Novel Design and Evaluation of a Machine Learning Approach for Staging and Prognosis of Lung Cancer through a Preliminary CT

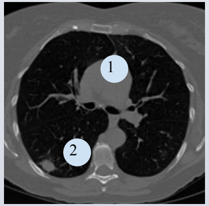
INTRODUCTION

Lung cancer is the **deadliest cancer,** affecting hundreds of thousands annually [17]

- 228,150 cases of lung cancer in 2019 (116,440 men and 111,710 women) [16]
- Oncologists are often wary of making useful prognostic claims because of a lack of accuracy and standardized methodology [15]
- Staging traditionally consists of 3 categories (**TNM**). Tumor (**T**) referring to the size and nature of the tumor, Lymph nodal (**N**) referring to the effective spread to lymph nodes, and **M** representing metastasis.
- Staging is an effective tool to predict the behavior of tumors (TNM) [5]
  - Potential for surgical resection
  - Assess possibility of further procedures such as lung biopsies
  - Metastatic potential
  - Clinical aggressivity
- Ethical applications of more accurate prognosis can be utilized to tailor therapeutic strategies to address quality of life (post disease) and timeliness of treatment [3]

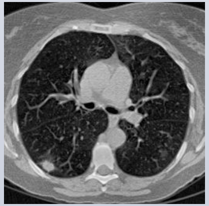
IMAGE PREPROCESSING

Before feeding CT slice images into our model, image preprocessing was done to accentuate key features in the slice such as the primary **tumor** or **nodule** and the mediastinal **lymph nodes**, both necessary parts of the TNM staging process.



1. Lymph Nodes
2. Malignant Nodule

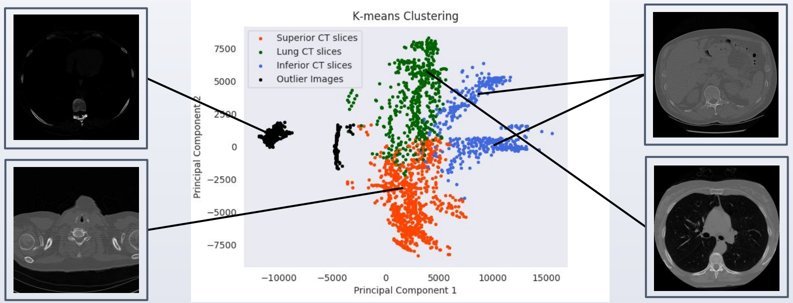
Both features are visible but not accentuated nor emphasized.



Contrast Limited Adaptive Histogram Equalization (CLAHE)

Histogram equalization will be employed for contrast enhancement. Adaptive equalization is applied through a windowing approach so as to not over-contrast certain parts of the image at the expense of others. Instead, grayscale values are stretched/flattened locally rather than globally [4].

IMAGE CLUSTERING



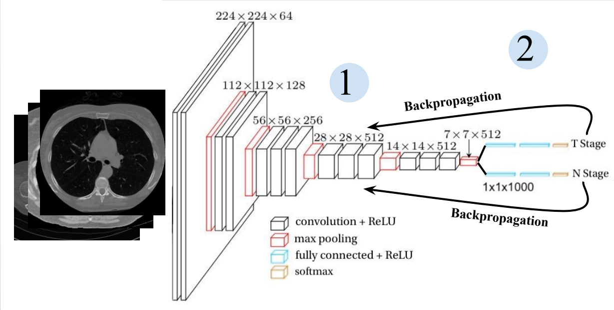
K-Means Clustering is an algorithm that clusters data into *k* classes depending on their component features or, for these images, their flattened pixel values. The algorithm was performed on the dataset in order to distinguish between slices the model cannot generalize to and those that clearly show lung lobes and pertinent lymph nodes.

MODEL ARCHITECTURE

The primary architecture consisted of two main steps.

1. **Deep Feature Extraction** using pre-trained weights of VGGNet [14], a deep CNN, trained on over 20,000 classes of everyday objects.
2. **Classification** via multi-layer perceptron (MLP) was done for each separate task on the extracted features.

Multi-task learning: Because both the Tumor (T) and Lymph Nodal (N) staging require similar holistic understandings of each input CT image, two MLPs stemming from the main CNN were employed. This process, known as multi-task learning, allows for loss functions to be added and backpropagation to be applied both separately in each MLP and jointly in the root CNN. In essence, the CNN learns to create a better representation of the input CT image for both specific tasks.



RESULTS

	Nodule Detection	Diagnosis	T Stage	N Stage	M Stage
VGG16 + MLP	97%	86%	72% 84% (one stage off)*	71% 80%*	**70%**
VGG16 + Multi-Task MLPs			**80%** **84%***	**76%** **84%***	
Physicians [5][18]		81%	71%	70%	

***Table 1*:** Performance comparisons on lung cancer diagnosis and staging between proposed models and physicians.

	Prognosis
Regression Model	74%
Classification (BCE)	79%
Classification (Custom Loss)	**85%**
Physicians [15]	74%

***Table 2*:** Performance comparisons on lung cancer prognosis between proposed models and losses and physicians.

ANALYSIS

- Preprocessing
  - K-means clustering and statistical based methods employed provided sufficient clarity for future models to prognose and diagnosed with physician level quality
- Staging and Diagnosis (***Table 1***)
  - Basic nodule detection occurs with almost complete certainty
  - T and N staging accuracy for the basic MLP model was in line with physician expertise (~70%), and was able to predict the M stage with marginally less accuracy than the other two stages, a task that usually requires follow-up procedures [5]
  - Once the multi-task architecture was added, the T and N stage surpassed physician accuracy by 6-9%
  - The multi-task architecture and the basic model achieved similar accuracy when allowing for a 1 stage window of error with the more complex model achieving 4% higher.
- Prognosis (***Table 2***)
  - The most basic regression model achieved a 74% prognosis accuracy, exactly in line with physician expertise [15].
  - The custom loss function outperformed both the traditional BCE classification loss function as well as physician accuracy, surpassing the latter **by more than 10%**. These results can serve as proof of concept for our loss function in generalized multi-label classification problems.

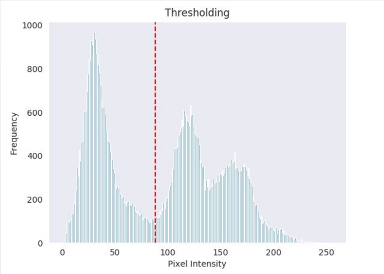
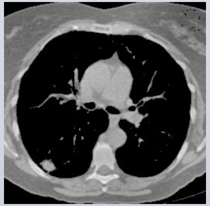


Fig. 1: Histogram of grayscale pixel density in the sample CT slice with red dashed line indicating threshold creating by ***Otsu's Method.***



Bilateral Filter and Thresholding

The Bilateral Filter is a modified Gaussian Filter which takes a kernel based on a normal distribution of pixel values and passes it convolutionally to remove speckle noise. The pixel intensities are then plotted and a threshold is produced through ***Otsu's Method***, as referenced in ***Fig. 1***, maximizing inter-class variance while minimizing intra-class variance.

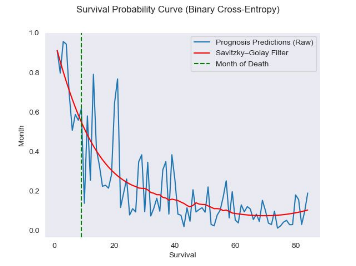
PROGNOSIS

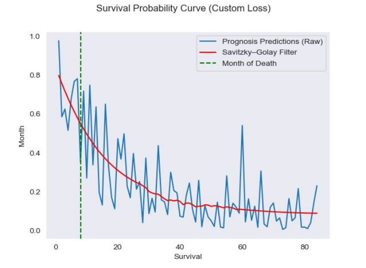
Three deep learning methods were employed by the proposed method to utilize extracted features for prognosis

1. **Regression** - As a baseline, prognosis was treated as a linear problem with one class. Feature extraction was performed and trained to produce one number as a result. Because the output was not a vector, but rather a 0 dimensional "death-date", no graph was produced.
2. **Classification (BCE)** - Feature extraction was then utilized, piping the feature map into a simple 84-class output layer (1 class for every month in a 7 year span). The loss function employed was an unmodified **Binary Cross-Entropy**. Unlike traditional multi-class classification problems, this is a a **multi-label** classification problem meaning a specific input can correspond to multiple classes or in this case survival months.
3. **Classification (Custom Loss)** - To ameliorate the main problem with Binary Cross-Entropy, being that the NN is unable to develop understanding of a general trend, a **custom, problem-specific loss function** is proposed as defined below which combines traditional BCE with the mean cubic difference of the previous prediction in order to incentivize a downward trend in survival probability.

$$-\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(p_i) + (1-y_i)\cdot\log(1-p_i)$$

$$-\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(p_i) + (1-y_i)\cdot\log(1-p_i) + \frac{1}{N}\sum_{i=1}^{N-1}(p_{i-1}-p_i)^3$$





Loss Functions are defined where p_i is the current prediction and y_i is the current ground truth

Savitzky-Golay Filter: A signal filter was applied to accentuate broader trends while smoothing out sharp micro movements. This filter performs convolutions on successive data subsets, applying the linear least square method to fit a line.

CONCLUSION

The proposed method achieved and in many cases surpassed benchmarks set in contextual literature

- K-means clustering successfully filtered undesirable data and allowed for future methods to calculate loss without distortion
- Feature extraction applied to VGGNet and subsequent classification MLP's resulted in accuracies on par with real physicians
- Furthermore, prognostic accuracy in all models exceeded that of physicians

All methods showed potential to form a tool to aid physicians in more accurate staging and prognostic work.

FUTURE WORK

The following areas have been identified as areas for potential future work

- Preprocessing
  - Remove extraneous features with more precision [4]
- Prognosis
  - Assess model in comparison with Recurrent NN models [13]
  - Adjust custom loss function to more accurately address micro-adjustments
- Test other classification algorithms on defined features outlined in the proposed method.

BIBLIOGRAPHY

1. Armato, Samuel G., III, and William F. Senakovik. "Automated Lung Segmentation for Thoracic CT: Impact on Computer-Aided Diagnosis." Academic Radiology, 11, no. 9 (September 2004). Accessed March 12, 2019. https://doi.org/ 10.1016/j.acra.2004.06.005.

2. Hua, Kai-Lung, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, and Yu-Jen Chen. "Computer-aided Classification of Lung Nodules on Computed Tomography Images via Deep Learning Technique." Dovepress: OncoTargets and Therapy, August 4, 2015. Accessed March 12, 2019. https://doi.org/10.2146/OTT.S80733.

3. Clément-Duchêne, Christelle, Charlotte Carnin, Francis Guillemin, and Yves Martinet. "How Accurate Are Physicians in the Prediction of Patient Survival in Advanced Lung Cancer?" The Oncologist, July 17, 2010, 782-89. Accessed March 12, 2019. https://doi.org/10.1634/theoncologist.2009-0149.

4. Bindu, Hima G., Prasad Reddy, and Ramakrishna G. Murty. "Image Preprocessing of Abdominal CT Scan to Improve Visibility of Any Lesions in Kidneys." Journal of Theoretical and Applied Information Technology 96, no. 8 (April 30, 2018): 2298-306. Accessed March 12, 2019.

5. Barker, John M., and Gerard A. Silvestri. "Lung Cancer Staging." Current Opinion in Pulmonary Medicine. Accessed March 12, 2019.

6. Vock, David M., Julian Wolfson, Patrick J. O'Connor, Sunayan Bandyopadhyay, Gediminas Adomavicius, Paul E. Johnson, and Gabriela Vazquez-Benitez. "Adapting Machine Learning Techniques to Censored Time-To-Event Health Record Data: A General-Purpose Approach Using Inverse Probability of Censoring Weighting." Journal of Biomedical Informatics 61 (June 2016): 119-31. Accessed March 12, 2019. https://doi.org/10.1016/j.jbi.2016.03.009.

7. He, Kaiming, et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi:10.1109/cvpr.2016.90.

8. Firmino, M., Angelo, G., Morais, H., Dantas, M. R., & Valentim, R. (2016). Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. BioMedical Engineering OnLine,15(1). doi:10.1186/s12938-015-0120-7

9. Sun, T., Wang, J., Li, X., Lv, P., Liu, F., Luo, Y., . . . Guo, X. (2013). Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. Computer Methods and Programs in Biomedicine,111(2), 519-524. doi:10.1016/j.cmpb.2013.04.016

10. Zacharaki, E. I., Wang, S., Chawla, S., Yoo, D. S., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic Resonance in Medicine,62(6), 1609-1618. doi:10.1002/mrm.22147

We thank the National Cancer Institute for access to NCI's data collected by the National Lung Screening Trial (NLST). The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

BIBLIOGRAPHY

11. Zhou, H., Dong, D., Chen, B., Fan, M., Cheng, Y., Gan, Y., . . . Tan, J. (2018). Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. Translational Oncology, 31-36.

12. Vijaya, G., & Suhasini, A. (2014). An Adaptive Preprocessing of Lung CT Images with Various Filters for Better Enhancement. Academic Journal of Cancer Research. doi:10.5829/idosi.ajcr.2014.7.3.84231

13. Katzman, Jared, Uri Shaham, Alexander Cloninger, and Jonathan Bates. "Deep Survival: A Deep Cox Proportional Hazards Network." ResearchGate. Accessed March 12, 2019. https://www.researchgate.net/publication/303812000_Deep_Survival_A_Deep_Cox_Proportional_Hazards_Network#pfb.

14. Andrew Zisserman and Karen Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv, last modified April 15, 2010, accessed March 12, 2019, https://arxiv.org/abs/1409.1556.

15. Jason Lambden et al., "Accuracy of Oncologists' Life-Expectancy Estimates Recalled by Their Advanced Cancer Patients: Correlates and Outcomes," Journal of Palliative Medicine, December 1, 2016, accessed March 12, 2019, https://doi/org/10.1089/jpm.2016.0121.

16. American Cancer Association, "Key Statistics for Lung Cancer," American Cancer Association, last modified January 8, 2019, accessed March 12, 2019, https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html.

17. "Lung Cancer," Mayo Clinic, accessed March 12, 2019, https://www.mayoclinic.org/diseases-conditions/lung-cancer/symptoms-causes/syc-20374620.

18. Gaber KA, DeBarr SE, Stevens S, Elkkari A, Bakere H, et al. (2012) Ability of Lung Specialist to Predict a Diagnosis of Lung Cancer during the First Visit to a Fast Track Clinic. J Pulmon Resp Med S5:004. doi:10.4172/2161-105X.S5-004

19. Kwasigroch A., Mikołajczyk A., Grochowski M. (2017) Deep convolutional neural networks as a decision support tool in medical problems – malignant melanoma case study. In: Mitkowski W., Kacprzyk J., Oprzędkiewicz K., Skruch P. (eds) Trends in Advanced Intelligent Control, Optimization and Automation. KKA 2017. Advances in Intelligent Systems and Computing, vol 577. Springer, Cham

20. Deng, Jia et al. "ImageNet: A large-scale hierarchical image database." 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009): 248-255.

21. Kadir, Timor and Fergus Gleeson. "Lung cancer prediction using machine learning and advanced imaging techniques" Translational lung cancer research vol. 7,3 (2018): 304-312.

22. Kanazawa, K., et al. "Computer-Aided Diagnosis for Pulmonary Nodules Based on Helical CT Images." Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170), 1998, doi:10.1109/icpr.1998.712045.