**SUPPLEMENTAL**

1. **MATERIAL AND METHODS**
   1. **Radiological Definition of Radiomics Features**

**A diagram of the human body

Description automatically generated**

**Supplemental Fig. S1.** Schematic illustration of the prostate gland anatomy, detailing all its components. DWI: Diffusion-weighted MRI imaging, DEC: Dynamic contrast-enhanced MRI imaging, T2WI: T2-weighted MRI imaging [1].

A collage of images of a human body

Description automatically generated

**Supplemental Fig. S2.** An example of MRI and prostate gland and lesion segmentations.

* 1. **Interpretable Machine Learning Algorithms**

**2.2.1. Enhancing model interpretability by Interpretable Feature Selection Algorithms.** Feature selection is crucial for making machine learning models more interpretable. By carefully choosing a relevant subset of features, the model's complexity is reduced, allowing for a clearer understanding of the patterns it identifies. This approach not only helps prevent overfitting—a frequent issue where models perform well on training data but struggle with new data—but also highlights and retains the most impactful features. This improved clarity supports a more transparent decision-making process and enables domain experts to extract actionable insights from the model's results [2]. Feature selection methods are categorized into three types: filter methods, which select subsets of variables independently of the prediction model; wrapper methods, which use a learning algorithm, as a black box, to assess and select subsets based on their predictive performance; and embedded methods, which incorporate feature selection directly into the model training process. [3] [4] indicated filter methods offer high interpretability by selecting features based on straightforward statistical criteria without relying on any model, whereas wrapper methods are less interpretable because they depend on model performance and involve complex feature interactions. Embedded methods, on the other hand, vary in interpretability; simpler models tend to be easier to understand, while more complex models pose greater challenges in interpretation. FSAs are elaborated in below.

**i) Chi-Square Test (CST)** is an interpretable feature selection method because it uses a simple and transparent statistical criterion to evaluate the relationship between categorical features and the target variable. By comparing the observed and expected frequencies in a contingency table, the CST determines whether a feature is significantly associated with the target variable. This straightforward process makes the CST easy to understand and explain, as it relies on basic statistical principles without involving complex model assumptions or interactions. Consequently, the resulting Chi-Square value provides a clear indication of the strength of the relationship between each feature and the target variable, making the feature selection process both transparent and interpretable. The CST works by first constructing a contingency table that displays the frequency distribution of the feature and target variable. It then calculates the expected frequencies under the assumption of independence. The Chi-Square statistic is computed by summing the squared differences between observed and expected frequencies, divided by the expected frequencies for all cells in the table. This statistic is then compared to a Chi-Square distribution to determine the p-value, which indicates the likelihood that the observed association occurred by chance. If the p-value is low, the feature is considered significantly associated with the target variable, aiding in effective feature selection [5] [6].

**ii) Correlation Coefficient (CCF)** feature selection algorithm is an interpretable method used to evaluate the linear relationship between a feature and the target variable. By calculating the correlation coefficient, typically Pearson’s correlation, the algorithm quantifies how changes in the feature correspond to changes in the target variable. A coefficient close to 1 or -1 indicates a strong positive or negative linear relationship, respectively, while a coefficient near 0 suggests no linear relationship. This clear and straightforward metric makes the CCF algorithm highly interpretable, as it allows for easy identification of features that significantly impact the target variable. Since the CCF method does not rely on complex models or assumptions, it offers a transparent and easily understandable process for feature selection, where features with high absolute correlation values are selected for further analysis, contributing to more efficient and accurate modeling [5] [6]..

**iii) Mutual Information (MIS)** feature selection algorithm is an interpretable method used to measure the dependency between a feature and the target variable, capturing both linear and non-linear relationships. MIS quantifies the amount of information gained about one variable through the knowledge of another by comparing the joint probability distribution of the feature and target variable with the product of their marginal distributions. Essentially, it measures how much knowing the value of a feature reduces the uncertainty about the target variable. The process involves estimating the joint probability distribution of the feature and the target variable, as well as their marginal distributions, and then calculating MIS by summing the differences between the joint probability and the product of the marginal probabilities for all possible values. A higher MIS value indicates a stronger dependency, making the feature more informative and relevant for the predictive model. MIS is considered interpretable because it provides a clear and intuitive metric for understanding the strength of relationships, directly reflecting how much information is shared between variables without relying on complex model structures or assumptions. This transparency in how MIS measures the relevance of features allows for easy understanding and justification of the selected features, making the feature selection process more transparent and accessible. By selecting features with high mutual information, the algorithm enhances the accuracy and robustness of the resulting model while ensuring that the feature selection process remains clear and comprehensible [5] [6].

**iv) Variance Threshold (VTS)** feature selection algorithm is an interpretable and straightforward method that removes features with low variance, operating on the assumption that such features do not significantly contribute to the model's predictive power. The algorithm works on the premise that a feature with low variance has little variability across different samples, meaning it carries minimal information that can distinguish between different classes or outcomes. The VTS algorithm calculates the variance of each feature and compares it against a predefined threshold, removing features with variance below this threshold from the dataset while retaining those above it for further analysis. VTS is considered interpretable because it uses a simple and clear criterion—variance—to assess feature relevance. The logic is easy to understand: if a feature shows little variation across samples, it likely offers little useful information for distinguishing between different outcomes, making it a candidate for removal. This simplicity makes VTS a transparent method, as the decision to keep or remove a feature is based on a well-defined and easily explainable metric. By focusing on features that exhibit sufficient variability, the VTS algorithm effectively reduces the dimensionality of the dataset, enhancing the model's efficiency without sacrificing important information [5] [6]**.**

**v) ANOVA F-test (AFT)** feature selection method is an interpretable statistical technique used to evaluate the significance of the differences in means between two or more groups for a given feature, making it particularly useful in classification tasks. The AFT method operates by calculating the ratio of the variance between the groups to the variance within the groups, known as the F-statistic. A higher F-statistic indicates that the feature is likely to have a significant impact on distinguishing between the different classes or outcomes. This method is considered interpretable because it uses a straightforward and well-established statistical test to assess feature relevance. The logic behind the AFT is simple: if the meaning of a feature across different groups vary significantly compared to the variability within each group, then the feature is likely informative for classification. The resulting F-statistics provides a clear metric that helps in deciding whether a feature should be included in the model. Features with high F-statistics are considered more relevant and are typically selected for further analysis. By focusing on the significance of the differences in means, the AFT ensures that only the most informative features are retained, reducing the complexity of the model and enhancing its performance. The method's reliance on basic statistical principles makes it easy to understand and justify, ensuring transparency and interpretability in the feature selection process [5] [6].

**vi) Information Gain (IGS)** feature selection method is an interpretable technique used to measure the reduction in uncertainty or entropy about the target variable when a feature is known. It quantifies how much knowing the value of a specific feature contributes to predicting the target variable. IG operates by calculating the difference between the entropy of the target variable before and after observing the feature. Features that result in a greater reduction of entropy—meaning they provide more information about the target—are considered more relevant and are thus selected for further analysis. Information Gain is considered interpretable because it uses a clear and intuitive measure—entropy reduction—to assess the importance of a feature. The logic is straightforward: if a feature significantly reduces uncertainty about the target variable, it likely contains valuable information that improves model performance. This clear-cut metric allows for easy understanding and justification of which features are selected or discarded, making the decision-making process transparent. IG is particularly useful in classification tasks, where it helps identify features that best split the data into classes, thereby improving the accuracy and efficiency of the model. The method's reliance on fundamental concepts from information theory ensures that the feature selection process is both scientifically sound and easy to interpret, contributing to the development of robust and transparent models [5] [6].

**viii) Univariate Feature Selection (UFS)** is an interpretable and straightforward method that evaluates each feature individually to determine its relevance to the target variable. The core idea behind UFS is to assess each feature independently, using statistical tests such as chi-square tests, t-tests, AFT, or mutual information scores, depending on the nature of the data and the target variable. Each feature is scored based on how well it predicts the target, and features with the highest scores are selected for inclusion in the model. UFS is considered interpretable because it uses clear, well-established statistical criteria to evaluate the importance of each feature. Since each feature is assessed independently, the process is transparent and easy to understand. The logic is simple: features that show a strong individual relationship with the target variable are likely to be useful predictors, and therefore, they are selected for further analysis. This transparency makes it easy to justify which features are retained and which are discarded. Moreover, UFS is particularly useful in scenarios where the goal is to quickly identify the most relevant features from a large dataset without getting into the complexity of multivariate interactions. By focusing on individual relationships, UFS simplifies the feature selection process, reducing dimensionality while maintaining the interpretability and effectiveness of the model. This method is especially valuable in the initial stages of analysis, where a quick and clear understanding of feature relevance is needed [5] [6].

**ix) Fisher Score (FSF)** feature selection algorithm is an interpretable method used to evaluate the importance of features by measuring how well they distinguish between different classes. The algorithm is based on the Fisher criterion, which assesses the ratio of the variance between classes to the variance within classes for each feature. A higher Fisher Score indicates that the feature provides a greater distinction between classes, making it more valuable for classification tasks. The Fisher Score is considered interpretable because it uses a straightforward and well-understood statistical principle to evaluate feature relevance. The logic behind FS is simple: if a feature shows a large variance between different classes while maintaining low variance within each class, it is likely a good discriminator and thus a useful feature for the model. This clear criterion allows for easy understanding and justification of the feature selection process. In practice, the Fisher Score is calculated by computing the mean and variance of the feature values within each class and then using these statistics to determine how well the feature separates the classes. Features with higher scores are considered more relevant and are selected for inclusion in the model. The simplicity and transparency of the Fisher Score make it a valuable tool in feature selection, especially in scenarios where interpretability and clarity are important. By focusing on how well features separate different classes, FS helps in building models that are both effective and easy to understand [5] [6].

**x) Least Absolute Shrinkage and Selection Operator (LAS)** is highly interpretable for feature selection because it automatically shrinks the coefficients of less important features to zero, effectively excluding them from the model. This results in a sparse model that includes only the most impactful features, making it easier to understand which features are driving the predictions. LAS works by adding an L1 regularization term to the traditional linear regression model, penalizing the absolute size of the coefficients. As the regularization parameter increases, more coefficients are driven to zero, leaving only the most significant features with non-zero coefficients. This transparency allows for direct interpretation of the selected features, showing exactly how each one contributes to the target variable. Additionally, LAS often aligns with domain knowledge, ensuring that the selected features are both relevant and credible, which enhances the interpretability and usefulness of the model in focusing on the most significant predictors. The use of cross-validation to select the optimal regularization parameter further ensures a balance between model complexity and predictive accuracy, making LAS a powerful tool for interpretable feature selection) [7] [8].

**2.2.2. Interpretable and complex Classification Algorithms.**

1. **Logistic Regression (LOR)** is widely regarded as an interpretable machine learning model due to its simplicity and the direct interpretability of its coefficients. Each coefficient in LR represents the relationship between a feature and the log-odds of the outcome, allowing for a clear understanding of how each feature influences the prediction. Positive coefficients indicate an increase in the probability of the positive class with higher feature values, while negative coefficients suggest the opposite. Additionally, LR provides a linear decision boundary and probabilistic output, making it easy to visualize and interpret the model's predictions. Its transparency and ease of interpretation make it suitable for applications where understanding the model's decision-making process is crucial [9].
2. **Decision Tree Classification (DTC**is widely regarded as an interpretable machine learning model due to its intuitive tree-like structure, which mirrors human decision-making processes. The model operates by splitting data into subsets based on feature values, with each path from the root to a leaf node representing a series of decisions that lead to a specific classification. This clear decision path allows users to easily trace the reasoning behind predictions, as the influence of each feature is transparently shown. Additionally, Decision Trees can be visually represented, further enhancing their interpretability by clearly illustrating how data is split and which features are most important. The simplicity of the decision-making process, which relies on straightforward conditions or thresholds, makes DTC a preferred choice in applications where understanding the model's logic is essential. DTC is ideal for scenarios where the reasoning behind predictions needs to be transparent and easily understandable, making it a valuable tool in fields that require clear and interpretable decision-making processes [10].
3. **Linear Discriminant Analysis (LDA)** is considered an interpretable machine learning model because it creates linear decision boundaries by finding a linear combination of features that best separates the classes. The model's coefficients directly represent the weight of each feature in distinguishing between classes, making it easy to understand how each feature contributes to the decision-making process. LDA is particularly useful in fields like medical diagnosis and social sciences, where understanding the relationships between features and class labels is crucial. Its outputs can be easily explained to non-technical stakeholders, and when used for dimensionality reduction, the resulting visualizations further enhance interpretability, making LDA a preferred choice in applications that require both accuracy and transparency. Additionally, by maximizing the ratio of between-class variance to within-class variance, LDA ensures that the classes are as distinct as possible in the projected space, which not only aids in clear classification but also provides meaningful insights into the underlying data structure. This process of creating linear discriminants helps identify key patterns and features, guiding decision-making and strategy development across diverse applications [11].
4. **Naive Bayes Classifier (NBC)** is considered an interpretable machine learning algorithm due to its simple and transparent probabilistic model based on Bayes' theorem. It calculates the probability of each class given the input features, allowing for easy examination of how predictions are made. NBC works by assuming feature independence, simplifying the computation while enhancing interpretability. During training, it determines the prior probability of each class and the likelihood of each feature given a class. For prediction, NBC calculates the posterior probability for each class by combining these factors, selecting the class with the highest probability. This process enables the direct tracing of each feature's contribution to the final decision, providing clear insights into the model's reasoning. These qualities make NBC particularly useful in applications where understanding the model's reasoning is crucial, such as in medical diagnosis or spam filtering [12].
5. **K-Nearest Neighbors (KNN)** algorithm is recognized as an interpretable machine learning method due to its simplicity and transparency. KNN classifies a new data point by identifying the "k" closest neighbors in the feature space and assigning the most common class among them, based on distance metrics like Euclidean or Manhattan distance. This straightforward decision-making process allows for easy understanding and tracing of how predictions are made, as each prediction is directly influenced by the nearest data points. Unlike many other algorithms, KNN makes no assumptions about the underlying data distribution, enhancing its versatility and interpretability. Additionally, its operations can be easily visualized, especially in lower-dimensional spaces, further aiding in comprehending the model’s behavior. These characteristics make KNN particularly useful in applications where transparency is essential, such as recommendation systems, image recognition, and medical diagnostics, where explaining the rationale behind classifications or predictions is crucial [12].
6. **RuleFit Classifier (RFC)** is an interpretable machine learning algorithm that combines the strengths of rule-based learning with linear models, making it ideal for capturing complex interactions while maintaining transparency. It works by generating decision rules from tree-based models, treating these rules as binary features, and then fitting a sparse linear model using both the rules and the original features. This approach allows RFC to provide intuitive, rule-based explanations for predictions, highlighting the most important features and rules contributing to the outcome. By balancing complexity and interpretability, RFC is particularly valuable in fields like credit scoring, healthcare, and marketing, where understanding and trusting the model's decisions are crucial [12].
7. **Random Forest Classifier (RFC)** is a complex machine learning algorithm that constructs an ensemble of decision trees to improve predictive performance. By employing techniques like bagging (bootstrap aggregation) and random feature selection at each node, RFC introduces randomness to ensure diversity among trees. This complexity enables it to capture intricate patterns and interactions within the data, making it highly effective for both classification and regression tasks. Its ensemble approach reduces overfitting and improves generalization by aggregating the outputs of multiple trees through voting or averaging. RFC is also equipped to handle high-dimensional data and assess feature importance, adding interpretability to its sophisticated structure [13].
8. **XGBoost Classifier (XGB)** is a structured, advanced gradient boosting algorithm that builds an ensemble of decision trees sequentially. Each tree is designed to correct errors made by the previous ones, optimizing a defined objective function during training. XGB incorporates structural enhancements such as regularization to control model complexity and prevent overfitting, and it uses split finding algorithms like weighted quantile sketch for efficient handling of missing values. Its design includes parallelized and distributed computations for faster model building. The structured training process and tunable hyperparameters allow it to adapt effectively to both simple and complex datasets, achieving high accuracy and robustness [14].
9. **LightGBM Classifier (LGB)** is a structured gradient boosting framework designed for efficiency and scalability in supervised learning tasks. It builds decision tree ensembles sequentially, focusing on reducing residual errors while optimizing a custom loss function. LGB uses a unique leaf-wise growth strategy, splitting the leaf with the largest loss reduction, which makes it faster and more accurate for complex datasets. It supports efficient handling of large datasets with high-dimensional features, leveraging histogram-based algorithms to speed up computation and reduce memory usage. Its structure incorporates regularization techniques to prevent overfitting and supports parallel learning. LGB is highly tunable and performs exceptionally well in classification and regression tasks, particularly with large and imbalanced datasets [14].
10. **CatBoost Classifier (CBC)** is a gradient boosting algorithm specifically designed to handle categorical features efficiently without extensive preprocessing. It builds an ensemble of decision trees sequentially, optimizing a loss function while reducing overfitting through ordered boosting and permutation-based techniques. CatBoost's structure uses a symmetric tree architecture, ensuring faster training and prediction with reduced memory consumption. It automatically handles categorical data, eliminating the need for one-hot encoding or label encoding, and is robust against overfitting. With its efficient processing of large datasets and superior performance on imbalanced data, CatBoost excels in both classification and regression tasks, making it ideal for real-world applications [14].
11. **Support Vector Machine (SVM)** is a supervised learning algorithm designed for classification and regression tasks. Its structure finds the optimal hyperplane that maximizes the margin between classes in a feature space, ensuring the best separation. For non-linear data, SVM uses kernel functions to map the data into higher dimensions, enabling the algorithm to create complex decision boundaries. SVM is effective in high-dimensional spaces and handles both linear and non-linear problems. It is robust to overfitting, especially in scenarios with a clear margin of separation, and supports regularization to balance complexity and accuracy. With proper tuning, SVM delivers high performance, particularly for small to medium-sized datasets [15].
12. **Stacking Classifier (STC)** is an ensemble learning technique that combines the predictions of multiple base classifiers to improve overall model performance. Its structure involves training various base models (e.g., decision trees, SVMs, or neural networks) on the same dataset, followed by a meta-model (often a simple model like logistic regression) that learns to combine their predictions. The meta-model uses the output of the base models as input features, effectively leveraging their strengths while mitigating individual weaknesses. STC is highly flexible, allowing for diverse base models and customization of the meta-model. It works well in complex datasets, providing robust and often superior predictive performance compared to single models [15].
13. **Multilayer Perceptron (MLP)** is a type of feedforward artificial neural network designed for supervised learning tasks like classification and regression. Its structure consists of an input layer, one or more hidden layers, and an output layer, with each layer comprising multiple neurons. Neurons are connected through weights and biases, and they use activation functions (e.g., ReLU or sigmoid) to introduce non-linearity, enabling MLP to model complex relationships. MLP learns by adjusting weights during backpropagation, guided by a loss function and an optimization algorithm like gradient descent. It is highly flexible and can approximate any continuous function given sufficient neurons and layers. MLP performs well on structured data but may require careful tuning of hyperparameters and regularization to prevent overfitting [16].
14. **RESULTS**

**Supplemental Table S1.** Comprehensive description of RFs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Feature Name** | **Abbiriviation** | Meaning | **Formula** |
| **First Order Features (FO)** | Energy (E) | FO\_E | A measure of the magnitude of voxel values in an image. A larger values implies a greater sum of the squares of these values. | A mathematical equation with numbers and symbols  Description automatically generated |
| Total Energy (TE) | FO\_TE | The value of Energy feature scaled by the volume of the voxel in cubic mm. |  |
| Entropy (En) | FO\_En | Specifies the uncertainty/randomness in the image values. It measures the average amount of information required to encode the image values. |  |
| Minimum (Min) | FO\_Min | The lowest pixel intensity value within the ROI. |  |
| The 10th Percentile (10P) | FO\_10P | The 10th percentile of X. |  |
| The 90th Percentile (90P) | FO\_90P | The 90th percentile of X. |  |
| Maximum Intensity (MaxI) | FO\_MaxI | The maximum gray level intensity within the ROI. |  |
| Mean Intensity (MI) | FO\_MI | The average gray level intensity within the ROI. |  |
| Median Intensity (MedI) | FO\_MedI | The median gray level intensity within the ROI. |  |
| Interquartile Range (IQR) | FO\_IQR | Here P25 and P75 are the 25th and 75th percentile of the image array, respectively. |  |
| Range (R) | FO\_R | The range of gray values in the ROI. |  |
| Mean Absolute Deviation (MAD) | FO\_MAD | The mean distance of all intensity values from the Mean Value of the image array. | A number and mathematical symbols  Description automatically generated with medium confidence |
| Robust Mean Absolute Deviation (rMAD) | FO\_rMAD | The mean distance of all intensity values from the Mean Value calculated on the subset of image array with gray levels in between, or equal to the 10th and 90th percentile. |  |
| Root Mean Square (RMS) | FO\_RMS | The square-root of the mean of all the squared intensity values.It is another measure of the magnitude of the image values. This feature is volume-confounded, a larger value of cincreases the effect of volume-confounding. | A math equations and symbols  Description automatically generated with medium confidence |
| Standard Deviation (SD) | FO\_SD | Measures the amount of variation or dispersion from the Mean Value. |  |
| Skewness (Sk) | FO\_Sk | Measures the asymmetry of the distribution of values about the Mean value. Depending on where the tail is elongated and the mass of the distribution is concentrated, this value can be positive or negative. |  |
| Kurtosis (Ku) | FO\_Ku | A measure of the ‘peakedness’ of the distribution of values in the image ROI.A higher kurtosis implies that the mass of the distribution is concentrated towards the tail(s) rather than towards the mean. A lower kurtosis implies the reverse: that the mass of the distribution is concentrated towards a spike near the Mean value. |  |
| Variance (V) | FO\_V | The the mean of the squared distances of each intensity value from the Mean value. This is a measure of the spread of the distribution about the mean. |  |
| Uniformity (Un) | FO\_Un | A measure of the sum of the squares of each intensity value. This is a measure of the homogeneity of the image array, where a greater uniformity implies a greater homogeneity or a smaller range of discrete intensity values. | A black and white math equation  Description automatically generated |
| **3D Shape Features (SF)** | Mesh Volume (Mv) | SF\_MV\_3D | The volume of the ROI VV is calculated from the triangle mesh of the ROI. |  |
| Voxel Volume (VV) | SF\_VV\_3D | The volume of the ROI Vvoxel is approximated by multiplying the number of voxels in the ROI by the volume of a single voxel Vk. | A mathematical equation with numbers and symbols  Description automatically generated |
| Surface Area (SA) | SF\_SA\_3D | To calculate the surface area, first the surface area Ai of each triangle in the mesh is calculated . The total surface area is then obtained by taking the sum of all calculated sub-areas. |  |
| Surface Area to Volume ratio (SAVR) | SF\_SAVR\_3D | Here, a lower value indicates a more compact (sphere-like) shape. This feature is not dimensionless, and is therefore (partly) dependent on the volume of the ROI. | A black and white text  Description automatically generated |
| Sphericity (Sp) | SF\_Sp\_3D | A measure of the roundness of the shape of the tumor region relative to a sphere.It is a dimensionless measure, independent of scale and orientation. |  |
| Compactness 1 (Com1) | SF\_Com1\_3D | A measure of how compact the shape of the tumor is relative to a sphere (most compact). It is therefore correlated to Sphericity and redundant. It is provided here for completeness. |  |
| Compactness 2 (Com2) | SF\_Com2\_3D | A measure of how compact the shape of the tumor is relative to a sphere (most compact). It is a dimensionless measure, independent of scale and orientation. | A black text on a white background  Description automatically generated |
| Spherical Disproportion (SpD) | SF\_SpD\_3D | The ratio of the surface area of the tumor region to the surface area of a sphere with the same volume as the tumor region, and by definition, the inverse of Sphericity. |  |
| Maximum 3D Diameter (Max3DD) | SF\_Max3DD | Defined as the largest pairwise Euclidean distance between tumor surface mesh vertices, Also known as Feret Diameter. |  |
| Maximum 2D diameter (Slice) | SF\_Max2DD (Slice) | The largest pairwise Euclidean distance between tumor surface mesh vertices in the row-column (generally the axial) plane. |  |
| Maximum 2D diameter (Column) | SF\_Max2DD (Column) | The largest pairwise Euclidean distance between tumor surface mesh vertices in the row-slice (usually the coronal) plane. |  |
| Maximum 2D diameter (Row) | SF\_Max2DD (Row) | the largest pairwise Euclidean distance between tumor surface mesh vertices in the column-slice (usually the sagittal) plane. |  |
| Major Axis Length (MAL) | SF\_MAL\_3D | This feature yield the largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λmajor. | A math equation with numbers  Description automatically generated |
| Minor Axis Length (MiAL) | SF\_MiAL\_3D | This feature yield the second-largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λminor. | A black and blue text  Description automatically generated with medium confidence |
| Least Axis Length (LAL) | SF\_LAL\_3D | This feature yield the smallest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λleast. In case of a 2D segmentation, this value will be 0. |  |
| Elongation (E) | SF\_E\_3D | The relationship between the two largest principal components in the ROI shape. For computational reasons, this feature is defined as the inverse of true elongation. |  |
| Flatness (F) | SF\_F\_3D | The relationship between the largest and smallest principal components in the ROI shape. For computational reasons, this feature is defined as the inverse of true flatness. |  |
| **2D Shape Features(SF)** | Mesh Surface (MS) | SF\_MS\_2D | To calculate the surface area, first the signed surface area Ai of each triangle in the mesh is calculated . The total surface area is then obtained by taking the sum of all calculated sub-areas , where the sign will ensure correct surface area, as the negative area of triangles outside the ROI will cancel out the surplus area included by triangles partly inside and partly outside the ROI. | A math equations and symbols  Description automatically generated |
| Pixel Surface (PS) | SF\_PS\_2D | The surface area of the ROI Apixel is approximated by multiplying the number of pixels in the ROI by the surface area of a single pixel Ak. This is a less precise approximation of the surface area. This feature does not make use of the mesh and is not used in calculation of other 2D shape features. |  |
| Perimeter (P) | SF\_P\_2D | To calculate the perimeter, first the perimeter Ai of each line in the mesh circumference is calculated. The total perimeter is then obtained by taking the sum of all calculated sub-areas. |  |
| Perimeter to Surface ratio (PSR) | SF\_PSR\_2D | Here, a lower value indicates a more compact (circle-like) shape. This feature is not dimensionless, and is therefore (partly) dependent on the surface area of the ROI. | A close-up of a text  Description automatically generated |
| Sphericity (Sp) | SF\_Sp\_2D | Sphericity is the ratio of the perimeter of the tumor region to the perimeter of a circle with the same surface area as the tumor region and therefore a measure of the roundness of the shape of the tumor region relative to a circle. It is a dimensionless measure, independent of scale and orientation. | A mathematical equation with black text  Description automatically generated |
| Spherical Disproportion (SpD) | SF\_SpD\_2D | The ratio of the perimeter of the tumor region to the perimeter of a circle with the same surface area as the tumor region, and by definition, the inverse of Sphericity. | A black text on a white background  Description automatically generated |
| Maximum 2D diameter (Max2DD) | SF\_Max2DD | The largest pairwise Euclidean distance between tumor surface mesh vertices. |  |
| Major Axis Length (MAL) | SF\_MAL\_2D | This feature yield the largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λmajor. |  |
| Minor Axis Length (MiAL) | SF\_MiAL\_2D | This feature yield the second-largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λminor. |  |
| Elongation (E) | SF\_E\_2D | The relationship between the two largest principal components in the ROI shape. For computational reasons, this feature is defined as the inverse of true elongation. |  |
| **Gray Level Co-occurrence Matrix (GLCM) Features** | Autocorrelation (AC) | GLCM\_AC | Returns the mean gray level intensity of the i distribution. | A mathematical equation with numbers and symbols  Description automatically generated |
| Joint Average (JA) | GLCM\_JA | Returns the mean gray level intensity of the i distribution. | A math equation with black lines  Description automatically generated with medium confidence |
| Cluster Prominence (CP) | GLCM\_CP | A measure of the skewness and asymmetry of the GLCM. A higher values implies more asymmetry about the mean while a lower value indicates a peak near the mean value and less variation about the mean. | A math symbols on a white background  Description automatically generated |
| Cluster Shade (CS) | GLCM\_CS | A measure of the skewness and uniformity of the GLCM. A higher cluster shade implies greater asymmetry about the mean. | A mathematical equation with numbers and symbols  Description automatically generated |
| Cluster Tendency (CT) | GLCM\_CT | A measure of groupings of voxels with similar gray level values. |  |
| Contrast (Co) | GLCM\_Co | A measure of the local intensity variation, favoring values away from the diagonal (i=j)(i=j). A larger value correlates with a greater disparity in intensity values among neighboring voxels. |  |
| Correlation (Corr) | GLCM\_Corr | A value between 0 (uncorrelated) and 1 (perfectly correlated) showing the linear dependency of gray-level values to their respective voxels in the GLCM. | A black and white math symbols  Description automatically generated with medium confidence |
| Difference Average (DA) | GLCM\_DA | Measures the relationship between occurrences of pairs with similar intensity values and occurrences of pairs with differing intensity values. |  |
| Difference Entropy (DiEn) | GLCM\_DiEn | A measure of the randomness/variability in neighborhood intensity value differences. | A mathematical equation with numbers and symbols  Description automatically generated |
| Joint Energy (JE) | GLCM\_JE | A measure of homogeneous patterns in the image. A greater Energy implies that there are more instances of intensity value pairs in the image that neighbor each other at higher frequencies. | A mathematical equation with numbers and symbols  Description automatically generated |
| Joint Entropy (JEn) | GLCM\_JEn | A measure of the randomness/variability in neighborhood intensity values. | A number of mathematical symbols  Description automatically generated with medium confidence |
| Informational Measure of Correlation (IMC1) | GLCM\_IMC1 | Assesses the correlation between the probability distributions of ii and jj (quantifying the complexity of the texture), using mutual information I(x, y): | A black text with a white background  Description automatically generated |
| Informational Measure of Correlation (IMC2) | GLCM\_IMC2 | Also assesses the correlation between the probability distributions of ii and jj (quantifying the complexity of the texture). | A black and white image of a mathematical equation  Description automatically generated |
| Inverse Difference Moment (IDM) | GLCM\_IDM | A measure of the local homogeneity of an image. IDM weights are the inverse of the Contrast weights (decreasing exponentially from the diagonal i=j in the GLCM). | A number and symbols of mathematical equations  Description automatically generated with medium confidence |
| Maximal Correlation Coefficient (MCC) | GLCM\_MCC | A measure of complexity of the texture and 0≤MCC≤10≤MCC≤1. | A math equation with numbers and symbols  Description automatically generated |
| Inverse Difference Moment Normalized (IDMN) | GLCM\_IDMN | A measure of the local homogeneity of an image. | A mathematical equation with numbers and symbols  Description automatically generated |
| Inverse Difference (ID) | GLCM\_ID | Another measure of the local homogeneity of an image. With more uniform gray levels, the denominator will remain low, resulting in a higher overall value. | A mathematical equation with numbers and symbols  Description automatically generated |
| Inverse Difference Normalized (IDN) | GLCM\_IDN | Another measure of the local homogeneity of an image | A mathematical symbols and equations  Description automatically generated with medium confidence |
| Inverse Variance (IV) | GLCM\_IV | Inverse variance. | A black and white text  Description automatically generated |
| Maximum Probability (MP) | GLCM\_MP | Occurrences of the most predominant pair of neighboring intensity values. |  |
| Sum Average (SA) | GLCM\_Sav | Measures the relationship between occurrences of pairs with lower intensity values and occurrences of pairs with higher intensity values. | A mathematical equation with numbers and symbols  Description automatically generated |
| Sum Entropy (SEn) | GLCM\_SEn | A sum of neighborhood intensity value differences. | A math formula with numbers and symbols  Description automatically generated with medium confidence |
| Sum of Squares (SQ) | GLCM\_SQ | A measure in the distribution of neigboring intensity level pairs about the mean intensity level in the GLCM. | A number of letters and numbers  Description automatically generated with medium confidence |
| **Gray Level Size Zone Matrix (GLSZM) Features** | Small Area Emphasis (SAE) | GLSZM\_SAE | A measure of the distribution of small size zones, with a greater value indicative of more smaller size zones and more fine textures. | A number of mathematical symbols  Description automatically generated with medium confidence |
| Large Area Emphasis (LAE) | GLSZM\_LAE | A measure of the distribution of large area size zones, with a greater value indicative of more larger size zones and more coarse textures. | A number of mathematical formulas  Description automatically generated with medium confidence |
| Gray Level Non-Uniformity (GLN) | GLSZM\_GLN | Measures the variability of gray-level intensity values in the image, with a lower value indicating more homogeneity in intensity values. | A math equations and symbols  Description automatically generated with medium confidence |
| Gray Level Non-Uniformity Normalized (GLNN) | GLSZM\_GLNN | Measures the variability of gray-level intensity values in the image, with a lower value indicating a greater similarity in intensity values. This is the normalized version of the GLN formula. | A black and white math equations  Description automatically generated with medium confidence |
| Size-Zone Non-Uniformity (SZN) | GLSZM\_SZN | Measures the variability of size zone volumes in the image, with a lower value indicating more homogeneity in size zone volumes. | A number of mathematical equations  Description automatically generated with medium confidence |
| Size-Zone Non-Uniformity Normalized (SZNN) | GLSZM\_SZNN | Measures the variability of size zone volumes throughout the image, with a lower value indicating more homogeneity among zone size volumes in the image. This is the normalized version of the SZN formula. | A number of mathematical equations  Description automatically generated with medium confidence |
| Zone Percentage (ZP) | GLSZM\_ ZP | Measures the coarseness of the texture by taking the ratio of number of zones and number of voxels in the ROI. | A mathematical equation with black letters  Description automatically generated |
| Gray Level Variance (GLV) | GLSZM\_GLV | Measures the variance in gray level intensities for the zones. | A math equations and symbols  Description automatically generated with medium confidence |
| Zone Variance (ZV) | GLSZM\_ZV | Measures the variance in zone size volumes for the zones. | A math equations and formulas  Description automatically generated with medium confidence |
| Zone Entropy (ZE) | GLSZM\_ZE | Measures the uncertainty/randomness in the distribution of zone sizes and gray levels. A higher value indicates more heterogeneneity in the texture patterns. | A close up of a number  Description automatically generated |
| Low Gray Level Zone Emphasis (LGLZE) | GLSZM\_LGLZE | Measures the distribution of lower gray-level size zones, with a higher value indicating a greater proportion of lower gray-level values and size zones in the image. | A black and white math symbol  Description automatically generated |
| High Gray Level Zone Emphasis (HGLZE) | GLSZM\_HGLZE | Measures the distribution of the higher gray-level values, with a higher value indicating a greater proportion of higher gray-level values and size zones in the image. | A number of mathematical equations  Description automatically generated with medium confidence |
| Small Area Low Gray Level Emphasis (SALGLE) | GLSZM\_SALGLE | Measures the proportion in the image of the joint distribution of smaller size zones with lower gray-level values. | A mathematical equation with numbers and lines  Description automatically generated |
| Small Area High Gray Level Emphasis (SAHGLE) | GLSZM\_SAHGLE | Measures the proportion in the image of the joint distribution of smaller size zones with higher gray-level values. | A black and white math equation  Description automatically generated with medium confidence |
| Large Area Low Gray Level Emphasis (LALGLE) | GLSZM\_LALGLE | Measures the proportion in the image of the joint distribution of larger size zones with lower gray-level values. | A math equation with numbers and symbols  Description automatically generated |
| Large Area High Gray Level Emphasis (LAHGLE) | GLSZM\_LAHGLE | Measures the proportion in the image of the joint distribution of larger size zones with higher gray-level values. | A number of mathematical equations  Description automatically generated with medium confidence |
| **Gray Level Run Length Matrix (GLRLM) Features** | Short Run Emphasis (SRE) | GLRLM\_SRE | A measure of the distribution of short run lengths, with a greater value indicative of shorter run lengths and more fine textural textures. | A black and white text  Description automatically generated with medium confidence |
| Long Run Emphasis (LRE) | GLRLM\_LRE | A measure of the distribution of long run lengths, with a greater value indicative of longer run lengths and more coarse structural textures. | A number of mathematical equations  Description automatically generated with medium confidence |
| Gray Level Non-Uniformity (GLN) | GLRLM\_GLN | Measures the similarity of gray-level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values. | A number of mathematical equations  Description automatically generated with medium confidence |
| Gray Level Non-Uniformity Normalized (GLNN) | GLRLM\_GLNN | Measures the similarity of gray-level intensity values in the image, where a lower GLNN value correlates with a greater similarity in intensity values. This is the normalized version of the GLN formula. | A close-up of mathematical equations  Description automatically generated |
| Run Length Non-Uniformity (RLN) | GLRLM\_RLN | Measures the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image. | A number of mathematical equations  Description automatically generated with medium confidence |
| Run Length Non-Uniformity Normalized (RLNN) | GLRLM\_RLNN | Measures the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image. This is the normalized version of the RLN formula. | A group of mathematical equations  Description automatically generated |
| Run Percentage (RP) | GLRLM\_RP | Measures the coarseness of the texture by taking the ratio of number of runs and number of voxels in the ROI. | A mathematical equation with numbers and symbols  Description automatically generated |
| Gray Level Variance (GLV) | GLRLM\_GLV | Measures the variance in gray level intensity for the runs. | A math formula with black text  Description automatically generated with medium confidence |
| Run Variance (RV) | GLRLM\_RV | A measure of the variance in runs for the run lengths. | A math equations and symbols  Description automatically generated with medium confidence |
| Run Entropy (REn) | GLRLM\_REn | Measures the uncertainty/randomness in the distribution of run lengths and gray levels. A higher value indicates more heterogeneity in the texture patterns. | A close-up of a number  Description automatically generated |
| Low Gray Level Run Emphasis (LGLRE) | GLRLM\_LGLRE | Measures the distribution of low gray-level values, with a higher value indicating a greater concentration of low gray-level values in the image. | A black and white math equation  Description automatically generated with medium confidence |
| High Gray Level Run Emphasis (HGLRE) | GLRLM\_HGLRE | Measures the distribution of the higher gray-level values, with a higher value indicating a greater concentration of high gray-level values in the image. | A number of mathematical equations  Description automatically generated with medium confidence |
| Short Run Low Gray Level Emphasis (SRLGLE) | GLRLM\_SRLGLE | Measures the joint distribution of shorter run lengths with lower gray-level values. | A mathematical equation with numbers and symbols  Description automatically generated |
| Short Run High Gray Level Emphasis (SRHGLE) | GLRLM\_SRHGLE | Measures the joint distribution of shorter run lengths with higher gray-level values. | A math equation with numbers and symbols  Description automatically generated |
| Long Run Low Gray Level Emphasis (LRLGLE) | GLRLM\_LRLGLE | Measures the joint distribution of long run lengths with lower gray-level values. | A mathematical equation with numbers and symbols  Description automatically generated |
| Long Run High Gray Level Emphasis (LRHGLE) | GLRLM\_LRHGLE | Measures the joint distribution of long run lengths with higher gray-level values. | A black and white math symbol  Description automatically generated |
| **Neighbouring Gray Tone Difference Matrix (NGTDM) Features** | Coarseness (Coar) | NGTDM\_Coar | A measure of average difference between the center voxel and its neighbourhood and is an indication of the spatial rate of change. A higher value indicates a lower spatial change rate and a locally more uniform texture. | A black and white text  Description automatically generated |
| Contrast (Con) | NGTDM\_Con | A measure of the spatial intensity change, but is also dependent on the overall gray level dynamic range. Contrast is high when both the dynamic range and the spatial change rate are high, i.e. an image with a large range of gray levels, with large changes between voxels and their neighbourhood. |  |
| Busyness (B) | NGTDM\_B | A measure of the change from a pixel to its neighbour. A high value for busyness indicates a ‘busy’ image, with rapid changes of intensity between pixels and its neighbourhood. |  |
| Complexity (Comp) | NGTDM\_Comp | An image is considered complex when there are many primitive components in the image, i.e. the image is non-uniform and there are many rapid changes in gray level intensity. |  |
| Strength (S) | NGTDM\_S | A measure of the primitives in an image. Its value is high when the primitives are easily defined and visible, i.e. an image with slow change in intensity but more large coarse differences in gray level intensities. | A math equation with a white background  Description automatically generated with medium confidence |
| **Gray Level Dependence Matrix (GLDM) Features** | Small Dependence Emphasis (SDE) | GLDM\_SDE | A measure of the distribution of small dependencies, with a greater value indicative of smaller dependence and less homogeneous textures. | A math equation with numbers and symbols  Description automatically generated with medium confidence |
| Large Dependence Emphasis (LDE) | GLDM\_LDE | A measure of the distribution of large dependencies, with a greater value indicative of larger dependence and more homogeneous textures. | A math equations and formulas  Description automatically generated with medium confidence |
| Gray Level Non-Uniformity (GLN) | GLDM\_GLN | Measures the similarity of gray-level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values. | A black and white math symbols  Description automatically generated with medium confidence |
| Dependence Non-Uniformity (DN) | GLDM\_DN | Measures the similarity of dependence throughout the image, with a lower value indicating more homogeneity among dependencies in the image. | A math equations and symbols  Description automatically generated with medium confidence |
| Dependence Non-Uniformity Normalized (DNN) | GLDM\_DNN | Measures the similarity of dependence throughout the image, with a lower value indicating more homogeneity among dependencies in the image. This is the normalized version of the DLN formula. | A number of mathematical equations  Description automatically generated with medium confidence |
| Gray Level Variance (GLV) | GLDM\_GLV | Measures the variance in grey level in the image. | A math equation with numbers  Description automatically generated |
| Dependence Variance (DV) | GLDM\_DV | Measures the variance in dependence size in the image. | A math equation with black text  Description automatically generated with medium confidence |
| Dependence Entropy (DeEn) | GLDM\_DeEn | Dependence entropy | A mathematical equation with numbers and symbols  Description automatically generated |
| Low Gray Level Emphasis (LGLE) | GLDM\_LGLE | Measures the distribution of low gray-level values, with a higher value indicating a greater concentration of low gray-level values in the image. | A math formula with numbers and symbols  Description automatically generated with medium confidence |
| High Gray Level Emphasis (HGLE) | GLDM\_HGLE | Measures the distribution of the higher gray-level values, with a higher value indicating a greater concentration of high gray-level values in the image. | A number of mathematical equations  Description automatically generated with medium confidence |
| Small Dependence Low Gray Level Emphasis (SDLGLE) | GLDM\_SDLGLE | Measures the joint distribution of small dependence with lower gray-level values. | A mathematical equation with numbers and symbols  Description automatically generated |
| Small Dependence High Gray Level Emphasis (SDHGLE) | GLDM\_SDHGLE | Measures the joint distribution of small dependence with higher gray-level values. |  |
| Large Dependence Low Gray Level Emphasis (LDLGLE) | GLDM\_LDLGLE | Measures the joint distribution of large dependence with lower gray-level values. | A black and white math equation  Description automatically generated with medium confidence |
| Large Dependence High Gray Level Emphasis (LDHGLE) | GLDM\_LDHGLE | Measures the joint distribution of large dependence with higher gray-level values. | A black and white math symbol  Description automatically generated |

**Supplemental Table S2.** Comprehensive description of PI-RADS element

|  |  |  |
| --- | --- | --- |
| **Tumour Zone** | **PI-RADS** | **Definition of Sequence in This Column** |
|
|  | PI-RADS 1=T2WI score 1 | T2WI score 1 = homogeneous, intermediate signal intensity, or a round, completely encapsulated  nodule (“typical nodule”) |
| PI-RADS 2 = T2WI score = 2 and DWI score =1-3 | T2WI score 2 = A mostly encapsulated nodule or a  homogeneous circumscribed nodule  without encapsulation  (“atypical nodule”) or a  homogeneous mildly hypointense area  between nodules |
| PI-RADS 3 = T2WI score = 3 and DWI score =1-4 or T2WI score =2 and DWI score =4-5 |
| DWI score = 1 : No abnormality (i.e. normal) on ADC and high b-value DWI |
| DWI score 2 = Linear/wedge shaped hypointense on ADC  and/or linear/wedge shaped hyperintense  on high b-value DWI Non-focal hypointense on ADC and/or  hyperintense on high b-value DWI |
| DWI score 3 = Focal hypointense on ADC  and/or focal hyperintense on high b-value  DWI; may be markedly hypointense on  ADC or markedly hyperintense on high b-value DWI, but not both |
| T2WI score 3 = Heterogeneous signal intensity with  obscured margins. Includes  others that do not qualify as 2, 4, or 5 |
| DWI score 1-3 has been explained above. |
| DWI score 4 = Focal, markedly hypointense on ADC and  markedly hyperintense on high b-value  DWI; <1.5cm in greatest dimension |
| DWI score 5 = Same as 4 but ≥1.5cm in greatest  dimension or definite extraprostatic  extension/invasive behavior |
| T2WI score 2 has been explained in PI-RADS 2. |
| PI-RADS 4 = (T2WI score 4 ) or ( T2WI score 3 and DWI score 5 ) | T2WI score 4 = Lenticular or non-circumscribed, homogeneous,  moderately hypointense, and <1.5 cm  in greatest dimension |
| T2WI score 3 and DWI score 5 has been explained in PI-RADS 3. |
| PI-RADS 5 | T2WI score 5 = Same as 4, but ≥ 1.5cm in greatest  dimension or definite  extraprostatic extension/invasive  behavior |
| **Tumour Zone** | **PI-RADS OF COLUMN** | **DEFINITION OF SEQUENCE IN THIS COLUMN** |
| **Peripheral Zone** | PI-RADS 1= DWI score 1 | DWI score 1 = No abnormality (i.e. normal) on ADC and high b-value DWI |
| PI-RADS 2 = DWI score 2 | DWI score 2 = Linear/wedge shaped hypointense on  ADC and/or linear/wedge shaped  hyperintense on high b-value DWI |
| PI-RADS 3 = DWI score 3 and DCE negative | DWI score 3 = Focal hypointense on ADC and/or  focal hyperintense on high b-value DWI;  may be markedly hypointense on ADC or  markedly hyperintense on high b-value  DWI, but not both. |
| DCE negative = No early or contemporaneous enhancement; or diffuse multifocal enhancement not corresponding to a focal finding  on T2W and/or DWI or focal enhancement corresponding to a lesion demonstrating features of  BPH on T2WI (including features of extruded BPH in the PZ) |
| PI-RADS 4 = (DWI score 4) or (DWI score 3 and DCE positive) | DWI score 4 is = Focal markedly hypointense on ADC and  markedly hyperintense on high b-value  DWI; <1.5cm in greatest dimension |
| DCE positive = Focal/  earlier than or contemporaneously with enhancement of adjacent normal  prostatic tissues; corresponds to suspicious finding on T2W and/or DWI |
| DWI score 3 has been explained in PI-RADS 3. |
| PI-RADS 5 = DWI score 5 | DWI score 5 = Same as 4 but ≥1.5cm in greatest  dimension or definite extraprostatic  extension/invasive behavior |

**Supplemental Table S3.** Comprehensive description of semantic features

|  |  |
| --- | --- |
| **Visual Semantic** | **Meaning** |
| Heterogeneous & Homogenous | Heterogeneous refers to a structure with dissimilar components or elements, appearing irregular or variegated. It is the antonym for homogeneous, meaning a structure with similar components. |
| Uniformity | It is the synonym for homogeneous, meaning a structure with similar components. |
| Hypo/Hyper/Intermediate signal intensity | Signal intensity in a lesion should be visually compared to the average signal of “normal” prostate tissue in the histologic zone in which it is located. |
| Focal | Discrete and different from the background |
| Markedly | Defined as a more pronounced signal change  than any other focus in the same zone. |
| Round | The shape of a circle or sphere |
| Lenticular | Having the shape of a double-convex lens, crescentic |
| Linear | In a line or band-like shape |
| Indistinct margin | Blurred margin |
| Wedge-shaped | Having the shape of a wedge, pie, or V-shaped (pi-rads document) |
| Encapsulated | Bounded by a distinct, uniform, smooth low-signal line (BPHnodule);  completely encapsulated nodule is entirely surrounded by a smooth  low-signal line in at least two imaging planes (“typical nodule”); almost  completely or incompletely encapsulated nodule is not entirely  surrounded by a smooth low-signal line (“atypical nodule”) |
| Circumscribed | Well-defined |
| Extra-prostatic Extension | definite extra-prostatic extension/invasive behavior |

**Supplemental Table S4.** Comprehensive radiological/biological dictionary of Radiomics Features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PI-RADS score** | **Semantic Features \_Based on PIRADS** | **Radiomics** | **Meaning of Radiomics** | **Relationship** |
| **PI-RADS 1** | ROUND Shape | Sphericity (3D) from the S Fcategory (SF\_Sp\_3D) | A measure of the roundness of the shape of the tumor region relative to a sphere. It is a dimensionless measure, independent of scale and orientation. The value range is 0<sphericity≤1 , where a value of 1 indicates a perfect sphere (a sphere has the smallest possible surface area for a given volume, compared to other solids). | Based on the meaning of the feature. |
| ROUND Shape | Compactness 1 (3D) from the SF category (SF\_Com1\_3D) | Similar to Sphericity. | Based on the meaning of the feature. |
| ROUND Shape | Compactness 2 (3D) from the SF category (SF\_Com2\_3D) | Similar to Sphericity and Compactness 1. | Based on the meaning of the feature. |
| ROUND Shape | Spherical Disproportion (3D) from the SF category (SF\_Sp\_3D) | It is the ratio of the surface area of the tumor region to the surface area of a sphere with the same volume as the tumor region, and by definition, the inverse of Sphericity. Therefore, the value range is spherical disproportion≥1, with a value of 1 indicating a perfect sphere. | Based on the meaning of the feature. |
| ROUND Shape | Surface Area to Volume ratio (3D) from the SF category (SF\_SAVR\_3D) | Here, a lower value indicates a more compact (sphere-like) shape. This feature is not dimensionless, and is therefore (partly) dependent on the volume of the ROI. | Surface area to volume ratio for particular shape. |
| ROUND Shape | Flatness (3D) from the SF category (SF\_F\_3D) | Shows the relationship between the largest and smallest principal components in the ROI shape. For computational reasons, this feature is defined as the inverse of true flatness. | In a round shape, the largest and smallest principal components are the same so flateness is equal to 1 or close to 1. |
| Margine (Circumbscribe & Encapsulated Border) | Gray Level Variance from the GLSZM category (GLSZM\_GLV) | Gray Level Variance (GLV) measures the variance in gray level intensities for the zones. | Captures the variability within the lesion, contrasting with the uniformity of the boundary, which can help delineate well-defined lesions. |
| Capsule | Long Run Low Gray Level Emphasis from the GLRLM category (GLRLM\_LRLGLE) | Measures the joint distribution of long run lengths with higher gray level values. | If a lesion has a well-defined capsule, it might also appear hypo-intensive compared to the surrounding tissue. LRLGLE would capture this by measuring the consistent low intensity runs within the lesion. |
| Uniformity | Range from the FO category (FO\_R) | The range of gray values in the ROI. | The very low range mean more similarity in intensities, leads to more homogeneity. |
| Uniformity | Coarseness from the NGTDM category (NGTDM\_Coar) | A measure of average difference between the center voxel and its neighbourhood and is an indication of the spatial rate of change. A higher value indicates a lower spatial change rate and a locally more uniform texture. | A higher value suggests a lower rate of spatial change and a more uniform texture within the region of interest (ROI). This means that areas with high coarseness have fewer abrupt intensity changes, reflecting a smoother and more homogeneous appearance. |
| Uniformity | Gray Level Non-Uniformity from the GLDM category (GLDM\_GLN) | Measures the similarity of gray level intensity values in the image, where a lower GLN value correlates with a greater similarity in intensity values. | Lower value correlates with greater similarity in intensity values, leading to higher homogeneity. |
| Uniformity | Cluster Tendency from the GLCM category (GLCM\_CT) | A measure of groupings of voxels with similar gray level values. | CT is related to uniformity in the sense that it measures the tendency of voxels in an image to form groupings based on similar gray level values. A higher Cluster Tendency value indicates a stronger tendency for clusters of voxels with similar gray level values to form within the image. This can imply that there are regions within the image where the gray level values are more uniform or consistent, leading to the formation of clustered areas with similar texture patterns.  Therefore, while Cluster Tendency does not directly measure uniformity, it is related to the concept of uniformity in terms of how it reflects the tendency of voxels to cluster together based on their gray level values, suggesting regions of the image where the values are more consistent or uniform. |
| Uniformity | Correlation from the GLCM category (GLCM\_Corr) | It is a value between 0 (uncorrelated) and 1 (perfectly correlated) showing the linear dependency of gray level values to their respective voxels. | In a highly uniform texture, pixel values are more predictable and show consistent patterns, which can result in higher correlation values as the gray levels of neighboring pixels are closely related. On the other hand, in a less uniform (more heterogeneous) texture, the correlation values would typically be lower because the pixel values are more variable and less predictable. |
| Uniformity | Joint Energy from the GLCM category (GLCM\_JE) | A measure of homogeneous patterns in the image. A greater Energy implies that there are more instances of intensity value pairs in the image that neighbour each other at higher frequencies. | Higher values indicate a greater frequency of neighboring intensity pairs, reflecting homogeneity. |
| Uniformity | Inverse Difference Moment from the GLCM category (GLCM\_IDM) | A measure of the local homogeneity of an image. IDM weights are the inverse of the Contrast weights (decreasing exponentially from the diagonal i=j). | This feature assesses the similarity of pixel intensities within a region, a high IDM value means the region has low variability and high local similarity in pixel intensities and high homogeneity and uniformity. |
| Uniformity | Inverse Difference Normalized from the GLCM category (GLCM\_IDN) | Another measure of the local homogeneity of an image. Unlike Homogeneity1, IDN normalizes the difference between the neighbouring intensity values by dividing over the total number of discrete intensity values. | The formula for IDN emphasizes smaller gray level differences while normalizing by the number of gray levels, making it a useful feature for assessing the smoothness and uniformity of textures in medical images. |
| Uniformity | Sum of Squares from the GLCM category (GLCM\_SQ) | A measure in the distribution of neigbouring intensity level pairs about the mean intensity level. | Sum of Squares or Variance is a measure of the spread or dispersion of pixel intensity values from the mean. High values indicate more variability and less uniformity in the texture. |
| Uniformity | Complexity from the NGTDM category (NGTDM\_Com) | An image is considered complex when there are many primitive components in the image, i.e. the image is non-uniform and there are many rapid changes in gray level intensity. | The Complexity feature captures the degree of variation in the image. A higher Complexity value indicates more heterogeneity, while a lower value indicates more uniformity. |

# **References**

|  |  |
| --- | --- |
| [1] | P. Wibulpolprasert, S. Raman and et al, "Influence of the location and zone of tumor in prostate cancer detection and localization on 3-T multiparametric MRI based on PI-RADS version 2," *American Journal of Roentgenology,* pp. 1101-1111, 2020. |
| [2] | D. Theng and K. Bhoyar, "Feature selection techniques for machine learning: a survey of more than two decades of research," *Knowledge and Information Systems,* vol. 66, no. 3, pp. 1575-1637, 2024. |
| [3] | A. Haury, P. Gestraud and J. Vert, "The Influence of Feature Selection Methods on Accuracy, Stability and Interpretability of Molecular Signatures," *PloS one,* vol. 6, no. 12, p. e28210, 2011. |
| [4] | C. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering,* vol. 40, no. 1, pp. 16-28, 2014. |
| [5] | A. Jović, K. Brkić and N. Bogunović, "A review of feature selection methods with applications," in *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015. |
| [6] | A. Bommert, X. Sun and et al, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis,* vol. 143, p. 106839, 2020. |
| [7] | F. Emmert-Streib and M. Dehmer, "High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection," *Machine Learning and Knowledge Extraction,* vol. 1, no. 1, pp. 359-383, 2019. |
| [8] | W. Kanyongo and A. Ezugwu , "Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives," *Informatics in Medicine Unlocked,* vol. 38, p. 101232, 2023. |
| [9] | S. Kost, O. Rheinbach and H. Schaeben, "Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling," *Geochemistry,* vol. 81, no. 4, p. 125826, 2021. |
| [10] | P. Linardatos, V. Papastefanopoulos and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy,* vol. 23, no. 1, p. 18, 2020. |
| [11] | Y. Chen , D. Wang and et al, "Gaussian assumptions-free interpretable linear discriminant analysis for locating informative frequency bands for machine condition monitoring," *Mechanical Systems and Signal Processing,* vol. 199, p. 110492, 2023. |
| [12] | M. Christoph, Interpretable machine learning: A guide for making black box models explainable, Leanpub, 2020. |
| [13] | M. Salmanpour, M. Hosseinzadeh and et al, "Prediction of TNM stage in head and neck cancer using hybrid machine learning systems and radiomics features," vol. 12033, pp. 662-667, 2022. |
| [14] | A. Fathi Jouzdani, A. Abootorabi and et al, "Impact of Clinical Features Combined with PET/CT Imaging Features on Survival Prediction of Outcome in Lung Cancer," *Journal of Nuclear Medicine,* vol. 65, no. supplement 2, pp. 242130-242130, 2024. |
| [15] | M. Salmanpour, G. Hajianfar and et al, "Multitask outcome prediction using hybrid machine learning and PET-CT fusion radiomics," vol. 62, pp. 1424-1424, 2021. |
| [16] | M. Salmanpour, A. Saberi and et al, "Optimal feature selection and machine learning for prediction of outcome in Parkinson’s disease," vol. 61, pp. 524-524, 2020. |
| [17] | h. gggg, a. ffgg and a. tttt, "ffffffffffffffffffffffffff," *ffffffff,* p. 12, 2321. |
| [18] | R. ElShawi, Y. Sherif and et al, "Interpretability in healthcare: A comparative study of local machine learning interpretability techniques," *Computational Intelligence,* vol. 37, no. 4, pp. 1633-1650, 2021. |
| [19] | N. Aslam, I. Ullah Khan and et al, "Interpretable Machine Learning Models for Malicious Domains Detection Using Explainable Artificial Intelligence (XAI)," *Sustainability ,* vol. 14, no. 12, p. 7375, 2022. |