaty ✉, *V.A. Pavlov*, *M.U. Surakov*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ shariaty3@gmail.com

Abstract. The article explores the problem of comparing and selecting radiomic and deep convolutional features extracted from CT images to enhance the accuracy of texture classification in CT diagnostics. By using the mRMR method, the study assesses the significance of these features in predicting genetic mutations in patients with lung cancer, highlighting their importance for refining diagnostic procedures. The developed predictive model demonstrates high classification accuracy of 92%, which indicates its high efficiency. Analysis of the results reveals that deep learning features effectively capture complex, high-level abstract textures that indicate the presence of pathologies. At the same time, radiomic features provide key information about the phenotypic characteristics of tumors, such as shape, texture, and intensity. This comprehensive approach not only improves the accuracy of non-invasive diagnostics, but also contributes significantly to personalized medicine by facilitating the development of more precise treatment strategies based on genetic profiles.

Keywords: radiomics, deep convolutional features, computed tomography, machine learning, feature selection

Acknowledgements: The research was supported by the Russian Science Foundation grant No. 24-25-00204 “Integrating genomic analysis and medical imaging to accurately predict and predict cancer characteristics and treatment outcomes in early-stage non-small cell lung cancer”. Available online: https://rscf.ru/project/24-25-00204/

Citation: Shariaty F., Pavlov V.A., Surakov M.U. Comparison and selection of radiomic and deep convolutional features for improving the accuracy of CT-image texture classification. Computing, Telecommunications and Control, 2024, Vol. 17, No. 2, Pp. 62–70. DOI: 10.18721/JCSTCS.17206

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17206
УДК 004.932.72

СРАВНЕНИЕ И ВЫБОР РАДИОМИЧЕСКИХ И ГЛУБОКИХ СВЕРТОЧНЫХ ПРИЗНАКОВ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ КЛАССИФИКАЦИИ ТЕКСТУР КТ-ИЗОБРАЖЕНИЙ

Ф. Шариати ✉ , *В.А. Павлов* , *М.У. Сураков*

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ shariaty3@gmail.com

Аннотация. В статье подробно рассматривается задача сравнения и выбора радиомических и глубоких сверточных признаков, извлекаемых из КТ-изображений, для повышения точности классификации текстур в рамках КТ-диагностики. Использование метода mRMR позволило оценить значимость этих признаков в контексте прогнозирования наличия генетических мутаций у пациентов с раком легкого, подчеркивая их важность для уточнения диагностических процедур. Разработанная модель показала высокую точность классификации – 92%, что свидетельствует о ее высокой эффективности. Анализ результатов выявил, что признаки, основанные на глубоком обучении, эффективно фиксируют сложные, высокоуровневые абстрактные текстуры, что указывает на наличие патологий. В то же время радиомические признаки обеспечивают ключевую информацию о детальных фенотипических характеристиках опухолей, включая форму, текстуру и интенсивность. Такой комплексный подход не только повышает точность неинвазивной диагностики, но и вносит значимый вклад в персонализированную медицину, способствуя разработке более точных стратегий лечения на основе генетических профилей.

Ключевые слова: радиомика, глубокие сверточные признаки, компьютерная томография, машинное обучение, отбор признаков

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 24-25-00204 «Интеграция геномного анализа и медицинской визуализации для точного прогнозирования и предсказания характеристик рака и результатов лечения при ранней стадии неклеточного рака легкого», https://rscf.ru/project/24-25-00204/

Для цитирования: Shariaty F., Pavlov V.A., Surakov M.U. Comparison and selection of radiomic and deep convolutional features for improving the accuracy of CT-image texture classification // Computing, Telecommunications and Control. 2024. Т. 17, № 2. С. 62–70. DOI: 10.18721/JCSTCS.17206

Introduction

Image processing and analysis now occupy an important place in many areas of science and technology. The ability to accurately analyze and interpret images is particularly critical in medical imaging, where it can directly impact diagnostic capabilities. The development of new feature extraction methods not only contributes to the accuracy of analysis, but also opens new opportunities for early diagnosis and personalized treatment approach [1].

Recent advances in radiomics and deep learning have opened new horizons in extracting and analyzing complex features from medical images. Radiomics involves extracting a large number of features from medical images [2]. On the other hand, deep learning, especially convolutional neural networks (CNN), has shown success in image classification tasks [3, 4].

Radiomic features, although informative, may not cover the full range of available information in the images under study. Deep learning models often act as “black boxes”, providing little information

about features that have a greater impact on decision making in image classification. In addition, the performance of these models can depend significantly on the quality and quantity of training images.

This paper proposes the results of a comparative analysis of radiomic and deep convolutional features on CT images to classify textures indicating the presence or absence of genetic mutations in patients with lung cancer. The aim of this study is to improve the accuracy of CT image texture classification by jointly using the most effective radiomic and deep convolutional features.

Materials and methods

Data collection and pre-processing

This study used dataset [5] which contains diagnostic and prognostic information such as: CT images, semantic annotations, gene mutation status information.

Deep learning features

Deep convolutional feature extraction was performed using the pre-trained ResNet18 [2], which was chosen because of its relatively simple structure compared to deeper models, facilitating faster training and image processing while still being able to extract complex features. The model was adapted for the task by removing the last classification layer, allowing it to function as a feature generator that generates a deep convolutional feature vector from the last convolutional layer immediately preceding the classification layer. The vector consists of 512 floating point numbers («Deep feature 1», ..., «Deep feature 512») and encompasses high-level representations of CT images learned by the model from a large number of images. These representations are assumed to include textures related to tumor morphology and possibly indirect markers of genetic mutations. The extracted features were used as part of an integrated feature set, contributing model-derived knowledge to the classification process.

Radiomic features extraction

Radiomic features play a key role in describing pulmonary nodule texture features on CT images. The first extraction step begins with segmentation of pulmonary nodules, using either manual annotation by expert radiologists or automated segmentation algorithms. This segmentation is then used for quantitative analysis, where radiomic features are systematically extracted and classified into three main groups [6–8]:

- **First-order statistics**:

– *Mean intensity*. Mean intensity represents the average of the pixels within a nodule image and is a basic measure of its overall brightness. Mathematically it is expressed as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

where x_i is intensity value of the i-th pixel, a N is the total number of pixels in the nodule. Mean intensity indicates average nodule density.

– *Standard deviation*. Standard deviation quantifies the variation of intensity values around the mean value, reflecting heterogeneity within the nodule. A larger standard deviation implies a wider range of intensity values, indicating variability in nodule composition. The standard deviation is calculated by the formula:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \mu\right)^2}.$$

This index gives an indication of the texture of the nodule.

– *Skewness*. Skewness measures the asymmetry of the intensity distribution around the mean value. Skewness is defined as:

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)}\right)^3}.$$

It can help determine whether a nodule contains predominantly high or low intensity pixels.

– *Kurtosis*. Kurtosis measures the sharpness of a peak in the intensity distribution. It is mathematically expressed as:

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)}\right)^4} - 3.$$

• **Shape-based features**:

– *Volume*. Volume of a pulmonary nodule is a direct indicator of its size. At the same time, large volumes often require closer scrutiny for potential malignancy. It is calculated by counting the total number of pixels (or voxels in 3D) that make up a segmented nodule and multiplying by the spatial pitch of the pixel (or voxel) to convert to physical units (e.g., cubic millimeters). The volume V is defined as:

$$V = N \times \left(d\right)^3,$$

where N is the number of pixels in a nodule, d is the pixel distance.

– *Surface area*. Surface area gives an indication of the complexity of the nodule's appearance. A more irregular surface area may indicate a higher likelihood of malignancy. The surface area A can be approximately calculated using the «marching cubes» algorithm or similar techniques that triangulate the surface of a segmented nodule:

$$A = \sum_{i=1}^{M} a_i,$$

where M is the total number of triangles approximating the nodule surface, a_i is the area of the i-th triangle. This approximation provides a quantitative measure of the external complexity of a nodule.

– *Sphericity*. Sphericity assesses how much the shape of a nodule resembles a sphere, which is often used to distinguish between regular and irregular nodules. Sphericity Ψ is defined as:

$$\psi = \frac{\pi^{\frac{1}{3}}\left(6V\right)^{\frac{2}{3}}}{A},$$

where V is the volume, A is the nodule surface area. Sphericity values close to 1 indicate a more spherical shape, while values far from 1 suggest irregular shapes.

– *Compactness*. Compactness measures the density of the nodule shape, indicating how densely packed the structure is. It is inversely proportional to sphericity and can be expressed as:

$$C = \frac{A}{V^{\frac{2}{3}}}.$$

A higher compactness value implies a more irregular or complex nodule shape, which may indicate malignancy.

• **Texture features**:

– *Entropy*. Entropy measures randomness or disorder in the intensity distribution of nodule pixels, serving as an indicator of texture irregularity. High entropy values suggest a complex texture with a high degree of heterogeneity, which is often observed in malignant tumors. Entropy is calculated by the formula:

$$H = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log_2 p(i,j),$$

where $p(i,j)$ is normalized co-occurrence matrix, representing the probability of neighboring intensity of pixel i next to intensity of pixel j, L is the number of possible intensity levels.

– *Contrast*. The contrast quantifies local variations in pixel intensity, emphasizing the presence of distinct edges or patterns in a nodule. It reflects the depth of the texture and the sharpness of image details. High contrast values are associated with textures containing significant differences between pixel intensities. Contrast is calculated by the formula:

$$C = \sum_{n=0}^{N-1}\left[n^2 \sum_{i,j|i-j|=n} p(i,j) \right],$$

where N is the number of different intensity levels, n is the difference in intensity levels being considered. This calculation emphasizes the contribution of pixel pairs with significantly different intensities.

– *Homogeneity*. Homogeneity measures the degree of consistency or evenness of the texture. High values of uniformity indicate a smooth, regular texture, while lower values indicate a variety of patterns and irregularities. Homogeneity is defined as:

$$\Gamma = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{p(i,j)}{1+|i-j|}.$$

This equation weights the elements of the co-occurrence matrix by the inverse of their distance from the diagonal, favoring homogeneous textures where pixel intensities are similar.

– *Correlation*. Correlation estimates the degree of linear dependence between pixel intensities in a given direction within the region of interest. It helps to identify oriented textures and structured patterns. A high correlation indicates a strong relationship between pixel intensity levels across the texture. The correlation is calculated by the formula:

$$\rho = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j},$$

where μ_i and μ_j are mean values, σ_i and σ_j are standard deviations of the sums of rows and columns of the co-occurrence matrix, respectively. This indicator provides information about the predictability and structure of texture patterns.

Fig. 1 shows a flowchart describing the process of generating, combining and selecting radiomic and deep convolutional features. The radiomics and deep convolution features are combined into a vector, which is fed to the algorithm to select the most important features. Based on the selected features, a new vector is formed which is used to determine the classification of image textures.

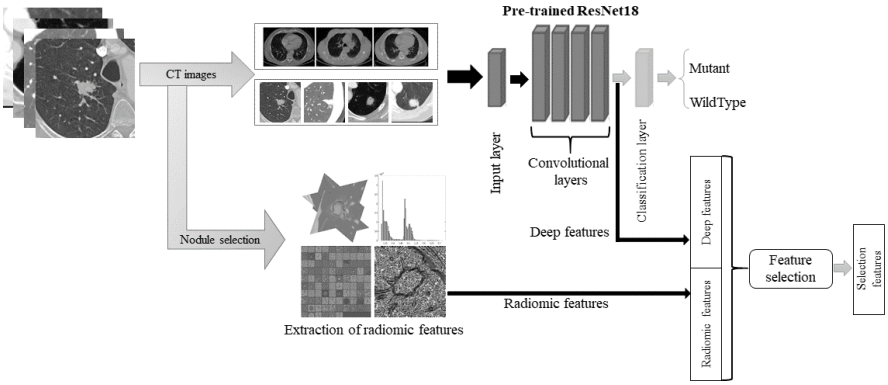

Fig. 1. Flowchart of the proposed method

Features integration

In this paper, a combined approach is applied to analyze radiomic and deep convolutional features using support vector method (SVM) and mRMR (minimum Redundancy Maximum Relevance) feature selection algorithm [9, 10]. These methods play a key role in improving the performance of predictive models by identifying the most significant features that contribute to the determination of the presence of mutations [11].

The mRMR algorithm is used for initial feature selection, minimizing redundancy and maximizing relevance with respect to the target variable. Features that have both maximum correlation with the target variable and minimum correlation with each other are selected for further analysis, thereby improving the informativeness of the data:

$$\max D(S,C) = \frac{1}{|S|}\sum_{x_i \in S} I(x_i;C),$$

where S is selected features set, C is target variable, $I(x_i;C)$ is mutual information between feature x_i and target variable.

Minimum redundancy is defined as minimizing the average mutual information between feature pairs in the selected set:

$$\min R(S) = \frac{1}{|S|^2}\sum_{x_i, x_j \in S} I(x_i;x_j),$$

where $I(x_i;x_j)$ is mutual information between features x_i and x_j.

Combining these two criteria, mRMR algorithm seeks to find the feature set S that optimizes the function:

$$\max \Phi(S) = D(S,C) - R(S),$$

which ensures the selection of features most useful for modeling while minimizing their redundancy. This approach improves the quality of classification, making the model more interpretable and effective in recognizing complex patterns.

The selected features are then analyzed using SVM to assess their ability to classify CT image textures for mutations. This step evaluates how radiomic and deep convolutional features affect model accuracy by providing quantitative performance measures such as *Accuracy, Precision and Recall*.

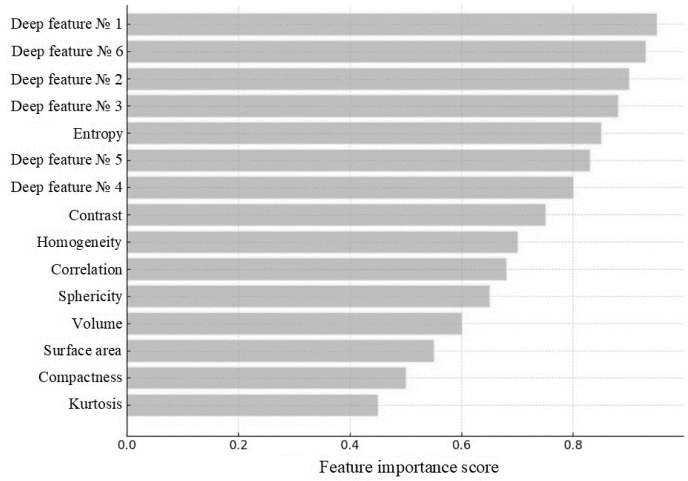

Fig. 2. Assessing the importance of features

This approach not only allows us to compare the performance of radiomic and deep convolutional features, but also to improve our understanding of their interactions and contributions to determining the presence of mutations. This methodology emphasizes the unique contributions of each type of feature and their potential to improve the diagnostic accuracy of models.

Results and discussion

Fig. 2 shows the interaction between traditional radiomic features, such as "Entropy" (0.85), "Contrast" (0.75) and "Sphericity" (0.65) and features obtained by deep learning, including "Deep Feature 1" (0.95) and "Deep Feature 6" (0.93) (the number denotes the position of the selected features in the original vector, with dimensionality 1×512), etc.

The feature importance analysis determined by the mRMR method provides insight into the predictive power of the proposed integrated model for classifying CT image texture for mutations. It can be concluded that deep learning features rank high in importance, emphasizing the depth of information they include about the underlying pathology. However, the significant ranking of radiomic features suggests that they also provide essential, unreliable information useful for the classification task.

Table

Comparison of the effectiveness of approaches for determining the presence of mutations in the EGFR gene

Method	Accuracy, %
Deep convolutional features + SVM	89
Radiomic features + SVM	87
Deep convolutional features + Radiomic features + SVM	92
"Mut-SeResNet" [14]	88

In this study, we compared the efficacy of different approaches to determine the presence of mutations in the EGFR gene using SVM [1, 12, 13]. The results are presented in Table. As can be seen from the table, the model developed and tested in our study showed the highest accuracy among the approaches considered, achieving an Accuracy of 92%. Compared to another study [14] using traditional approaches and specialized models, our approach shows improved results, which may contribute to a more accurate and efficient diagnosis.

Conclusion

The paper considers the problem of comparing and selecting radiomic and deep convolutional features extracted from CT images in order to improve the accuracy of CT image texture classification. Using the mRMR algorithm, the effectiveness of radiomics features in classification tasks was demonstrated. Then, a classifier was developed that showed high accuracy in classifying the presence of genetic mutations among the considered approaches, achieving an Accuracy of 92%. Analysis of the results showed that deep learning features reveal high-level abstract textures indicative of underlying pathologies, while radiomic features provide essential information about tumor phenotypic features such as shape, texture, and intensity.

REFERENCES

1. **Skalunova M., Shariaty F., Rozov S., Radmard A.R.** Personalized Chemotherapy Selection for Lung Cancer Patients Using Machine Learning and Computed Tomography. International Conference on Electrical Engineering and Photonics (EExPolytech), 2023, pp. 128–131. DOI: 10.1109/EExPolytech58658.2023.10318700

2. **Taranova D., Shariaty F.** Radiomic analysis for prediction of T stage parameter (T1–T2) in lung cancer patients. Proceedings of the All-Russian Conference "Science Week IET", 2022, pp. 77–80.

3. **Pavlov V.A., Zavjalov S.V., Pervunina T.M., Orooji M., Shariaty F.** Application of deep learning for COVID-19 classification and pulmonary edema on pulmonary CT images. Voprosy radioelektroniki. seriya: tekhnika televideniya [Issues of radio electronics. Series: Television technology], 2022, no. 1, pp. 65–72.

4. **Zavjalov S.V., Pavlov V.A., Fyodorov S.A., Shariaty F., Pervunina T.M.** Automatic segmentation and classification of COVID-19 on CT images. Voprosy radioelektroniki. seriya: tekhnika televideniya [Issues of radio electronics. Series: Television technology], 2023, no. 4, pp. 76–83.

5. **Bakr S., Gevaert O., Echegaray S. et al.** A radiogenomic dataset of non-small cell lung cancer. Scientific data, 2018, Vol. 5, no. 1, Article no. 180202. DOI: 10.1038/sdata.2018.202

6. **Amiri S., Akbarabadi M., Abdolali F., Nikoofar A., Esfahani A.J., Cheraghi S.** Radiomics analysis on CT images for prediction of radiation-induced kidney damage by machine learning models. Computers in Biology and Medicine, 2021, Vol. 133, Article no. 104409. DOI: 10.1016/j.compbiomed.2021.104409

7. **Reiazi R., Abbas E., Famiyeh P., Rezaie A., Kwan J.Y.Y., Patel T., Bratman S.V., Tadic T., Liu F.F., Haibe-Kains B.** The impact of the variation of imaging parameters on the robustness of Computed Tomography radiomic features: A review. Computers in Biology and Medicine, 2021, Vol. 133, Article no. 104400. DOI: 10.1016/j.compbiomed.2021.104400

8. **Shiri I., Sorouri M., Geramifar P., Nazari M., Abdollahi M., Salimi Y., Khosravi B., Askari D., Aghaghazvini L., Hajianfar G., Kasaeian A., Abdollahi H., Arabi H., Rahmim A., Radmard A.R., Zaidi H.** Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. Computers in Biology and Medicine, 2021, Vol. 132, Article no. 104304. DOI: 10.1016/j.compbiomed.2021.104304

9. **Wang Z. et al.** Segmentalized mRMR features and cost-sensitive ELM with fixed inputs for fault diagnosis of high-speed railway turnouts. IEEE Transactions on Intelligent Transportation Systems, 2023, Vol. 24, no. 5, pp. 4975–4987. DOI: 10.1109/TITS.2023.3239636

10. **Lanjewar M.G., Parab J.S., Shaikh A.Y. et al.** CNN with machine learning approaches using ExtraTreesClassifier and MRMR feature selection techniques to detect liver diseases on cloud. Cluster Computing, 2023, Vol. 26, no. 6, pp. 3657–3672. DOI: 10.1007/s10586-022-03752-7

11. **Ding C., Peng H.** Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology, 2005, Vol. 3, no. 2, pp. 185–205. DOI: 10.1142/s0219720005001004

12. **Pavlov V.A., Shariaty F., Orooji M., Velichko E.N.** Application of Deep Learning Techniques for Detection of COVID-19 Using Lung CT Scans: Model Development and Validation. Springer Proceedings in Physics, 2022, Vol. 268, pp. 85–96. DOI: 10.1007/978-3-030-81119-8_9

13. **Sun L., Dong Y., Xu S., Feng X., Fan X.** Predicting Multi-Gene Mutation Based on Lung Cancer CT Images and Mut-SeResNet. Applied Sciences, 2023, Vol. 13, no. 3, Article no. 1921. DOI: 10.3390/app13031921

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Shariaty Faridoddin
Шариати Фаридоддин
E-mail: shariaty3@gmail.com
ORCID: https://orcid.org/0000-0002-7060-8826

Pavlov Vitalii A.
Павлов Виталий Александрович
E-mail: pavlov_va@spbstu.ru
ORCID: https://orcid.org/0000-0003-0726-6613

Surakov Murat U.
Сураков Мурат Уралович
E-mail: murat.surakov@gmail.com

Submitted: 15.04.2024; Approved: 06.07.2024; Accepted: 19.07.2024.
Поступила: 15.04.2024; Одобрена: 06.07.2024; Принята: 19.07.2024.