Interpretable Spiculation Quantification for Lung Cancer Screening

Wookjin Choi*, Saad Nadeem*, Sadegh Riyahi1, Joseph O Deasy1, Allen Tannenbaum2, and Wei Lu1

1 Department of Medical Physics, Memorial Sloan Kettering Cancer Center
2 Departments of Computer Science & Applied Mathematics, Stony Brook University

Abstract. Spiculations are spikes on the surface of pulmonary nodule and are important predictors of malignancy in lung cancer. In this work, we introduced an interpretable, parameter-free technique for quantifying this critical feature using the area distortion metric from the spherical conformal (angle-preserving) parameterization. The conformal factor in the spherical mapping formulation provides a direct measure of spiculation which can be used to detect spikes and compute spike heights for geometrically-complex spiculations. The use of the area distortion metric from conformal mapping has never been exploited before in this context. Based on the area distortion metric and the spiculation height, we introduced a novel spiculation score. A combination of our spiculation measures was found to be highly correlated (Spearman’s rank correlation coefficient $\rho = 0.48$) with the radiologist’s spiculation score. These measures were also used in the radiomics framework to achieve state-of-the-art malignancy prediction accuracy of 88.9% on a publicly available dataset.

1 Introduction

Lung cancer is the leading cause of cancer death in the United States [12]. The National Lung Cancer Screening Trial showed a clear survival benefit for screening with a low-dose computed tomography (CT) in current and former smokers. Recently radiomics studies [1, 2, 4], which extract a large number of quantitative features from medical images and subsequently perform data mining, have been proposed for various clinical applications. For instance, radiomics has been studied for the prediction of tumor responses and patient outcomes, resulting in more accurate prediction of local control and overall survival.

Lung cancer screening using radiomics has also been studied. Hawkins et al. [4] proposed a random forest classifier using 23 stable radiomic features. Buty et al. [1] developed a random forest classifier using 4096 appearance features extracted with a pre-trained deep neural network and 400 shape features extracted with spherical harmonics; spherical harmonics are a decomposition of frequency-space basis for representing functions defined over the sphere. The decomposition is applicable to describe overall shape of the object, but it cannot

* The first two authors contributed equally to this work.
provide local features for a given region on a shape (e.g. spiculation). However, the area distortion metric from the spherical conformal (angle-preserving) parameterization can accurately provide local spiculation features. Kumar et al. developed a deep neural network model using 5000 features. Liu et al. proposed a linear classifier based on 24 image traits visually scored by physicians. Choi et al. developed a model using support vector machine and least absolute shrinkage and selection operator (SVM-LASSO) to predict malignancy of pulmonary nodules (PNs) with only two CT radiomic features (shape and texture). Although these radiomics studies have improved prediction accuracy, there still remains a limitation with respect to the lack of clinical/biological interpretation of the features.

Radiographic edge characteristics of a PN, especially spiculation, influence the probability of malignancy. Typically, benign nodules have well-defined and smooth boundaries while malignant nodules have blurred and irregular boundaries. The Lung Imaging Reporting and Data System (Lung-RADS) was developed by the American College of Radiology (ACR) to standardize the screening of lung cancer on CT images using size, appearance type and calcification. For more accurate prediction, spiculation was suggested as an additional image finding that increases the suspicion of malignancy. The McWilliams model was introduced to compute the probability of lung cancer. This model uses nine variables, such as age, sex, family history of lung cancer, emphysema, size, type, location, count, and spiculation. PN size and spiculation were the major malignancy predictors in both models.

Spiculation quantification has been previously studied in a number of papers, but not in the context of malignancy prediction. Niehaus et al. developed a computer-aided diagnosis (CAD) system and investigated the size dependence of shape features for quantifying spiculation. Dhara et al. reported a differential geometry-based technique for quantifying spiculation using the binary mask of the segmented nodule.

In this work, we presented a comprehensive method to quantify spiculation using spherical parameterization and evaluate its importance in malignancy prediction. The contributions are as follows:

1. A novel interpretable spiculation feature is introduced, computed using the area distortion metric from spherical conformal parameterization;
2. Improved malignancy prediction using radiomics (shape and texture) + proposed spiculation quantification with accuracy 88.9%;
3. Achieved higher Spearman’s correlation, $\rho = 0.48$ between our spiculation measures and radiologist’s spiculation score than previous methods.

2 Method

In this section, we introduce some background for using area distortion metric and spherical conformal parameterization in spiculation quantification, followed by a novel spiculation score based on the quantified measures. Finally, we talk
Fig. 1. Radiologist’s spiculation score, $s_r$, for different PNs (top row) and our spiculation quantification pipeline (bottom row). (a) First non-trivial eigenfunction of the Laplace-Beltrami operator is computed for a given mesh. The zeroth-level set (red curve) of this eigenfunction is used to divide the mesh into two topological disks, which are conformally welded and stereographically projected to a sphere (b), which in angle-preserving spherical parameterization [10]. (c) The area distortion metric of the spherically parameterized surface is used to detect apex (red x’s), and compute heights (yellow curves) for each spike/spiculation.

2.1 Conformal mappings and area distortion

We first give a theoretical overview of the distortion of area in conformally mapping a genus zero Riemannian surface $S$ to the unit sphere $S^2$ to motivate the pipeline in spiculation; see [10] for the relevant mathematical references. By the *Theorema Egregium* of Gauss, one cannot find a diffeomorphism from $S$ with non-constant Gaussian curvature to $S^2$ which preserves both angles and area. Further, by a general result in complex analysis (uniformization), $S$ and $S^2$ are conformally equivalent, that is, there exists a diffeomorphism $\phi: S \to S^2$ that preserves angles. $\phi$ is unique up to Möbius transformation on $S^2$. By re-scaling, we may assume that the surface area of $S$ is $4\pi$. This is the *spherical parameterization* of a compact genus 0 surface for which we want to measure area distortion.

The first approach does not use the explicit mapping $\phi$. Namely, let $g_0$ be the Riemannian metric on $S$ with corresponding Gaussian curvature $K_0$. Let $K_u$ be the curvature on the conformally equivalent surface with metric $g_u = e^{2u}g_0$. Then it is well-known that $\Delta u + K_ue^{2u} = K_0$. This equation gives a specific measure of the distortion in area in any spherical parameterization procedure. Indeed, for the unit sphere $K_u = 1$, and thus $u$ satisfies the Poisson equation $\Delta u = K_0 - e^{2u}$. $u$ is called the *conformal distortion factor*, and $e^{2u}$ measures the distortion in

about spiculation classification and malignancy prediction in the radiomics analysis subsection.
area in going from the surface $S$ to the sphere $S^2$. If one examines the latter Poisson equation, one qualitatively sees that the more $K_0(x)$ varies, the greater the variation in $u$, and from the maximum principle, **spikes/spiculations may be identified by the greatest negative variation in area distortion**.

The second approach explicitly employs the conformal mapping $\phi : S \to S^2$, to give a measure of area distortion. (Via Gauss-Bonnet it is equivalent to the first approach.) We can quantify the change of area as a density function $\mu$ at each point of the sphere $S^2$, so that the integral on the unit sphere will give us the area measure of the original surface: $\int_{\phi^{-1}(U)} dy = \int_U \mu(x) dx$, $\forall U \subset S^2$ measurable. By change of variables, it is easy to see that this density function is the determinant of the Jacobian of $\phi^{-1}$, i.e., $\mu = \text{det}(\nabla \phi^{-1})$.

### 2.2 Spiculation quantification pipeline

In the paper [10], the above program is carried out in a discrete setting with respect to a triangulated surface $S = (V, E, F)$, where $V$ denotes the vertices, $E$ the edges, and $F$ the faces. Here, one may measure the area distortion on each triangle. Using this discrete version of spherical parameterization, the pipeline of spiculation detection (with height and width detection; see Fig. 1), is as follows:

1. Compute conformal (angle-preserving) spherical parameterization [10]: First non-trivial eigenfunction of the Laplace-Beltrami operator is computed for a given mesh (Fig. 1a). The zeroth-level set (red curve in Fig. 1a) of this eigenfunction is used to divide the mesh into two topological disks, which are conformally welded and stereographically projected to a sphere (Fig. 1b).
2. Compute the normalized area distortion. For each vertex $v_i$, the area distortion is defined as
   \[ \epsilon_i := \log \frac{\sum_{j,k} A([\phi(v_i), \phi(v_j), \phi(v_k)])}{\sum_{j,k} A([v_i, v_j, v_k])} \]
   where $A(.)$ represents the area of a triangle, and $[v_i, v_j, v_k]$ is the triangle formed by $v_i, v_j, v_k$.
3. Find all the baselines where normalized area distortion is zero.
4. Recursively traverse closed curves toward the negative distortion values until an apex with the most negative area distortion vertex is reached. During the recursion, the closed curves can break into multiple closed curves and move towards different spikes. The spikes and the corresponding closed curves are assigned unique IDs to track their progression and for height computation in the next step.
5. The sum of the distances between the successive centroids of the traversed closed curves give the height of the spike from the baseline. The spike width is the largest distance between the baseline vertices.

**NOTE:** Comparisons with other state-of-the-art spherical mapping algorithms are provided in [10]. The lowest angle distortion achieved in [10] provides a unique opportunity to exploit the corresponding area distortion metric effectively in our context.
2.3 Spiculation score

Here, we described a novel spiculation score normalized by height, $s_1$, which summarizes sharpness and height of spikes for each PN,

$$s_1 = \frac{\sum_i \text{mean}(\epsilon_{p(i)}) \cdot h_{p(i)}}{\sum_i h_{p(i)}}$$

where $p(i)$ is spike $i$, $h_{p(i)}$ is height of spike $p(i)$, and mean($\epsilon_{p(i)}$) is the mean of the area distortion of all the vertices (sharpness) in $p(i)$. Figures 2 show the results of our spiculation quantification, its measures (no. of spikes $N_p$ and spiculation score $s_1$) and radiologist’s score ($s_r$). We compared our spiculation score with Dhara’s spiculation scores ($s_a$ and $s_b$) in [3],

$$s_a = \sum_i e^{-\omega_{p(i)}} h_{p(i)}; s_b = \frac{\sum_i h_{p(i)} \cos \omega_{p(i)}}{\sum_i h_{p(i)}}$$

where $\omega_{p(i)}$ is the solid angle subtended at apex of spike $p(i)$.

2.4 Spiculation classification & Malignancy prediction

We evaluated our spiculation measures ($N_p$ and $s_1$) and radiomic features for classifying spiculation because the current clinical standard (Lung-RADS and McWilliams model) uses binary classification of spiculation. Dhara’s spiculation scores ($s_a$ and $s_b$) were also evaluated. We extracted 103 radiomic features from each PN to quantify its intensity, shape, and texture [2]. Moreover, we extracted features from the mesh model, such as shape features (size - average of longest
and its perpendicular diameters, volume, equivalent volume sphere’s diameter, and roundness) and statistical features of the area distortion metric $\epsilon$. We performed a univariate analysis using Wilcoxon rank-sum test, the area under the receiver operating characteristic curve (AUC), and Spearman’s correlation coefficient $\rho$ to evaluate the significance of each feature to classify spiculation. $p$-values were adjusted by Bonferroni correction, because we tested multiple features for a single outcome.

For multivariate analysis, we developed a binary classification model using a radial basis kernel SVM ($\gamma = 0.001$ and cost=64). We divided our data into three groups (training 60%, validation 20%, and test 20%), and the details of model construction are as follows:

1. Training set: distinctive features were identified based on hierarchical clustering [2], and then fed to the SVM classifier training.
2. Validation set: The best feature set was selected by a backward selection (recursive feature elimination) to optimize the SVM classification which was trained using the training set.
3. Test set: The performance of the optimized binary classification model was evaluated.

To predict malignancy of PNs, we applied the SVM-LASSO model [2] which uses two CT radiomic features: Bounding Box Dimension of Anterior Posterior axis (BB$_{AP}$), and Standard Deviation of Inverse Difference Moment (SD$_{IDM}$). We evaluated the original SVM-LASSO model and combinations of the model and spiculation classifications using our spiculation measures and radiologist’s spiculation score ($s_r$), respectively.

## 3 Results

The Lung Image Database Consortium image collection (LIDC-IDRI) contains 1018 cases with low-dose screening thoracic CT scans and marked-up annotated lesions [8]. Four experienced thoracic radiologists performed contouring and image annotation including spiculation, lobulation, texture, margin and malignancy. Spiculation scoring ranged between 1 (non-spiculated) and 5 (highly spiculated). We binarized the score (1 as 0, non-spiculated and 2 to 5 as 1, spiculated) because the current clinical standard uses binary classification. 883 cases (585 non-spiculated vs. 298 spiculated) in the dataset have PNs with contours.

We divided the dataset into two subsets depending on whether diagnostic data (pathological malignancy) was available (72 cases) or not (811 cases). The 811 cases (266 spiculated) were used for training of the spiculation classification, and the 72 cases (32 spiculated) were used for validation of the classification. On the other hand, the 72 cases were used for training of the malignancy prediction, and the 811 cases were used for validation of the prediction. For the validation, we used malignancy score (1 highly unlikely, 2 moderately unlikely, 3 indeterminate likelihood, 4 moderately suspicious, and 5 highly suspicious for cancer) determined radiologically since pathological malignancy is not available.
3.1 Spiculation classification

We generated synthetic PNs with spiculations (2 mm and 5 mm height) to validate spiculation quantification as shown in Figure 3. Three of four isolated 2 mm and all four 5 mm synthetic spiculations were detected. The measured average heights were 3.8 and 6.5 mm respectively.

In univariate analysis, 23 features were identified as significant features (adjusted $p$-value<0.05) to classify spiculation. Two roundness features from mesh model and voxel-based mask image respectively were the best features. All statistical features of $\epsilon$ (minimum, variance, maximum, mean, skewness, kurtosis, and median) and most spiculation scores ($s_1$ and $s_b$) were significant. In addition, five shape features and seven texture features were significant. Table 1 shows the univariate analysis results of the ten most significant features. Our spiculation score ($s_1$: 0.67 AUC, $P=4.29E-08$, $\rho=-0.33$) outperformed Dhara’s ($s_a$: 0.59 AUC, $P=0.068$, $\rho=0.22$ and $s_b$: 0.66 AUC, $P=1.47E-07$, $\rho=0.29$). Figure 4 compares our method and Dhara’s mean curvature based method. Mean curvature based method detects all critical points on the surface first and then find the baseline of each critical point to detect spikes. So it generated too many spike candidates at a single spike and baseline detection was not accurate. On the other hand, the proposed method can directly detect spikes as a whole.

Table 1. Ten most significant features in univariate analysis for Spiculation classification.

| Rank | Feature name          | AUC  | Corr | P            |
|------|-----------------------|------|------|--------------|
| 1    | Roundness (Mesh)      | 0.72 | -0.42| 1.82E-14     |
| 2    | Roundness (Voxel)     | 0.70 | -0.40| 2.64E-11     |
| 3    | Minimum $\epsilon$    | 0.69 | -0.37| 2.93E-10     |
| 4    | Long Run Emphasis     | 0.67 | -0.28| 9.40E-09     |
| 5    | Variance $\epsilon$   | 0.67 | 0.33 | 1.48E-08     |
| 6    | Maximum $\epsilon$    | 0.67 | 0.32 | 2.22E-08     |
| 7    | $s_1$                 | 0.67 | -0.33| 4.29E-08     |
| 8    | $s_b$                 | 0.66 | 0.29 | 1.47E-07     |
| 9    | Mean $\epsilon$       | 0.66 | -0.33| 1.64E-07     |
| 10   | 2D Roundness          | 0.65 | -0.33| 1.33E-06     |
Fig. 4. Comparison with mean curvature method. (a,d) original model, (b,e) spiculation apex and height results computed via area distortion metric, and (c,f) spiculation apex and height results computed via mean curvature method [3].

Table 2. Spiculation classification results

|                  | Sensitivity | Specificity | Accuracy | AUC |
|------------------|-------------|-------------|----------|-----|
| Training on 811 cases | 73.3%       | 77.9%       | 76.7%    | 0.80|
| Validation on 72 cases  | 75.0%       | 72.5%       | 73.6%    | 0.82|

The number of spikes $N_p$ and spiculation score $s_1$ were selected to classify PNs into spiculated or non-spiculated in multivariate analysis, and the classification performance is shown in Table 2. The spiculation classification achieved an accuracy of 76.70%, and the validation on the 72 cases showed comparable results (73.61% accuracy). Texture and intensity features were not good predictors. Statistical features of area distortion metric showed similar performance to our spiculation measures.

3.2 Malignancy Prediction

As shown in Table 2, our spiculation measures (in addition to shape and texture features [2]) improved the malignancy prediction. The prediction accuracy was comparable to using radiologist’s spiculation score (Spearman’s $\rho = 0.48$) not only on the 72 cases (10×10 fold cross-validation) but also on the 811 cases (independent validation of the model trained by the 72 cases). Malignancy was defined as moderately suspicious to highly suspicious (malignancy score $>3$) on radiologic readings. Many moderately suspicious PNs were mis-classified as benign (false negative) and thus the low sensitivity 47%. When malignancy was defined as highly suspicious (malignancy score $>4$), which was closer to true pathological malignancy, the sensitivity increased to 73.7%.
Table 3. Malignancy prediction by SVM-LASSO radiomics model (shape and texture) [2] and combining with predicted spiculations by radiologist’s score ($s_r$) and our measures ($N_p + s_1$)

|                      | Sensitivity | Specificity | Accuracy | AUC  |
|----------------------|-------------|-------------|----------|------|
| **Training on 72 cases** |             |             |          |      |
| Shape+Texture        | 87.2%       | 81.2%       | 84.6%    | 0.89 |
| Shape+Texture+Radiologist’s score ($s_r$) | 87.8%       | 87.1%       | 87.5%    | 0.91 |
| Shape+Texture+Our measures ($N_p + s_1$) | 92.7%       | 83.9%       | 88.9%    | 0.92 |
| **Validation on 811 cases (malignancy score>3)** |             |             |          |      |
| Shape+Texture        | 34.6%       | 93.3%       | 74.2%    | 0.81 |
| Shape+Texture+Radiologist’s score ($s_r$) | 40.3%       | 93.8%       | 76.5%    | 0.81 |
| Shape+Texture+Our measures ($N_p + s_1$) | 47.5%       | 89.8%       | 76.1%    | 0.81 |
| **Validation on 811 cases (malignancy score>4)** |             |             |          |      |
| Shape+Texture        | 63.2%       | 82.1%       | 79.4%    | 0.79 |
| Shape+Texture+Radiologist’s score ($s_r$) | 73.7%       | 64.1%       | 65.5%    | 0.77 |
| Shape+Texture+Our measures ($N_p + s_1$) | 73.7%       | 80.9%       | 79.9%    | 0.82 |

Table 3 shows the comparisons with recently reported lung cancer screening radiomic studies. Our method showed better or comparable performance, and its sensitivity, specificity, accuracy, and AUC were well balanced. The malignancy prediction performance was improved when combining spiculation quantification into the radiomics model (shape and texture) [2].

Table 4. Comparison with lung cancer screening radiomic studies

|                       | Dataset                                      | Sensitivity | Specificity | Accuracy | AUC  |
|-----------------------|----------------------------------------------|-------------|-------------|----------|------|
| Hawkins et al. (2016) [1] | Baseline CT scans of 261 patients in NLST | 51.7%       | 92.9%       | 80.0%    | 0.83 |
| Buty et al. (2016) [1] | LIDC 2054 PNs                                 |             |             |          |      |
|                       | Ground-truth by radiologists assessment      |             |             |          |      |
| Kumar et al. (2015) [5] | LIDC 97 patients                              | 79.1%       | 76.1%       | 77.5%    |      |
|                       | Including metastatic tumors                  |             |             |          |      |
| Liu et al. (2016) [4] | LIDC 172 patients                             | 71.4%       | 83.7%       | 80.0%    | 0.81 |
|                       | Two independent cohorts 102 and 70 patients   |             |             |          |      |
| Choi et al. (2018) [2] | LIDC 72 patients                              | 87.2%       | 81.2%       | 84.6%    | 0.89 |
| Proposed Method      | LIDC 72 patients                              | 86.6%       | 84.5%       | 88.9%    | 0.92 |

4 Conclusion and Future Work

We presented a novel method for quantification of pulmonary nodule spiculation in lung cancer using the spherical conformal parameterization. The quantitative
spiculation measures were found to be highly correlated with the radiologist’s spiculation score and lead to state-of-the-art malignancy prediction results with accuracy of 88.9%. A current limitation of our work is the use of manual segmentations from the LIDC datasets, which do not precisely delineate the spiculations. We plan to use semi-automatic segmentation to extract more accurate and reliable mesh models for spiculation quantification. Moreover, we will test this new measure in breast cancer datasets where spiculation again is a good malignancy predictor.

References

1. Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., Mollura, D.J.: Characterization of lung nodule malignancy using hybrid shape and appearance features. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 662–670. Springer (2016)
2. Choi, W., Oh, J.H., Riyahi, S., Liu, C.J., Jiang, F., Chen, W., White, C., Rimmer, A., Mechakalos, J.G., Desay, J.O., Lu, W.: Radiomics analysis of pulmonary nodules in low-dose ct for early detection of lung cancer. Medical Physics (2018)
3. Dhara, A.K., Mukhopadhyay, S., Saha, P., Garg, M., Khandelwal, N.: Differential geometry-based techniques for characterization of boundary roughness of pulmonary nodules in CT images. International Journal of Computer Assisted Radiology and Surgery 11(3), 337–349 (2016)
4. Hawkins, S., Wang, H., Liu, Y., Garcia, A., Stringfield, O., Krever, H., Li, Q., Cherezov, D., Gatenby, R.A., Balagurunathan, Y., et al.: Predicting malignant nodules from screening ct scans. Journal of Thoracic Oncology 11(12), 2120–2128 (2016)
5. Kumar, D., Shafiee, M.J., Chung, A.G., Khalvati, F., Haider, M.A., Wong, A.: Discovery radiomics for computed tomography cancer detection. arXiv preprint arXiv:1509.00117 (2015)
6. Liu, Y., Balagurunathan, Y., Atwater, T., Antic, S., Li, Q., Walker, R.C., Smith, G., Massion, P.P., Schabath, M.B., Gillies, R.J.: Radiological image traits predictive of cancer status in pulmonary nodules. Clinical Cancer Research pp. clincares–3102 (2016)
7. McKee, B.J., Regis, S.M., McKee, A.B., Flacke, S., Wald, C.: Performance of acr lung-rads in a clinical ct lung screening program. Journal of the American College of Radiology 12(3), 273 – 276 (2015)
8. McNitt-Gray, M.F., Armato, S.G., Meyer, C.R., Reeves, A.P., et al.: The lung image database consortium (lidc) data collection process for nodule detection and annotation. Academic Radiology 14(12), 1464–1474 (2007)
9. McWilliams, A., Tammemagi, M.C., Mayo, J.R., Roberts, H., et al.: Probability of cancer in pulmonary nodules detected on first screening ct. New England Journal of Medicine 369(10), 910–919 (2013)
10. Nadeem, S., Su, Z., Zeng, W., Kaufman, A., Gu, X.: Spherical parameterization balancing angle and area distortions. IEEE Transactions on Visualization and Computer Graphics 23(6), 1663–1676 (2017)
11. Niehaus, R., Raicu, D.S., Furst, J., Armato, S.: Toward understanding the size dependence of shape features for predicting spiculation in lung nodules for computer-aided diagnosis. Journal of Digital Imaging 28(6), 704–717 (2015)
12. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. CA: A Cancer Journal for Clinicians 66(1), 7–30 (2016)

13. Swensen, S.J., Silverstein, M.D., Istrup, D.M., Schleck, C.D., Edell, E.S.: The probability of malignancy in solitary pulmonary nodules: Application to small radiologically indeterminate nodules. Archives of Internal Medicine 157(8), 849–855 (1997)