

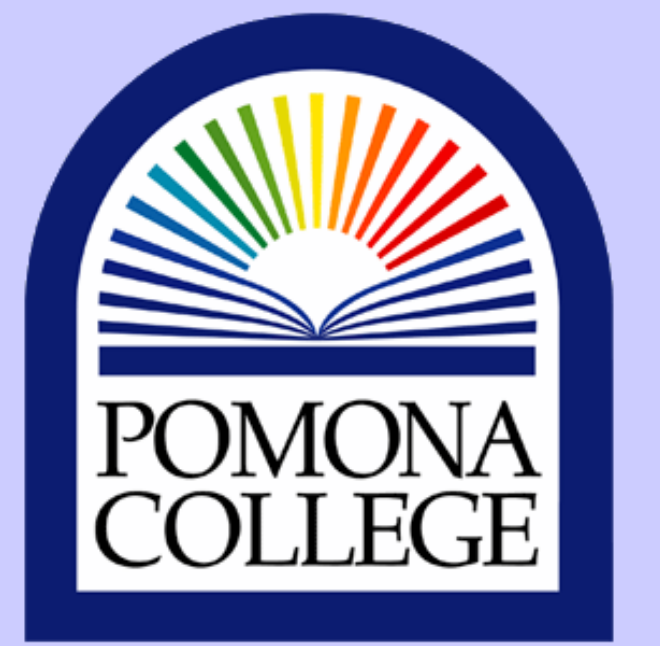
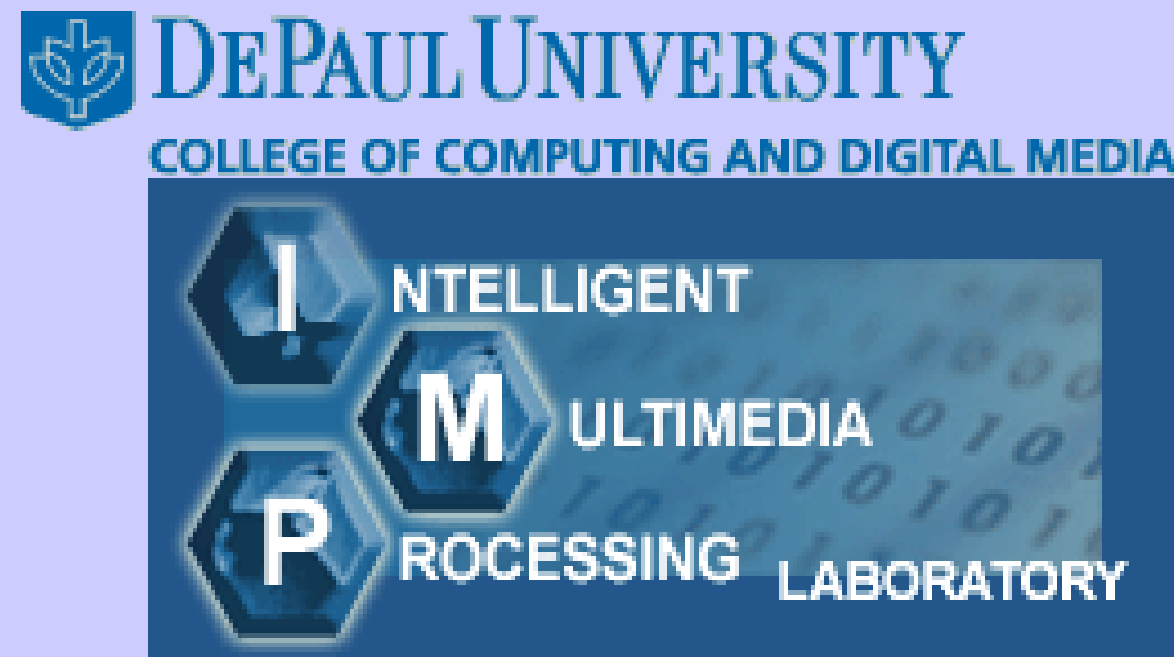
Learning Lung Nodule Similarity Using a Genetic Algorithm

¹Kerry A. Seitz, Jr., ²Anne-Marie Giuca, ³Jacob D. Furst, ³Daniela S. Raicu

¹Trinity University, One Trinity Place, San Antonio, Texas, USA 78212

²Pomona College, 333 North College Way, Claremont, California, USA 91711

³DePaul University, 243 South Wabash Avenue, Chicago, Illinois, USA 60604



Introduction

This project demonstrates that by using a genetic algorithm (GA), the image features used in content-based image retrieval (CBIR) of lung computed tomography (CT) scans can be optimized to include only the features most relevant for accurate image retrieval.

Dataset

CT images from the Lung Image Database Consortium (LIDC) were used in this study. Image features were extracted from the nodules based on texture, size, shape, and intensity. Each nodule received both a radiologist-predicted and computer-predicted malignancy rating. **Figures 1 and 2** show an example CT image and a CBIR example using nodules from the LIDC, respectively.

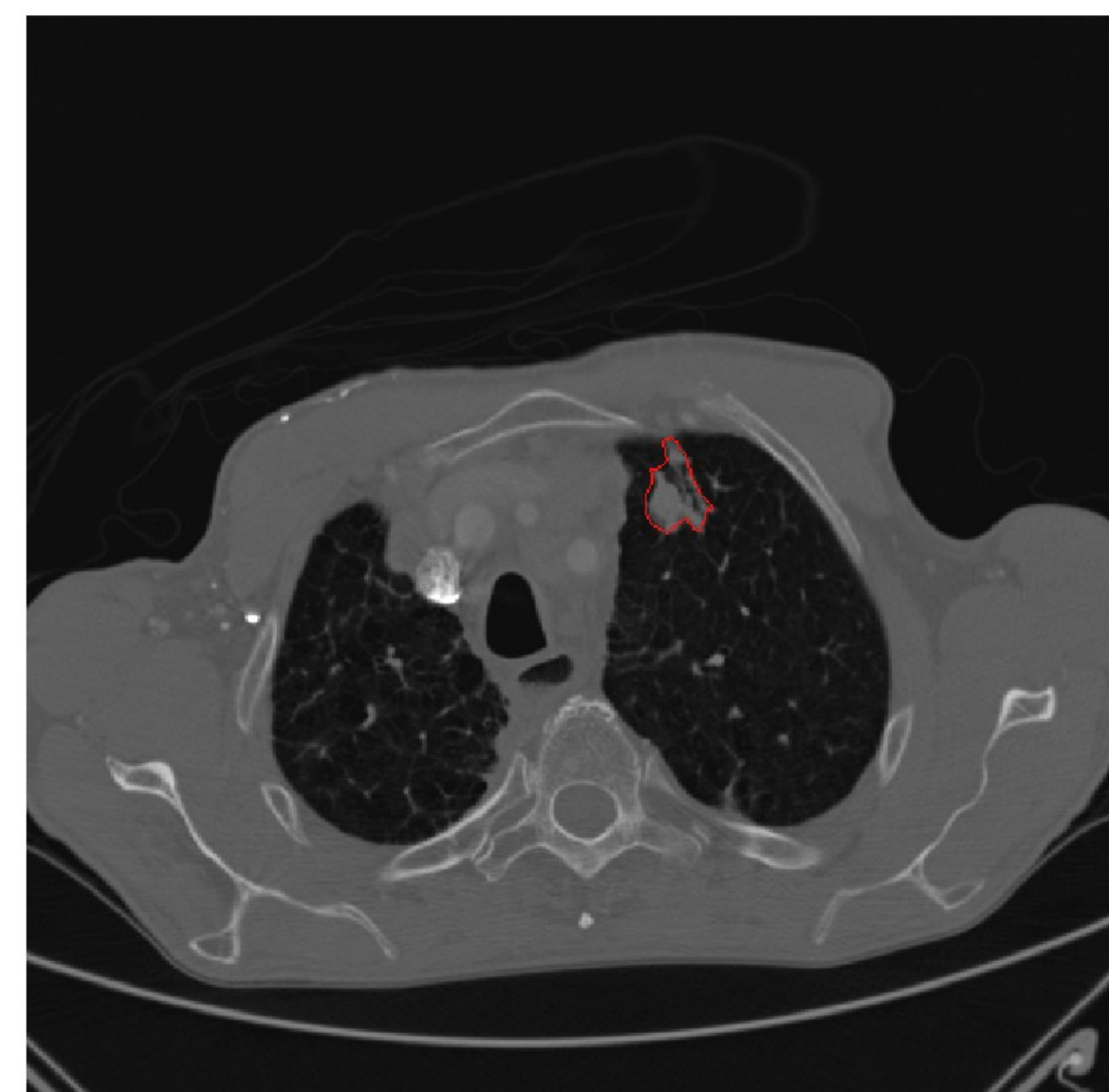
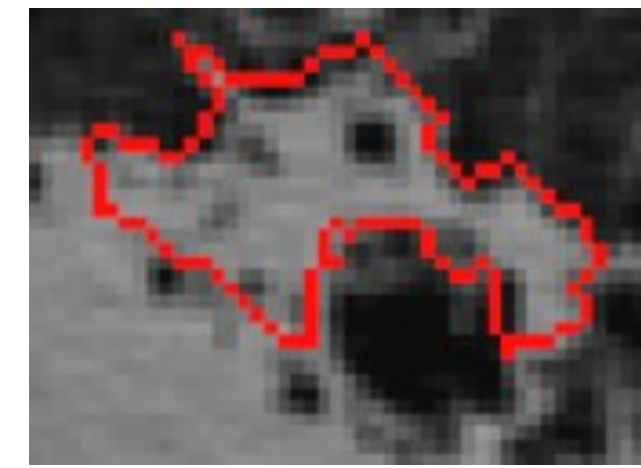


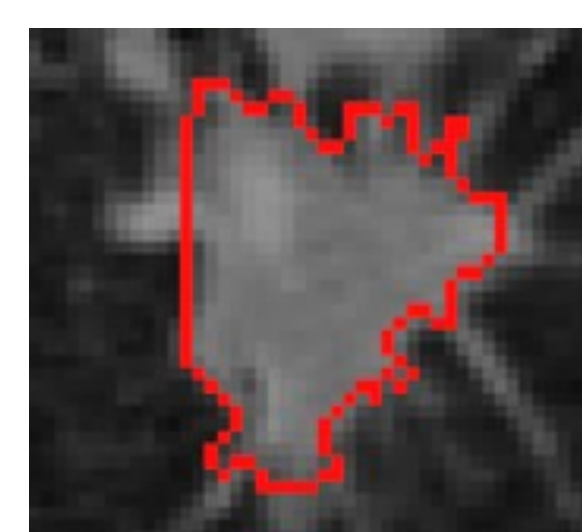
Figure 1. A CT image from the LIDC database.

Query

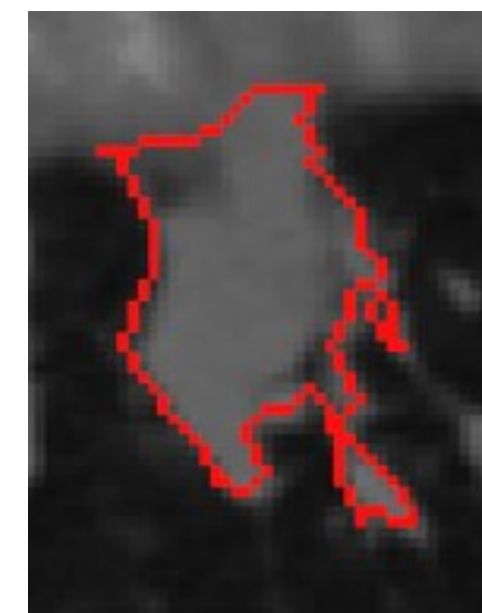


No. 367

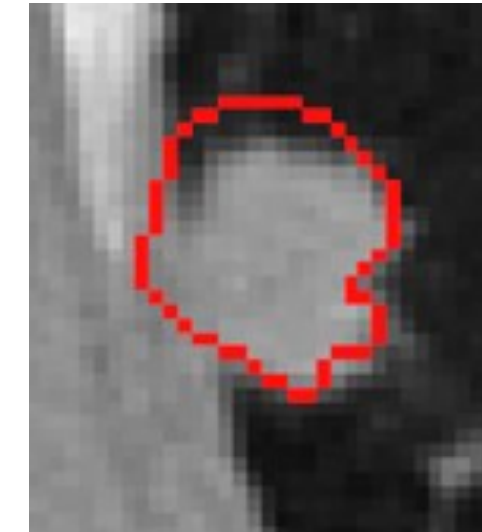
Retrieved



No. 528



No. 266



No. 561

Figure 2. A CBIR image retrieval example. The image features from the query image (top) were used to retrieve the other three images (bottom). Numbers indicate the nodule number as listed in the LIDC dataset.

The Genetic Algorithm

Individuals in the population represented subsets of image features to use when retrieving images using CBIR. Successful individuals were those that resulted in high precision after three images retrieved. The algorithm then mutated and combined the best individuals to create the next generation. **Table 1** shows the configurations used with the GA, which varied both in nodules used as well as malignancy ground truth. The following GA parameters were used: mutation rate — 0.1, crossover rate — 0.9, population size — 500, and number of generation — 50.

Table 1. Configurations used to test the CBIR system.

Configuration	Include Unknown Malignancy Nodules ^b	Exclude Unknown Malignancy Nodules ^b
Radiologist-Predicted	1	2
Computer-Predicted	3	4

^a Indicates whether the radiologist-predicted or computer-predicted malignancy was used as ground truth.

^b Indicates whether the nodules that were rated 'unknown' with respect to malignancy were used in the dataset.

Results

Baseline and GA precision results are shown in **Tables 2 and 3**, respectively. CBIR using the GA-reduced set of features outperformed CBIR with all 63 image features. All precision increases are significant except after 20 images retrieved with configurations 1 and 4 (pair-wise one-tail t-test, $p < 0.05$).

Table 2. Baseline precision results for the CBIR system.

Configuration	P _{AVG@3} ^a	P _{AVG@5} ^a	P _{AVG@10} ^a	P _{AVG@20} ^a	P _{AVG@50} ^a
1	51.57%	51.23%	49.78%	50.26%	48.58%
2	82.77%	82.57%	81.94%	80.90%	79.23%
3	65.43%	65.43%	64.15%	62.78%	59.84%
4	84.24%	83.20%	82.76%	82.07%	79.75%

^a Average precision after 3, 5, 10, 20, or 50 images retrieved using all 63 image features.

Table 3. GA precision results for the CBIR system.

Configuration	P _{AVG@3} ^a	P _{AVG@5} ^a	P _{AVG@10} ^a	P _{AVG@20} ^a	P _{AVG@50} ^a
1	55.76%	53.70%	52.11%	50.69%	49.12%
2	85.95%	84.10%	83.53%	81.95%	80.37%
3	71.30%	69.74%	67.28%	65.45%	62.39%
4	86.91%	85.32%	84.13%	82.58%	80.90%

^a Average precision after 3, 5, 10, 20, or 50 images retrieved using a reduced set of image features.

Conclusion

Advantages of optimized feature set for CBIR based computer-aided diagnosis (CAD) systems:

- More accurate image retrieval
- Improved efficiency

The features selected by the GA include texture, size, shape, and intensity features, indicating that all of these categories are important in determining nodule similarity. The results of this study also suggest that in the absence of radiologist-predicted malignancy ratings, computer-predicted ratings can be used, as the CBIR system performed similarly when using each of these malignancy predictions.

References

- Armato, S. G., McLennan, G., McNitt-Gray, M. F., Meyer, C. R., Yankelevitz, D., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Reeves, A. P., Croft, B. Y., and Clarke, L. P., "Lung image database consortium: developing a resource for the medical imaging research community," *Radiology* **232**, 739–748 (2004).
- Lam, M., Disney, T., Pham, M., Raicu, D., and Furst, J., "Content-based image retrieval for pulmonary computed tomography nodule images," *Proc. SPIE* **6516**, 65160N (2007).
- Zinovev, D., Raicu, D., Furst, J., and Armato, S., "Predicting radiological panel opinions using a panel of machine learning classifiers," *Algorithms* **2**, 1473–1502 (2009).
- <http://code.google.com/p/jiva-ng>

Acknowledgments

This work was supported in part by NSF award 1062909.



Want more info?

Scan this QR code with your smartphone!

