Biomedical Signal Processing and Control 123 (2026) 110560

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Improving deep learning explainability by elucidating texture-based learning patterns in medical image segmentation

Saad Abdullah [a,1], Md Masum Billah [b,1], Victor Armando Canales-Lima [a], Pragati Manandhar [a], Lameya Islam [a], Alexis Gbeckor-Kove [a], Sarosh Krishan [a], Hergys Rexha [a,*], Sebastien Lafond [a], Kurt Benke [c], Sepinoud Azimi [b], Janan Arslan [d,e]

[a] Faculty of Science and Engineering, Åbo Akademi University, Turku, Finland
[b] Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands
[c] School of Engineering, University of Melbourne, Parkville, Australia
[d] Sorbonne Université, Institut du Cerveau–Paris Brain Institute–ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France
[e] Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia

ARTICLE INFO

Keywords:
Explainability
Image segmentation
Texture analysis

ABSTRACT

The black-box nature of deep learning (DL) models presents a significant challenge for their adoption in clinical settings. The field of explainable artificial intelligence (XAI) has emerged to improve the transparency and interpretability of models. However, current techniques do not adequately describe the reasoning underpinning DL models. This study replicates and extends previous research on the use of texture analysis to improve interpretability in clinically geared segmentation tasks. We evaluate Law's Texture Energy Measures (LTEMs) in the learning and decision-making processes of different DL architectures. We extend the work to include breast cancer, skin lesion, and gastrointestinal polyp datasets, as well as CLAHE-enhanced datasets to identify any divergence in learning. Experimental results reiterate that LTEMs, specifically level-edge convolution masks, are highly influential across multiple DL architectures. Additionally, Gray-Level Co-occurrence Matrix (GLCM) analysis highlights autocorrelation as a key descriptor. The results confirm that texture-based representations, learned primarily in the early layers of the network, are sufficient for robust learning. Through LTEMs, we can characterize the patterns learned in DL and associate these patterns with verbal descriptions and clinically objective measures, thus translating the DL learning into human terms. This psychophysical approach eases the clinical interpretability of DL models. Code availability: https://github.com/xrai-lib/xai-texture.

1. Introduction

The advent of Artificial Intelligence (AI) in healthcare, including deep learning (DL) models, has revolutionized medical diagnostics. AI-driven techniques have shown superior performance compared to traditional statistical methods, particularly in oncology, where they have been used to assist in cancer detection, classification, and prognosis [1–4]. Examples of successful AI applications in breast cancer include [2,5], renal cell carcinoma, and lung cancer diagnoses [6,7]. Despite its success, particularly in image segmentation, the black-box nature of AI prevents its full integration into real clinical settings. Concerns about AI's reliability, trustworthiness, and interpretability, especially in high-stakes healthcare applications, hinder its full integration. Transparency is essential for clinical adoption [1–4,8]. To address these concerns, Explainable AI (XAI) has emerged as a crucial field that aims to enhance the interpretability of AI models without sacrificing predictive power. Several XAI methods exist, such as Class Activation Maps (CAM) and Shapley Additive Explanations (SHAP), to name a few [9,10]. In XAI, SHAP is a method for computing Shapley values to explain a prediction by assessing the contribution of each feature, whereas CAM is a visualization method providing a heat map. That is, SHAP provides input feature attribution, whereas CAM provides spatially based explanations; thus, they are to some extent complementary. However, these methods do not directly provide the reasoning used by DL models. Interpretation of results could be improved by verbal interpretation of visual attributes of image data through a psychophysical approach.

A potential XAI method to improve interpretation is texture analysis. Traditionally, texture analysis is used to quantify texture features,

* Corresponding author.
E-mail address: hergys.rexha@abo.fi (H. Rexha).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bspc.2026.110560
Received 13 August 2025; Received in revised form 5 November 2025; Accepted 4 May 2026
Available online 12 May 2026

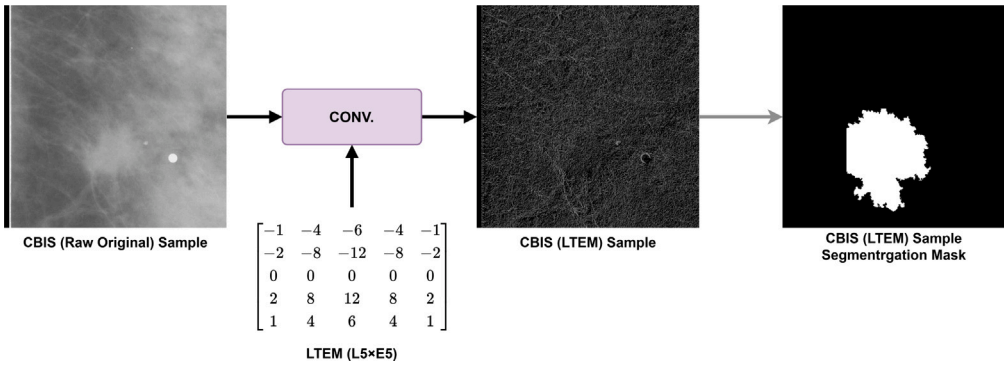

Fig. 1. Example of LTEM mask (Level-Edge) applied on a sample image of the CBIS-DDSM dataset.

such as coarseness or directionality in spatial patterns, to understand image heterogeneity [1]. Texture analysis could bridge feature quantification and interpretation, given that we have prior knowledge of various texture patterns and how to interpret these textures in layman's terms. Thus, feature maps in DL models can be correlated with known texture features, which can be interpreted in the context of a medical application.

In a previous study, we initially validated this theory by exploring the integration of Law's Texture Energy Measure (LTEM) with DL segmentation models to improve explainability in breast cancer imaging [11]. This work found that texture features, particularly the "Level" (L5) and "Edge" (E5) patterns, are significantly correlated with the segmentation process. In particular, the study has shown that DL models trained on LTEM-transformed data achieved segmentation performance comparable to models trained on raw images while improving interpretability through texture-driven insights. Other studies have shown that the L5E5 features are useful for capturing tumor heterogeneity in PET images and analyzing bone texture in osteoporosis, which explains why these patterns are widely included in radiomic analyses [12–14]. Furthermore, Gray-Level Co-occurrence Matrix (GLCM) analysis highlighted common learning patterns between models [11]. These findings suggest that DL models in medical image segmentation respond strongly to texture periodicity and edge structures. Building on these insights, this article aims to extend the analysis to a wider range of medical images beyond breast cancer. In addition, we expand the scope of the analysis by increasing the number of DL architectures under investigation from three to six. This extended evaluation aims to (1) replicate and validate the original findings by broadening the scope of the experiment and (2) verify whether DL processes are consistent or deviate when preprocessing techniques for image enhancement are applied. This study revealed that texture analysis can be used to interpret traditional black-box models using verbal descriptions associated with LTEMs.

2. Methods

Experiments included the evaluation of several DL models, the training of these models using different dataset configurations (e.g., using Contrast-Limited Adaptive Histogram Equalization [CLAHE] for preprocessing) to find patterns of similarities or divergence, and the influence of these configurations on the features learned by the DL models, with a focus on LTEMs and GLCMs.

2.1. Datasets

This study used three different medical image datasets: a subset of 904 images in the CBIS-DDSM dataset, an updated and standardized version of the Digital Database for Screening Mammography [15]; HAM10000, a comprehensive collection of 10,015 dermatoscopic images (skin surface microscopy) that encompass a diverse range of pigmented skin lesions (with seven distinct diagnostic categories [16]); and POLYP, an open-access, manually annotated dataset of 1000 gastrointestinal polyp images (endoscopy) [17]. To enhance image quality, we applied CLAHE as a preprocessing step on all datasets. It is common in DL pipelines to apply preprocessing methods, such as CLAHE, to improve image quality and thus segmentation. Various models were trained on the raw and CLAHE-equivalent datasets, since CLAHE improves local contrast while reducing noise artifacts, which is particularly beneficial for medical images with varying intensity distributions [18].

2.2. Deep learning architectures

We extend previous work by evaluating six segmentation architectures: three from the original study (DeepLabV3 [19], FCN [20], and U-Net [21]) plus three additional models (FPN [22], LinkNet [23], and HRNet [24]). The extension was designed to determine whether the preliminary results were repeatable, reproducible, and even consistent, given that DeepLabV3 employs dilated convolutions and Atrous Spatial Pyramid Pooling to capture multiscale context. Fully Convolutional Networks (FCN) use only convolutional layers with skip connections to preserve spatial information. U-Net applies a contracting-expanding path architecture [21,25,26]. Feature Pyramid Network (FPN) constructs feature pyramids through bottom-up and top-down pathways. LinkNet uses trained encoders for feature extraction coupled with a decoder module for classification. HRNet maintains high-resolution representations throughout the network, preserving spatial details for better boundary delineation [25,27].

2.3. Texture analysis

Law's Texture Energy Measure (LTEM): LTEMs serve as our primary texture feature extraction method, employing a systematic approach to quantify texture variations within fixed-size windows. The technique uses convolution masks derived from three fundamental vectors of length three. Through self-convolution and cross-convolution of these vectors, five distinct vectors of length five are generated, representing fundamental image properties: Level (L5), Edge (E5), Spot (S5), Ripple (R5), and Wave (W5). These base vectors are then combined through the outer products of column and row vectors to create 5×5 convolution masks, each designed to capture specific textural characteristics (see Fig. 1) [28]. This process enables the creation of nine distinct texture-enhanced versions of our datasets, each emphasizing different structural properties.

Gray Level Co-occurrence Matrix (GLCM): GLCM is a second-order statistical texture analysis method that captures spatial relationships between pixel intensities by computing the frequency of all

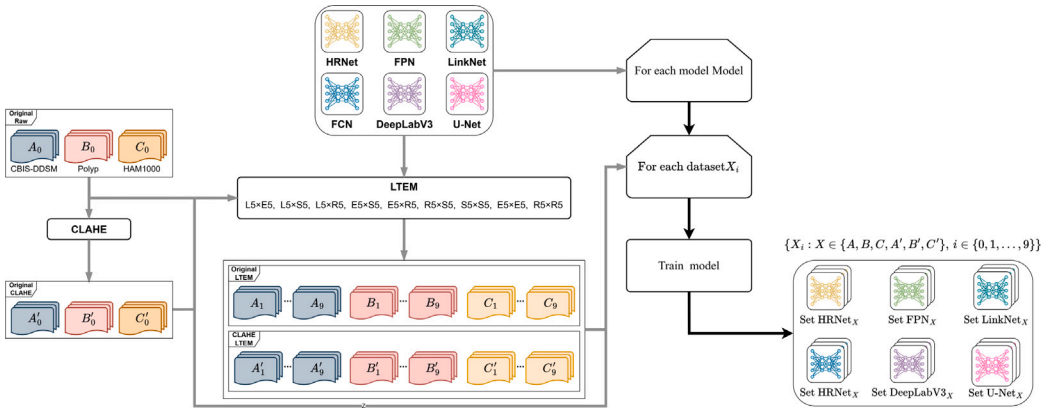

Fig. 2. Complete pipeline, including the evaluation of raw and CLAHE-enhanced datasets across six DL models. All models were processed to evaluate the nine LTEMs and their influence on DL learning processes.

possible gray-level value combinations at specified distances and directions [29]. Although GLCMs can be computed in eight directions, we employed the conventional approach of combining opposite directions to generate symmetric matrices, effectively reducing dimensionality while preserving essential textural information [30]. From these matrices, we derived a comprehensive suite of statistical descriptors, including Angular Second Moment (ASM), Contrast, Correlation, Variance, Inverse Difference Moment (IDM), Sum Average, Sum Entropy, Entropy, Difference Entropy, and Information Measures of Correlation (IMC1 and IMC2), as well as Autocorrelation. We applied this analysis to both the original image data and the feature representations generated by the terminal layers of our neural networks.

2.4. Experimental design

The experimental design followed a systematic approach to evaluate the impact of texture-based features on the performance and interpretability of the model. For each of the six DL architectures, we created multiple training configurations combining raw images, CLAHE-enhanced images, and nine LTEM-transformed versions of each dataset. This resulted in a total of 60 distinct dataset configurations (3 datasets × [raw + CLAHE + 9 LTEM of raw + 9 LTEM of CLAHE] = 60).

A model instance of each architecture was trained for each dataset configuration (6 architectures × 60 dataset configurations), as shown in Fig. 2. We evaluated segmentation performance using the Intersection over Union (IoU) metric, allowing us to rank the effectiveness of different LTEM-based models against those trained on original datasets.

Cosine similarity was used to compare the representation of the feature map generated by models trained on different dataset configurations. This analysis focused particularly on understanding how different LTEM transformations influenced the model's internal representations compared to those generated from raw image training. Layer-wise analysis was performed across all models to examine how texture-based transformations affected feature extraction at different network depths.

3. Results and discussion

Among the LTEMs, L5E5/E5L5 emerged as the most critical feature for accuracy and model performance. As shown in Table 1, these features consistently ranked in the top three positions in 35 of 36 tests, outperforming all other LTEM features. For more detailed results on the rankings of other texture features please refer to Table A.1. The L5 (Level) filter captures smooth transitions and gradual intensity variations, while the E5 (Edge) filter emphasizes high-frequency details, such as boundaries and abrupt intensity changes. In terms of their translation (specifically for medical settings), L5E5/E5L5 could be used to describe density-related characteristics of clinical outcomes. For example, in breast cancer imaging, it could be effective in identifying high-density regions (as seen in Fig. 1). Thus, this feature could potentially be used to describe fibroglandular density, a known risk factor for breast cancer [31].

Other LTEM features, including E5E5 and L5S5, also showed strong performance, ranking in the top three positions 30 and 28 times, respectively. However, their influence was not as consistently dominant as L5E5/E5L5. Regardless of the DL model that is trained, a clear preference for these features emerges, which reinforces the idea that texture-based representations are crucial for DL learning. This has significant implications for the interpretation of medical image segmentation and classification models.

Our analysis using GLCM revealed that autocorrelation is the most significant texture feature across all datasets and models, as shown in Fig. 3. This agrees with previous findings where GLCM autocorrelation provided key predictive information, such as in MRI-based studies of gliomas [32] and cardiac risk stratification [33]. It has also been associated with tumor genomic risk scores in breast MRI [34]. Autocorrelation measures the linear dependency of the pixel intensities within a texture, effectively capturing repetitive patterns and spatial uniformity. The prevalence of autocorrelation across various architectures and datasets indicates that DL models inherently rely on spatial texture information to segment complex medical images. These findings support the use of texture-based explainability techniques to interpret AI-driven medical models.

Edge detection filters and early convolutional neural network layers typically capture low-level visual concepts such as boundaries and simple textures. Zhou et al. [35] showed that convolutional units act as detectors for both low-level and high-level concepts, but it remains difficult to determine the specific contribution of each unit to different categories. Similarly, attention maps indicate where the model focuses in an image but do not explain the precise structural patterns that drive decisions. By explicitly isolating texture features through LTEM transformations and GLCM analysis, this study provides more direct evidence of the types of patterns, such as autocorrelation and L5E5 textures, that are consistently used by the models. This moves beyond general localization to identify which structural characteristics are critical for segmentation.

Although textures may appear as abstract features in images, their patterns can be explicitly described in quantitative analysis. The LTEM masks, for example, provide pattern matches that correspond to known spatial properties such as edge sharpness, density, or periodic structures. This eases the interpretability from model to layman's language. Furthermore, texture can be related to established clinical descriptors,

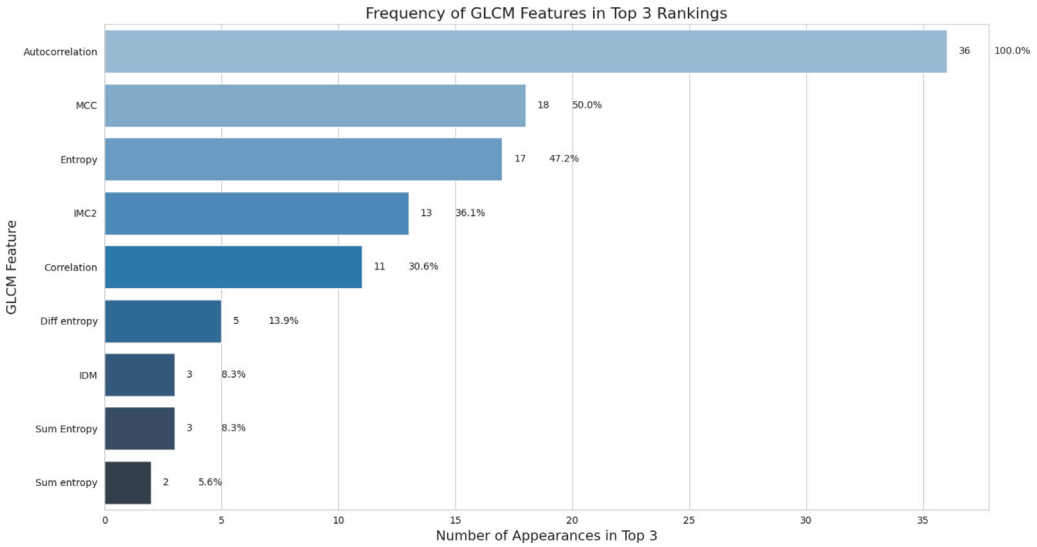

Fig. 3. Frequency of GLCM features across the evaluated models for both the raw-original and CLAHE-enhanced versions of the three datasets used (CBIS-DDSM, HAM10000, Polyp).

Table 1
Comparison of LTEM model accuracies using IoU across various architectures and datasets, including CLAHE-enhanced versions. L5E5/E5L5 features consistently achieved top ranking.

Dataset : curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8699	0.8485	0.8249	0.8145	0.6456	0.6209
Texture	L5E5/E5L5: 0.8258	L5E5/E5L5: 0.8178	L5E5/E5L5: 0.7714	E5E5: 0.6707	L5E5/E5L5: 0.6679	L5E5/E5L5: 0.5637
Dataset : CLAHE - curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8389	0.8285	0.7974	0.8243	0.7243	0.6413
Texture	L5E5/E5L5: 0.7636	L5E5/E5L5: 0.7842	L5E5/E5L5: 0.7616	L5E5/E5L5: 0.7562	L5E5/E5L5: 0.6559	S5S5: 0.4035
Dataset : HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8895	0.8905	0.9089	0.9058	0.8408	0.8586
Texture	L5E5/E5L5: 0.8892	L5E5/E5L5: 0.8855	L5E5/E5L5: 0.8887	L5E5/E5L5: 0.8798	L5E5/E5L5: 0.8378	L5E5/E5L5: 0.8229
Dataset : CLAHE - HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8909	0.8975	0.8939	0.9013	0.8804	0.8317
Texture	L5E5/E5L5: 0.8815	L5E5/E5L5: 0.8738	L5E5/E5L5: 0.8699	L5E5/E5L5: 0.8541	L5E5/E5L5: 0.8264	L5E5/E5L5: 0.7936
Dataset : polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8823	0.8836	0.8794	0.8566	0.7759	0.7202
Texture	L5E5/E5L5: 0.8421	L5E5/E5L5: 0.8427	L5E5/E5L5: 0.8209	L5E5/E5L5: 0.8242	L5E5/E5L5: 0.6987	L5E5/E5L5: 0.5825
Dataset : CLAHE -polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8640	0.8900	0.8550	0.8807	0.8139	0.7165
Texture	E5E5: 0.8230	L5E5/E5L5: 0.8204	L5E5/E5L5: 0.8081	L5E5/E5L5: 0.8052	L5E5/E5L5: 0.7196	L5S5/S5L5: 0.5655

such as BI-RADS categories in breast imaging or visual assessments of fibroglandular density. Showing that textures are fundamental to how models make predictions makes it easier to relate the results to established clinical criteria.

3.1. Consistency of texture-driven learning across models

A significant outcome of this study is the demonstration that DL models can achieve stable performance when trained exclusively on LTEM features. Despite variations in model architectures and preprocessing techniques (e.g., CLAHE enhancement), all models showed near-identical learning patterns when relying solely on LTEM-derived representations. This consistency suggests that DL models fundamentally focus on a set of primary core features, at least in the context of segmentation.

To further evaluate the effectiveness of the LTEM features, we conducted 324 pairwise comparisons using six DL models across six

Table 2
Average cosine similarities of feature maps between original models and LTEM-enhanced models across layers. High similarity scores, particularly for L5E5/E5L5 features, indicate strong behavioral alignment and preservation of critical information.

Dataset : curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3444	L5E5/E5L5: 0.3812	L5E5/E5L5: 0.4858	E5E5: 0.4399	R5R5: 0.6658	L5E5/E5L5: 0.3323
Dataset : CLAHE - curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.4092	L5R5/R5L5: 0.3959	L5E5/E5L5: 0.5017	E5E5: 0.4616	R5R5: 0.6749	E5E5: 0.3027
Dataset : HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.2932	L5E5/E5L5: 0.36704	L5E5/E5L5: 0.4193	L5E5/E5L5: 0.4402	L5E5/E5L5: 0.6724	S5S5: 0.3047
Dataset : CLAHE - HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.2912	L5E5/E5L5: 0.3459	L5E5/E5L5: 0.4131	R5R5: 0.3971	L5E5/E5L5: 0.6949	L5E5/E5L5: 0.2815
Dataset : polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3361	E5E5: 0.3841	L5E5/E5L5: 0.5021	E5E5: 0.4495	E5S5/S5E5: 0.6938	E5E5: 0.3476
Dataset : CLAHE - polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3888	E5E5: 0.4149	E5E5: 0.5023	E5E5: 0.48828	L5E5/E5L5: 0.6974	E5E5: 0.3629

datasets, including CLAHE-enhanced versions. The results, shown in Table 2, show the highest similarity score feature, where the L5E5/E5L5 models consistently ranked among the top performing configurations, appearing in the top three 34 out of 36 times. For more detailed information about similarity scores of other features, please refer to Table A.2 in the Appendix. Specifically, these models secured the first position 20 times, the second position 8 times, and the third position 6 times, reinforcing their role as the most influential LTEM features. Furthermore, the L5S5/S5L5 features, typically acquired during the initial training stages, were identified as among the most prominent features on 19 occasions, underscoring their persistent importance.

The ability of DL models to maintain robust segmentation and classification performance through the use of texture-based features suggests that LTEM can streamline model architectures without compromising accuracy. This finding has substantial implications for resource-constrained environments, where reducing computational complexity is crucial. By using LTEM-based representations, models can sustain strong decision-making capabilities while improving explainability and efficiency in AI-driven medical diagnostics.

Although processing LTEM filter transformations individually, rather than as combined multi-channel inputs, reduces the representational dimensionality of the original images, models trained on these texture-enhanced inputs consistently achieve segmentation performance comparable to models trained on raw data. When using LTEM, each convolution mask derived captures a specific textural property. If individual filter responses are used, the remaining texture components are omitted, inevitably leading to information loss; from an information-theoretic standpoint, a single LTEM filter response cannot fully reconstruct the original image. This reduction is not detrimental but rather deliberate, as it enables the isolation of the most informative LTEM masks. It is important to note that minimal perturbations can cause failures in DL models, so the continuity of learned representations cannot be assumed [36]. The stable and high performance observed in all of our experiments suggests that the texture features captured by the LTEM masks represent the core structural elements on which the models are based, highlighting the significance of these textures for prediction and interpretability.

4. Conclusions and future work

This study underscores the importance of texture analysis in enhancing the explainability of DL models in medical image segmentation. Building on our previous research, we validate our earlier findings by applying them to additional open-source datasets and DL architectures. We demonstrate that LTEMs, particularly the Level (L5) and Edge (E5) features, are crucial for model learning and decision-making. Furthermore, GLCM analysis confirms that autocorrelation is a key descriptor across multiple architectures, reinforcing the notion that DL models inherently rely on essential features in textural patterns, especially in the early layers. This reliance is further evident when preprocessing techniques such as CLAHE are employed. We can leverage this information to interpret the texture-based features used in DL models and associate them with clinical descriptors. For example, L5E5 has been shown to correlate with pattern density, which, in the context of breast cancer, could serve as an indicator of fibroglandular density.

In parallel, future studies will focus on broadening the scope of texture-based explainability beyond the current segmentation framework. This includes exploring how such approaches can be integrated with multimodal data sources, such as radiomics and pathology reports, to enrich clinical context, as well as assessing their applicability in real-time clinical environments where computational efficiency and interpretability are essential. Another direction involves extending texture-driven reasoning to related tasks, including classification and prognosis, to evaluate generalization across diverse diagnostic settings. These aspects are currently being investigated in a separate ongoing study aimed at advancing the clinical translation of texture-aware explainable artificial intelligence (XAI) methods.

CRediT authorship contribution statement

Saad Abdullah: Writing – original draft, Visualization, Validation, Software, Investigation, Data curation. **Md Masum Billah:** Writing – original draft, Visualization, Validation, Software, Investigation, Data curation, Conceptualization. **Victor Armando Canales-Lima:** Writing – review & editing, Writing – original draft, Visualization,

Software, Investigation, Data curation. **Pragati Manandhar:** Validation, Investigation, Data curation. **Lameya Islam:** Software, Investigation. **Alexis Gbeckor-Kove:** Software, Investigation. **Sarosh Krishan:** Software, Methodology, Investigation, Data curation. **Hergys Rexha:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Sebastien Lafond:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition. **Kurt Benke:** Writing – review & editing, Formal analysis, Conceptualization. **Sepinoud Azimi:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Janan Arslan:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used notebooklm and editGPT in order to search for relevant literature and proofread the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work has been partially supported by the funding of the EMJMD master's program in Engineering of Data-Intensive Intelligent Software Systems (EDISS), issued by the European Union's Education, Audiovisual and Culture Executive Agency (grant number 619819)

Appendix. Detailed results of the texture feature rankings among models and datasets

See Tables A.1 and A.2.

Data availability

The code used for the experiments is shared in GitHub.

Table A.1
Comparison of LTEM model accuracies using IoU across various architectures and datasets, including CLAHE-enhanced versions. Ranking of the accuracies achieved by different texture features. L5E5/E5L5 features consistently achieved top ranking.

Dataset : curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8699	0.8485	0.8249	0.8145	0.6456	0.6209
1	L5E5/E5L5: 0.8258	L5E5/E5L5: 0.8178	L5E5/E5L5: 0.7714	E5E5: 0.6707	L5E5/E5L5: 0.6679	L5E5/E5L5: 0.5637
2	E5E5: 0.7359	E5E5: 0.7453	E5E5: 0.7186	E5S5/S5E5: 0.6224	L5S5/S5L5: 0.4957	S5S5: 0.4113
3	L5S5/S5L5: 0.7329	L5S5/S5L5: 0.7290	E5S5/S5E5: 0.7137	L5E5/E5L5: 0.6218	E5E5: 0.4804	L5S5/S5L5: 0.4021
Dataset : CLAHE - curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8389	0.8285	0.7974	0.8243	0.7243	0.6413
1	L5E5/E5L5: 0.7636	L5E5/E5L5: 0.7842	L5E5/E5L5: 0.7616	L5E5/E5L5: 0.7562	L5E5/E5L5: 0.6559	S5S5: 0.4035
2	E5E5: 0.7177	E5E5: 0.7121	E5E5: 0.7246	E5E5: 0.7181	L5S5: 0.4764	E5E5: 0.3929
3	E5S5/S5E5: 0.7129	E5S5/S5E5: 0.7078	L5S5/S5L5: 0.7012	E5S5/S5E5: 0.7056	E5E5: 0.4710	E5S5/S5E5: 0.2663
Dataset : HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8895	0.8905	0.9089	0.9058	0.8408	0.8586
1	L5E5/E5L5: 0.8892	L5E5/E5L5: 0.8855	L5E5/E5L5: 0.8887	L5E5/E5L5: 0.8798	L5E5/E5L5: 0.8378	L5E5/E5L5: 0.8229
2	L5S5/S5L5: 0.8506	E5E5: 0.8381	E5E5: 0.8368	L5S5/S5L5: 0.8290	L5S5/S5L5: 0.7739	L5S5/S5L5: 0.7351
3	E5E5: 0.8437	L5S5/S5L5: 0.8198	L5S5/S5L5: 0.8291	E5S5/S5E5: 0.8119	E5E5: 0.7413	E5E5: 0.7024
Dataset : CLAHE - HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8909	0.8975	0.8939	0.9013	0.8804	0.8317
1	L5E5/E5L5: 0.8815	L5E5/E5L5: 0.8738	L5E5/E5L5: 0.8699	L5E5/E5L5: 0.8541	L5E5/E5L5: 0.8264	L5E5/E5L5: 0.7936
2	L5S5/S5L5: 0.8359	L5S5/S5L5: 0.8294	L5S5/S5L5: 0.8064	E5E5: 0.8112	E5E5: 0.7378	E5E5: 0.6999
3	E5E5: 0.8903	E5E5: 0.8270	E5E5: 0.8059	L5S5/S5L5: 0.7973	L5S5/S5L5: 0.7226	L5S5/S5L5: 0.6990
Dataset : polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8823	0.8836	0.8794	0.8566	0.7759	0.7202
1	L5E5/E5L5: 0.8421	L5E5/E5L5: 0.8427	L5E5/E5L5: 0.8209	L5E5/E5L5: 0.8242	L5E5/E5L5: 0.6987	L5E5/E5L5: 0.5825
2	L5S5/S5L5: 0.8276	E5E5: 0.8112	L5S5/S5L5: 0.7955	E5E5: 0.8040	E5S5: 0.6619	L5S5/S5L5: 0.5676
3	E5E5: 0.8167	L5S5/S5L5: 0.8078	E5E5: 0.7934	E5S5/S5E5: 0.7826	L5S5/S5L5: 0.6614	E5E5: 0.5385
Dataset : CLAHE -polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
Original	0.8640	0.8900	0.8550	0.8807	0.8139	0.7165
1	E5E5: 0.8230	L5E5/E5L5: 0.8204	L5E5/E5L5: 0.8081	L5E5/E5L5: 0.8052	L5E5/E5L5: 0.7196	L5S5/S5L5: 0.5655
2	L5S5/S5L5: 0.8075	L5S5/S5L5: 0.8102	L5S5/S5L5: 0.7952	E5E5: 0.7966	L5S5/S5L5: 0.6677	L5E5/E5L5: 0.5645
3	L5E5/E5L5: 0.8065	E5E5: 0.8020	E5S5: 0.7325	L5S5/S5L5: 0.7862	E5E5: 0.6222	E5E5: 0.5413

Table A.2
Average cosine similarities of feature maps between original models and LTEM-enhanced models across layers. The ranking of similarity scores for the three top texture features. L5E5/E5L5 features indicate strong behavioral alignment and preservation of critical information.

Dataset : curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3444	L5E5/E5L5: 0.3812	L5E5/E5L5: 0.4858	E5E5: 0.4399	R5R5: 0.6658	L5E5/E5L5: 0.3323
2	L5R5/R5L5: 0.3191	L5R5/R5L5: 0.3694	L5R5/R5L5: 0.4292	L5S5/S5L5: 0.4165	L5E5/E5L5: 0.6653	E5E5: 0.2971
3	L5S5/S5L5: 0.2919	L5S5/S5L5: 0.3325	E5E5: 0.4259	L5R5/R5L5: 0.4089	E5E5: 0.6646	L5R5/R5L5: 0.2899
Dataset : CLAHE - curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS DDSM)						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.4092	L5R5/R5L5: 0.3959	L5E5/E5L5: 0.5017	E5E5: 0.4616	R5R5: 0.6749	E5E5: 0.3027
2	E5E5: 0.3660	L5S5/S5L5: 0.388	E5E5: 0.4669	L5S5/S5L5: 0.4473	L5E5/E5L5: 0.6702	L5S5/S5L5: 0.3003
3	L5R5/R5L5: 0.3628	L5E5/E5L5: 0.3781	L5R5/R5L5: 0.4596	L5R5/R5L5: 0.4285	E5S5/S5E5: 0.6613	L5E5/E5L5: 0.2926
Dataset : HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.2932	L5E5/E5L5: 0.36704	L5E5/E5L5: 0.4193	L5E5/E5L5: 0.4402	L5E5/E5L5: 0.6724	S5S5: 0.3047
2	L5S5/S5L5: 0.2791	R5R5: 0.3202	L5S5/S5L5: 0.4023	L5S5/S5L5: 0.3989	S5S5: 0.6503	E5R5/R5E5: 0.3004
3	R5R5: 0.2614	S5S5: 0.3034	E5S5/S5E5: 0.3629	E5E5: 0.3824	E5S5/S5E5: 0.6439	L5E5/E5L5: 0.2899
Dataset : CLAHE - HAM10000 dataset, dermatoscopic images of common pigmented skin lesion						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.2912	L5E5/E5L5: 0.3459	L5E5/E5L5: 0.4131	R5R5: 0.3971	L5E5/E5L5: 0.6949	L5E5/E5L5: 0.2815
2	R5R5: 0.2553	L5S5/S5L5: 0.3299	L5S5/S5L5: 0.3896	E5S5/S5E5: 0.3859	R5R5: 0.6582	R5R5: 0.2796
3	L5S5/S5L5: 0.2517	E5S5/S5E5: 0.3297	R5R5: 0.3739	L5E5/E5L5: 0.3839	E5E5: 0.6549	R5S5/S5R5: 0.2726
Dataset : polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3361	E5E5: 0.3841	L5E5/E5L5: 0.5021	E5E5: 0.4495	E5S5/S5E5: 0.6938	E5E5: 0.3476
2	E5E5: 0.3232	L5E5/E5L5: 0.3749	E5E5: 0.4839	L5E5/E5L5: 0.4449	L5E5/E5L5: 0.6929	L5S5/S5L5: 0.3437
3	L5S5/S5L5: 0.3099	L5S5/S5L5: 0.3653	L5R5/R5L5: 0.4598	L5R5/R5L5: 0.4372	E5E5: 0.6776	L5E5/E5L5: 0.3219
Dataset : CLAHE - polyp dataset for Computer-Aided Gastrointestinal Disease Detection						
Rank	DeepLabv3	FCN	FPN	LinkNet	HRNet	U-Net
1	L5E5/E5L5: 0.3888	E5E5: 0.4149	E5E5: 0.5023	E5E5: 0.48828	L5E5/E5L5: 0.6974	E5E5: 0.3629
2	L5S5/S5L5: 0.3399	L5E5/E5L5: 0.3993	L5E5/E5L5: 0.4897	L5E5/E5L5: 0.4769	S5S5: 0.6874	L5S5/S5L5: 0.3544
3	E5E5: 0.3358	L5R5/R5L5: 0.3579	L5S5/S5L5: 0.4830	R5R5: 0.4525	L5S5/S5L5: 0.6769	L5E5/E5L5: 0.3420

References

[1] G. Corrias, G. Micheletti, L. Barberini, J.S. Suri, L. Saba, Texture analysis imaging "what a clinical radiologist needs to know", Eur. J. Radiol. 146 (2022) 110055.

[2] R. Karthiga, K. Narasimhan, R. Amirtharajan, et al., Review of AI & XAI-based breast cancer diagnosis methods using various imaging modalities, Multimedia Tools Appl. (2024) 1–52.

[3] A. Nicolaou, M. Pantzaris, C.P. Loizou, A. Kakas, C.S. Pattichis, An explainable AI model in the assessment of multiple sclerosis using clinical data and brain MRI lesion texture features, in: 2023 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI, 2023, pp. 1–4.

[4] L. Zou, H.L. Goh, C.J.Y. Liew, J.L. Quah, G.T. Gu, J.J. Chew, M.P. Kumar, C.G.L. Ang, A.W.A. Ta, Ensemble image explainable AI (XAI) algorithm for severe community-acquired pneumonia and COVID-19 respiratory infections, IEEE Trans. Artif. Intell. 4 (2) (2022) 242–254.

[5] H. Dihmani, A. Bousselham, O. Bouattane, A new computer aided diagnosis for breast cancer detection of thermograms using metaheuristic algorithms and explainable AI, Algorithms 17 (10) (2024) 462.

[6] S.T.H. Shah, S.A.H. Shah, I.I. Khan, A. Imran, S.B.H. Shah, A. Mehmood, S.A. Qureshi, M. Raza, A. Di Terlizzi, M. Cavaglia, et al., Data-driven classification and explainable-AI in the field of lung imaging, Front. Big Data 7 (2024) 1393758.

[7] K. Wu, P. Wu, K. Yang, Z. Li, S. Kong, L. Yu, E. Zhang, H. Liu, Q. Guo, S. Wu, A comprehensive texture feature analysis framework of renal cell carcinoma: pathological, prognostic, and genomic evaluation based on CT images, Eur. Radiol. (2022) 1–11.

[8] T. Telecan, I. Andras, N. Crisan, L. Giurgiu, E.D. Căta, C. Caraiani, A. Lebovici, B. Boca, Z. Balint, L. Diosan, et al., More than meets the eye: using textural analysis and artificial intelligence as decision support tools in prostate cancer diagnosis—a systematic review, J. Pers. Med. 12 (6) (2022) 983.

[9] S.M. Lundberg, S. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems 30, 2017, pp. 4765–4774.

[10] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 618–626.

[11] M. Billah, P. Manandhar, S. Krishan, A. Cedillo, H. Rexha, S. Lafond, K. Benke, S. Azimi, J. Arslan, Explainability in deep learning segmentation models for breast cancer by analogy with texture analysis, in: Medical Imaging with Deep Learning (MIDL 2024).

[12] S. Karacavus, B. Yılmaz, A. Tasdemir, Ö. Kayaaltı, E. Kaya, S. İçer, O Ayyıldız, Can laws be a potential PET image texture analysis approach for evaluation of tumor heterogeneity and histopathological characteristics in NSCLC? J. Digit. Imaging 31 (2018) 210–223.

[13] M. Rachidi, A. Marchadier, C. Gadois, E. Lespessailles, C. Chappard, C. Benhamou, Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis, Skelet. Radiol. 37 (2008) 541–548.

[14] V. Parekh, M. Jacobs, Radiomics: a new application from established techniques, Expert. Rev. Precis. Med. Drug Dev. 1 (2016) 207–226.

[15] R.S. Lee, F. Gimenez, A. Hoogi, K.K. Miyake, M. Gorovoy, D.L. Rubin, A curated mammography data set for use in computer-aided detection and diagnosis research, Sci. Data 4 (1) (2017) 1–9.

[16] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018) 180161.

[17] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H.D. Johansen, Kvasir-SEG: A segmented polyp dataset, in: MultiMedia Modeling, 2020, pp. 451–462.

[18] A. Mishra, Contrast limited adaptive histogram equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints, 2021, CoRR. abs/2109.00886. https://arxiv.org/abs/2109.00886.

[19] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587.

[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3431–3440.

[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, 2015, arXiv preprint arXiv:1505.04597.

[22] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 2117–2125.

[23] A. Chaurasia, E. Culurciello, LinkNet: Exploiting encoder representations for efficient semantic segmentation, in: IEEE Visual Communications and Image Processing, VCIP, 2017, pp. 1–4.

[24] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, B. Xiao, Deep high-resolution representation learning for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2021) 3349–3364.
[25] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2022) 3523–3542.
[26] Z. Xiao, T. Chai, N. Li, X. Shen, T. Guan, J. Tian, X. Li, Research advances in deep learning for image semantic segmentation techniques, IEEE Access 12 (2024) 175715–175741.
[27] S. Schutte, J. Uddin, Deep segmentation techniques for breast cancer diagnosis, BioMedInformatics 4 (2) (2024) 921–945.
[28] H. Singh, V. Sharma, D. Singh, Comparative analysis of proficiencies of various textures and geometric features in breast mass classification using k-nearest neighbor, Vis. Comput. Ind. Biomed. Art 5 (1) (2022) 3.
[29] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621.
[30] L.-K. Soh, C. Tsatsoulis, Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, IEEE Trans. Geosci. Remote Sens. 37 (2) (1999) 780–795.
[31] A.L. Brown, C. Vijapura, M. Patel, A. De La Cruz, R. Wahab, Breast cancer in dense breasts: Detection challenges and supplemental screening opportunities, RadioGraphics 43 (10) (2023).
[32] S. Kihira, N. Tsankova, A. Bauer, Y. Sakai, K. Mahmoudi, N. Zubizarreta, J. Houldsworth, F. Khan, N. Salamon, A. Hormigo, et al., Multiparametric MRI texture analysis in prediction of glioma biomarker status: added value of MR diffusion, Neuro-Oncol. Adv. 3 (2021) vdab051.
[33] A. Amyar, D. Al-Deiri, J. Sroubek, A. Kiang, F. Ghanbari, S. Nakamori, J. Rodriguez, D. Kramer, W. Manning, D. Kwon, et al., Radiomic cardiac MRI signatures for predicting ventricular arrhythmias in patients with nonischemic dilated cardiomyopathy, JACC: Adv. 4 (2025) 101684.
[34] S. Huang, B. Franc, R. Harnish, G. Liu, D. Mitra, T. Copeland, V. Arasu, J. Kornak, E. Jones, S. Behr, et al., Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis, NPJ Breast Cancer 4 (2018) 24.
[35] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[36] V. Antun, F. Renna, C. Poon, B. Adcock, A. Hansen, On instabilities of deep learning in image reconstruction and the potential costs of AI, Proc. Natl. Acad. Sci. 117 (2020) 30088–30095.