ISSN 2219-9365

https://doi.org/10.31891/2219-9365-2026-86-52
UDC 004.8

KYRYCHENKO Oleksandr
Khmelnytskyi National University
https://orcid.org/0009-0007-7061-631X
e-mail: kyrychenkoom@khmnu.edu.ua

OPTIMAL BINARIZATION METHOD FOR IBSI FEATURES IN INTERPRETABLE BRAIN TUMOR DIAGNOSIS

This paper presents an optimal binarization method for standardized IBSI features in creating interpretable diagnostic models for brain tumors. The growing adoption of radiomics in medical imaging has generated extensive quantitative features describing texture, morphology, and intensity properties of pathological formations. However, integration of these continuous numerical characteristics into interpretable rule-based diagnostic systems requires effective discretization approaches that preserve diagnostic information while maintaining clinical interpretability. The proposed approach is based on mutual information maximization criterion considering local diagnostic context and ensures optimal balance between feature informativeness and compatibility with logical rule architecture. An adaptive threshold determination procedure has been developed through discrete search over candidate set with composite criterion application that accounts for both information value and distribution balance of binarized values. The methodology incorporates percentile-based thresholds and statistically grounded values to form a comprehensive candidate set, enabling robust feature transformation across diverse clinical scenarios. Experimental validation on MRI dataset with four pathology classes (glioma, meningioma, pituitary tumor, and no tumor) containing 64 IBSI-standardized features showed that optimized thresholds provide average mutual information of 0.342 bits compared to 0.287 bits for fixed median thresholds, representing 19.2% improvement with statistical significance ($p < 0.001$). The proposed method ensures generation of more stable and clinically relevant diagnostic rules through preservation of medical meaningfulness of selected features during their transformation to binary format. Integration with Decision Rules Network architecture demonstrated 7.3% accuracy improvement and achieved 89.3% local consistency with base VGG-16 model. The research addresses fundamental challenge of feature discretization for rule-based interpretable systems in medical imaging by developing theoretically grounded optimization framework that maintains both mathematical rigor and clinical applicability of resulting binary features, facilitating transparent and trustworthy AI-assisted diagnostic decision-making in clinical practice.

Keywords: feature binarization, IBSI standard, mutual information, interpretable models, radiomic features, medical diagnostics.

КИРИЧЕНКО Олександр
Хмельницький національний університет

ОПТИМАЛЬНИЙ МЕТОД БІНАРИЗАЦІЇ ОЗНАК IBSI ДЛЯ ІНТЕРПРЕТОВАНОЇ ДІАГНОСТИКИ ПУХЛИН ГОЛОВНОГО МОЗКУ

У статті представлено оптимальний метод бінаризації стандартизованих ознак IBSI для побудови інтерпретованих діагностичних моделей пухлин головного мозку. Зростання використання радіоміки в медичній візуалізації зумовило формування великих наборів кількісних ознак, що описують текстурні, морфологічні та інтенсивнісні характеристики патологічних утворень. Проте інтеграція цих неперервних числових характеристик в інтерпретовані діагностичні системи, засновані на правилах, потребує ефективних підходів до дискретизації, які забезпечують збереження діагностичної інформативності та клінічної зрозумілості результатів. Запропонований підхід базується на критерії максимізації взаємної інформації з урахуванням локального діагностичного контексту та забезпечує оптимальний баланс між інформативністю ознак і сумісністю з архітектурою логічних правил. Розроблено адаптивну процедуру визначення порогових значень шляхом дискретного пошуку серед множини кандидатів із використанням комплексного критерію, який враховує як інформаційну цінність ознаки, так і збалансованість розподілу бінаризованих значень. Методологія передбачає використання порогів на основі процентилів і статистично обґрунтованих значень для формування розширеної множини кандидатів, що забезпечує надійну трансформацію ознак у різних клінічних сценаріях. Експериментальна перевірка на наборі МРТ-зображень із чотирма класами патологій (гліома, менінгіома, пухлина гіпофіза та відсутність пухлини), який містив 64 стандартизовані ознаки IBSI, показала, що оптимізовані порогові значення забезпечують середню взаємну інформацію на рівні 0,342 біта порівняно з 0,287 біта для фіксованих медіанних порогів, що відповідає покращенню на 19,2 % при статистичній значущості ($p < 0,001$). Запропонований метод забезпечує формування більш стабільних і клінічно релевантних діагностичних правил завдяки збереженню медичної змістовності відібраних ознак під час їх перетворення у бінарний формат. Інтеграція з архітектурою мережі правил прийняття рішень (Decision Rules Network) продемонструвала підвищення точності класифікації на 7,3 % та досягнення 89,3 % локальної узгодженості з базовою моделлю VGG-16. Дослідження розв'язує фундаментальну проблему дискретизації ознак для інтерпретованих систем, заснованих на правилах, у сфері медичної візуалізації шляхом розроблення теоретично обґрунтованого оптимізаційного підходу, який поєднує математичну строгість із практичною клінічною придатністю бінарних ознак, сприяючи підвищенню прозорості та довіри до систем підтримки діагностичних рішень на основі штучного інтелекту.

Ключові слова: бінаризація ознак, стандарт IBSI, взаємна інформація, інтерпретовані моделі, радіомічні ознаки, медична діагностика.

Стаття надійшла до редакції / Received 02.03.2026
Прийнята до друку / Accepted 14.04.2026
Опубліковано / Published 31.05.2026

PROBLEM STATEMENT

The development of interpretable artificial intelligence systems for medical diagnosis requires the transformation of complex quantitative characteristics of medical images into a representation compatible with decision-making logic that is understandable to clinicians. This challenge is particularly acute in the diagnosis of brain tumors, where radiomic analysis of MRI images generates a large number of continuous numerical features describing the textural, morphological, and intensity-based properties of pathological lesions.

The standardization of radiomic features through the Image Biomarker Standardization Initiative (IBSI) [1] has ensured the unification of computational procedures and terminology; however, it has left open the question of how these features should be optimally represented for use in interpretable models. IBSI-compliant features, such as Gray Level Co-occurrence Matrix characteristics, Gray Level Run Length Matrix parameters, and morphological descriptors, exhibit wide value ranges and complex nonlinear relationships with diagnostic classes.

Conventional approaches to the use of radiomic features in machine learning systems rely on their direct utilization in continuous form. Such methods are effective for complex models, including gradient boosting algorithms and deep neural networks, which are capable of automatically capturing nonlinear dependencies within the data. However, for interpretable systems-particularly those based on logical rules-continuous feature representations introduce substantial limitations.

Rule-based reasoning requires explicit conditions of the form "IF a feature satisfies a condition, THEN a conclusion follows." Continuous features complicate the formulation of such conditions, as they necessitate the specification of precise numerical thresholds that may lack intuitive meaning for clinicians when presented without appropriate context. Moreover, rule-based decision systems have limited capacity to handle a large number of possible feature values. Each distinct value of a continuous feature can potentially generate a new branch in the decision space, leading to an exponential increase in model complexity. This renders the training and interpretation of rules impractical while maintaining acceptable diagnostic accuracy.

Clinical practice naturally operates with categorical concepts. Physicians reason in terms such as "high/low," "present/absent," or "normal/pathological," rather than exact numerical values. Binary feature representations align naturally with this clinical reasoning paradigm, facilitating comprehension and adoption of diagnostic rules [2].

The binarization of continuous features is therefore a critical task that directly affects the quality of the resulting interpretable model. The selection of threshold values for each feature determines how effectively the binary representation preserves the diagnostic information contained in the original data. Suboptimal thresholds may lead to the loss of clinically relevant information or to the introduction of spurious relationships.

Existing binarization methods typically rely on simple heuristics, such as the median, mean value, or distribution percentiles. These approaches do not account for the specifics of the diagnostic task or the class distribution within the data. Median-based binarization ensures a balanced split of values but may separate observations that are clinically similar while grouping observations that are substantially different in diagnostic terms.

Alternative approaches, including supervised entropy-based discretization methods [3], take into account the relationship between features and the target variable but often produce multiple intervals rather than a binary split. This increases rule complexity and reduces interpretability. Furthermore, such methods generally fail to incorporate the local context of individual diagnostic cases.

Preserving the clinical relevance of binarized features is of particular importance. Threshold values should not only optimize statistical class separation criteria but also correspond to medical knowledge regarding the characteristics of different tumor types. Automatically derived thresholds may be mathematically optimal yet lack clinical meaning, thereby undermining trust in the system.

An additional challenge arises from the need to adapt thresholds to the local diagnostic context. Different clinical cases may require distinct binarization strategies depending on pathology characteristics, image quality, and patient demographic factors. Fixed global thresholds are unable to capture this variability.

The problem is further compounded by the interdependence between threshold selection and the structure of the interpretable model. Different binarization thresholds lead to different sets of logical rules, which may vary in diagnostic accuracy and complexity. Thresholds must therefore be optimized in a way that ensures the resulting model is simultaneously accurate, simple, and clinically meaningful.

In light of these challenges, there is a clear need to develop a method for the optimal binarization of IBSI-standardized features that maximizes the preservation of diagnostic information during transformation into a binary format, accounts for the local context of specific clinical cases, maintains the clinical interpretability of threshold values, and integrates seamlessly with the architecture of interpretable rule-based models.

ANALYSIS OF RECENT RESEARCH

The problem of discretization and binarization of continuous features for machine learning systems continues to develop actively in medical imaging, particularly in oncology. Recent studies emphasize the integration of information-theoretic criteria, such as mutual information, with deep learning approaches in order to improve the stability and interpretability of radiomic features.

Standardization of Radiomic Features and the Image Biomarker Standardization Initiative (IBSI). Study [1] presented a comprehensive overview of radiomic feature standardization through IBSI, introducing a six-step checklist

and an implementation framework for reproducible imaging biomarkers. This framework encompasses standardized feature definitions, preprocessing procedures, and reporting guidelines for clinical studies. In [2], reference values for 169 commonly used radiomic features were established through multi-center validation by independent research teams, achieving strong consensus for 95.1% of features in Phase I and 90.6% in Phase II, thereby enabling reliable calibration of radiomics software. The standardization of brain MRI images across different scanners and acquisition protocols was investigated in [4], which evaluated the impact of three intensity normalization methods (Nyul, WhiteStripe, and Z-score), combined with two gray-level discretization strategies, on first- and second-order radiomic features. This work established a unified methodological basis for future brain radiomics studies.

In [5], a radiological–biological dictionary of radiomic features was proposed, linking Lung-RADS semantic descriptors with quantitative metrics extracted in compliance with IBSI standards. This approach enhanced model explainability in oncology and demonstrated a 79% improvement in classification accuracy when combined with SHAP analysis. Within the field of neuro-oncology, [6] integrated radiomics with tumor biomarkers to construct interpretable machine learning models, showing that a limited set of clinically meaningful features can achieve performance comparable to complex ensemble models while emphasizing explainability. Furthermore, a framework for responsible collaboration between clinicians and AI systems was proposed in [7], focusing on bias reduction and trust enhancement through decision visualization in oncological diagnostics.

Discretization techniques and mutual information-based feature selection were investigated in [8], where the impact of different mutual information estimators on the performance of the mRMR method was evaluated. The study compared Parzen window estimation, equal-width binning, and bias-corrected estimators in classification tasks involving both linear and nonlinear dependencies. In [9], two novel feature discretization techniques were introduced: the first combines the Linde–Buzo–Gray quantization algorithm with a relevance criterion and supports unsupervised, supervised, and semi-supervised discretization; the second operates in a supervised mode by maximizing mutual information between discretized features and class labels, achieving superior accuracy compared to existing approaches on high-dimensional datasets.

IBSI-based standardization remains a central component of radiomics research. In [5], this standardization was extended through the creation of an explainable AI dictionary for personalized medicine, linking radiomic features with image semantics to improve interpretability in tumor diagnostics. Study [10] proposed a hybrid radiomics and machine learning approach for brain tumor classification, integrating GLCM and Curvelet features to achieve accuracy up to 95%, with particular emphasis on feature selection. Method [11] combined Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Pattern (LBP) features for quantitative analysis of tumor images (glioma, meningioma, pituitary tumor), constructing interaction features via the outer product of GLCM and LBP feature vectors, thereby significantly enhancing discriminative power. A comprehensive system for brain cancer grading [12] utilized histopathological images and employed multiple feature extraction algorithms, including GLCM, LBP, multi-LBGLCM, GLRLM, color moment features, and RSHD, together with a hybrid ensemble classifier incorporating decision trees, radial basis function SVMs, and fast large-margin classifiers.

A comprehensive review [13] analyzed the application of explainable AI (XAI) methods in medical imaging systems based on deep learning, identifying key trends, techniques, and evaluation methodologies, while highlighting both the advantages and challenges of XAI in medical imaging. Study [14] proposed a novel XAI framework specifically designed for medical image analysis, integrating statistical, visual, and rule-based explanations. This framework employed a two-stage feature selection process, decision trees, and RuleFit models for classification and extraction of human-interpretable rules, validated across five medical imaging datasets and confirmed by medical experts. Federated learning for brain tumor classification was addressed in [15], which introduced an interpretable and collaborative federated learning model combining XAI with decentralized training to enhance trust and interpretability in medical imaging while preserving patient data privacy.

Mutual information-based feature selection methods demonstrated advantages in [16], where radiomic feature selection for breast cancer diagnosis combined with SHAP explanations resulted in performance improvements of up to 15% over baseline methods. In [17], mutual information was applied to feature discretization in glioblastoma radiogenomics, integrating imaging phenotypes with molecular markers for survival prediction.

Recent studies have increasingly employed explainable artificial intelligence (XAI) techniques for the interpretation of radiomic models in the classification of glioma malignancy grades [18], in the differential diagnosis of glioblastoma versus primary central nervous system lymphomas [19], and in multitask classification frameworks for multiple brain tumor entities based on hybrid radiomic features derived from the Gray Level Co-occurrence Matrix (GLCM) and the Curvelet transform [10].

Interpretability and reproducibility have also become central research directions, particularly in overall survival prediction for glioblastoma patients [20], the development of hybrid radiomic–biomarker models [5, 6], and the prediction of molecular tumor characteristics (Ki-67) using radiomics-based nomograms [21].

Particularly noteworthy are recent attempts to integrate quantum-based approaches with radiomics in order to enhance model explainability in the differentiation between large brain metastases and high-grade gliomas [22]. Study [23] developed a radiomic signature combined with clinical factors and molecular biomarkers (IDH mutation status and MGMT promoter methylation) for predicting overall survival in patients with gliomas, achieving a

concordance index (C-index) of 0.774 in the training cohort and 0.776 in an external validation cohort. These results outperformed both the clinical–molecular model (0.719 and 0.689, respectively) and the standalone radiomic signature (0.771 and 0.711), demonstrating the superiority of integrated prognostic modeling.

A comprehensive approach proposed in [24] integrated multiregional radiomic features extracted from multiparametric magnetic resonance imaging, including diffusion and perfusion maps, to predict survival in patients with isocitrate dehydrogenase (IDH) wild-type glioblastoma. The model achieved an area under the receiver operating characteristic curve (AUC) exceeding 0.65 across all evaluated time points through the analysis of 25 statistically significant features selected from an initial set of 1,862 radiomic characteristics. Furthermore, s tudy [25] presented a stable and externally validated prognostic model based on T2-weighted MRI for 652 glioma patients across three independent cohorts, demonstrating that 14 radiomic features were significantly associated with the tumor immune response, particularly macrophage infiltration, and exhibited high prognostic power for overall survival.

Early prediction of treatment efficacy in glioblastoma and brain metastases has also been extensively investigated. A systematic review of existing studies [26] demonstrated that artificial intelligence–based approaches, particularly deep learning techniques such as convolutional neural networks (CNNs), achieve superior performance in predicting treatment outcomes, with reported AUC values ranging from 0.72 to 0.99 when AI-derived MRI features are combined with clinical variables. At the same time, the review highlights persistent challenges related to limited model interpretability and restricted inter-institutional generalizability, issues that the Image Biomarker Standardisation Initiative (IBSI) explicitly seeks to address. Interpretable rule-based models have further evolved, with study [27] extending the Local Interpretable Model-agnostic Explanations (LIME) framework to provide localized explanations for MRI-based brain tumor diagnosis, thereby improving model consistency and clinical trust.

A critical gap analysis reveals the absence of adaptive binarization methods for IBSI-standardized radiomic features that explicitly incorporate local clinical context. Existing studies [5, 7, 28] emphasize the need for comprehensive optimization criteria that simultaneously account for informational relevance, distributional balance, and clinical meaningfulness of the resulting feature representations.

The objective of this study is to develop a method for optimal binarization of IBSI-standardized radiomic features for the construction of interpretable models for brain tumor diagnosis, ensuring maximal preservation of diagnostically relevant information during transformation into a binary representation while explicitly incorporating local clinical context and the medical interpretability of threshold values.

To achieve this objective, the study focuses on the formulation of an optimality criterion for threshold selection based on the maximization of mutual information under distributional balance constraints, the development of an adaptive procedure for generating candidate threshold sets grounded in the statistical properties of feature distributions, the proposal of a locally adaptive binarization algorithm tailored to individual diagnostic cases, the design of a validation strategy for assessing the clinical relevance of the derived threshold values, and the experimental validation of the proposed approach on real-world MRI datasets of brain tumors.

The proposed method of optimal binarization of IBSI features is based on an information-theoretic framework that explicitly accounts for the specific requirements of medical applications. The fundamental concept of the method lies in maximizing the mutual information between a binarized feature and the diagnostic class, while preserving both the balance of the resulting distribution and the medical interpretability of the threshold values.

For a feature f with continuous values, a threshold value θ is defined to transform the feature into a binary form. The binarized feature is expressed as

$$f^{bin} = I\left(f > \theta\right) \tag{1}$$

$$\theta^* = \arg\max_{\theta}\left[I\left(f^{bin}(\theta), y\right) \cdot W(\theta)\right]$$

where $I\left(f_{bin}(\theta), y\right)$ denotes the indicator function and the diagnostic class y. The optimal threshold value $W(\theta) = 2 \cdot min(P(f_{bin} = 0), P(f_{bin} = 1))$ is determined by maximizing a modified mutual information criterion:

$$I(f^{bin}, y) = \sum_b \sum_c P(f^{bin} = b, y = c) \cdot \log \frac{P(f^{bin} = b, y = c)}{P(f^{bin} = b) \cdot P(y = c)} \tag{2}$$

where the summation is performed over all possible values of the binary feature $b \in \{0,1\}$ and the diagnostic classes $c \in \{1,2,3,4\}$.

The procedure for constructing the candidate threshold set is designed such that, for each IBSI feature, a set of potential threshold values Θ is generated. This set includes both percentile-based thresholds derived from the empirical feature distribution and statistically justified thresholds based on descriptive statistics. Such a combination ensures a balance between sensitivity to the empirical data distribution and adherence to theoretically grounded statistical characteristics.

Formally, the candidate thresholds are defined as

$$t_k = \text{percentile}(f, p_k), \quad p_k \in \{5,10,15,...,95\} t_j = \mu_f \pm j \cdot \sigma_f, \quad j \in \{0.5,1.0,1.5\} \tag{3}$$

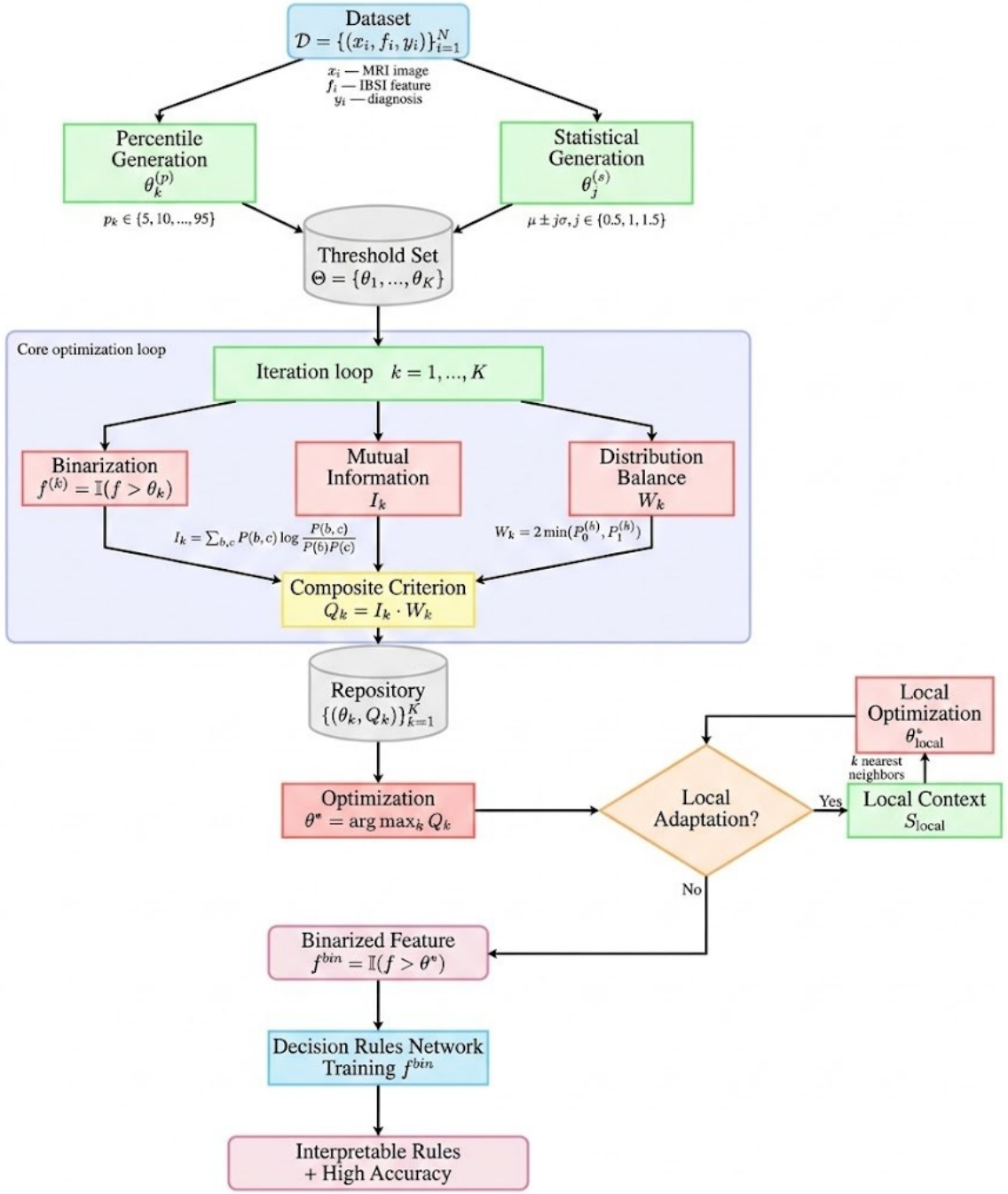

Fig. 1. Method of optimal binarization of IBSI-compliant features

For each candidate feature, feature binarization is performed, followed by the computation of mutual information based on the empirical class distribution in the training set, the calculation of a weighting coefficient, the computation of a composite criterion, and the selection of the optimal threshold:

$$\begin{aligned} f^{bin}(\theta_k) &= I(f > \theta_k); \\ I_k &= I(f^{bin}(\theta_k), y); \\ W_k &= 2 \cdot \min\left(\frac{n_0}{N}, \frac{n_1}{N}\right); \\ Q_k &= I_k \cdot W_k; \\ \theta^* &= \arg\max_k Q_k. \end{aligned} \tag{4}$$

Local adaptive binarization is performed as follows. For a given diagnostic case $x_{current}$, the method provides for the adaptation of threshold values to the local context. A local set is constructed that includes the current case and its nearest neighbors in the feature space:

$$S_{local} = \{x_{current}\} \cup \{x_i : \text{rank}(d(x_{current}, x_i)) \le k\}, \quad (5)$$

where $d(x_i, x_j)$ denotes the Euclidean distance between cases in the IBSI feature space. Threshold optimization is performed on the local set using local class $P_{local}(f_{bin} = b, y = c)$ distributions.

Modification of the criterion for the local context:

$$\theta^*_{local} = \arg\max_\theta [I_{local}(f^{bin}(\theta), y) \cdot W_{local}(\theta) \cdot R_{global}(\theta)], \quad (6)$$

where I_{local} denotes the mutual information computed on the local set, W_{local} is the local weighting coefficient, and R_{global} is the regularization coefficient:

$$R_{global}(\theta) = \exp\left(-\frac{|\theta - \theta^*_{global}|^2}{2\sigma^2_{threshold}}\right). \quad (7)$$

This coefficient ensures a balance between adaptation to the local context and the preservation of global consistency of threshold values.

To validate clinical relevance, the identified threshold values are assessed for compliance with medical criteria through an expert evaluation procedure. For each binarized feature, a clinical relevance index is computed:

$$CR_i = \alpha \cdot \text{Corr}(f_i, \text{marker}_{primary}) + \beta \cdot \text{Corr}(f_i, \text{marker}_{secondary}), \quad (8)$$

where $marker_{primary}$ and $marker_{secondary}$ are the expert-defined primary and secondary radiological markers for each tumor type, $\alpha = 0.7$ and and $\beta = 0.3$ are weighting coefficients. Features with an index $CR_i < 0.3$ value below the predefined threshold is labeled as potentially irrelevant and are subject to further expert analysis.

The proposed method was validated on a dataset containing brain MRI images with four classes: glioma, meningioma, pituitary tumor, and no tumor. For each image, 64 IBSI-standardized features were extracted, including texture features derived from the Gray Level Co-occurrence Matrix (GLCM, 22 features), Gray Level Run Length Matrix (GLRLM, 16 features), Gray Level Size Zone Matrix (GLSZM, 16 features), and morphological descriptors (10 features).

The proposed optimal binarization (OB) method was compared with four baseline approaches: median binarization (MB), mean binarization (MeanB), fixed percentile binarization (FP), and entropy-based discretization (EBD).

The average mutual information between binarized features and diagnostic classes for the different methods was as follows: optimal binarization (OB): 0.342 ± 0.087 bits; entropy-based discretization (EBD): 0.318 ± 0.091 bits; median binarization (MB): 0.287 ± 0.094 bits; mean binarization (MeanB): 0.279 ± 0.098 bits; and fixed percentile binarization (FP): 0.283 ± 0.096 bits. The proposed OB method demonstrated an improvement of 19.2% compared to median binarization (paired t-test, $p < 0.001$) and a 7.5% improvement compared to the EBD method.

The average value of the weighting coefficient W for the different methods was as follows: optimal binarization: $W = 0.948 \pm 0.031$; median binarization: $W = 1.000$ (by definition); mean binarization: $W = 0.891 \pm 0.112$; and EBD: $W = 0.827 \pm 0.145$. The OB method ensures a well-balanced feature distribution while preserving maximal informativeness.

For GLCM features, the average mutual information achieved by the OB method was 0.387 bits compared to 0.312 bits for MB. For GLRLM features, the improvement was 0.328 versus 0.281 bits. For GLSZM features, the OB method yielded 0.315 bits compared to 0.267 bits for MB. Morphological descriptors exhibited the smallest difference, with 0.298 versus 0.265 bits.

Additional experiments involving local adaptive thresholding (with $k = 20$ nearest neighbors) demonstrated an average improvement in mutual information of 8.4% compared to global optimization. The most pronounced effect was observed for complex cases near class boundaries, where local adaptation yielded improvements of up to 15%.

Interpretable models based on decision rule networks (DRN) were trained using different binarization methods. The classification accuracy of the DRN with optimal binarization reached 76.4%, compared to 71.2% for median binarization (an improvement of 7.3%, $p < 0.001$). Local consistency with the baseline VGG-16 model improved from 84.7% to 89.3%.

An analysis of the variability of the generated rules across different data splits showed that the OB method resulted in a 34% lower structural variability of rules compared to MB (Jaccard coefficient for the sets of activated rules: 0.82 versus 0.61). Expert evaluation of the generated rules indicated that 87% of rules derived from optimally binarized features were assessed as clinically meaningful, compared to 73% for median binarization.

CONCLUSIONS

A method for the optimal binarization of IBSI-standardized radiomic features has been developed to facilitate the creation of interpretable diagnostic models for brain tumors. The method is based on maximizing the mutual information between binarized features and diagnostic classes, while accounting for the balance of value distribution.

The proposed optimality criterion yields a 19.2% improvement in average mutual information compared to traditional median binarization, supported by statistically significant results ($p < 0.001$).

Locally adaptive binarization, tailored to the specifics of individual diagnostic cases, provides an additional 8.4% improvement. This is particularly effective for complex cases at the class boundaries, thereby increasing the reliability of personalized explanations.

Integration of the method with the Decision Rules Network architecture increased the accuracy of the interpretable model by 7.3% and achieved local consistency with the primary VGG-16 model of 89.3%. These results confirm the practical value of the proposed approach for developing robust explainable AI (XAI) systems in medical diagnostics.

References

1. Aguirre-Meneses H., Stoehr-Muñoz P., Molina-Gonzalez M., Nuñez-Gaona M.-A., Roldan-Valadez E. Radiomics and the Image Biomarker Standardisation Initiative (IBSI): A Narrative Review Using a Six-Question Map and Implementation Framework for Reproducible Imaging Biomarkers. Cureus. 2025. Vol. 17, No. 10. Pp. e95335. URL: https://doi.org/10.7759/cureus.95335

2. Zwanenburg A., Vallières M., Abdalah M. A., Aerts H. J. W. L., Andrearczyk V., Apte A., Löck S., та ін. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020. Vol. 295, No. 2. Pp. 328–338. URL: https://doi.org/10.1148/radiol.2020191145

3. Green R. W., Fischerová D., Testa A. C., Franchi D., Frühauf F., Lindqvist P. G., Legge A. di, Cibula D., Fruscio R., Haak L. A., Opolskiene G., Vidal Urbinati A. M., Timmerman D., Bourne T., Bosch T. van den, Epstein E. Sonographic, Demographic, and Clinical Characteristics of Pre- and Postmenopausal Women with Endometrial Cancer; Results from a Post Hoc Analysis of the IETA4 (International Endometrial Tumor Analysis) Multicenter Cohort. Diagnostics. 2024. Vol. 14, No. 1. Pp. 1. URL: https://doi.org/10.3390/diagnostics14010001

4. Carré A., Klausner G., Edjlali M., Lerousseau M., Briend-Diop J., Sun R., Ammari S., Reuzé S., Alvarez Andres E., Estienne T., Niyoteka S., Battistella E., Vakalopoulou M., Dhermain F., Paragios N., Deutsch E., Oppenheim C., Pallud J., Robert C. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. Scientific Reports. 2020. Vol. 10, No. 1. Pp. 12340. URL: https://doi.org/10.1038/s41598-020-69298-z

5. Jouzdani A. F., Taeb S., Maghsudi M., Gorji A., Rahmim A., Salmanpour M. R. Towards Interpretable AI in Personalized Medicine: A Radiological-Biological Radiomics Dictionary Connecting Semantic Lung-RADS and imaging Radiomics Features; Dictionary LC 1.0. 2025. URL: https://doi.org/10.48550/arXiv.2512.24529

6. Brocki L., Chung N. C. Integration of Radiomics and Tumor Biomarkers in Interpretable Machine Learning Models. Cancers. 2023. Vol. 15, No. 9. Pp. 2459. URL: https://doi.org/10.3390/cancers15092459

7. Nasarian E., Alizadehsani R., Acharya U. R., Tsui K.-L. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. Information Fusion. 2024. Vol. 108. Pp. 102412. URL: https://doi.org/10.1016/j.inffus.2024.102412

8. Papaioannou N., Myllis G., Tsimpiris A., Vrana V. The Role of Mutual Information Estimator Choice in Feature Selection: An Empirical Study on mRMR. Information. 2025. Vol. 16, No. 9. Pp. 724. URL: https://doi.org/10.3390/info16090724

9. Ferreira A. J., Figueiredo M. A. T. Feature Discretization with Relevance and Mutual Information Criteria / Pattern Recognition Applications and Methods, Cham, Springer International Publishing, 2015. Pp. 101–118. URL: https://doi.org/10.1007/978-3-319-12610-4_7

10. Jafari M., Rahmani M., Alibabaei S., Fatahiasl J., Razmjoo S., Birgani M. J. T., Tahmasbi M. Hybrid Radiomics and Machine Learning for Brain Tumors Multi-Task Classification: An Exploratory Study on Integrating GLCM and Curvelet-Based Features for Enhanced Accuracy. Health Science Reports. 2025. Vol. 8, No. 8. Pp. e71195. URL: https://doi.org/10.1002/hsr2.71195

11. Dheepak G., J. A. C., Vaishali D. Brain tumor classification: a novel approach integrating GLCM, LBP and composite features. Frontiers in Oncology. 2024. Vol. 13. URL: https://doi.org/10.3389/fonc.2023.1248452

12. Elazab N., Gab Allah W., Elmogy M. Computer-aided diagnosis system for grading brain tumor using histopathology images based on color and texture features. BMC Medical Imaging. 2024. Vol. 24, No. 1. Pp. 177. URL: https://doi.org/10.1186/s12880-024-01355-9

13. Houssein E. H., Gamal A. M., Younis E. M. G., Mohamed E. Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. Cluster Computing. 2025. Vol. 28, No. 7. Pp. 469. URL: https://doi.org/10.1007/s10586-025-05281-5

14. Ullah N., Guzmán-Aroca F., Martínez-Álvarez F., De Falco I., Sannino G. A novel explainable AI framework for medical image classification integrating statistical, visual, and rule-based methods. Medical Image Analysis. 2025. Vol. 105. Pp. 103665. URL: https://doi.org/10.1016/j.media.2025.103665

15. Mastoi Q., Latif S., Brohi S., Ahmad J., Alqhatani A., Alshehri M. S., Al Mazroa A., Ullah R. Explainable AI in medical imaging: an interpretable and collaborative federated learning model for brain tumor classification. Frontiers in Oncology. 2025. Vol. 15. URL: https://doi.org/10.3389/fonc.2025.1535478

16. Oladimeji O. O., Ayaz H., McLoughlin I., Unnikrishnan S. Mutual information-based radiomic feature selection with SHAP explainability for breast cancer diagnosis. Results in Engineering. 2024. Vol. 24. Pp. 103071. URL: https://doi.org/10.1016/j.rineng.2024.103071

17. Buzdugan S., Mazher M., Puig D. Radiogenomics for Glioblastoma Survival Prediction: Integrating Radiomics, Clinical, and Genomic Features Using Artificial Intelligence. Journal of Imaging Informatics in Medicine. 2025. URL: https://doi.org/10.1007/s10278-025-01692-3

18. Shi Y., Jiang W., Zhou L., Shi Y., Ji X., Zheng W., Zhao H., Yao N. Radiomics-based machine learning for glioma grade classification: a multicenter study with SHapley Additive exPlanations interpretability analysis. Translational Cancer Research. 2025. Vol. 14, No. 10. Pp. 7329–7346. URL: https://doi.org/10.21037/tcr-2025-2024

19. Xia X., Tan Q., Xie Y., Wu W., Gou Q. Interpretable radiomics-based machine learning model for differentiating glioblastoma from primary central nervous system lymphoma using contrast-enhanced T1-weighted imaging. Scientific Reports. 2025. Vol. 15, No. 1. Pp. 38598. URL: https://doi.org/10.1038/s41598-025-22391-7

20. Duman A., Sun X., Thomas S., Powell J. R., Spezi E. Reproducible and Interpretable Machine Learning-Based Radiomic Analysis for Overall Survival Prediction in Glioblastoma Multiforme. Cancers. 2024. Vol. 16, No. 19. Pp. 3351. URL: https://doi.org/10.3390/cancers16193351

21. Zhao E., Yang Y.-F., Bai M., Zhang H., Yang Y.-Y., Song X., Lou S., Yu Y., Yang C. MRI radiomics-based interpretable model and nomogram for preoperative prediction of Ki-67 expression status in primary central nervous system lymphoma. Frontiers in Medicine. 2024. Vol. 11. URL: https://doi.org/10.3389/fmed.2024.1345162

22. Felefly T., Roukoz C., Fares G., Achkar S., Yazbeck S., Meyer P., Kordahi M., Azoury F., Nasr D. N., Nasr E., Noël G., Francis Z. An Explainable MRI-Radiomic Quantum Neural Network to Differentiate Between Large Brain Metastases and High-Grade Glioma Using Quantum Annealing for Feature Selection. Journal of Digital Imaging. 2023. Vol. 36, No. 6. Pp. 2335–2346. URL: https://doi.org/10.1007/s10278-023-00886-x

23. Hao M., Yan J., Wang X., Tan Y., Zhang H., Yang G. Survival prediction in gliomas based on MRI radiomics combined with clinical factors and molecular biomarkers. PeerJ. 2025. Vol. 13. Pp. e19906. URL: https://doi.org/10.7717/peerj.19906

24. Styliara E. I., Astrakas L. G., Alexiou G., Xydis V. G., Zikou A., Kafritsas G., Voulgaris S., Argyropoulou M. I. Survival Outcome Prediction in Glioblastoma: Insights from MRI Radiomics. Current Oncology. 2024. Vol. 31, No. 4. Pp. 2233–2243. URL: https://doi.org/10.3390/curroncol31040165

25. Li G., Li L., Li Y., Qian Z., Wu F., He Y., Jiang H., Li R., Wang D., Zhai Y., Wang Z., Jiang T., Zhang J., Zhang W. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. Brain. 2022. Vol. 145, No. 3. Pp. 1151–1161. URL: https://doi.org/10.1093/brain/awab340

26. Moreau N. N., Valable S., Jaudet C., Dessoude L., Thomas L., Hérault R., Modzelewski R., Stefan D., Thariat J., Lechervy A., Corroyer-Dulmont A. Early characterization and prediction of glioblastoma and brain metastasis treatment efficacy using medical imaging-based radiomics and artificial intelligence algorithms. Frontiers in Oncology. 2025. Vol. 15. URL: https://doi.org/10.3389/fonc.2025.1497195

27. Kim S. H., Wihl J., Schramm S., Berberich C., Rosenkranz E., Schmitzer L., Serguen K., Klenk C., Lenhart N., Zimmer C., Wiestler B., Hedderich D. M. Human-AI collaboration in large language model-assisted brain MRI differential diagnosis: a usability study. European Radiology. 2025. Vol. 35, No. 9. Pp. 5252–5263. URL: https://doi.org/10.1007/s00330-025-11484-6

28. Chen Y., Tan L., Cao J., Fu X., Xue W., Chen J., Zhu Q. Deep-radiomics and explainable AI for asthma severity assessment: A multi-center CT imaging study with nomogram and decision curve analysis. Journal of Radiation Research and Applied Sciences. 2025. Vol. 18, No. 3. Pp. 101676. URL: https://doi.org/10.1016/j.jrras.2025.101676