THESIS

AI FOR PERSONALIZED MEDICINE

Submitted by

Aidan Michael Lewis

Department of Environmental & Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2023

Master's Committee:

Advisor: Susan M. Bailey

Tom J. LaRocca
Erin O. Nishimura

ABSTRACT

AI FOR PERSONALIZED MEDICINE

In 2021, Americans spent an estimated $4.3 trillion dollars on healthcare, an extraordinary amount for treatment that is often less effective than care in other developed nations (*1-3*). Precision, or personalized, medicine represents a new frontier in healthcare that promises treatment plans and optimized health strategies tailored to an individual (*4*) thereby making medicine more effective and less costly. Contemporary Artificial Intelligence (AI) and Machine Learning (ML) approaches have tremendous potential to advance the field of precision medicine by leveraging the technology's power of deciphering patterns in the data to make predictions about an individual's health outcomes (*3, 5-8*). However, many developing AI/ML approaches to precision medicine have not proven particularly successful in making accurate predictions and conclusions mostly due to the limited availability of high-quality medical data for input. The Wake Forrest University Non-Human Primate Radiation Late Effects Cohort (NHP RLEC) provides an unprecedented opportunity to test AI's ability to be trained on a comprehensive human health analog data set in an experimentally irradiated Rhesus monkey cohort with extraordinary life-time records of biomarker levels and health outcomes. Here, we test prevalent, scalable ML models to improve accuracy of predictions specifically related to radiation biomarkers, dose, and health outcomes. We find that CatBoost, ElasticNet, and XGBoost models can accurately predict lymphocyte counts for both monkey populations and individual monkeys. Furthermore, these models can accurately predict radiation dose and

biomarker levels using only five other features within the models. Although the models were only marginally successful at predicting lymphocyte counts using radiation-related data alone, and at predicting the health outcomes of the monkeys, these findings and this unique dataset represent key steps toward refining the combinations of factors necessary for the successful use of AI/ML models in precision medicine for humans.

BIBLIOGRAPHY

1. (Peter G. Peterson Foundation, https://www.pgpf.org/blog/2022/07/how-does-the-us-healthcare-system-compare-to-other-countries, 2022).

2. (U.S. Centers for Medicare & Medicaid Services, 2022), chap. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.

3. (Peter G Peterson Foundation, 2023), chap. https://www.pgpf.org/blog/2023/01/why-are-americans-paying-more-for-healthcare#:~:text=How%20Much%20Does%20the%20United,to%20about%20%2412%2C900%20per%20person.

4. L. H. Goetz, N. J. Schork, Personalized medicine: motivation, challenges, and progress. *Fertility and Sterility* **109**, 952-963 (2018).

5. N. S. Abul-Husn, E. E. Kenny, Personalized Medicine and the Power of Electronic Health Records. *Cell* **177**, 58-69 (2019).

6. L. M. O. Briganti G. (Front Med (Lausanne). , 2020 Feb), vol. 5;7:27.

7. P. Rajpurkar, E. Chen, O. Banerjee, E. J. Topol, AI in health and medicine. *Nature Medicine* **28**, 31-38 (2022).

8. E. J. Fertig, E. M. Jaffee, P. Macklin, V. Stearns, C. Wang, Forecasting cancer: from precision to predictive medicine. *Med* **2**, 1004-1010 (2021).

ACKNOWLEDGMENTS

The Wake Forest Non-Human Primate Radiation Late Effects Cohort (NHP RLEC) is supported by NIH/NIAID (U01 AI150578) and (U19 AI 67798); funding from NASA is also gratefully acknowledged (80NSSC19K0434).

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Background

Context

The laboratory of Dr. Susan Bailey focuses on using AI to advance fields related to radiation health. As such, the work in this thesis was inspired by the former research of Dr. Jared Luxton who used ML to begin classifying how prostate cancer patients (*9*) and astronauts (*10*) respond to radiation therapy and space radiation respectively, using telomere length and chromosome aberrations as informative biomarkers of exposure and potential health outcomes. Although confined to the context of the specific research studies, Dr. Luxton's work provided a valuable framework for the current studies.

This thesis represents next steps and provides highly tested, generalizable, precision medicine ML models to be used within the radiation health field. While the AI developed here needs to be evolved before clinical use, this research builds a foundational platform by providing proof of concept using ML to make personalized health predictions. The limits and uses of different ML models are illustrated in this work, and commentary on how medical AI should be best developed for the radiation health field is posited based on the research findings. Future uses of the AI developed here can be analytical by testing how different variables affect a predicted outcome and assessing which features correlate to specific end states. Future uses can be clinical by adapting the models for diagnostic and mitigative uses in the radiation health field. And future uses can also be technical by providing a benchmark and platform for other AI models to be built and tested against.

Therefore, my goal and hope is that this work provides the Bailey lab (and the rest of the research community) easily accessible tools for implementing AI to further their research objectives. I also hope that organizations such as NASA will use the work here to advance their predictive, personalized medical capabilities to grant future generations the opportunities and wonder of space.

Space – the next frontier

With the increasing popularity and availability of commercial space flight, space offers a promising future not only for exploration and discovery, but also for industry and planetary colonization. However, all such possibilities require prolonged periods of time living and working in the extreme environment of space. Currently, long duration space flight and habitation is precluded by deleterious and largely unknown health effects on the human body. While NASA has declared that the "number one" obstacle for achieving lunar and Martian missions is exposure to space radiation (*11*) there other numerous, well annotated health effects associated with space travel are numerous and include various maladies such as deformation of the eyes, muscular atrophy, bone density loss, increased inflammatory response, neurological fluid redistribution, skin injuries, immune dysregulation, and countless internal biological changes (*12-14*).

While some of the physiological changes to the human body are well appreciated, there are a multitude of molecular changes that occur during spaceflight, as well. The NASA Twins Study and Telomeres investigations (*13*) (*10, 14*) represent foundational studies for evaluating molecular changes associated with long-duration spaceflight, which will be continued with the

sentinel Complement of Integrated Protocols for Human Exploration Research (CIPHER) study. Among the molecular changes that occurred due to space, several well-known biomarkers such as telomere length also respond to spaceflight, sometimes in unexpected and profound ways (*10*). The long-term health implications of such biomarker alterations are still unknown, and the confounding nature of such changes introduces entirely new variables into the risk mitigation strategies of NASA and other space companies when attempting to plan for long duration space missions. Therefore, being able to accurately predict relevant health biomarker dynamics of astronauts, particularly in response to ionizing radiation (IR) exposure, will be critical to unlocking the possibilities that the next generation of space travel has to offer for exploration, habitation, and business opportunities.

Individual radiosensitivity

The difficulty with predicting how a person's health might be affected by exposure to IR is that individuals respond very differently (*12*). Meaning, that with the same IR exposure, two people could manifest very different health effects like cancer, immune system dysfunction, skin injuries, etc. (*11, 12, 15*). Even in cases where IR is being used for therapeutic benefit (e.g., radiation therapy for cancer), patients' response to radiation can vary greatly (*9, 16, 17*). Individualistic responses to IR exposure, and physicians' inability to accurately predict a patient's response, increases the risk of developing radiation late effects (e.g., secondary cancers, cardiopulmonary disease) following radiation therapy (*16-18*); thus, better predictions of a person's individual radiosensitivity is of vital importance to effective treatment planning (*9*). Similarly, in cases of accidental environmental exposures such as those associated with

Three Mile Island, Fukushima, Chernobyl, or other nuclear disasters, accurate radiation dose estimation is critical for effective treatment strategies for exposed individuals (*19*). Occupational radiation exposures (e.g., mining, airline pilots and crews, and astronautics) are also a concern. In all cases, effective treatment measures and mitigation strategies to reduce the long-term health effects of radiation exposure would be better informed by knowing what biomarkers to look for in order to more accurately predict individual radiosensitivity and/or the dose they received.

Traditional statistical methods (such as mixed linear modeling) have been used to try and predict a person's response to radiation. However, to date, these methods have had limited success and are often confined to the context of the study they are implemented in. New generations of computational and analytical power, specifically artificial intelligence (AI) machine learning (ML), provides promise that personalized, precise tools can be developed to better predict how a person will respond to radiation exposure, thus allowing more effective and more powerful mitigation strategies to be developed for exposed populations, including astronauts.

Artificial intelligence

The use and prevalence of AI, or the simulation of human intelligence processes by machines, is rapidly becoming a foundational tool in science as has seen application in a variety of fields including medicine, genetics, and biomarker prediction (*6, 20-23*). The power of AI particularly within the medical field is based on AI's ability to decipher patterns within data that would otherwise go unnoticed by the obtuse speculations of traditional analytical techniques

including the human eye. This property is extremely promising for advancing fields such as personalized medicine, the goal of which is to personalize treatment strategies based on how a particular individual is predicted to respond to difference therapies and health effectors. However, AI has yet to be used for predicting how a person will respond to radiation exposure and is still unused by the space health community.

Many of the recent advancements in AI utilize foundational models such as neural networks (*24-26*) (generally, deep learning models) that require involved software engineering and are based on processes that are still largely unknown, making them less than ideal for many life science research applications. However, there are ML techniques, particularly gradient boosting, that have proven easier to use and are still effective within biological and medical research (*9, 23, 27, 28*). Gradient boosting models are also easier to grasp conceptually, easier to use practically, and easier to scale for people with minimal computer science knowledge. Therefore, while I test and compare different forms of AI within this research project, I focus on the production of robust gradient boosting models for general use and scalability for future generations of research effort and applications of AI in space health.

Reductively, gradient boosting works by making ensembles of decision trees from the data that iteratively step towards predicting data points with increasingly smaller variance from the true values. There are several useful features of gradient boosting that allow for more precise study of how AI can be used within aerospace medicine and health. Notably, these models can describe how important different features within the data are for making accurate predictions, and the analytical parameters that the model uses are easily tuned and modified

for greater accuracy. This reduces the "black box" effect of many AI tools, empowering researchers to refine the data use and computational necessities of AI for specific applications.

The greatest barrier to using any AI within the context of precision medicine, and space health in particular, is that AI models are only as good as the data used to train them (*29*). In its simplest form, AI takes data, finds patterns within that data, and uses those patterns to perform different calculations such as predicting, classifying, and even generating text or images. There are several choke points within this pipeline framework, primarily the amount and quality of data that the AI can use to find patterns. For example, imagine that your friend tells you about a new type of animal that they found in the woods, and you are trying to draw a picture of that animal for National Geographic. If the person telling you about the animal is extremely descriptive, your drawing will be better than if the person can only tell you crude details about the animal. Similarly, if an AI model is given incomplete or very sparse data, the end products it will produce will be equivalently incomplete or inaccurate.

This critical issue is exacerbated when trying to apply AI within the context of human health data, which is highly protected, siloed, and generally not available for robust use by artificial intelligence (*6, 30*). Furthermore, health data from high profile individuals, like astronauts, is even more limited as the population is small and health data is highly protected by NASA and HIPAA regulations.

Non-Human Primate Radiation Late Effects Cohort

The extreme limitations associated with obtaining human health data in general, and the relatively small populations of humans exposed to high doses of ionizing radiation, make it

imperative to build models on data sets from exposed human analogs. While it is recognized that human health analogues are limited in their ability to translate to human health precisely, developing and testing AI strategies on different data sources to inform evolutions of models that directly deal with human data is a powerful approach that is described as one-shot or few-shot learning (*31*).

Thus, the AI models described within this research use a robust, comprehensive, and highly contiguous dataset analogous to human radiation response to build base models that can be elaborated upon with other data (including human health and astronaut data) to make more powerful predictions about human health on Earth and in space. Specifically, the data used for this study is a compilation of rhesus monkey health records and blood panel collections from Wake Forest University's Non-Human Primate Radiation Late Effects Cohort (NHP RLEC). Monkeys in the cohort have had frequent blood analytics performed (approximately every month) and meticulous health records kept while housed at Wake Forrest University. Additionally, the monkeys are of different sexes, ages, species, have experienced various IR exposures at different places prior to their arrival, and they are observed for their entire lifespan. This dataset is unique as it includes both detailed dose information, continuous health assessments, and health outcomes throughout the monkey's lives. Thus, this remarkably rich dataset is ideal for development and testing of AI models and sets a benchmark for future evolutions of medical AI for humans.

Hypothesis

My overall hypothesis is that artificial intelligence (AI) models can be developed to greatly improve predictions of radiation dose, biomarkers, and health outcomes of exposed rhesus monkeys, which will serve as a proof-of-principle demonstration of the translational potential and future application of AI to exposed human populations.

Rational & Intent

The comprehensive nature of the monkey dataset together with the powerful analytical and predictive properties of AI will be used to catalyze creation of AI tools capable of making accurate predictions regarding radiation dose, biomarker dynamics, and health outcomes of the exposed monkeys. As the monkey dataset is analogous to human health within the context of radiation exposure, AI models built upon this extraordinary resource will provide the foundation for further evolutions with potential to direct human health datasets and AI for personalized medicine. This research will provide a proof of principle demonstration of the efficacy of AI within the overall context of personalized medicine, which will have future application to exposed human populations on Earth, as well as astronauts on exploration missions, living and working in space in pursuit of capitalizing on the opportunities of space, and ultimately colonizing other planets. To directly test the hypothesis and develop powerful AI approaches and models for predicting radiation effects, the NHP RLEC data set will be utilized to:

Aim1: Predict IR-induced blood biomarker levels

ML models will be developed to predict the blood biomarkers of both a subpopulation of monkeys, and of a randomly chosen individual monkey. Accurate prediction of the blood biomarker levels would be a substantial step towards predicting the internal health statuses of the monkeys. Predicting blood biomarkers offers a cost effective and highly replicable use of AI within the context of radiation exposure health effects. Being able to predict blood biomarkers lays the groundwork for expanding other biomarkers such as cancer biomarkers, aging biomarkers, and other cellular biomarkers like telomere length.

Aim2: Predict IR-induced health outcomes

ML models will be developed to classify the occurrence of different health outcomes for the individual monkeys. Accurate prediction of the health effects experienced by the monkeys as a result of their radiation exposure paves the groundwork for being able to plan treatment and mitigation strategies for other exposed populations, including humans. These findings and the AI tools developed will also provide a framework for improving and informing powerful personalized medicine approaches. Accurate prediction of IR-induced health outcomes such as cancer, cardiovascular disease (CVD), and neurological diseases, greatly increases the potential effectiveness of prevention, treatment, and early mitigation efforts of otherwise devastating diseases.

Aim3: Predict radiation dose

ML models will be developed to predict radiation dose that a monkey was exposed to using the blood biomarker data. Development of AI tools capable of accurately predicting the

known dose of radiation an individual monkey was exposed to provides a means not only of validating results from other aims of this research, but also establishes a means for quickly generating critical insight into any given IR exposure, whether accidental, intentional, or occupational, thereby informing triage, treatment, and prognosis strategies.

BIBLIOGRAPHY

1. J. Chancellor, G. Scott, J. Sutton, Space Radiation: The Number One Risk to Astronaut Health beyond Low Earth Orbit. *Life* **4**, 491-510 (2014).

2. K. E. Applegate *et al.*, Individual response of humans to ionising radiation: governing factors and importance for radiological protection. *Radiation and Environmental Biophysics* **59**, 185-209 (2020).

3. F. E. Garrett-Bakelman *et al.*, The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science* **364**, (2019).

4. E. Afshinnekoo *et al.*, Fundamental Biological Features of Spaceflight: Advancing the Field to Enable Deep-Space Exploration. *Cell* **184**, 6002 (2021).

5. J. J. Luxton *et al.*, Telomere Length Dynamics and DNA Damage Responses Associated with Long-Duration Spaceflight. *Cell Rep* **33**, 108457 (2020).

6. J.-K. Kang, S. Seo, Y. W. Jin, Health Effects of Radon Exposure. *Yonsei Medical Journal* **60**, 597 (2019).

7. J. J. Luxton *et al.*, Telomere Length Dynamics and Chromosomal Instability for Predicting Individual Radiosensitivity and Risk via Machine Learning. *Journal of Personalized Medicine* **11**, 188 (2021).

8. Y. Zhang *et al.*, Replication Stress: A Review of Novel Targets to Enhance Radiosensitivity-From Bench to Clinic. *Front Oncol* **12**, 838637 (2022).

9. A. M. Buckley, N. Lynam-Lennon, H. O'Neill, J. O'Sullivan, Targeting hallmarks of cancer to enhance radiosensitivity in gastrointestinal cancers. *Nat Rev Gastroenterol Hepatol* **17**, 298-313 (2020).

10. K. Wang, J. E. Tepper, Radiation therapy-associated toxicity: Etiology, management, and prevention. *CA: A Cancer Journal for Clinicians* **71**, 437-454 (2021).

11. J. W. Burnham, J. Franco, Radiation. *Crit Care Clin* **21**, 785-813, vii-viii (2005).

12. L. M. O. Briganti G. (Front Med (Lausanne). , 2020 Feb), vol. 5;7:27.

13. A. C. Holly *et al.*, Towards a gene expression biomarker set for human biological age. *Aging Cell* **12**, 324-326 (2013).

14. G. Hannum *et al.*, Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell* **49**, 359-367 (2013).

15. J. G. Fleischer *et al.*, Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biology* **19**, (2018).

16. C. S. Bang, J. J. Lee, G. H. Baik, Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. *Journal of Medical Internet Research* **22**, e21983 (2020).

17. N. Kriegeskorte, T. Golan, Neural network models and deep learning. *Current Biology* **29**, R231-R236 (2019).

18. H. Yang *et al.*, Lead federated neuromorphic learning for wireless edge artificial intelligence. *Nature Communications* **13**, (2022).

19. Y. Xu, H. Zhang, Convergence of deep convolutional neural networks. *Neural Networks* **153**, 553-563 (2022).

20. M. A. Rguibi, N. Moussa, A. Madani, A. Aaroud, K. Zine-Dine, Forecasting Covid-19 Transmission with ARIMA and LSTM Techniques in Morocco. *SN Computer Science* **3**, (2022).

21. I. Ghosh, P. Dragan, Can financial stress be anticipated and explained? Uncovering the hidden pattern using EEMD-LSTM, EEMD-prophet, and XAI methodologies. *Complex & Intelligent Systems*, (2022).

22. Z. Xiong *et al.*, Facing small and biased data dilemma in drug discovery with enhanced federated learning approaches. *Science China Life Sciences* **65**, 529-539 (2022).

23. I. Dayan *et al.*, Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine* **27**, 1735-1743 (2021).

24. S. Mishra, P. Zhu, V. Saligrama, Interpretable Compositional Representations for Robust Few-Shot Generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-18 (2022).

CHAPTER 2:

USING AI TO PREDICT RADIATION BIOMARKERS, DOSE, AND HEALTH OUTCOMES IN EXPOSED RHESUS MONKEYS[1]

Overview

Individual radiosensitivity, or a person's response to intentional or incidental ionizing radiation (IR) exposure varies considerably (*1*). This personalized response to IR makes it difficult to form treatment strategies for people exposed to IR (such as in radiation therapy, or occupationally as an astronaut) and for those who have already been exposed to radiation (such as an atomic bomb survivor). However, contemporary Artificial Intelligence (AI) and Machine Learning (ML) approaches in the field of precision medicine have potential to help clinicians to formulate radiation treatment and mitigation plans by leveraging AI/ML's power of deciphering patterns in the data to make predictions about an individual's health outcomes (*2-6*). However, many developing AI/ML approaches to precision medicine have not proven particularly successful in making accurate predictions and conclusions mostly due to the limited availability of high-quality medical data for input. The Wake Forrest University Non-Human Primate Radiation Late Effects Cohort (NHP RLEC) provides an unprecedented opportunity to

1.This is a draft of a manuscript being submitted to the Radiation Research Journal as part of a special issue on the NHP RLEC cohort. Author information: Aidan M. Lewis1,2*, Yuiko Chino1*, Thomas E. Johnson1, John D. Olson3, George W. Schaaf3, J. Mark Cline3, Susan M. Bailey1,2†. 1Department of Environmental & Radiological Health Sciences and 2Cell and Molecular Biology Program, Colorado State University, Fort Collins, CO USA. 3Wake Forest School of Medicine, Winston-Salem, NC USA. *Radiation Research Society Scholars in Training. †Corresponding Author

test AI's ability to be trained on a comprehensive human health analog data set in an experimentally irradiated Rhesus monkey cohort with extraordinary life-time records of biomarker levels and health outcomes. Here, we test prevalent, scalable ML models to improve accuracy of predictions specifically related to radiation biomarkers, dose, and health outcomes. We find that standalone CatBoost, ElasticNet, and XGBoost models can accurately predict lymphocyte counts for monkey populations and for individual monkeys with accuracy up to >99%. Furthermore, these models can accurately predict radiation dose and biomarker levels using only five other features within the models with accuracy up to >99%. Although the models were only marginally successful at predicting lymphocyte counts using radiation-related data alone, and at predicting the health outcomes of the monkeys, these findings and this unique dataset represent key steps toward refining the combinations of factors necessary for the successful use of AI/ML models in radiation-related precision medicine for humans.

Introduction

In 2021, Americans spent an estimated $4.3 trillion dollars on healthcare, an extraordinary amount for treatment that is often less effective than care in other developed nations (*4, 7, 8*). Precision, or personalized, medicine represents a new frontier in healthcare that promises treatment plans and optimized health strategies tailored to an individual (*9*) thereby making medicine more effective and less costly. A promising application of precision medicine is prediction and treatment tailoring related to individual radiosensitivity (*1*). Such adverse responses to ionizing radiation (IR) exposure can range in severity from skin injuries and immune system dysfunction, to increased risk of cardiovascular disease and cancer (*1, 10,*

11). The high doses associated with radiation therapy are particularly problematic, as individual differences in radiosensitivity are often unknown prior to treatment and can preclude physicians and oncologists from making rapid diagnoses and effective treatment plans for individual patients.

It is also challenging to predict radiosensitivity when considering occupational exposures and predicting the extent of mitigation required to minimize health effects to individuals. Although extreme, the growing number and diversity of astronauts living and working in the space radiation environment for extended periods of time represents one such category. Indeed, as plans to return to the Moon and to venture on to Mars become reality in the not so distant future, NASA has stated that IR exposure is the "number one risk to astronaut health beyond low Earth orbit (*11*).

A potential solution to the imprecise and unpredictability of IR exposure to humans may be found in the emerging fields of personalized and predicative medicine, which aims to forecast the health outcomes an individual or group of people will sustain based on their exposures, propensities, demographics, etc. (*6*). Artificial Intelligence and Machine Learning (AI/ML) has advanced the personalized and predictive medicine fields by finding correlative patterns in patient data to make highly accurate predictions about the susceptibilities and prognoses of an individual. The predictive properties of AI/ML make these technologies promising candidates for creating risk mitigation strategies and treatment plans for IR exposed patients. For example, we have previously demonstrated the utility of AI/ML tools in applications such as making predictions of how individual prostate cancer patients respond to radiation therapy (*12*). Using telomere length and chromosome aberrations as radiosensitivity

biomarkers, this study demonstrated that XGBoost ML models can accurately predict the biomarker dynamics of radiation therapy patients over the course of their treatment. Here, we expand on our previous studies, to intensify and broaden the capabilities of AI/ML in radiosensitivity health predictions.

However, one of the largest barriers to developing effective AI/ML applications for advancing personalized medicine is that the models require large amounts of high-quality data to make accurate predictions. Medical data for human patients is so highly protected that obtaining enough human health data to create powerful AI/ML models is extremely difficult, especially for early model development and for people/organizations without immense amounts of funding. Here, we overcome current data challenges and to begin exploring the capabilities of powerful AI/ML tools in personalized and predictive medicine by using comprehensive health data from the Non-Human Primate Radiation Late Effects Cohort (NHP RLEC) currently housed at Wake Forrest University. Referring to the cover paper of this special edition, this unique Rhesus monkey dataset is highly valuable for advancing medical AI as there are over 300 monkeys in the cohort with frequent blood analytics (appx. every month) and meticulous health records. Additionally, the monkeys are of different sexes, ages, species, have come from different places before arriving at Wake Forrest, and are observed for their entire lifespan. Furthermore, ~79% of the monkeys have been acutely exposed to full body IR at different ages and doses, creating a health-effector exposure that AI/ML models can attempt to predict the health outcome of.

As reported in Yuiko Chino's paper within this package, concurrent studies have shown that to make informative analyses from the NHP RLEC data using traditional statistical methods

(such as mixed linear modeling) one must drastically reduce the number of monkeys and features to glean any insight into how the monkeys may be responding to the radiation. Additionally, statistical modeling was able to determine the longitudinal changes of only a few biomarkers within the reduced cohort. These methods are not able to make meaningful analyses with large amounts of monkeys, with monkeys of varying age or sex, or with any missing and/or non-numeric variables. And importantly, these methods are unsuccessful at making accurate personalized assessments of a monkey's health and biomarkers. This significant limitation of information going into and coming out of the statistical models is adequate for extremely minute analyses. But to make fast, scalable, individualized insights into the health of monkey/patient, such reductive methods of investigation are infeasible. However, AI/ML ameliorates many challenges of traditional statistical analyses within a health context by offering larger data integration, multivariate analysis, and high throughput for individualized applications.

Therefore, this study tests if CatBoost, ElasticNet, and XGBoost AI models can accurately predict the radiation biomarker levels, the radiation dose, and the health outcome of the NHP RLEC monkeys, as these models have been used successfully previously in related applications (*12-14*), they are effective with minimal engineering, and they are promising contenders for scalable, effective precision medicine AI/ML model. This study verifies the proof of concept that AI/ML can be used for personalized health predictions related to individual radiosensitivity within a structured, small population context. We show that these early models are exceptional at biomarker and dose predictions with accuracy up to >99% and but may need further refinement of health outcome classification before clinical application despite accuracy

and precision as high as >95% in some cases. These developments represent the promising potential of AI/ML in personalized medicine within the radiation field and are foundational for further development and for future advancement towards achieving powerful AI/ML models for human precision medicine.

Methods

Data Processing

The Blood Biomarker data for the monkeys included a complete blood panel for each monkey over frequent timepoints throughout their lifespan. This data spanned from April of 2011 until November of 2021 and data on living monkeys continues to be collected. The demographic breakdown of the monkeys as well as the 18-feature blood panel are shown in (**Figure 1A**). This dataset includes 272 monkeys with 4552 total blood panel collections across all monkeys at all time points. There were 970 collections for non-irradiated monkeys and 3582 collections for irradiated monkeys (**Figure 1B**). The final data was refined by removing collections and/or features with very few data points and removing any monkeys from the data set that showed possibly erroneous information (for example, showing a dose of radiation when marked as a non-irradiated monkey). While the data used here is more structured and complete than what would be expected in real world scenarios, the main objective of the current study was to evaluate the efficacy of the models, which was most readily accomplished with a clean dataset.

For the Health Outcomes dataset, the intent was to reduce data to what could potentially be gleaned from a patient that had been exposed to radiation and the health

outcomes thereafter. Therefore, as seen in **Figure 1A**, features included dose, time since exposure, age at an exposure, basic demographic features of the individual monkey, 18 different health outcomes, and whether the monkey was living or deceased. The health outcome data included 152 different monkeys, 120 that were exposed to radiation and 32 that were not. Any time series features, or natural language notes were removed for simplicity. The different health outcomes were then turned into binary features (i.e. has an individual monkey experienced a particular health effect: yes/no) for the ML models to classify in this preliminary form.

This structured, comprehensive data enables the benchmarking of model performance in this initial stage of testing. Furthermore, the completeness of the data gives us confidence that the variation in model performance is not due to the data, but due to the models themselves, allowing for controlled experimentation of how different ML models perform predicting radiation biomarkers, dose, and health outcomes under different technical conditions.

Machine Learning Models

In this study, we compare the predictive efficacy of three different ML models: CatBoost, XGBoost, and ElasticNet. The data for all predictions were split into 70/30% train/test datasets in unsupervised pooling. Datetime data (date of birth, radiation date, sample date) were omitted for regression and classification predictions. Categorical features for each model were declared to the CatBoost models prior to training and were either converted to binary

(0/1) or omitted for the ElasticNet and XGBoost models depending on the feasibility of conversion.

CatBoost: We developed CatBoost gradient boosting ML models as CatBoost has been shown to calculate more accurate predictions, operates faster, requires less parameter tuning, and uniquely allows for the implementation of non-numerical data (such as sociodemographic data) in the models (*15*). To predict blood biomarker levels in the monkeys, we utilized the CatBoostRegressor method as the biomarker levels, such as lymphocyte percentages, were measured as a continuous variable. For health outcome predictions, we utilized the CatBoostClassifier method as we transformed the health outcome data into binary class data. The learning rate for both the CatBoostRegressor and CatBoostClassifier models were set to 0.5 and iterations ranged from 1 – 1000. All other parameters were left at default unless using Optuna optimization.

ElasticNet: ElasticNet was developed to enhance the learning efficacy of regression ML, and as such, ElasticNet has previously shown to make robust biomarker predictions (*16-21*). However, ElasticNet does not have an endogenous classification feature. Therefore, ElasticNet was not used to make predict health outcomes but was used to predict biomarker levels and dose. The models used an alpha of 1.0 and l1_ratio of 0.5 which are recommended for problems such as these (*22*).

XGBoost: To predict blood biomarker levels in the monkeys, we utilized the XGBoostRegressor method as the biomarker levels are measured as a continuous variable. For health outcome predictions, we utilized the XGBoostClassifier method as we transformed the health outcome data into binary class data. The n_estimators parameter for both the

XGBoostRegressor and XGBoostClassifier models was set to 10 and "seed" parameter set to 123. For regression, the "objective" was set to "reg:squarederror" and for classification, the "objective" was set to "binary:logistic" as these parameters are common to use for similar problems out of the box (*23*). All other parameters were left at default.

Feature Importance

To fine-tune and reduce the "black box" effect of these models, we utilized CatBoost's internal feature importance function (*24*) to calculate which feature(s) influenced the predictions the most. This function ranks the features based on how much, on average, the predictions change if the given feature is changed. The sum of the feature importance values is normalized, and the total of all feature importance values sum to 100. While similar information can be gleaned from XGBoost and ElasticNet, such calculations with these models would require transforming the categorical features which would skew the determined importance for said features. Therefore, we used CatBoost feature importance to get the most accurate assessment for all features in the datasets (**Sup. Fig. 1-4).** By conducting feature importance analysis, we were able to assess how the data for the models can be refined but produce accurate models. In a clinical or field setting, this reduction of features for faster, cheaper diagnostics and screening. Additionally, by observing the correlative patterns within the feature importance and predictions, one can begin to test the biological mechanisms of the prediction outcomes.

Model Evaluation

For regression, models were evaluated on r-squared scores (R2) and mean squares error (MSE). R2 scores are commonly used for regressions in ML and in statistics as they describe the variance within the data (*25*), where the more variance in these predictions means the less accurate the models are performing. MSE scores are widely used in the ML field (*26*) as they describe on average how much error there is between the predicted data and actual data. MSE also accounts for greater deviances from the true value; the lower the MSE scores, the more accurate the model.

For classification, the models were evaluated based on their Accuracy and Prediction scores, which are widely used in classification ML problems (*27*). Accuracy is defined as how often the models are correct, where precision describes how effective the models are at classifying the different outcomes.

Model Optimization

For autonomous parameter optimization, we utilized Optuna with our CatBoost models. CatBoost recommends using the Optuna as they support endogenous integration with Optuna, and Optuna has been used extensively to tune ML models (*28-30*). In summary, hyperparameter tuners such as Optuna operate by testing different combinations of parameters until the most effective parameter set is found. Optuna was set to 50 n_trials and to timeout after 600 seconds. The parameters given to search through can be found on the GitHub repository for this research. As the optimized CatBoost models did not outperform the hand tuned models (**Fig. 2B**) and because all models were performing with comparable high

accuracy, the XGBoost and ElasticNet models were left unoptimized with the parameters stated above.

Results

Basic Observations

We found no significant differences in biomarker distributions among the different populations of monkeys (**Figure 1B, 1C**). If there had been a drastic, population-wide impact of radiation, sex, age, or time since irradiation on the different biomarkers, we would have expected significantly different patterns in the biomarker distributions. These patterns were also seen in concurrent studies which struggled to find significant differences between biomarker distributions between monkey demographics using traditional statistical methods as reported by Chino et al.

Due to the homogenous distributions of biomarkers between monkey populations, ML was chosen to make biological predictions about these monkeys as ML focuses on assessing the deeper patterns within the data to make individualistic predictions that other analysis techniques may struggle to do.

Multi Monkey Predictions

When trained on the entirety of the Blood Biomarker dataset excluding the 30% “test” lymphocyte data, regression ML was able to accurately predict the test lymphocyte percentages. While these initial models make little commentary on the biological mechanisms driving the changes in biomarker levels, these results reaffirm previous observations that

lymphocyte counts are correlated to IR response. Although, based on the feature importance analysis of these predictions (**Sup. Fig. 1-4**) one can begin to visualize patterns of biomarker responses to begin testing how biomarkers are interlinked within individual radiosensitivity. While the models can perform very well in predicting biomarker levels, the models need to be tuned to make accurate predictions. For example, holding a learning rate of 0.5, CatBoost regression models perform with R2s of 0.64, 0.98, and 0.99 respective to 1 iteration, 10 iterations, and 1000 iterations, also showing MSEs of 86.36, 4.74, and 2.30 (**Figure 2A**).

The best performing out of the box CatBoost model had a learning rate of 0.5 and was trained on 100 iterations, achieving an R2: 1.0 (0.99+) and MSE: 0.84 (Figure 1B). We also attempted to optimize the model tuning with the Optuna library due to its prevalent use and seamless integration with CatBoost. Using automated parameter optimization, the CatBoost model was able to predict multi monkey lymphocyte counts with an R2: 0.99 and MSE: 1.58. Therefore, due to simplicity and for the slightly better performance, CatBoost models were trained with learning rates of 0.5 and for 1000 iterations with early stopping rounds of 5 iterations (meaning, if the model started performing worse before 1000 iterations were complete, it would stop training 5 iterations after performance began to get worse) to get the best performance while not overfitting the models.

ElasticNet models were able to accurately predict multi monkey lymphocyte counts accurately with an R2; 0.97 and MSE: 6.83. XGBoost models were also able to accurately predict multi monkey lymphocyte counts with R2: 0.99 and MSE: 2.49 (**Figure 2C, 2D**). These data show that even with minimal optimization tuning, ML models make highly accurate biomarker

predictions even when performing slightly lower than the CatBoost models for these predictions.

Also of note, we tested the model accuracy with predicting all features included in the Blood Biomarkers dataset and while the models performed highly accurate for many of the features, the variation in accuracy between the feature predictions was evident (**Sup. Fig. 1-6**). Interestingly, the three models performed differently between the feature predictions, and showed variance in which features they performed best on. This indicates that while models can be tuned to be generally good at predicting many different biomarkers, being able to apply specific models for different predictions may prove to be the most effective means of biomarker prediction.

Single Monkey Predictions

Personalized medicine AI models will need to be trained on many individuals' data while predicting the outcomes of a single patient. Therefore, in addition to predicting the biomarker of from an entire test data set, we also attempted to predict the biomarkers of a single, randomly chosen monkey. Of clinical importance, the models were able to predict a single monkey's lymphocyte counts when the training data excluded the target monkey's data (**Figure 2B, 2C, 2D**). CatBoost models were able to predict the single monkey lymphocytes with an R2: 0.99 and MSE: 0.77 (Figure 2B). ElasticNet models were able to predict the single monkey lymphocytes with an R2: 0.99 and MSE: 0.54 (**Figure 2C**). XGBoost models were able to predict the single monkey lymphocytes with an R2: 0.98 and MSE: 0.82 (**Figure 2D**). Analyzing the minute differences in performance between the three models would only prove to be

speculative. The key takeaway is that in a preemptive, personalized approach to risk mitigation, these models can be extremely effective.

Reduced Features

It is important to know which features in the data set influence predictions the most both for biological and clinical relevance. Therefore, using the built-in feature importance calculation tool within CatBoost, we elucidated the most important features for making lymphocyte predictions both for all the monkeys and for the single monkey using the Blood Biomarkers dataset (**Sup. Fig. 1-4**). Of note, feature analysis of the single monkey's predictions showed some similarities to that of the entire monkey cohort, but also differences both in feature importance and magnitude of importance which is to be expected due to individuality of radiation response, emphasizing the need for continued research and development of AI for personalized medicine applications.

For the multi monkey predictions, the five most important features were neutrophil counts (NEUT_(%)), the absolute lymphocyte counts (ABS_LYMPH_(/uL)), the absolute monocyte levels (ABS_MONO_(/uL)), the eosinophil counts (EOS_(%)), and hemoglobin levels (Hemoglobin_(HGB_g/dL)). However, lymphocyte percentage counts directly rely on absolute lymphocyte counts, so the sixth most important feature, monocyte counts (MONO_(%)), was included in the training instead. These top features reinforce the correlation of leukocytes in IR response (especially between neutrophils and lymphocytes), but these findings also suggest that there are other, more subtle patterns between lymphocytes other cell types like eosinophils and monocytes that influence radiosensitivity. While beyond the scope of this

study, these results suggest a central mechanism differentially affecting leukocytes in the radiation response pathway, some more similarly than others.

As seen in (**Figure 3A**), the models were still able to make accurate lymphocyte count predictions using just five features. The CatBoost model achieved an R2: 0.99 and MSE: 2.67, the ElasticNet model achieved an R2: 0.97 and MSE: 7.44, and the XGBoost model achieved an R2: 0.99 and MSE: 2.97.

For the single monkey predictions, it was found the five most important features were neutrophil counts (NEUT_(%)), the type of study the monkeys came from (TYPE), the absolute lymphocyte counts (ABS_LYMPH_(/uL)), the eosinophil counts (EOS_(%)), and red blood cell counts (RBC_(M/uL)). However, the absolute lymphocyte counts feature was replaced with the sixth most important feature, hematocrit levels (Hemocrit_(HCT_%)) Additionally, because the type of study the monkeys came from is a categorical feature the ElasticNet and XGBoost models could not be trained on that information, and thus, the seventh most important feature, Mean_corpuscular_hemoglobin_concentration_(MCHC_g/dL) was included instead. The difference between the multi monkey and single monkey most important features for lymphocyte prediction reaffirm that while individuals may have similar characteristics to a larger population's radiation response, it is critical to understand an individual's radiosensitivity. For example, the multi monkey top 5 features included many leukocytes, whereas the single monkey top 5 features included more erythrocyte variables. While beyond the scope of this study, these findings suggest that there may be separate IR response mechanisms between leukocytes and erythrocytes that may help account for individual radiosensitivity.

Even with only 5 features, the models were able to make accurate single monkey lymphocyte predictions (**Figure 3B**). The CatBoost model achieved an R2: 0.74 and MSE: 18.46, the ElasticNet model achieved an R2: 0.98 and MSE: 1.29, and the XGBoost model achieved an R2: 0.97 and MSE: 2.16.

Additionally, we tested to see if biomarkers could be predicted using only minimal demographic and radiation exposure data. The only training data used were the monkey's ID, whether or not they had been irradiated (IRRAD), their age in months at the time of exposure (MONTHS_AT_IRRAD), the dose (DOSE_(Gy)), age (AGE), sex (SEX), days since exposure (DAYS_SINCE_IRRAD), months since exposure (MONTHS_SINCE_IRRAD) and lymphocyte counts (LYMPHOCYTES_(%)).

We found that the models could not make accurate lymphocyte count predictions using only radiation-related data as the best model was with XGBoost achieving an R2: 0.2 and MSE: 192.54 (**Figure 3C)**. However, this result also reinforces that the models are not overfitting by learning the different monkeys' radiation regimens. If this were the case, these models would perform with high accuracy in this instance. These results are also expected due to individual radiosensitivity. That is, if all monkeys had lymphocytes that responded to radiation similarly, then the models would be able to perform well with only radiation related data. Thus, these findings reiterate that some minimal context of an individual's IR response is needed for these models to perform well, perhaps using a subset of the top 5 features described previously.

Radiation Dose Predictions

Both to address another clinical application of these models and to provide a method of orthogonal verification, we tested the models' efficacies predicting the radiation dose (Gy) monkeys were exposed to. The models performed variably well with CatBoost models able to achieve an R2: 0.95 and MSE: 0.4. XGBoost models were able to achieve a comparable but slightly less accurate R2:0.93 and MSE: 0.52. However, ElasticNet models could not make accurate dose predictions and achieved an R2: 0.28 and MSE: 5.7 (**Figure 3D**).

The difference between the performance of ElasticNet and the other two models is most likely due to ElasticNet's inability to use features such as the age of monkeys at a radiation (MONTHS_AT_IRRAD) and other radiation related data as these features include None-type variables. CatBoost and XGBoost can handle such data whereas ElasticNet cannot without imputation that would skew the interpretability of the results. This is important because the top features of these predictions (**Sup. Fig. 1**) require features that involve None-type data (such as age and time since exposure) giving ElasticNet an immediate disadvantage on predictions. Interestingly, the most accurate dose predictions showed that erythrocyte features (e.g. mean corpuscular hemoglobin, red blood cell count, etc.) were more important for predictions than leukocyte features, contradicting a predominant belief in the field. While out of the scope of this study, based on the most important features for dose predictions, there may be temporal interchange between leukocyte and erythrocyte response to radiation.

Health Status Classification

In addition to predicting radiation biomarkers, there is great clinical significance in predicting increased risk and/or occurrence of disease, including cancer. With radiation

exposure, the onset (or reoccurrence) of tumorigenesis is of great concern and having the ability to predict if an exposed person will develop cancer would significantly enhance the efficacy of proactive mitigation strategies.

To test machine learnings ability to make health outcome predictions, Catboost and XGBoost classification models were developed to train on the Health Outcomes dataset and predict if the monkeys experienced different health effects. To examine the versatility of the models and data, predictions were made when trained on the entire data set, only with the top five important features, and only radiation-related data (**Figure 4**).

These models were somewhat effective, with accuracy up to 80% but with precision as low as 14%. Based on these insights, the models are better at classifying when a monkey will not develop tumors than when they will. These finding may not be immediately beneficial in a clinical or risk mitigation context. However, these models are solely based on the health outcomes the monkeys' experienced, radiation data, and the age of the monkeys', making the training data very limited and not entirely representative of all the data available for these models. That is, it is possible that combining the Blood Biomarker and Health Outcome dataset would provide more complete data for these models to train on, thus improving their efficacy.

Of note, we use these models to classify every health outcome of the monkeys and the predictions varied based on the health outcome with up to 96% accuracy and up to 95% precision (**Sup. Fig. 7**). These results show that even at a minimalistic level ML models can make accurate health outcome predictions. Although, it is recommended that models be custom tuned for the specific function and outcome one is attempting to predict.

Discussion

The fact that these ML models were able to make accurate predictions underscores the importance of high-quality data for developing medical AI. Even in this limited study, being able to make health-related predictions with a "patient" cohort of 272 is significantly smaller than the hundreds of petabytes of data some commercial AI companies are amassing for the same purpose. This means, small cohorts of individuals and organizations – such as NASA – can feasibly create medical AI that makes health predictions within the context of risk mitigation and treatment planning without accumulating massive amounts of exogenous data.

As such, the importance of the current research is in creating the framework not only for retrospective estimation of radiation exposure and dose, but also for prospective prediction of radiation biomarkers and long-term health effects/outcomes for exposed individuals. Relevant applications include radiotherapy patients, accidentally exposed individuals (e.g., Three Mile Island, Fukushima, and Chernobyl accidents), and occupationally exposed individuals, including astronauts. For example, imagine you want to become an astronaut. To ensure you are not susceptible for detrimental radiosensitivity, an organization like NASA can use the concepts and models here to predict how your biomarkers will respond to the radiation you may be exposed to as a proxy of your overall health. Alternatively, if you lived near the Fukushima Daiichi Nuclear Power Plant accident, you were likely exposed to radiation, and knowing how much, if any, IR you sustained is critical for a physician's treatment planning, By collecting a few highly predictive radiation biomarkers such as the top 5 features in this study, and running them through AI models based on those above a clinician could gain a good sense of how much radiation you were exposed to, making any treatment triage more effective.

The difference between model performance in this study was oftentimes negligible. However, focusing on the scalability of the models, ElasticNet's inability to process None-type values and categorical features makes the model extremely specific for uses like the biomarker-related predictions in this study. XGBoost can process None-type values which expands its scalability considerably, but in a medical setting, many of the features of a model will be categorical in nature, and thus, XGBoost would be limited in its utility due to the model's inability to process categorical data (while there are workarounds for this issue, the effort and convolution involved to implement these loopholes on a large-scale is significant). CatBoost therefore offers the best solution of the three models tested here as it performs well, can process the data provided, and offers the best route to scale.

In subsequent iterations of this work, assembling other data into these models that represent a more holistic view of an individual will be an important aspect. Potentially powerful data to include would be open-source data particularly curated by government agencies such as NASA (if studying how space radiation effects health), the NIH (if observing radiations effects on tumorigenesis and cancer treatment), or other organizations like the Department of Energy (if using these models to predict the effects of occupational radiation exposure). Specifically, including other potent radiation biomarkers like telomere length would let the models learn how the different biomarkers interplay in radiation response, thus increasing model performance. Additionally, including as much human data as possible and selecting for data that has overlapping features with the NHP RLEC datasets to observe the efficacy of translating these health-analog trained AI systems to human health would also improve getting closer to the ideal data and use cases. While the combination of data would need to be carefully

considered and vetted, the potential broadening and enrichment of insights gained by these models far outweigh any potential pitfalls that may occur in the data confluence.

A potential way to create these ensembles of powerful models, and to also test the efficacy of next generation edge computing in the medical field, would be to implement a federated learning framework for these models. Federated learning allows for faster, more energy efficient, and more secure ways of training AI/ML models on data (*31*), thus providing solutions to the computational and privacy considerations of using AI for personalized medicine. As the models are currently established, the computational requirements and data acquisition barriers would make the described approach to developing medical AI/ML infeasible for robust clinical application.

Furthermore, the models described here are highly accurate for post hoc and ad hoc applications, but their power to make predictions in the future (i.e. forecasting) is relatively nonexistent. Forecasting the potential health outcomes a person might experience is an extremely powerful technological advancement, and therefore, a future elaboration upon the current study is to incorporate time series forecasting models ((S)ARIMA(X) (*32*), Prophet (*33*), etc.) to test the limits and potential uses of forecasting health outcomes.

Another potential enrichment of these models would be incorporating different tools to assess causal inference (*34*) such as the DoWhy, Causallib, or CRISP libraries. While the models illustrated are excellent at pattern recognition and prediction/classification, they make very little commentary on the causality of the radiation exposures or other factors in the end health outcomes. Thus, to probe the significance of the effects of certain environmental exposures

such as radiation on the end health outcomes a patient experiences it might be beneficial to incorporate such causal inference assessments.

Lastly, while the described models were used for their known success and simplicity of use, it is possible that the scaling of medical AI tools to address the challenges exhibited here would require more cutting-edge techniques. Such examples could be deep learning models (*35*) and transformer type models made famous by ChatGPT (*36*). While these methods are relatively untested in the medical field, both have proven their efficacy and other areas outside of radiological medicine and could further the field of medical AI as well. Overall, the intent of the previously delineated models and research reported here is to provide the groundwork for the plethora of future directives and evolutions of AI in personalized medicine for human health applications.

BIBLIOGRAPHY

1. K. E. Applegate *et al.*, Individual response of humans to ionising radiation: governing factors and importance for radiological protection. *Radiation and Environmental Biophysics* **59**, 185-209 (2020).

2. N. S. Abul-Husn, E. E. Kenny, Personalized Medicine and the Power of Electronic Health Records. *Cell* **177**, 58-69 (2019).

3. L. M. O. Briganti G. (Front Med (Lausanne). , 2020 Feb), vol. 5;7:27.

4. (Peter G Peterson Foundation, 2023), chap. https://www.pgpf.org/blog/2023/01/why-are-americans-paying-more-for-healthcare#:~:text=How%20Much%20Does%20the%20United,to%20about%20%2412%2C900%20per%20person.

5. P. Rajpurkar, E. Chen, O. Banerjee, E. J. Topol, AI in health and medicine. *Nature Medicine* **28**, 31-38 (2022).

6. E. J. Fertig, E. M. Jaffee, P. Macklin, V. Stearns, C. Wang, Forecasting cancer: from precision to predictive medicine. *Med* **2**, 1004-1010 (2021).

7. (Peter G. Peterson Foundation, https://www.pgpf.org/blog/2022/07/how-does-the-us-healthcare-system-compare-to-other-countries, 2022).

8. (U.S. Centers for Medicare & Medicaid Services, 2022), chap. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.

9. L. H. Goetz, N. J. Schork, Personalized medicine: motivation, challenges, and progress. *Fertility and Sterility* **109**, 952-963 (2018).

10. J.-K. Kang, S. Seo, Y. W. Jin, Health Effects of Radon Exposure. *Yonsei Medical Journal* **60**, 597 (2019).

11. J. Chancellor, G. Scott, J. Sutton, Space Radiation: The Number One Risk to Astronaut Health beyond Low Earth Orbit. *Life* **4**, 491-510 (2014).

12. J. J. Luxton *et al.*, Telomere Length Dynamics and Chromosomal Instability for Predicting Individual Radiosensitivity and Risk via Machine Learning. *Journal of Personalized Medicine* **11**, 188 (2021).

13. J. G. Fleischer *et al.*, Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biology* **19**, (2018).

14. N. Safaei *et al.*, E-CatBoost: An efficient machine learning framework for predicting ICU mortality using the eICU Collaborative Research Database. *PLOS ONE* **17**, e0262895 (2022).

15. J. T. Hancock, T. M. Khoshgoftaar, CatBoost for big data: an interdisciplinary review. *Journal of Big Data* **7**, (2020).

16. A. C. Holly *et al.*, Towards a gene expression biomarker set for human biological age. *Aging Cell* **12**, 324-326 (2013).

17. Y. Zhang *et al.*, DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nature Communications* **8**, 14617 (2017).

18. C. Xu *et al.*, A novel strategy for forensic age prediction by DNA methylation and support vector regression model. *Scientific Reports* **5**, 17788 (2015).

19. S. Horvath, DNA methylation age of human tissues and cell types. *Genome Biology* **14**, R115 (2013).

20. M. J. Peters *et al.*, The transcriptional landscape of age in human peripheral blood. *Nature Communications* **6**, 8570 (2015).

21. G. Hannum *et al.*, Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell* **49**, 359-367 (2013).

22. Brownlee, Jason. (Machine Learning Mastery, 2020), chap. https://machinelearningmastery.com/elastic-net-regression-in-python/.

23. (dmlc XGBoost, 2022), chap. https://xgboost.readthedocs.io/en/stable/get_started.html.

24. CatBoost. (2023). https://catboost.ai/

25. D. Jain. (Geeks for Geeks, 2021), chap. https://www.geeksforgeeks.org/ml-r-squared-in-regression-analysis/.

26. J. Brownlee. (Machine Learning Mastery, 2021), chap. https://machinelearningmastery.com/regression-metrics-for-machine-learning/.

27. R. Agarwal. (Towards Datascience, 2019), chap. https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226.

28. J.-P. Lai *et al.*, Tree-Based Machine Learning Models with Optuna in Predicting Impedance Values for Circuit Analysis. *Micromachines* **14**, 265 (2023).

29. L. Lin, J. Zhang, N. Zhang, J. Shi, C. Chen, Optimized LightGBM Power Fingerprint Identification Based on Entropy Features. *Entropy* **24**, 1558 (2022).

30. P. Lacerda, B. Barros, C. Albuquerque, A. Conci, Hyperparameter Optimization for COVID-19 Pneumonia Diagnosis Based on Chest CT. *Sensors* **21**, 2174 (2021).

31. H. Yang *et al.*, Lead federated neuromorphic learning for wireless edge artificial intelligence. *Nature Communications* **13**, (2022).

32. M. A. Rguibi, N. Moussa, A. Madani, A. Aaroud, K. Zine-Dine, Forecasting Covid-19 Transmission with ARIMA and LSTM Techniques in Morocco. *SN Computer Science* **3**, (2022).

33. I. Ghosh, P. Dragan, Can financial stress be anticipated and explained? Uncovering the hidden pattern using EEMD-LSTM, EEMD-prophet, and XAI methodologies. *Complex & Intelligent Systems*, (2022).

34. P. K. Parida, T. Marwala, S. Chakraverty, A multivariate additive noise model for complete causal discovery. *Neural Netw* **103**, 44-54 (2018).

35. N. Kriegeskorte, T. Golan, Neural network models and deep learning. *Current Biology* **29**, R231-R236 (2019).

36. Z. Dai, B. Cai, Y. Lin, J. Chen, Unsupervised Pre-Training for Detection Transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-11 (2022).

CHAPTER 3: CONCLUSIONS & FUTURE DIRECTIONS

Findings

This research tested the ability of CatBoost, ElasticNet, and XGBoost machine learning (ML) models to predict radiation biomarker levels, dose, and health outcomes in exposed monkeys within the Wake Forest University Non-Human Primate Radiation Late Effects Cohort (NHP RLEC). In brief, these ML models made highly accurate biomarker predictions both for a multi-monkey test group and for a single monkey. The models could also predict each biomarker of the dataset in turn, which revealed that while the models generally perform well on this type of data, each model performed better with different biomarkers.

Additionally, the CatBoost and XGBoost models could accurately predict the radiation dose monkeys were exposed to, but the ElasticNet model could not. However, the CatBoost and XGBoost models could not make health outcome predictions with precision and accuracy high enough for clinical relevance. This shortcoming may be ameliorated by the incorporation of more comprehensive data about the monkeys in the RLEC, which underscores the intent of developing these models as proof of concept for using artificial intelligence (AI) within a human health context. From a developmental standpoint, CatBoost offers the best platform for scaling AI for medical purposes as it allows for the incorporation of non-numeric variables, None-type variables, and performs comparably if not better than ElasticNet and XGBoost. Importantly, this research shows that AI/ML can surpass traditional statistical modeling (like mixed linear modeling) within the predictive and personalized medicine fields. Such strength comes from ML's ability to elucidate intricate patterns within greater amounts data, multi-variate data, and seemingly homogenous data.

As such, the clinical relevance of this research lies in creating a powerful framework for identifying and refining relevant biomarkers of radiation exposure, predicting individual radiosensitivity, predicting health outcomes of radiation exposure, and estimating radiation exposure dose. Thus, accurate AI models would inform and improve effective treatment measures and mitigation strategies to reduce the long-term health effects of radiation exposure, whether medical or occupation related. AI strategies would be especially useful in instances of accidental or environmental radiation exposures (e.g., Three Mile Island, Fukushima, Chernobyl, or other nuclear disasters), as accurate dose estimation is critical for effective triage and treatment strategies for exposed individuals.

The difference between model performance in this study was oftentimes negligible. However, focusing on the scalability of the models, ElasticNet's inability to process None-type values and categorical features makes the model extremely specific for uses like the biomarker-related predictions in this study. XGBoost can process None-type values which expands its scalability considerably, but in a medical setting, many of the features of a model will be categorical in nature, and thus, XGBoost would be limited in its utility due to the model's inability to process categorical data (while there are workarounds for this issue, the effort and convolution involved to implement these loopholes on a large-scale is significant). CatBoost therefore offers the best solution of the three models tested here as it performs well, can process the data provided, and offers the best route to scale.

Hypothesis and Aims Achievement

My overall hypothesis was that AI models can be developed to greatly improve predictions of radiation dose, biomarkers, and health outcomes of exposed rhesus monkeys, and thereby provide a proof of principle demonstration of the potential and translatability of AI to human personalized medicine. To test this hypothesis, I focused on developing the ML models to achieve the following aims:

Aim1: Predict IR-induced blood biomarker levels

All the ML models were able to predict blood biomarker levels with high accuracy. Both multi monkey and single monkey predictions were successful using both full-feature data and using the top five most important feature data. However, using only the radiation-related data did not provide accurate predictions, which may be enhanced by combining radiation data and a few important blood biomarkers into the training data set.

Aim2: Predict IR-induced health outcomes

The CatBoost an XGBoost models were able to make health outcome predictions with marginal accuracy and precision. The models could be enhanced by integrating more data (such as the blood biomarker data) to give the models more contextual learning features to make more accurate predictions.

Aim3: Predict radiation dose

The dose predictions made by CatBoost and XGBoost models were accurate, but the ElasticNet predictions were not. However, all the predictions could be made more accurate by

reducing the features in the training set and incorporating more data into the models, which will likely be the most efficient way to increase model accuracy.

Conclusions

AI models can be developed to accurately predict health effects in exposed rhesus monkeys. However, the efficacy of the models varied between the use cases and will likely be improved by incorporating more data, and different types of data into the models. Importantly, this research did achieve the objective of laying the groundwork for future developments using AI for personalized medicine in humans.

Future Directions

There are three main takeaways from this research: 1) AI can be developed to make accurate, high throughput, individualized health and radiation predictions; 2) a relatively small amount of data can be used to make accurate health predictions within a defined context; and 3) while promising, the current models can be greatly improved.

On the first point, this research shows that AI can be used for both personalized and population-wide health predictions. This enables physicians and public health officials to make more effective treatment and prevention plans as the correct resources can be applied to remedying an evident or predicted health ailment. As such, this research forges the foundation for ML models to integrate greater data to pinpoint personalized therapies, develop risk mitigation strategies, and screen individuals in high-risk occupations such as hospital workers, nuclear power plant operators, and astronauts. For example, if a person is exposed to an

unknown dose of radiation, AI can predict the dose so that physicians can make faster, more effective decisions to help the patient mitigate any health effects from the exposure.

A major finding of this research is that within a focused context, such as the NASA Astronaut Corps, ML models do not need copious amounts of data to make powerful predictions. This research demonstrates that even with only 272 "patients" AI can make accurate health predictions both for individuals within and outside of the dataset. Thus, for applications with reduced scope – such as astronautics, airline piloting, military screening, etc. – the data that such organizations currently possess might be enough to train AI models that can make powerful predictions within those contexts.

However, as was seen with the marginally accurate and precise health outcome predictions, these models need further refinement before being used for certain applications. Even so, and even in their current forms, being able to make health outcome predictions with limited, binary data is unprecedented within the field. From a development standpoint, there are several additions to these models that can make them more effective for clinical use. The first is incorporating different tools to assess causal inference (*1*) (such as the DoWhy, Causallib, or CRISP libraries). The models in this research make very little commentary on the causality of the radiation exposure or other factors in producing the end health outcomes. Additionally, these models are not effective at making predictions in the future (i.e., forecasting). Thus, integrating time series forecasting models (such as (S)ARIMA(X) (*2*), Prophet (*3*), etc.) will test the limits and potential uses of forecasting health outcomes.

While the models tested here are effective uses of ML in a medical context, the models themselves are not the cutting edge of AI. Therefore, this work also lays the foundation to test

the efficacy of emerging technologies such as edge computing (like federated learning (*4*) and transformer type models made famous by ChatGPT (*5*). If successful, such advancements would drastically expand the power of AI in medicine – and bring new promise for improving personalized human medicine strategies and healthcare outcomes.

BIBLIOGRAPHY

1. P. K. Parida, T. Marwala, S. Chakraverty, A multivariate additive noise model for complete causal discovery. *Neural Netw* **103**, 44-54 (2018).

2. M. A. Rguibi, N. Moussa, A. Madani, A. Aaroud, K. Zine-Dine, Forecasting Covid-19 Transmission with ARIMA and LSTM Techniques in Morocco. *SN Computer Science* **3**, (2022).

3. I. Ghosh, P. Dragan, Can financial stress be anticipated and explained? Uncovering the hidden pattern using EEMD-LSTM, EEMD-prophet, and XAI methodologies. *Complex & Intelligent Systems*, (2022).

4. H. Yang *et al.*, Lead federated neuromorphic learning for wireless edge artificial intelligence. *Nature Communications* **13**, (2022).

5. Z. Dai, B. Cai, Y. Lin, J. Chen, Unsupervised Pre-Training for Detection Transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-11 (2022).

APPENDIX I

List of Figures

A

	Blood Biomarker Dataset
0	ID
1	Species
2	SAMPLE_DATE
3	IRRAD
4	IRRAD_DATE
5	DOB
6	SEX
7	TYPE
8	DAYS_SINCE_IRRAD
9	MONTHS_SINCE_IRRAD
10	MONTHS AT IRRAD
11	DOSE_(Gy)
12	AGE
13	WBC_(K/uL)
14	RBC_(M/uL)
15	Hemaglobin_(HGB_g/dL)
16	Hemocrit_(HCT_%)
17	Mean_corpuscular_volume_(MCV_fL)
18	Mean_corpuscular_hemaglobin_(MCH_pg)
19	Mean_corpuscular_hemaglobin_concentration_(MCH...
20	NEUT_(%)
21	LYMPHOCYTES_(%)
22	MONO_(%)
23	EOS_(%)
24	BASO_(%)
25	ABS_NEUT_(/uL)
26	ABS_LYMPH_(/uL)
27	ABS_MONO_(/uL)
28	ABS_EOS_(/uL)
29	ABS_BASO_(/uL)
30	PLATLET_COUNT_(K/uL)

	Health Outcome Dataset
0	ID
1	Time_Under_Observation_(yrs)
2	Dose_(Gy)
3	Irrad_Age_(months)
4	Years_Since_Irrad
5	Current_Age_(years)
6	Animal_Status
7	Diabetes_(Y/N)
8	Cataracts_(Y/N)
9	Arthritis_(Y/N)
10	Periodontitis_(Y/N)
11	GI_(Y/N)
12	Behavior_(Y/N)
13	Tumor_(Y/N)
14	Heart_(Y/N)
15	Hypertension_(Y/N)
16	Testicular Atrophy_(Y/N)
17	Ovarian Dysfunction_(Y/N)
18	Hepatic_(Y/N)
19	Kidney_(Y/N)
20	Lung_(Y/N)
21	Bone_(Y/N)
22	Overweight_(Y/N)
23	Underweight_(Y/N)
24	Brain Lesion_(Y/N)

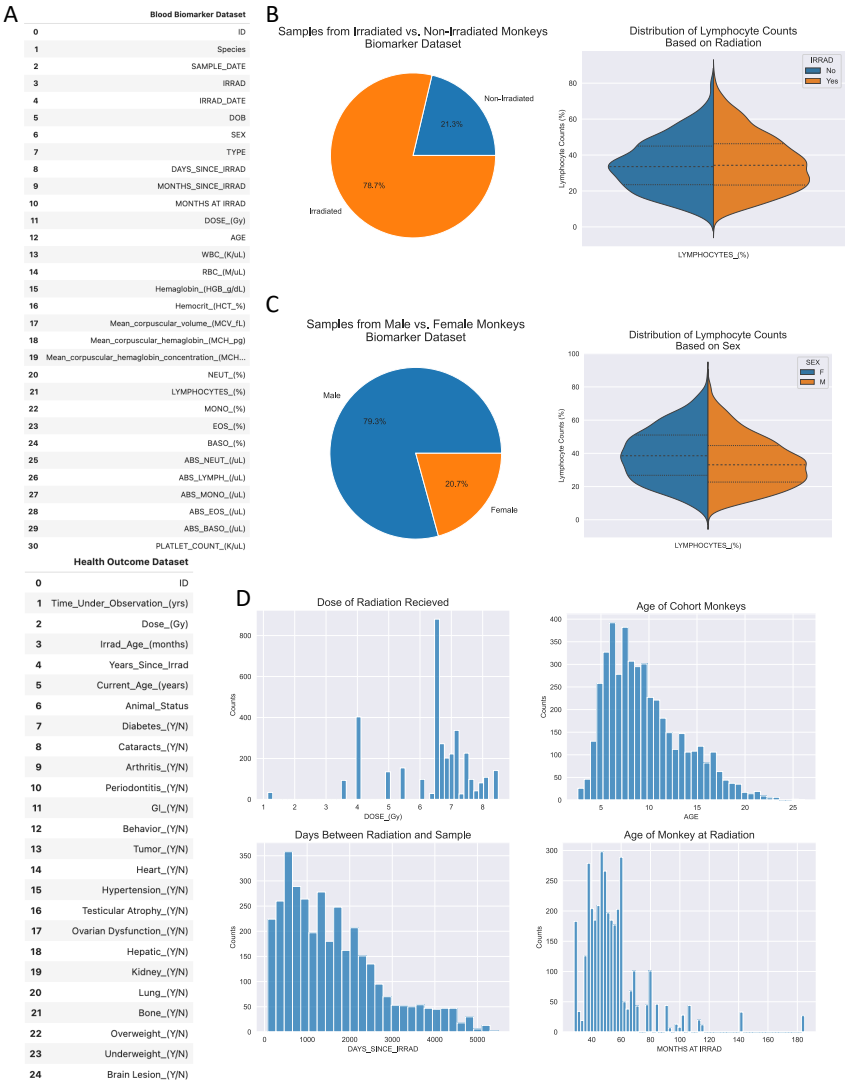


Figure 1: Dataset Features; No Population-level Differences in Biomarker Distributions
(A) Features of the Blood Biomarker & Health Outcome Datasets. The Blood Biomarker dataset includes 4552 samples across 272 monkeys observing 30 different features. The Health Outcome dataset includes 18 different health outcome observations for 152 monkeys and includes radiation-related data and age of the monkeys.
(B) Proportions of samples that are taken from irradiated monkeys (78.7%) and non-irradiated monkeys (21.35%) and the similar distribution of lymphocyte counts between irradiated and non-irradiated samples.
(C) Proportions of samples that are taken from male monkeys (79.3%) and female monkeys (20.7%) and the similar distribution of lymphocyte counts between male and female samples. (D) Histograms of the dose of radiation monkeys were exposed to, the age of the monkeys when samples were taken, the days since exposure that samples were taken at, and the different ages of the monkeys at exposure samples were collected from.

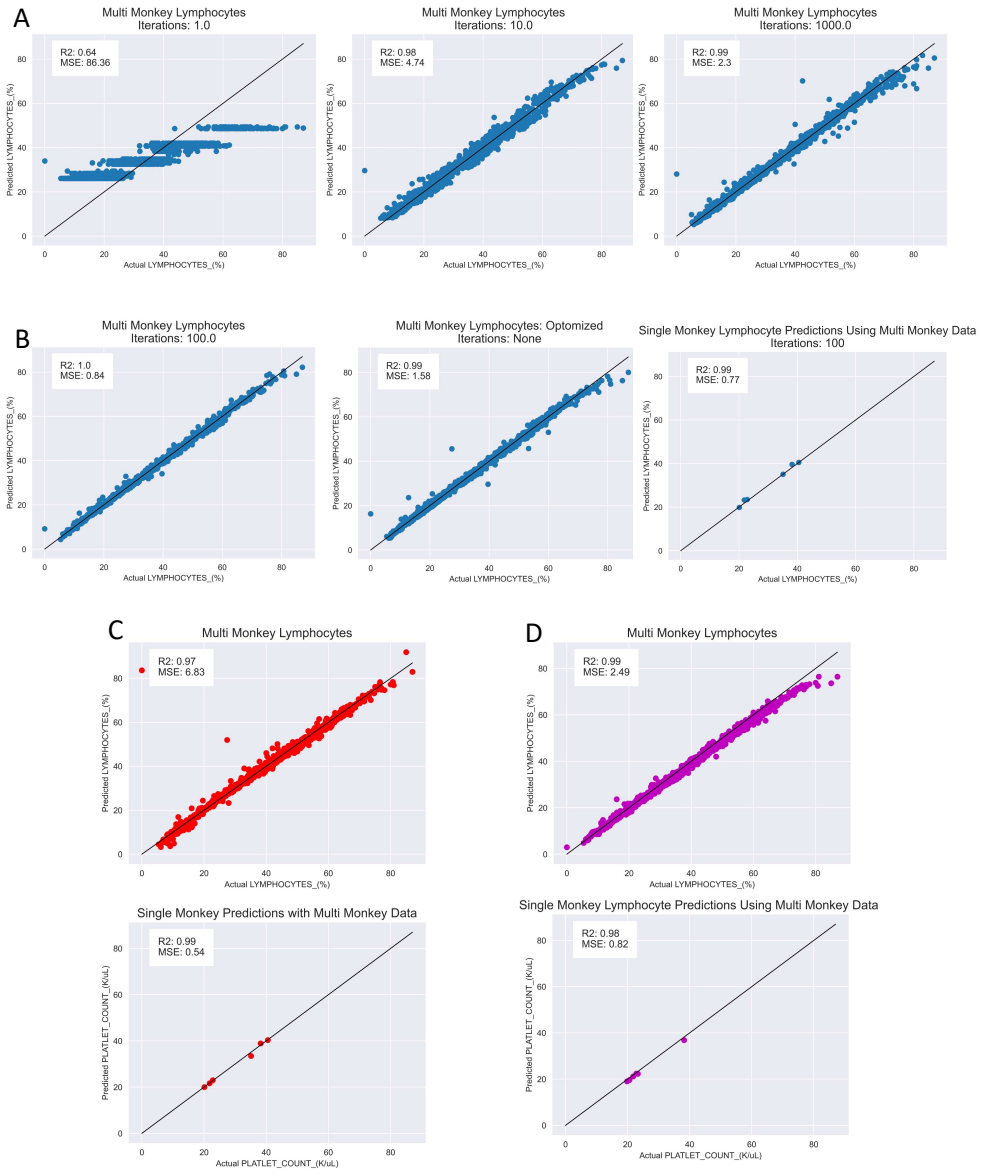


Figure 2: Machine Learning Models Accurately Predict Lymphocyte (%) for a Single and Multiple Monkeys
(A) ML models can be tuned to make more accurate biomarker predictions. Holding other parameters constant, increasing the iterations a CatBoost model trains on (1-1000) increases the R2 score (0.64-0.99) and decreases the MSE (86.36-2.3)(B) The most accurate multi-monkey CatBoost models were trained on 100 iterations and/or were optimized with Optuna, showing R2 of 1.0 and 0.99 and MSE of 0.84 and 1.56. ML models can also predict a single monkey's biomarkers when the training data included all monkeys except the target monkey (R2: 0.99, MSE: 0.77). (C) ElasticNet models were also able to make accurate multi monkey biomarker predictions (R2: 0.97, MSE: 6.63) and single monkey biomarker predictions (R2: 0.99, MSE: 0.54). Performing comparable to CatBoost and XGBoost models out of the box. (D) XGBoost models were also able to make accurate multi monkey biomarker predictions (R2: 0.99, MSE: 2.49) and single monkey biomarker predictions (R2: 0.98, MSE: 0.82). Performing comparable to CatBoost and ElasticNet models out of the box.

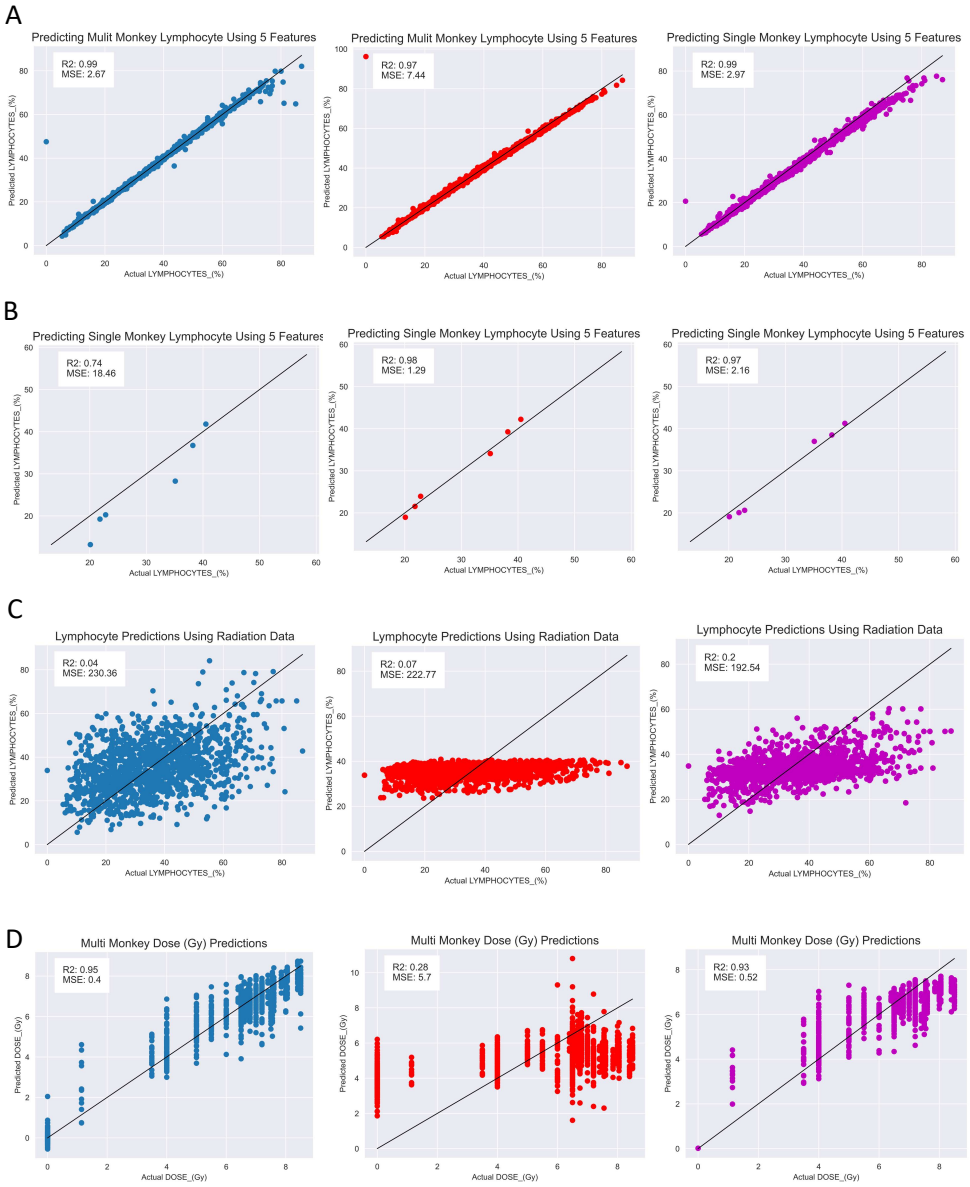


Figure 3: ML Models can Predict Biomarkers Using Reduced Data; Models can also Predict Dose Exposure
(A) CatBoost (1st, blue) models, ElasticNet (2nd, red) models, and XGBoost (3rd, magenta) models accurately predict multi monkey lymphocyte counts using only the top 5 most important and mutually exclusive features. CatBoost (R2: 0.99, MSE: 2.67), ElasticNet (R2: 0.97, MSE: 7.44), XGBoost (R2: 0.99, MSE: 2.97). (B) CatBoost (1st, blue) models, ElasticNet (2nd, red) models, and XGBoost (3rd, magenta) models accurately predict single monkey lymphocyte counts using only the top 5 most important and mutually exclusive features. CatBoost (R2: 0.74, MSE: 18.46), ElasticNet (R2: 0.98, MSE: 1.29), XGBoost (R2: 0.97, MSE: 2.16). (C) CatBoost (1st, blue) models, ElasticNet (2nd, red) models, and XGBoost (3rd, magenta) models were not able to make accurate multi-monkey lymphocyte predictions using only the radiation-related data. CatBoost (R2: 0.04, MSE: 230.36), ElasticNet (R2: 0.07, MSE: 222.77), XGBoost (R2: 0.2, MSE: 192.54). (D) CatBoost (1st, blue) models, ElasticNet (2nd, red) models, and XGBoost (3rd, magenta) models variable make accurate radiation dose predictions. CatBoost (R2: 0.95, MSE: 0.40), ElasticNet(R2: 0.28, MSE: 5.70), XGBoost (R2: 0.93, MSE: 0.52).

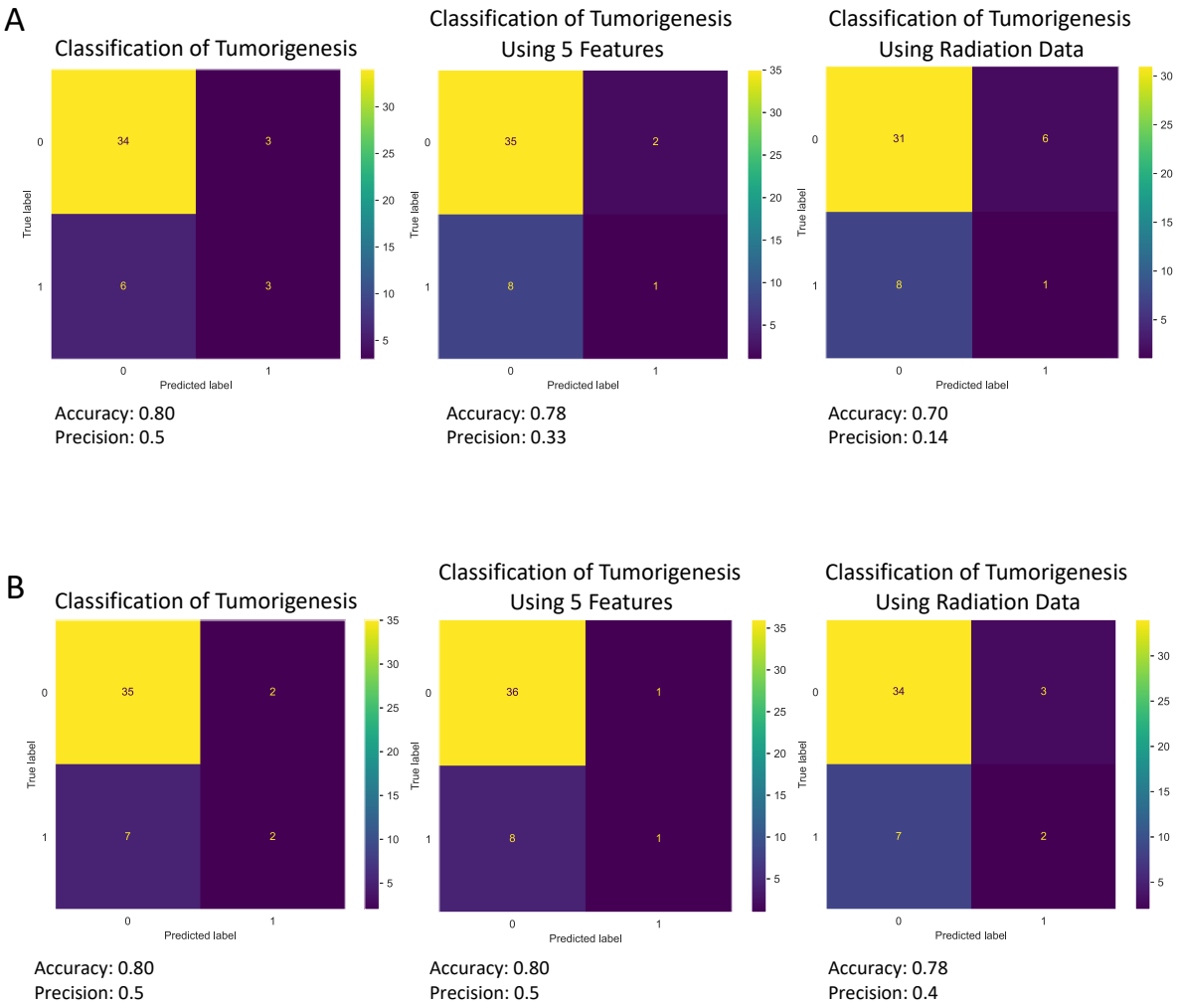
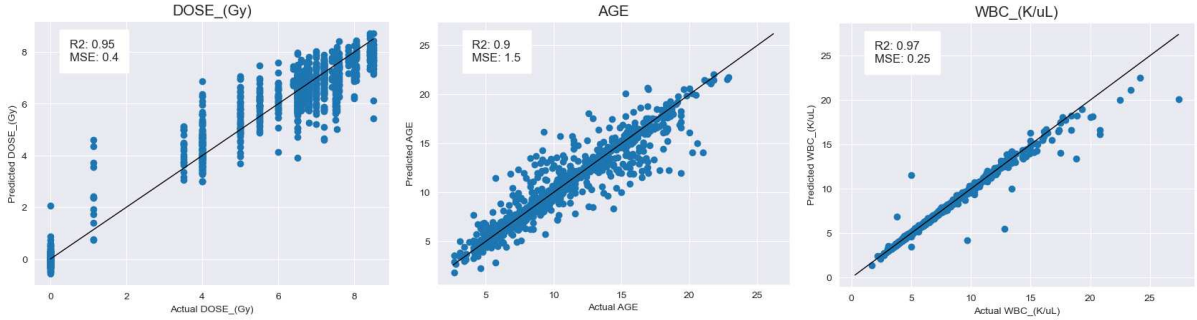


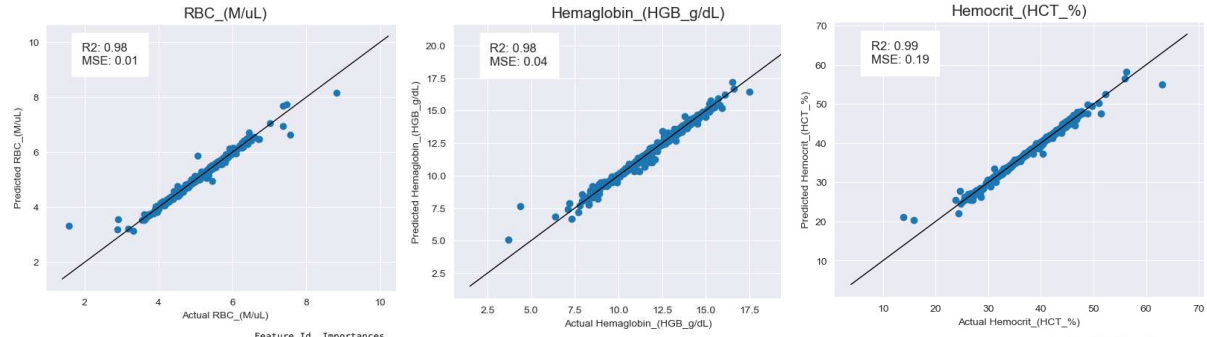
Figure 4: ML Models can Classify Tumorigenesis with Accuracy but not Precision
Confusion matrices show the actual occurrence of tumorigenesis ("True Label", Yes = 1, No = 0) vs. the predicted occurrence of tumorigenesis ("Predicted Label", Yes = 1, No = 0).
(A) CatBoost models can make accurate and somewhat precise classification of tumorigenesis using all the health outcome dataset (Accuracy: 0.80, Precision: 0.50), using the top 5 most important features (Accuracy:0.78, Precision: 0.33), and using just radiation-related data (Accuracy: 0.70, Precision: 0.14). (B) XGBoost models can make accurate and somewhat precise classification of tumorigenesis using all the health outcome dataset (Accuracy: 0.80, Precision: 0.5), using the top 5 most important features (Accuracy: 0.80, Precision: 0.5), and using just radiation-related data (Accuracy: 0.78, Precision: 0.4).



	Feature Id	Importances
0	MONTHS_SINCE_IRRAD	52.838634
1	MONTHS AT IRRAD	25.466664
2	DAYS_SINCE_IRRAD	5.794861
3	SEX	2.094320
4	Mean_corpuscular_hemaglobin_(MCH_pg)	1.550753
5	AGE	1.516966
6	RBC_(M/uL)	1.430748
7	Mean_corpuscular_volume_(MCV_fL)	1.420440
8	PLATLET_COUNT_(K/uL)	0.984636
9	Species	0.953250
10	WBC_(K/uL)	0.778311
11	ABS_EOS_(/uL)	0.727336
12	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.615141
13	ABS_LYMPH_(/uL)	0.535817
14	LYMPHOCYTES_(%)	0.456690
15	Hemocrit_(HCT_%)	0.427133
16	EOS_(%)	0.421901
17	ABS_NEUT_(/uL)	0.399634
18	MONO_(%)	0.381481
19	NEUT_(%)	0.328807
20	ABS_MONO_(/uL)	0.326953
21	Hemaglobin_(HGB_g/dL)	0.304843
22	ABS_BASO_(/uL)	0.139949
23	TYPE	0.060258
24	BASO_(%)	0.044473
25	IRRAD	0.000000

	Feature Id	Importances
0	MONTHS AT IRRAD	30.976083
1	DAYS_SINCE_IRRAD	13.112349
2	DOSE_(Gy)	12.407258
3	MONTHS_SINCE_IRRAD	12.031168
4	Species	10.967439
5	MONO_(%)	2.408904
6	PLATLET_COUNT_(K/uL)	2.148329
7	SEX	2.101515
8	Hemocrit_(HCT_%)	1.917453
9	ABS_LYMPH_(/uL)	1.582021
10	Mean_corpuscular_hemaglobin_concentration_(MCH...	1.224726
11	Hemaglobin_(HGB_g/dL)	1.050046
12	Mean_corpuscular_hemaglobin_(MCH_pg)	1.015374
13	ABS_MONO_(/uL)	0.919227
14	EOS_(%)	0.889778
15	WBC_(K/uL)	0.860899
16	RBC_(M/uL)	0.784481
17	Mean_corpuscular_volume_(MCV_fL)	0.717220
18	LYMPHOCYTES_(%)	0.637392
19	ABS_NEUT_(/uL)	0.594320
20	NEUT_(%)	0.590216
21	ABS_EOS_(/uL)	0.565299
22	ABS_BASO_(/uL)	0.312849
23	BASO_(%)	0.088245
24	TYPE	0.062535
25	IRRAD	0.034875

	Feature Id	Importances
0	ABS_NEUT_(/uL)	67.112243
1	ABS_LYMPH_(/uL)	28.707886
2	ABS_MONO_(/uL)	1.116364
3	DAYS_SINCE_IRRAD	0.935306
4	ABS_EOS_(/uL)	0.891282
5	DOSE_(Gy)	0.263144
6	AGE	0.172229
7	NEUT_(%)	0.118084
8	LYMPHOCYTES_(%)	0.102123
9	RBC_(M/uL)	0.092796
10	MONO_(%)	0.084673
11	ABS_BASO_(/uL)	0.083364
12	MONTHS AT IRRAD	0.054770
13	PLATLET_COUNT_(K/uL)	0.045744
14	Mean_corpuscular_volume_(MCV_fL)	0.044155
15	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.033019
16	EOS_(%)	0.030444
17	Hemocrit_(HCT_%)	0.028011
18	Mean_corpuscular_hemaglobin_(MCH_pg)	0.027300
19	MONTHS_SINCE_IRRAD	0.026309
20	BASO_(%)	0.024123
21	Hemaglobin_(HGB_g/dL)	0.003986
22	TYPE	0.001738
23	SEX	0.000524
24	Species	0.000383
25	IRRAD	0.000000



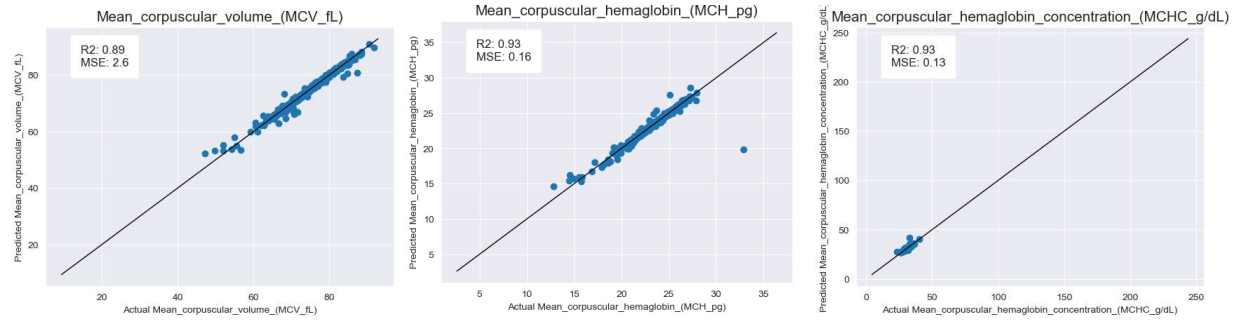
	Feature Id	Importances
0	Hemaglobin_(HGB_g/dL)	33.928279
1	Hemocrit_(HCT_%)	32.666497
2	Mean_corpuscular_volume_(MCV_fL)	20.758052
3	Mean_corpuscular_hemaglobin_(MCH_pg)	10.972927
4	DOSE_(Gy)	0.403739
5	AGE	0.388849
6	PLATLET_COUNT_(K/uL)	0.172791
7	WBC_(K/uL)	0.149662
8	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.140953
9	ABS_EOS_(/uL)	0.067737
10	MONTHS_SINCE_IRRAD	0.065116
11	MONTHS AT IRRAD	0.052553
12	ABS_NEUT_(/uL)	0.045116
13	LYMPHOCYTES_(%)	0.034646
14	ABS_LYMPH_(/uL)	0.032893
15	ABS_MONO_(/uL)	0.028755
16	DAYS_SINCE_IRRAD	0.017495
17	NEUT_(%)	0.016218
18	MONO_(%)	0.015758
19	ABS_BASO_(/uL)	0.013153
20	BASO_(%)	0.011285
21	EOS_(%)	0.010818
22	SEX	0.002793
23	TYPE	0.002587
24	Species	0.001326
25	IRRAD	0.000000

	Feature Id	Importanc
0	Hemocrit_(HCT_%)	60.3093
1	Mean_corpuscular_hemaglobin_concentration_(MCH...	16.3133
2	RBC_(M/uL)	12.1089
3	Mean_corpuscular_volume_(MCV_fL)	6.4489
4	Mean_corpuscular_hemaglobin_(MCH_pg)	2.0192
5	AGE	0.7339
6	WBC_(K/uL)	0.3825
7	DAYS_SINCE_IRRAD	0.3665
8	ABS_NEUT_(/uL)	0.2063
9	MONTHS AT IRRAD	0.1906
10	DOSE_(Gy)	0.1593
11	ABS_MONO_(/uL)	0.1411
12	PLATLET_COUNT_(K/uL)	0.1250
13	ABS_LYMPH_(/uL)	0.1186
14	MONO_(%)	0.1132
15	LYMPHOCYTES_(%)	0.1104
16	ABS_EOS_(/uL)	0.0361
17	EOS_(%)	0.0360
18	MONTHS_SINCE_IRRAD	0.0303
19	NEUT_(%)	0.0251
20	ABS_BASO_(/uL)	0.0215
21	BASO_(%)	0.0011
22	Species	0.0009
23	TYPE	0.0004
24	SEX	0.0003
25	IRRAD	0.0000

	Feature Id	Importances
0	Hemaglobin_(HGB_g/dL)	74.458778
1	Mean_corpuscular_hemaglobin_concentration_(MCH...	15.344311
2	RBC_(M/uL)	6.804278
3	Mean_corpuscular_volume_(MCV_fL)	2.195428
4	AGE	0.329975
5	WBC_(K/uL)	0.226687
6	PLATLET_COUNT_(K/uL)	0.176423
7	ABS_EOS_(/uL)	0.125808
8	DAYS_SINCE_IRRAD	0.082516
9	NEUT_(%)	0.040769
10	MONTHS_SINCE_IRRAD	0.029258
11	ABS_MONO_(/uL)	0.025102
12	ABS_NEUT_(/uL)	0.023787
13	ABS_LYMPH_(/uL)	0.023593
14	EOS_(%)	0.021354
15	Mean_corpuscular_hemaglobin_(MCH_pg)	0.020457
16	MONTHS AT IRRAD	0.020289
17	MONO_(%)	0.015892
18	LYMPHOCYTES_(%)	0.015722
19	BASO_(%)	0.006825
20	ABS_BASO_(/uL)	0.006398
21	DOSE_(Gy)	0.003958
22	Species	0.001330
23	TYPE	0.000731
24	SEX	0.000331
25	IRRAD	0.000000

Supplemental Figure 1: CatBoost Predictions & Feature Importance I

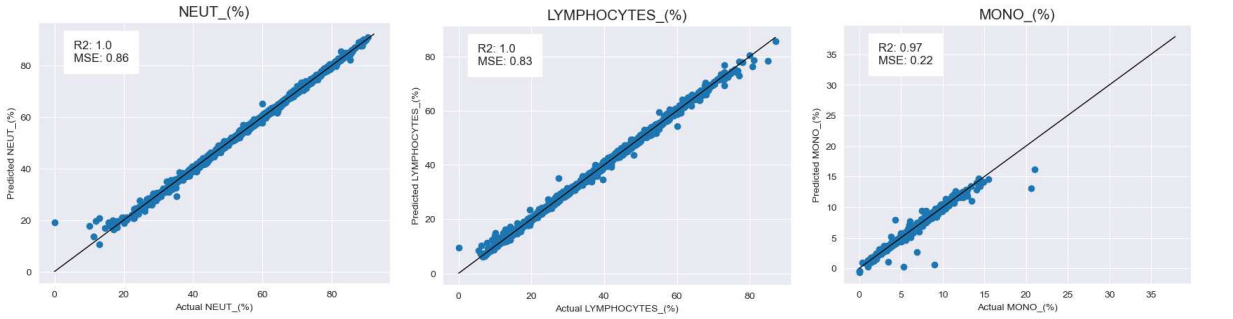
CatBoost regression model predictions of different features in the Blood Biomarker dataset. The calculated feature importance rankings for each feature predicted are below the respected scatterplots.



	Feature Id	Importances
0	Mean_corpuscular_hemaglobin_(MCH_pg)	57.118326
1	Mean_corpuscular_hemaglobin_concentration_(MCH...	34.332994
2	Hemaglobin_(HGB_g/dL)	4.045787
3	Hemocrit_(HCT_%)	1.572033
4	MONTHS_SINCE_IRRAD	1.262768
5	RBC_(M/uL)	0.406412
6	WBC_(K/uL)	0.256374
7	DOSE_(Gy)	0.215242
8	ABS_EOS_(/uL)	0.157035
9	PLATLET_COUNT_(K/uL)	0.114088
10	DAYS_SINCE_IRRAD	0.088020
11	AGE	0.069218
12	ABS_NEUT_(/uL)	0.058818
13	MONO_(%)	0.051697
14	ABS_LYMPH_(/uL)	0.041676
15	LYMPHOCYTES_(%)	0.039685
16	MONTHS AT IRRAD	0.038378
17	ABS_MONO_(/uL)	0.030798
18	BASO_(%)	0.028152
19	ABS_BASO_(/uL)	0.028052
20	EOS_(%)	0.027618
21	NEUT_(%)	0.014569
22	Species	0.001132
23	TYPE	0.000903
24	SEX	0.000223
25	IRRAD	0.000000

	Feature Id	Importances
0	Mean_corpuscular_volume_(MCV_fL)	59.945230
1	Mean_corpuscular_hemaglobin_concentration_(MCH...	35.776738
2	Hemaglobin_(HGB_g/dL)	1.053294
3	WBC_(K/uL)	0.744174
4	SEX	0.734496
5	AGE	0.467668
6	LYMPHOCYTES_(%)	0.363818
7	ABS_BASO_(/uL)	0.174553
8	Hemocrit_(HCT_%)	0.099139
9	RBC_(M/uL)	0.094960
10	MONO_(%)	0.081559
11	PLATLET_COUNT_(K/uL)	0.079134
12	ABS_MONO_(/uL)	0.067147
13	DAYS_SINCE_IRRAD	0.055925
14	EOS_(%)	0.037609
15	ABS_NEUT_(/uL)	0.035132
16	MONTHS_SINCE_IRRAD	0.033546
17	ABS_LYMPH_(/uL)	0.032678
18	NEUT_(%)	0.032262
19	DOSE_(Gy)	0.027057
20	ABS_EOS_(/uL)	0.021088
21	MONTHS AT IRRAD	0.020170
22	BASO_(%)	0.019313
23	TYPE	0.001773
24	Species	0.001538
25	IRRAD	0.000000

	Feature Id	Importances
0	Hemocrit_(HCT_%)	49.841852
1	Mean_corpuscular_hemaglobin_(MCH_pg)	23.117302
2	Mean_corpuscular_volume_(MCV_fL)	16.384109
3	Hemaglobin_(HGB_g/dL)	4.311097
4	BASO_(%)	0.638094
5	MONTHS AT IRRAD	0.619195
6	ABS_EOS_(/uL)	0.594382
7	AGE	0.530520
8	ABS_NEUT_(/uL)	0.410965
9	RBC_(M/uL)	0.380743
10	MONTHS_SINCE_IRRAD	0.380173
11	PLATLET_COUNT_(K/uL)	0.378571
12	EOS_(%)	0.377290
13	LYMPHOCYTES_(%)	0.367362
14	MONO_(%)	0.359183
15	WBC_(K/uL)	0.301040
16	ABS_BASO_(/uL)	0.202352
17	ABS_LYMPH_(/uL)	0.181857
18	DAYS_SINCE_IRRAD	0.179330
19	DOSE_(Gy)	0.176076
20	NEUT_(%)	0.165244
21	ABS_MONO_(/uL)	0.090628
22	TYPE	0.007666
23	SEX	0.003790
24	Species	0.001178
25	IRRAD	0.000000

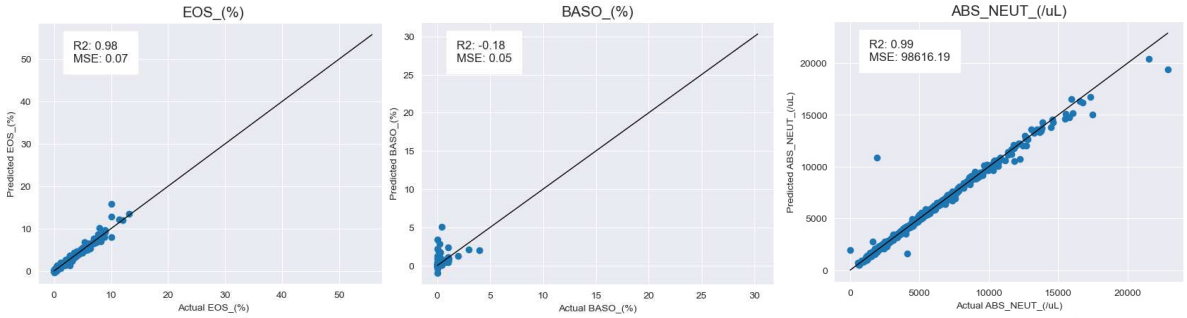


	Feature Id	Importances
0	LYMPHOCYTES_(%)	84.157079
1	MONO_(%)	4.944647
2	EOS_(%)	3.406394
3	ABS_NEUT_(/uL)	2.826048
4	WBC_(K/uL)	1.625402
5	ABS_EOS_(/uL)	1.075533
6	ABS_MONO_(/uL)	0.816814
7	AGE	0.362174
8	DOSE_(Gy)	0.141238
9	RBC_(M/uL)	0.118923
10	ABS_LYMPH_(/uL)	0.103752
11	BASO_(%)	0.093236
12	DAYS_SINCE_IRRAD	0.044528
13	MONTHS_SINCE_IRRAD	0.042983
14	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.039252
15	PLATLET_COUNT_(K/uL)	0.036247
16	Hemaglobin_(HGB_g/dL)	0.032973
17	Mean_corpuscular_volume_(MCV_fL)	0.032823
18	Mean_corpuscular_hemaglobin_(MCH_pg)	0.027593
19	Hemocrit_(HCT_%)	0.022973
20	ABS_BASO_(/uL)	0.022948
21	SEX	0.012982
22	MONTHS AT IRRAD	0.011657
23	TYPE	0.001096
24	Species	0.000706
25	IRRAD	0.000000

	Feature Id	Importances
0	NEUT_(%)	78.454520
1	ABS_LYMPH_(/uL)	9.256811
2	MONO_(%)	3.240007
3	ABS_EOS_(/uL)	1.929405
4	ABS_NEUT_(/uL)	1.816751
5	ABS_MONO_(/uL)	1.517071
6	WBC_(K/uL)	1.318517
7	DAYS_SINCE_IRRAD	0.857753
8	EOS_(%)	0.678479
9	AGE	0.251058
10	RBC_(M/uL)	0.123101
11	DOSE_(Gy)	0.097230
12	MONTHS_SINCE_IRRAD	0.081349
13	BASO_(%)	0.072396
14	MONTHS AT IRRAD	0.065867
15	PLATLET_COUNT_(K/uL)	0.053544
16	Mean_corpuscular_volume_(MCV_fL)	0.041264
17	Mean_corpuscular_hemaglobin_(MCH_pg)	0.039020
18	Hemocrit_(HCT_%)	0.036941
19	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.029147
20	Hemaglobin_(HGB_g/dL)	0.020638
21	ABS_BASO_(/uL)	0.013953
22	TYPE	0.003244
23	Species	0.001041
24	SEX	0.000894
25	IRRAD	0.000000

	Feature Id	Importances
0	ABS_MONO_(/uL)	62.972024
1	WBC_(K/uL)	32.363759
2	AGE	1.662824
3	ABS_NEUT_(/uL)	0.866357
4	NEUT_(%)	0.271485
5	Species	0.271306
6	DOSE_(Gy)	0.258465
7	ABS_BASO_(/uL)	0.127397
8	RBC_(M/uL)	0.117118
9	ABS_LYMPH_(/uL)	0.111969
10	Mean_corpuscular_volume_(MCV_fL)	0.110601
11	PLATLET_COUNT_(K/uL)	0.107173
12	MONTHS AT IRRAD	0.097644
13	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.097183
14	ABS_EOS_(/uL)	0.087184
15	Hemocrit_(HCT_%)	0.084644
16	MONTHS_SINCE_IRRAD	0.078233
17	LYMPHOCYTES_(%)	0.068009
18	SEX	0.060617
19	DAYS_SINCE_IRRAD	0.046367
20	Mean_corpuscular_hemaglobin_(MCH_pg)	0.044245
21	Hemaglobin_(HGB_g/dL)	0.042346
22	EOS_(%)	0.038201
23	BASO_(%)	0.014439
24	TYPE	0.000412
25	IRRAD	0.000000

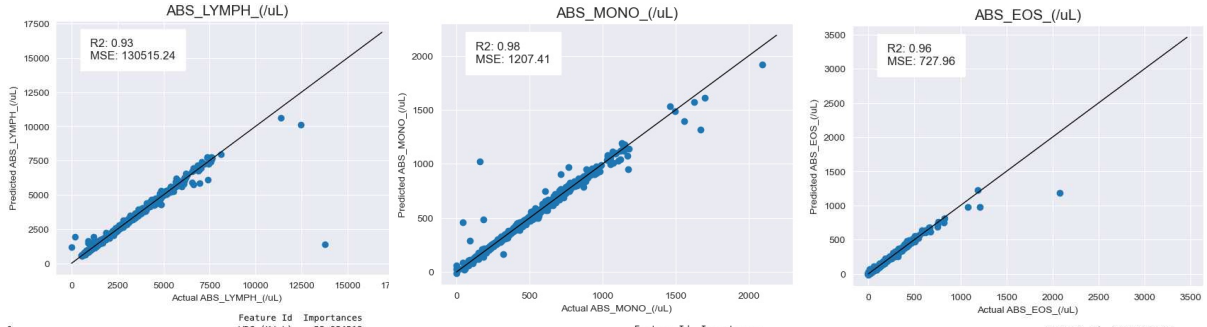
Supplemental Figure 2: CatBoost Predictions & Feature Importance II
CatBoost regression model predictions of different features in the Blood Biomarker dataset. The calculated feature importance rankings for each feature predicted are below the respected scatterplots.



	Feature Id	Importances
0	ABS_EOS_(/uL)	80.375764
1	WBC_(K/uL)	9.122040
2	ABS_NEUT_(/uL)	3.921808
3	ABS_LYMPH_(/uL)	1.434566
4	AGE	1.101122
5	NEUT_(%)	0.719840
6	RBC_(M/uL)	0.535518
7	DAYS_SINCE_IRRAD	0.401518
8	Hemocrit_(HCT_%)	0.317128
9	DOSE_(Gy)	0.299257
10	Mean_corpuscular_hemaglobin_(MCH_pg)	0.285813
11	LYMPHOCYTES_(%)	0.249607
12	Mean_corpuscular_volume_(MCV_fL)	0.243487
13	MONTHS AT IRRAD	0.225493
14	PLATLET_COUNT_(K/uL)	0.176924
15	ABS_BASO_(/uL)	0.132704
16	MONO_(%)	0.130011
17	Hemaglobin_(HGB_g/dL)	0.101311
18	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.081088
19	ABS_MONO_(/uL)	0.066358
20	MONTHS_SINCE_IRRAD	0.047861
21	BASO_(%)	0.016880
22	SEX	0.011565
23	TYPE	0.001573
24	Species	0.000762
25	IRRAD	0.000000

	Feature Id	Importances
0	ABS_BASO_(/uL)	23.519591
1	NEUT_(%)	15.443949
2	Mean_corpuscular_hemaglobin_(MCH_pg)	14.072947
3	ABS_NEUT_(/uL)	11.625690
4	DOSE_(Gy)	6.230773
5	ABS_EOS_(/uL)	6.188964
6	ABS_LYMPH_(/uL)	5.317845
7	LYMPHOCYTES_(%)	5.092551
8	DAYS_SINCE_IRRAD	4.784890
9	RBC_(M/uL)	1.512198
10	AGE	1.171227
11	WBC_(K/uL)	1.073993
12	MONTHS AT IRRAD	0.820091
13	PLATLET_COUNT_(K/uL)	0.812185
14	Hemocrit_(HCT_%)	0.777678
15	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.436667
16	EOS_(%)	0.356103
17	ABS_MONO_(/uL)	0.267792
18	Mean_corpuscular_volume_(MCV_fL)	0.260523
19	MONO_(%)	0.083597
20	MONTHS_SINCE_IRRAD	0.065131
21	Species	0.053737
22	SEX	0.016816
23	Hemaglobin_(HGB_g/dL)	0.014386
24	TYPE	0.000675
25	IRRAD	0.000000

	Feature Id	Importances
0	WBC_(K/uL)	58.107533
1	NEUT_(%)	23.665958
2	LYMPHOCYTES_(%)	16.241049
3	MONO_(%)	0.520269
4	ABS_LYMPH_(/uL)	0.313448
5	AGE	0.292920
6	ABS_EOS_(/uL)	0.259329
7	SEX	0.121124
8	DAYS_SINCE_IRRAD	0.062961
9	EOS_(%)	0.059141
10	PLATLET_COUNT_(K/uL)	0.051357
11	Mean_corpuscular_hemaglobin_(MCH_pg)	0.046711
12	RBC_(M/uL)	0.042807
13	ABS_MONO_(/uL)	0.040784
14	MONTHS_SINCE_IRRAD	0.032366
15	Mean_corpuscular_volume_(MCV_fL)	0.031339
16	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.023422
17	DOSE_(Gy)	0.022880
18	Hemocrit_(HCT_%)	0.022751
19	ABS_BASO_(/uL)	0.015316
20	Hemaglobin_(HGB_g/dL)	0.009474
21	MONTHS AT IRRAD	0.008673
22	BASO_(%)	0.008008
23	TYPE	0.000238
24	Species	0.000140
25	IRRAD	0.000000

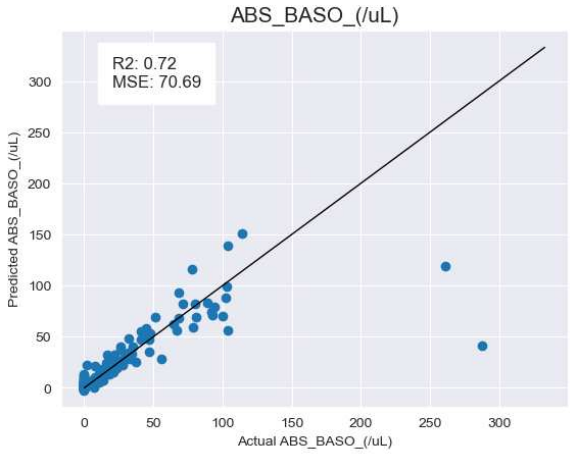


	Feature Id	Importances
0	WBC_(K/uL)	52.054512
1	LYMPHOCYTES_(%)	39.314732
2	NEUT_(%)	7.029048
3	ABS_NEUT_(/uL)	0.320864
4	DOSE_(Gy)	0.245355
5	AGE	0.226577
6	MONO_(%)	0.142299
7	ABS_EOS_(/uL)	0.087647
8	ABS_MONO_(/uL)	0.081421
9	Mean_corpuscular_hemaglobin_(MCH_pg)	0.072081
10	MONTHS_SINCE_IRRAD	0.060372
11	RBC_(M/uL)	0.055500
12	DAYS_SINCE_IRRAD	0.054481
13	Mean_corpuscular_volume_(MCV_fL)	0.049366
14	ABS_BASO_(/uL)	0.042677
15	PLATLET_COUNT_(K/uL)	0.033040
16	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.030731
17	EOS_(%)	0.023927
18	MONTHS AT IRRAD	0.020396
19	Hemocrit_(HCT_%)	0.018931
20	SEX	0.017125
21	Hemaglobin_(HGB_g/dL)	0.008599
22	BASO_(%)	0.006266
23	Species	0.002204
24	TYPE	0.001850
25	IRRAD	0.000000

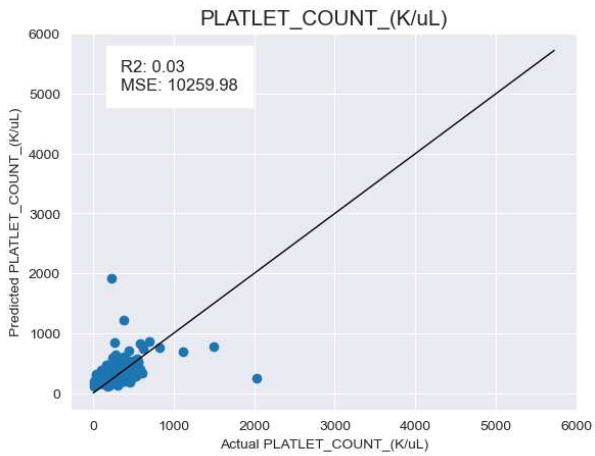
	Feature Id	Importances
0	MONO_(%)	49.914585
1	WBC_(K/uL)	47.418970
2	DAYS_SINCE_IRRAD	1.529896
3	ABS_LYMPH_(/uL)	0.203981
4	ABS_NEUT_(/uL)	0.155217
5	NEUT_(%)	0.118630
6	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.104966
7	PLATLET_COUNT_(K/uL)	0.072132
8	RBC_(M/uL)	0.066443
9	MONTHS_SINCE_IRRAD	0.060124
10	ABS_EOS_(/uL)	0.053077
11	AGE	0.051777
12	Mean_corpuscular_volume_(MCV_fL)	0.046774
13	EOS_(%)	0.036908
14	Hemocrit_(HCT_%)	0.035786
15	Mean_corpuscular_hemaglobin_(MCH_pg)	0.027563
16	LYMPHOCYTES_(%)	0.021835
17	MONTHS AT IRRAD	0.018673
18	Hemaglobin_(HGB_g/dL)	0.016315
19	BASO_(%)	0.015884
20	ABS_BASO_(/uL)	0.013957
21	DOSE_(Gy)	0.009353
22	SEX	0.004974
23	TYPE	0.001776
24	Species	0.000403
25	IRRAD	0.000000

	Feature Id	Importances
0	EOS_(%)	78.754164
1	WBC_(K/uL)	16.704329
2	ABS_MONO_(/uL)	1.375121
3	DOSE_(Gy)	0.923140
4	ABS_NEUT_(/uL)	0.400725
5	AGE	0.302803
6	ABS_LYMPH_(/uL)	0.186249
7	MONTHS_SINCE_IRRAD	0.183655
8	Mean_corpuscular_volume_(MCV_fL)	0.174997
9	RBC_(M/uL)	0.145147
10	Hemocrit_(HCT_%)	0.142975
11	LYMPHOCYTES_(%)	0.139715
12	MONO_(%)	0.137237
13	DAYS_SINCE_IRRAD	0.094536
14	Mean_corpuscular_hemaglobin_(MCH_pg)	0.072697
15	NEUT_(%)	0.064852
16	Hemaglobin_(HGB_g/dL)	0.049518
17	ABS_BASO_(/uL)	0.045850
18	Mean_corpuscular_hemaglobin_concentration_(MCH...	0.034417
19	PLATLET_COUNT_(K/uL)	0.034214
20	MONTHS AT IRRAD	0.025838
21	BASO_(%)	0.004053
22	TYPE	0.002188
23	Species	0.001302
24	SEX	0.000277
25	IRRAD	0.000000

Supplemental Figure 3: CatBoost Predictions & Feature Importance III
CatBoost regression model predictions of different features in the Blood Biomarker dataset. The calculated feature importance rankings for each feature predicted are below the respected scatterplots.

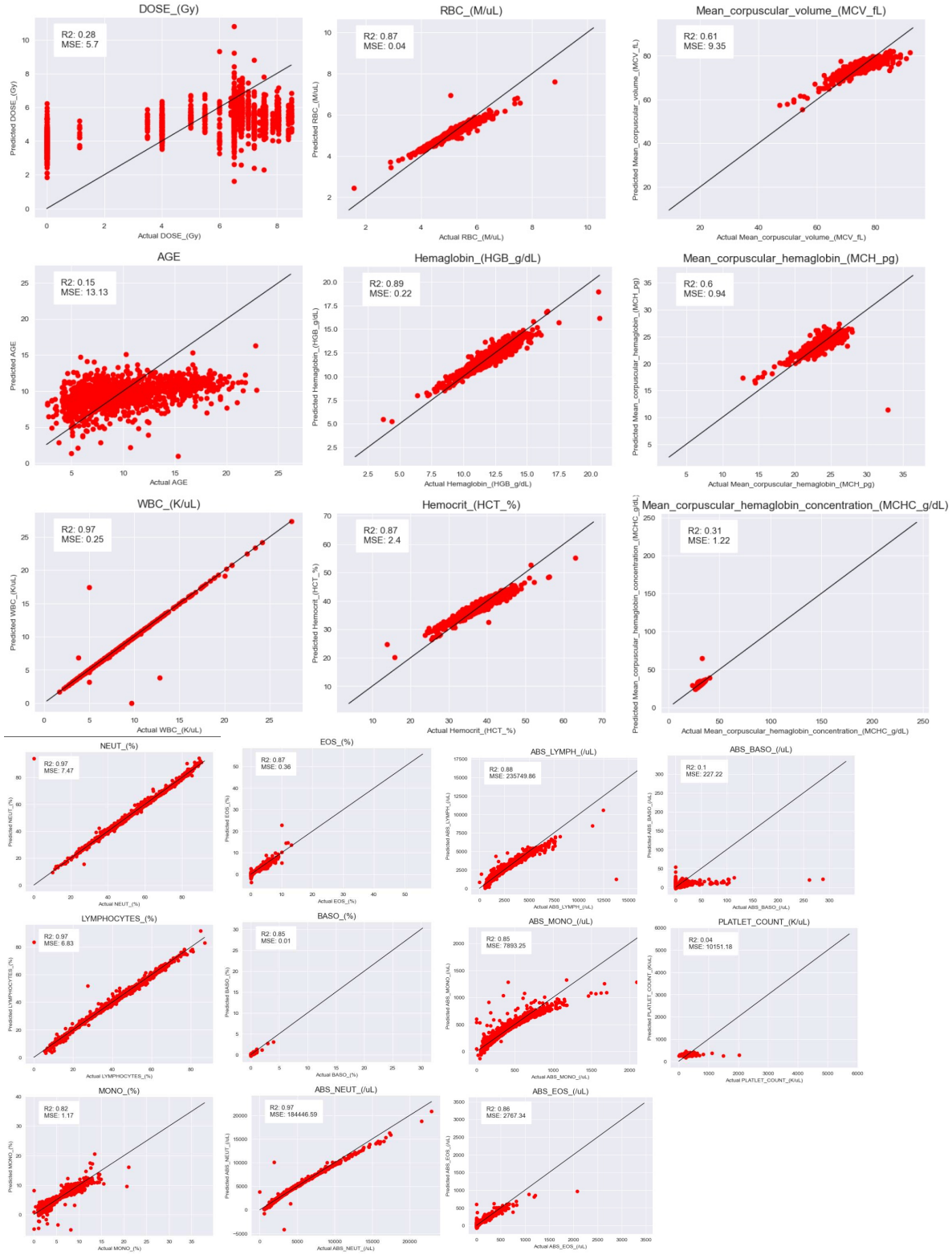


	Feature Id	Importances
0	BASO_(%)	70.250363
1	WBC_(K/uL)	9.924362
2	ABS_MONO_(/uL)	3.539048
3	LYMPHOCYTES_(%)	2.362997
4	ABS_NEUT_(/uL)	1.704159
5	RBC_(M/uL)	1.592002
6	Mean_corpuscular_hemaglobin_concentration_(MCH...	1.321336
7	ABS_EOS_(/uL)	1.257676
8	PLATLET_COUNT_(K/uL)	0.972146
9	EOS_(%)	0.883321
10	Mean_corpuscular_volume_(MCV_fL)	0.876703
11	AGE	0.857875
12	DAYS_SINCE_IRRAD	0.752017
13	Hemaglobin_(HGB_g/dL)	0.677283
14	MONTHS AT IRRAD	0.537684
15	Mean_corpuscular_hemaglobin_(MCH_pg)	0.521143
16	NEUT_(%)	0.471311
17	Hemocrit_(HCT_%)	0.388734
18	MONTHS_SINCE_IRRAD	0.366645
19	ABS_LYMPH_(/uL)	0.315123
20	MONO_(%)	0.305725
21	DOSE_(Gy)	0.111344
22	Species	0.005266
23	TYPE	0.004737
24	SEX	0.000997
25	IRRAD	0.000000

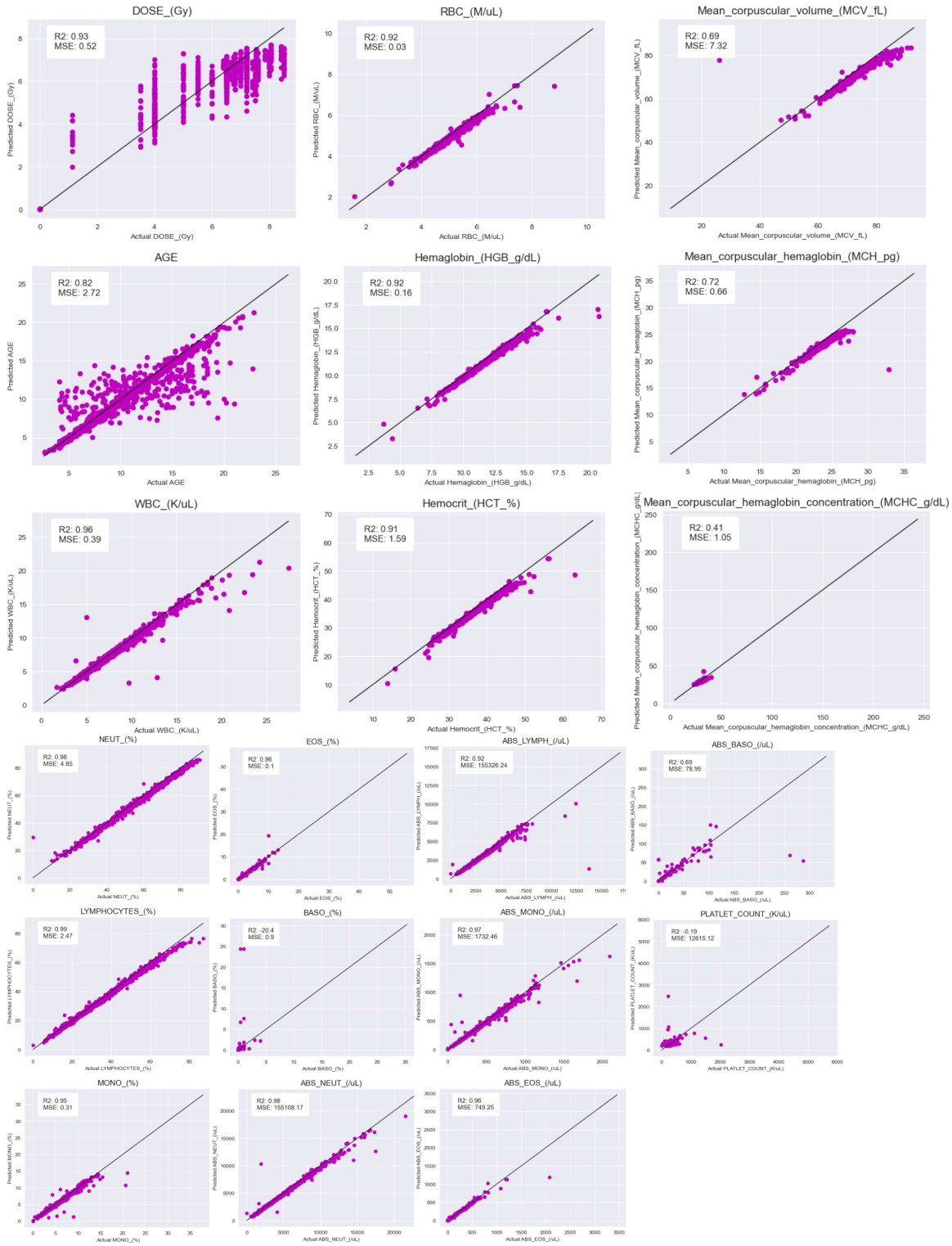


	Feature Id	Importances
0	AGE	12.834489
1	MONO_(%)	11.186112
2	MONTHS AT IRRAD	8.334823
3	DOSE_(Gy)	6.591006
4	Mean_corpuscular_hemaglobin_(MCH_pg)	6.353025
5	ABS_MONO_(/uL)	5.602487
6	Mean_corpuscular_hemaglobin_concentration_(MCH...	5.379462
7	ABS_LYMPH_(/uL)	5.254545
8	Mean_corpuscular_volume_(MCV_fL)	5.214640
9	DAYS_SINCE_IRRAD	4.906594
10	Hemaglobin_(HGB_g/dL)	4.704848
11	MONTHS_SINCE_IRRAD	3.355676
12	RBC_(M/uL)	3.007297
13	WBC_(K/uL)	2.955332
14	Hemocrit_(HCT_%)	1.890984
15	ABS_EOS_(/uL)	1.886817
16	SEX	1.792575
17	EOS_(%)	1.662829
18	NEUT_(%)	1.633126
19	ABS_NEUT_(/uL)	1.557730
20	ABS_BASO_(/uL)	1.302898
21	LYMPHOCYTES_(%)	0.995982
22	TYPE	0.986505
23	Species	0.419206
24	BASO_(%)	0.191014
25	IRRAD	0.000000

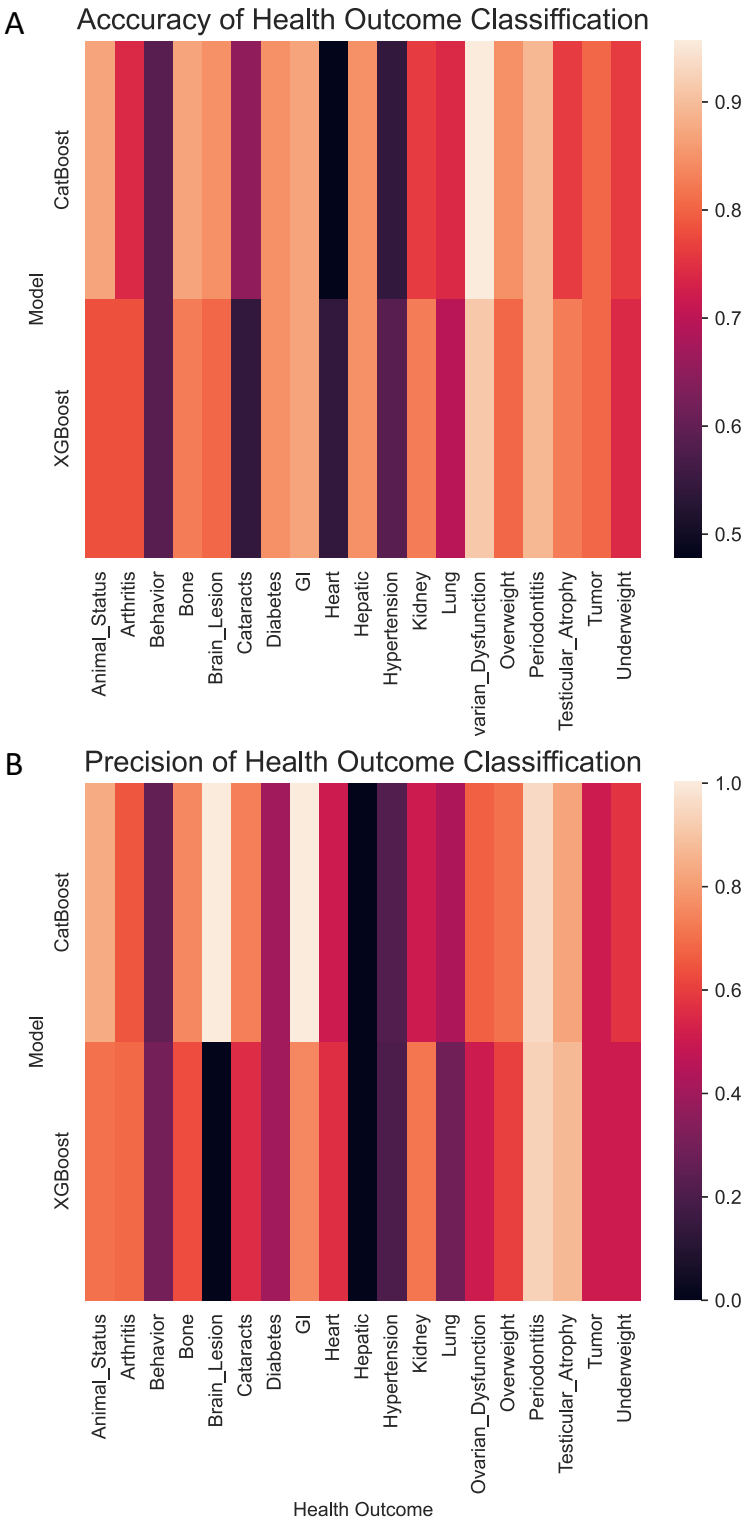
Supplemental Figure 4: CatBoost Predictions & Feature Importance IV
CatBoost regression model predictions of different features in the Blood Biomarker dataset. The calculated feature importance rankings for each feature predicted are below the respected scatterplots.



Supplemental Figure 5: ElasticNet Predictions of all Blood Biomarker features
ElasticNet model predictions of different features in the Blood Biomarker dataset. The model had varying success with the different feature predictions with the most accurate being white blood cell count (WBC_(K/uL)) with R2: 0.97 and MSE: 0.25 and the least accurate being (PLATELET_COUNT_(K/uL)) with R2: 0.04 and MSE: 10151.18.



Supplemental Figure 6: XGBoost Predictions of all Blood Biomarker features
XGBoost regression model predictions of different features in the Blood Biomarker dataset. The model had varying success with the different feature predictions with the most accurate being lymphocyte counts (LYMPHOCYTES_%)) with R2: 0.99 and MSE: 2.47 and the least accurate being (PLATLET_COUNT_(K/uL)) with R2: -0.19 and MSE: 12615.12.



Supplemental Figure 7: CatBoost and XGBoost Accuracy & Precision Scores For All Health Outcome Classifications

(A) Heatmap showing the accuracy of CatBoost and XGBoost classifying the different health outcomes.

(B) A. Heatmap showing the Precision of CatBoost and XGBoost classifying the different health outcomes.

APPENDIX II

Access to Code

The code for the analyses, model development, and preliminary application can be found at https://github.com/aidan-lew/monkey_health. Detailed descriptions can be found in the repository's Wiki pages.