Risk Classification of Stereotactic Body Radiation Therapy applied to Thoracic Cancers

(Technical Paper)

Analysis of Systemic Bias Introduced by Machine Learning Applications

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

John Fishbein

Fall, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____ Date _____

John Fishbein

Approved _____ Date _____

Dr. Alfred Weaver, Department of Computer Science

Approved _____ Date _____

Sean Ferguson, Department of Engineering and Society

Introduction

In practice today, current radiation oncologists treat early-stage thoracic cancers using a technique known as Stereotactic Body Radiation Therapy (SBRT). While definitive treatments with SBRT are curing many patients of their lung cancers, there are associated cardiovascular side effects of radiation that can impact a patient's overall survival rate, quality of life, and healthcare costs. When performing SBRT, an oncologist can create a patient-specific treatment plan in which the dose of radiation to each coordinate of the patient's body is known. My supervisor and I believe that this data is extremely relevant to controlling a radiation treatment's unwanted/potentially fatal side effects. My technical project objective is to apply machine learning to classify the risk of a given treatment plan. Ultimately, this would assist oncologists in evaluating whether or not to proceed with a given treatment plan.

In the subfield of machine learning, a big challenge is overcoming the bias inherent in the data that is used. On a high level, machine learning works by providing large amounts of data to a model so that an algorithm can learn the patterns and draw conclusions from it. The conclusions reached are highly dependent on the diversity and quantity of data used in the training process. If the given data contains bias, the model's predictions will as well. This exact issue has appeared in my technical work. At first glance, we saw predictions with a very high percentage accuracy. However, after careful evaluation, we realized the presence of bias in our data and that the numbers were not representative of the model's actual predictability. Having encountered this first-hand, we saw how easily someone could carelessly ignore the bias and publish or rely on flawed results.

These issues of bias in the results of machine learning algorithms are a significant problem especially given the increase in their availability and use. This technology is powerful, and few experts fully understand it. Thus, this combination automatically makes machine learning dangerous. Data and the misuse of data is the heart of the problem. In my STS research project, I will present these issues of systemic bias and discuss how we can better prepare professionals to handle and avoid these problems.

Technical Topic

SBRT has been used extensively by oncologists for over 20 years and has proven to be effective at treating thoracic cancers. Essentially, SBRT works by delivering a massive dose of radiation, often more than 50Gy, to the Planning Target Volume (PTV) to eradicate the tumor. Unfortunately, it is not possible to truly target only the tumors, and thus adjacent normal tissue inevitably receives some radiation. When considering thoracic cancers, by definition, the tumors appear in the central region of the body. Therefore, the adjacent tissue is the patient's vital organs. The dose and volume of radiation to these normal structures strongly impacts the side-effects and is directly dependent on the exact location of the primary tumor. The development of a strategy to limit cardiac toxicity in SBRT is a complex issue. This is primarily due to incomplete toxicity information with hypo-fractionated radiation, uncertainty in tolerance of radiation to different anatomical cardiac areas and adjacent major vessels, inter and intra-fraction cardiac motion, and patient-specific radiation toxicity susceptibility factors (Xue et al., 2016). There are currently no models available to evaluate the overall cardiac toxicity of a lung SBRT plan.

Over the past year, my supervisor and I have had two major goals in my technical project. The first objective was to preprocess the data to achieve a normalized coordinate system to be used with each patient's dose map. The second objective is to use the treatment dosage data and other relevant patient data to create a predictive model of the risk associated with a given treatment plan.

A given treatment plan consists of numerous CT scans, the dose plan, the dose files, and the organ structure contours. When these files are fed through the current data pipeline, we have the specific dose of radiation given to each voxel (3-D pixel) of the patient's body. In order for this data to be useful in a generalizable way, we need to account for variations in patient height, shape, organ size, and other differences. In its current state, the same coordinates in two different patients represent two entirely different locations. Additionally, organ motion can affect radiation absorption (Kataria et al., 2016).

In order to normalize the dose data across the entire patient population, we plan to implement the Normalized Thoracic Coordinate System(NTCS) proposed by Dr. Hongkai Wang and his colleagues. (Kataria et al., 2016). In this research, Wang investigates variations in skeletal structure and organ position between a group of patients and propose a new set of coordinate axes that best accounts for these subtle shifts. This coordinate system is a solid step towards a generalizable model.

Once we normalize the coordinate system used in the patient dose maps, we will move on to machine learning. We plan to generate a multi-dimensional model of patient risk using organ-specific cardiovascular toxicity from the SBRT based on: 1) dosimetry characteristics of the treatment plan, 2) elapsed time between RT treatment and event of interest, and 3) patient-specific risk factors (Darby et al.,

2013). A machine learning model will therefore be trained to determine risk using the following inputs: Organ-specific V25Gy, D0.5cc, D1cc, D4cc, and Dmax from the SBRT plan; 3D coordinates of the maximum dose point in the organ (determined using the NTCS system); the presence or absence of cardiovascular risk factors such as smoking history, prior cardiac events, etc. ; and patient-specific criteria such as age, gender, race, and genetic mutations/variations (Wennstig et al., 2017).

We will utilize two independent machine learning approaches to help develop this model: 1) a Random Forest Classification model (RF), 2) a Support Vector Classification model (SVC), both of which have been proven to successfully classify multi-dimensional data. Each model will be trained to predict the class probability whether or not the given SBRT plan will result in a cardiovascular event in the patient. Effectively, the positive class probability will represent the risk to the patient associated with the treatment plan.

My role in the technical project has been as the primary developer. Each week, I am responsible for adding to the existing codebase and working towards the goals described above. I am required to meet with my project supervisor, Dr. Krishni Wijesooriya, often to discuss results and frame new short-term goals.

STS Topic

The advancements in machine learning in the past few decades have led to increased availability of powerful machine learning models. Using open-source software packages such as Scikit-learn, TensorFlow, or Keras, a developer with minimal experience can employ powerful machine learning models to any data and

problem they choose. With this availability comes the opportunity for vast misuse. Both the data and the methods of analysis used in production-level machine learning need to be carefully understood; these powerful, hyper-technical tools have the potential to be misinterpreted by society in a myriad of critical ways.

At its core, machine learning is the intersection of computer science with statistics and applied to big data. Modern algorithms are used to create powerful models that can do a wide variety of things. Take, for example, the convolutional neural network. This machine learning technique is most often used in the task of image recognition and is used in practice for anything from classifying food images to modern facial recognition software. With the help of software packages, anyone with a minimal coding background can create their own convolutional neural network in under 20 lines of code. The potential for this technology given the open access to it, is incredible. However, because of its widespread accessibility, those who use this technology without fully understanding what is going on "under the hood", can create a product with results that mislead the consumers.

Bias- or the overemphasis on one variable - is a major concern in machine learning. At the highest level, machine learning works by "training" the algorithm using known data so that the model can in turn predict unknown data. Bias occurs when the "known" data that is used for training is flawed in some way, and then the resulting predictions become flawed. For example, certain facial recognition algorithms used in production have been shown to contain statistical bias. As discussed in Buolamwini's article in Time Magazine, facial recognition software was created using a convolutional neural network (Buolamwini, 2019). However, this

model exhibited a significantly higher error rate in identifying women versus men. A similar discrepancy was noted on the error rates of identifying minorities. After investigation, this flaw was due to the fact that the known data used to train the model consisted almost exclusively of men.

When a problem like this is seen in production, it is due to the developer's irresponsibility or ignorance. The developer needs to carefully consider the data being used and thoroughly investigate the test results for bias. This problem needs to be fundamental to the curriculum of every introductory machine learning course because of its potential for dangerous consequences, politically and socially. Ruha Benjamin, in her article *Race After Technology: Abolitionist Tools for the New Jim Code*, presents the implications that follow from "cultural coding [that is] embedded into the technical coding of software programs" (Benjamin, 2019). It is far-fetched to think that a prejudiced engineer is sitting somewhere and intentionally writing maliciously discriminatory code. The much more likely scenario is that biased data is used in the machine learning process and skewed results are present in production. Furthermore, since these models are visually nothing more than complex black boxes giving output, it is easy for the results to be misinterpreted by those who do not fully understand their inner workings – the algorithm and the data.

This misinterpretation is multi-faceted and not always easy to detect. In her article, Whitman (2020) discusses the prevalence and use of predictive modeling in institutions-- specifically schools/universities dedicated to higher education. Her evaluation addressed the biased results that higher learning institutions are applying to nudge students into greater academic success. She talks about the

approach of breaking down data for these models into two categories: "attributes" and "behaviors"—where attributes are inherently unchangeable, and behaviors are determined by what students have control over. In this discussion, she argues that behavior modifications being achieved using the predictive models are not necessarily the university's desired outcome. The correlation between a given student's actions and success is subject to change depending on how administrators define it; this directly influences the reliability of the results when used to assist students. This highlights the foundational concept of why there is a need for the creators of such predictive models to better understand their data.

In the last 25 years, incredible breakthroughs in the field of Artificial intelligence have created immense value for society. It is increasingly evident that systemic bias can be unintentionally introduced into influential technology through machine learning applications. I seek to understand how successful machine learning applications manage the societal bias present in data. In addition to the cases discussed briefly above, I will investigate other applications of machine learning, both positive and negative. Ultimately, I will analyze how engineers can successfully identify and prevent this bias.

Next Steps

Systemic bias is present in today's world and has detrimental consequences on society as a whole. In the following thesis, I will analyze the presence of this bias in machine learning algorithms and its implications in society. Professionals in the field of machine learning need to be cognizant of this problem and how to prevent it. I will investigate and present different curriculums exercised in

introductory machine learning courses. I believe that a shift is needed in how computer scientists and society view machine learning. It is a powerful tool, but data and the context in which it is applied must be carefully understood with respect to its implications in society. I plan to conduct interviews of professionals, including the UVA professor of machine learning, and use the results to formulate a discussion of handling bias. Furthermore, I will investigate successful implementations of machine learning using unbiased data to analyze the key aspects that reduce bias. I plan to focus on presenting the key ideas that must be incorporated by any developer setting out to apply machine learning in the real world.

References

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. 258.

Buolamwini, J. (2019, February 7). *Artificial Intelligence Has a Problem With Gender and Racial Bias*. Time. https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Darby, S. C., Ewertz, M., McGale, P., Bennet, A. M., Blom-Goldman, U., Brønnum, D., Correa, C., Cutter, D., Gagliardi, G., Gigante, B., Jensen, M.-B., Nisbet, A., Peto, R., Rahimi, K., Taylor, C., & Hall, P. (2013). Risk of ischemic heart disease in women after radiotherapy for breast cancer. *The New England Journal of Medicine*, *368*(11), 987–998. https://doi.org/10.1056/NEJMoa1209825

Kataria, T., Bisht, S. S., Gupta, D., Abhishek, A., Basu, T., Narang, K., Goyal, S., Shukla, P., Bansal, M., Grewal, H., Ahlawat, K., Banarjee, S., & Tayal, M. (2016). Quantification of coronary artery motion and internal risk volume from ECG gated radiotherapy planning scans. *Radiotherapy and Oncology*, *121*(1), 59–63. https://doi.org/10.1016/j.radonc.2016.08.006

Murphy, H. (2017, October 9). *Why Stanford Researchers Tried to Create a 'Gaydar' Machine—The New York Times*. https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html

Spetz, J., Moslehi, J., & Sarosiek, K. (2018). Radiation-Induced Cardiovascular Toxicity: Mechanisms, Prevention, and Treatment. *Current Treatment Options in Cardiovascular Medicine*, *20*(4), 31. https://doi.org/10.1007/s11936-018-0627-x

Wang, H., Bai, J., & Zhang, Y. (2008). A normalized thoracic coordinate system for atlas mapping in 3D CT images. *Progress in Natural Science*, *18*(1), 111–117. https://doi.org/10.1016/j.pnsc.2007.08.004

Wennstig, A.-K., Garmo, H., Hållström, P., Nyström, P. W., Edlund, P., Blomqvist, C., Sund, M., & Nilsson, G. (2017). Inter-observer variation in delineating the coronary arteries as organs at risk. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology*. https://doi.org/10.1016/j.radonc.2016.11.007

Whitman, M. (2020). "We called that a behavior": The making of institutional data. *Big Data & Society*, *7*(1), 2053951720932200. https://doi.org/10.1177/2053951720932200

Xue, J., Kubicek, G., Patel, A., Goldsmith, B., Asbell, S. O., & LaCouture, T. A. (2016). Validity of Current Stereotactic Body Radiation Therapy Dose Constraints for Aorta and Major Vessels. *Seminars in Radiation Oncology*, *26*(2), 135–139. https://doi.org/10.1016/j.semradonc.2015.11.001