**Machine intelligence for radiation science: summary of the Radiation Research Society 67th annual meeting symposium**

Lydia J Wilsona,1, Frederico C Kifferb, Daniel C. Berriosc, Abigail Bryce-Atkinsond, Sylvain V. Costesc, Olivier Gevaerte,f, Bruno F. E. Matarèseg,h, Jack Millerc,i, Pritam Mukherjeee,j, Kristen Peachk, Paul N. Schofieldl, Luke T. Slaterm-p, and Britta Langenq

aDepartment of Radiation Oncology, St. Jude Children’s Research Hospital, 262 Danny Thomas Place, Memphis, TN, 38105

bDepartment of Anesthesia and Critical Care Medicine, The Children’s Hospital of Philadelphia Research Institute, Philadelphia, PA, 19146

cNASA Ames Research Center

dDivision of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

eStanford Center for Biomedical Informatics Research, Department of Medicine, 1265 Welch Road, Stanford, CA 94305

fDepartment of Biomedical Data Science, Stanford University, Palo Alto, CA, USA

gThe Cavendish Laboratory, University of Cambridge

hDept of Haematology, School of Clinical Medicine, University of Cambridge

iKBR, NASA Ames Research Center

jRadiology and Imaging Sciences, NIH Clinical Center, 10 Center Dr, Bethesda, MD 20892

kBionetics, NASA Ames Research Center. Moffett Field, CA, USA

lDept. Of Physiology Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, UK

mInstitute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

nInstitute of Translational Medicine, University Hospitals Birmingham, NHS Foundation Trust, Birmingham, UK

0MRC Health Data Research UK (HDR UK) Midlands, UK

pUniversity Hospitals Birmingham NHS Foundation Trust, Edgbaston, Birmingham, UK

qDepartment of Radiation Oncology, Section of Molecular Radiation Biology, UT Southwestern Medical Center, Dallas, TX, 75390

1All correspondence to Lydia J Wilson, PhD, Department of Radiation Oncology, St. Jude Children’s Research Hospital, 262 Danny Thomas Place, MS210, Memphis, TN 38105, USA; tel: +1 901 595 1895; email: [lydia.wilson@stjude.org](mailto:lydia.wilson@stjude.org)

**Running head**: Machine intelligence for radiation science

**Keywords:** Machine learning, artificial intelligence, radiobiology, radiotherapy, lung cancer, voxel-based analysis, ontology

**Abstract**

The era of high-throughput techniques created big data in the medical field and research disciplines. Machine intelligence (MI) approaches can overcome critical limitations on how those large-scale data sets are processed, analyzed, and interpreted. The 67th Annual Meeting of the Radiation Research Society featured a symposium on MI approaches to highlight recent advancements in the radiation sciences and their clinical applications. This article summarizes three of those presentations regarding recent developments for metadata processing and ontological formalization, data mining for radiation outcomes in pediatric oncology, and imaging in lung cancer.

**Introduction**

Recent decades have seen machine intelligence (MI) approaches become more widely used in basic research and clinical applications. This rapid advancement was made possible through the convergence of several factors, most notably easier access to information technologies, increase in availability of machine learning (ML) algorithms across all popular programming languages, breakthroughs in instrumentation for data acquisition, decline in wet lab and computational hardware costs, and a new generation of scientists and medical professionals trained in bioinformatics techniques. The terms MI and artificial intelligence (AI) are often used interchangeably and describe a technology that enables a machine to (convincingly) mimic human behavior, AI being a broader definition applicable to any system, whereas MI is specific to computational mimicry. ML is a specific subset of this field characterized by algorithms that process large quantities of data to build an analytical model that improves with experience (iteration), thereby “learning” from datasets. This means that the model-building itself is data-driven and adaptive, rather than programmed a priori. In this article, the term MI is used to encompass both ML and AI. Distinctions between specific sub-categories of ML are made where necessary. Artificial neural networks are the ML computational approaches that are most quickly increasing in popularity due to their high inference accuracy achieved via reinforcement learning, which is inapplicable to other MI computational approaches, and their network adaptability to relatively modest datasets. Neural networks involve an input layer, a few hidden layers, and an output layer, loosely resembling the processing architecture of cortical neurons. For ML applications involving large, complex datasets, so-called ‘deep learning’ algorithms, which instead use neural networks with many hidden layers, are recommended. Despite its overwhelming advantages over classical data modeling approaches, ML is limited to the characteristics of datasets, and depending on application, the similarity between the data used to train algorithms vs the data used to test or infer outcomes from trained algorithms.

Within the field of radiation research, some of the key challenges for the successful integration of MI approaches reside in how the breadth of multi-disciplinary data is organized and presented to the algorithms. Moreover, radiation therapy and related research cover virtually all cancer types across all sex and age groups while considering a myriad of normal tissues for risk assessment and toxicity studies. Apart from the various disease and patient characteristics, the field also covers basic physics, chemistry, and epidemiology–all of which create additional layers of data (features) that could be incorporated into MI/ML approaches.

The Online 67th Annual Meeting of the Radiation Research Society featured the Scholars-In-Training (SIT) Hot Topic Symposium titled "Machine intelligence for radiation science" to highlight recent developments and solutions to current challenges. The invited speaker Dr. Joseph O. Deasy outlined the fundamentals of MI methods in the context of radiation research. Dr. Cynthia McCollough was also invited to present the use of MI for radiation dose management in medical imaging. Dr. Paul N. Schofield presented the Radiation Biology Ontology (RBO) effort to support data harmonization, discovery, and integration of MI in radiation research. Dr. Lydia J. Wilson addressed challenges of method-transfer between adult and pediatric cohorts, specifically how to apply voxel-based analysis approaches to study radiation outcomes across a large age range in children. Dr. Pritam Mukherjee, a Scholar-In-Training invited speaker, presented applications of MI for opportunistic diagnosis of small-cell lung cancer (SCLC) using CT scans, and how scan-based data can inform prognostication of lung cancer patients. This review summarizes the contributions by Drs. Schofield, Wilson, and Mukherjee. Certain talks were not included here at the authors’ discretion.

**Summary of included talks**

***Dr. Paul Schofield: Formal ontology to support data discovery, integration and analysis using artificial intelligence***

The application of ML and MI to problems in radiobiology requires the development and mobilization of extremely large, coherent, and interoperable datasets for training and testing. Some such datasets are already available, and new ones are being produced at pace [(Schofield et al. 2019)](https://paperpile.com/c/33loEO/u7M3h). We face two challenges: identifying and integrating these very large datasets, and using them to support analysis using MI. Formal ontologies (such as the major bioinformatics initiative Gene Ontology, https://geneontology.org), provide powerful metadata models, and can be used to annotate data subjects with categorical descriptors (e.g. pathology or anatomy terms), or to qualify numeric data as to data type, units etc. This semantic annotation supports the creation of large knowledge graphs and integration of background knowledge, facilitating data discovery and knowledge synthesis across distributed databases.

Creation of an ontology to support these types of applications in radiation biology presents significant scientific and logistical challenges. Like many interdisciplinary sciences, data producers and consumers in the field of radiation biology use a wide variety of terminologies to describe their experiments, data, and findings. Furthermore, space exploration systems and technologies are rapidly evolving, and a shared understanding and common terminology for these is also lacking. In response to these challenges, we have developed the Radiation Biology Ontology (RBO) (**Fig. 1)**. The RBO is an effort to comprehensively cover all aspects of the domain from radiochemistry, environmental studies, space research, and mechanistic studies, to ethico-legal and psychosocial research. It combines consensus terms from existing ontologies describing relevant scientific specialties together with new concepts developed using our own expertise.

The ontology was developed with open-source software (the Ontology Development Kit, Protégé and WebProtégé) as recommended by the Open Biological and Biomedical (OBO) Foundry [(Jackson et al. 2021)](https://paperpile.com/c/33loEO/SvHQb), which also requires that OBO ontologies implement a set of development principles and practices for ontology consistency, uniformity, and accountability. The content coverage for the RBO was largely driven by two diverse use cases; the need to annotate data deposited in the STORE database (<https://www.storedb.org/>) for the European Commission–funded RadoNorm project (<https://www.radonorm.eu/>), which has a very broad frame of reference including both medical and environmental radiation exposure, and NASA’s GeneLab ‘omics database (<https://genelab.nasa.gov>), which contains data from experiments and measurements conducted in space and terrestrial space-like environments [(Berrios et al. 2021)](https://paperpile.com/c/33loEO/ee8e0). We have previously used versions of the mouse pathology (MPATH [(Schofield et al. 2013)](https://paperpile.com/c/33loEO/BydoO)) and adult mouse anatomy (MA; [(Hayamizu, Baldock, and Ringwald 2015)](https://paperpile.com/c/33loEO/Di2GH)) ontologies for annotations of the legacy datasets in the European Radiobiological Archive (ERA) containing legacy data from the JANUS and other US, European and Japanese programs [(Birschwilks et al. 2011)](https://paperpile.com/c/33loEO/CWGvs). This initial phase of concept modeling has yielded an RBO that defines more than 300 concepts, with another >3500 concepts imported from other OBO Foundry ontologies to provide critical, meaningful context.

During this first phase we focused on concepts for annotating samples, environments, exposures, and measurements. The next phase will focus on supporting annotation of results and findings, such as concept models of molecular, cellular and tissue effects of radiation exposure.

*FAIR data; discovery and integration*

The efficiency of research organizations engaged in data acquisition and curation can be enhanced by standardizing metadata through the use of knowledge resources. Sophisticated models such as formal ontologies enable more automated data-acquisition processes and more accurate meta-analysis through more efficient and complete data discovery and retrieval.

These tasks are inseparable from the wider challenges of making more data publicly available. The past decade has seen an increasing awareness of the importance of public data availability and reliability, resulting in 2016 in the publication of the FAIR (Findable, Accessible, Interoperable, and Reusable) framework, which is now widely regarded as normative, but not yet widely or fully implemented. A critical part of the FAIR requirements is that metadata should be standardized and based on community-accepted and interoperable ontologies [(Wilkinson et al. 2016, 2018)](https://paperpile.com/c/33loEO/unN49+lprGf), which systems may then use for more accurate and/or automated data discovery (e.g., through application programming interfaces) and to encode distributed queries of multiple data sources. Tools are now being developed for the assessment of FAIRness, e.g. [(D. J. B. Clarke et al. 2019; Wilkinson et al., 2018)](https://paperpile.com/c/33loEO/c6NGA+lprGf) and the NIH in its Bridge2 MI initiative has recently recognized the importance of formal data standardization and discoverability (<https://commonfund.nih.gov/bridge2ai>) including the development of domain ontologies like the RBO.

*The RBO ontology and MI*

Formal ontologies can be used in many analytical contexts as has been summarized recently in extensive reviews [(Kulmanov et al. 2021; Robinson and Haendel 2020)](https://paperpile.com/c/33loEO/KuFzC+QkSo8). While ontologies support applications of MI both in the senses of MI and ML (for discussion of the distinction, see Robinson and Haendel 2020), the vast majority of applications currently being explored are of the latter type.

The background knowledge expressed in ontologies, particularly by axiomatic class definitions, can be used to create and enrich knowledge graphs. This may either be accomplished through direct annotation with, for example, disease or phenotype classes, but also through complex semantic descriptions of instances.

Semantic annotation of data with bio-ontologies facilitates analysis that can provide unexpected insights into biomedicine. For example, gene set enrichment has been used routinely for some time to evaluate the results of Genome-Wide Association Studies, uncovering the associations between biological functions and genetic variants. Ontologies can also be leveraged in exploratory or unsupervised ML to unmask unsuspected relationships between entities and concepts.

Ontology classes (terms) can be logically defined using other ontologies (axioms) and this allows the capture of deep background knowledge across multiple domains in a form amenable to computation. For example, a term in a phenotype ontology might be represented using terms from a biological process ontology, an anatomy ontology and a trait ontology: a *process* in an *anatomical location* is *abnormal*. This then allows the ontologies all to be formally connected and background knowledge contained within them made explicit, i.e. computable. In this example the process term would come from the Gene Ontology (GO), the anatomical term from, say, the Mouse Anatomy (MA) ontology, and the qualifier term from the Phenotype and Trait ontology (PATO). Axiomatisation would allow the relationship between the abnormal process captured in this phenotype class and closely-related processes asserted in the biological\_process arm of the Gene Ontology (Gene Ontology Consortium, 2021; Smaili et al 2020). For example, in a mouse strain with a phenotype annotated to the term *conus arteriosus formation*, the axiom tells us that this term relates to part of the *heart*. This will be helpful in establishing the relationship between two strains of mice, for example, which have different, but related, cardiac developmental process abnormalities. In a sense the algorithm “knows” that there is a relationship between *conus arteriosus formation* and other heart phenotypes and can use that background knowledge to establish a computable relationship between defects in *conus arteriosus formation* and other cardiac phenotypes, such as *atrial septal defect* and *tetralogy of Fallot*. An example of the implementation of this use of the deep knowledge in cancer research can be found in Althubaiti et al (2019), where we learn the characteristics of known tumor driver genes using annotations made with such interlinked ontologies, and then use this to predict more than a hundred new tumor drivers together with recovery of most of the existing genes.

We are entering a new era in the techniques of MI. Although traditionally statistical MI makes use of numeric data for machine learning, neuro-symbolic MI also incorporates symbolic rather than just numeric data in the generation of models. Symbolic data would for example include terms describing color, texture, branching defects etc. The advantage of this emerging area of MI is that it is able to combine the logical operations, such as inference, and rules associated with symbolic logic, with the powerful statistical approach of traditional MI using very large datasets. Combining symbolic logic and learning makes the MI more human-like, and improves both model explainability and human interpretability and, importantly, allows the incorporation of deep background knowledge. This background knowledge comes from the use of ontologies as the symbolic representation of concepts in the input data. As explained above, ontologies may be used to generate domain-specific knowledge graphs of relationships between concepts across a wide range of multi-modal data including phenotype annotations, ‘omics data and, through text mining, public literature (e.g. PubMed abstracts, PubMed Central texts), electronic health records, and institutional reports.

In order to capture deep semantic information from a graph such as an ontology we use embeddings that can be considered as compressed numeric representations of the structure of the graph. More easily, this can be thought of as taking multiple random walks between the nodes (terms) in the ontology graph along its edges (relationships). These walks are done many times from different starting points to capture the structure of the ontology graph, and can then be used as input to a neural network that learns a numeric representation (a vector) for each node or term. Each subject, say a person or a tree, that is annotated with multiple ontology terms then becomes annotated with multiple vectors which capture the whole depth of the relationships within the ontology, including axioms. This is a sparse representation of the knowledge in the ontology that is suitable for deep-learning applications. These approaches have been applied, among others, to rare-disease diagnosis [(Decherchi et al. 2021)](https://paperpile.com/c/33loEO/g0D1Z) and gene-disease association [(Chen, Althagafi, and Hoehndorf 2021; Smaili, Gao, and Hoehndorf 2020)](https://paperpile.com/c/33loEO/YJsBl+ASWXT). Other novel approaches seek to use ontologies as a semantic framework for neuro-symbolic integration of data into neural networks [(Althubaiti et al. 2019)](https://paperpile.com/c/33loEO/74zn8). This approach combines the connectionism of neural networks with high-level symbolic representations of a problem that are human-readable.

The RBO contains knowledge that can facilitate these applications, providing new ways of analyzing and exploiting datasets to gain hitherto inaccessible insights into the underlying biology of radiation effects.

The value of the RBO will be determined in part by our ability to engage the community in its development, and we have established a Radiobiology Informatics Consortium with unrestricted membership to encourage investigators, system owners and others to join us in this effort. Anyone can report issues or request new terms or other features for the RBO directly on GitHub (RBO has a Github site on: https[://github.com/Radiobiology-Informatics-Consortium/RBO](https://github.com/Radiobiology-Informatics-Consortium/RBO)). Using the BioPortal application programming interface, systems can pose dynamic queries to the latest version of the RBO for information on individual classes or entire hierarchies; this design eliminates the need for systems to be updated in order to use newer versions of the RBO. We hope to contribute to the advancement of open radiobiological science through the continued, open development of the RBO, that will provide more precise, machine-interpretable descriptions of investigations, as well as support data meta-analysis through ML or other MI methods.

***Dr. Lydia Wilson: Image-based data mining for radiation outcomes research robustly applies to data from children***

Improvements in cancer detection and treatment have enabled 5-year cancer survival rates to steadily increase worldwide [(Torre et al., 2016)](https://paperpile.com/c/33loEO/kQP6u). Radiation therapy is an important tool contributing to these high survival rates [(M. Clarke et al. 2005; Jebsen et al. 2008)](https://paperpile.com/c/33loEO/LoSRq+9NPiB) but, unfortunately, it is also associated with toxicities that can detract from a cancer survivor’s length and quality of life [(Suh et al. 2020)](https://paperpile.com/c/33loEO/LeQkI). As a result, radiation therapy research increasingly focuses on managing long-term morbidity and mortality of curative treatments. Personalized treatment and survivorship-care plans aim to improve patient outcomes by considering the relation between radiation exposures and their biologic effects. Critically, this relation is often uncertain or unknown.

Image-based data mining (IBDM) is a voxel-based method developed to identify functional sub-units driving radiation toxicity. The method comprises two primary steps: spatial normalization, comprising first rigid/affine registration and second deformable registration of individual patient data to a reference anatomy, followed by a statistical dose comparison in each voxel. In its simplest implementation (used in this work), the statistical dose comparison involves a permutation test of per-voxel Student’s t-tests over 1000 permutations of the binary effect labels. The full IBDM pipeline has only been tested in and applied to data collected from adults [(Beasley et al. 2018; Heemsbergen et al. 2010; McWilliam et al. 2017; Palma et al. 2016)](https://paperpile.com/c/33loEO/QQkR9+amf9a+zsotY+PnT56). Its applicability to data from children is unclear because of the large age-related differences in anatomic size that are characteristic of childhood and the unique histologies and side effects associated with childhood cancers and their treatments. The goal of this work (Wilson et al. 2022) was to test the feasibility of applying IBDM methods to CT imaging data collected from children and simulated doses and effect labels representative of those relevant to pediatric radiotherapy research (**Fig. 2**). The use of simulated dose distributions gave us complete control over every characteristic of the dose distributions and allowed us to assess the extent to which specific dosimetric characteristics (e.g., noise level, the magnitude of the dose difference between groups, the volume over which that dose difference exists) affect IBDM performance. The use of simulated health-effect labels gave us complete control over the population characteristics (i.e., the relative proportionalities and specific patients classified as affected or unaffected by a hypothetical effect of interest).

We used CT images from 167 children (age 10 months to 20 years) who previously received radiotherapy for primary brain tumors at St. Jude Children’s Research Hospital to test the applicability of IBDM to data from children. We first tested the feasibility of spatial normalization using data from four reference patients selected from among those in the cohort on the basis of representing the median age (male and female), target volume, and brain volume. We assessed the accuracy of these spatial normalizations via contour-distance and center-of-mass–deviation analyses.

These tests revealed that discrepancies between the spatially normalized individual subject anatomies were small with average center-of-mass deviations within 6 mm and average contour distances within 3 mm. These results were statistically insensitive to the selected reference anatomy with only a slight improvement for spatial normalization to the reference anatomy selected on the basis of the median brain volume.

Next, we probed for the performance limits of the IBDM statistical dose comparison using simulated dose distributions and health effects. The IBDM dose comparison aims to identify significant differences between the dose distributions of groups of patients who experienced some differential health effect (i.e., those who experienced some effect of interest in one group and those who were free from it in the other). IBDM achieves this via permutation testing with the null hypothesis that there is no difference between the average dose distributions of the two groups and a post-hoc voxel-wise test to identify the specific voxels within which dose significantly differs. We performed these methods using the previously published pipeline known as the Manchester Tool Kit [(Beasley et al. 2018)](https://paperpile.com/c/33loEO/QQkR9).

We randomly assigned the simulated health-effect labels at rates representative of those associated with cranial irradiation in children and reshuffled our groupings five times to test the sensitivity of IBDM to the patient groupings. We generated the simulated dose distributions to be relevant to pediatric late-effects research (i.e., out-of-field dose calculations and volumes of potential pediatric functional sub-units) and to include a systematic dose difference correlated with the simulated side effects. With this, we could test IBDM’s ability to identify the systematic dose difference that we had introduced. We assessed the performance via the Dice similarity coefficient comparing the volume identified as significant by IBDM to the volume in which the simulated dose distributions systematically differed.

To test all possible combinations of dosimetric characteristics, simulated effect rates, simulated effect groupings, and reference patients, we performed 270 IBDM analyses. Of these tests, 81% correctly identified the presence of a difference between simulated dose distributions of the effect and no-effect groups (p < 0.01). The tests that failed to identify the introduced dose differences tended to involve the smallest test volume, which mimicked a small sensitive sub-volume of interest, and the smallest magnitude of an introduced dose difference, which mimicked small differences in out-of-field dose deposition. Dice analyses of voxel-wise results revealed that Dice scores were generally high (peak frequency > 0.9) but, IBDM’s performance was also sensitive to the magnitude of random noise in the simulated dose distribution, which mimicked Monte Carlo out-of-field dose calculations run for decreasing numbers of histories, resulting in increasing statistical noise.

This work, for the first time, comprehensively tested the full IBDM pipeline with data obtained from children and revealed that it is feasible to use IBDM to identify a sensitive sub-volume under conditions consistent with those expected in pediatric late-effects research. Together, these tests suggest that spatial normalization can be achieved at similar uncertainty levels to those observed in adult [(Beasley et al. 2018)](https://paperpile.com/c/33loEO/QQkR9) and other pediatric spatial normalizations [(Veiga et al. 2021)](https://paperpile.com/c/33loEO/ijgaE), but dose comparisons may have limited sensitivity to small dose discrepancies, especially when localized in small sensitive sub-volumes. Additionally, every effort should be made to decrease the noise in out-of-field dose calculations, for example, by increasing the number of histories in Monte Carlo simulations.

The IBDM method enables an unbiased consideration of the spatial dose-effect relation in children. The results reported in this presentation may inform cohort selection and study design for future investigations of pediatric radiotherapy dose responses. A deeper understanding of this relationship can inform treatment planning and survivorship care to improve long-term health outcomes for childhood cancer survivors.

***Dr. Pritam Mukherjee: Artificial intelligence for imaging in lung cancer***

Lung cancer is the most common fatal malignancy in adults worldwide, with more than 1.8 million deaths worldwide attributed to it every year [(Sung et al. 2021)](https://paperpile.com/c/33loEO/EPtdt). Computed tomography (CT) scans are routinely used in the clinic for lung-cancer screening, diagnosis, and assessing treatment response. These medical images are also readily amenable to quantitative analysis with modern machine-learning techniques and have the potential to facilitate faster, cheaper, and more accurate clinical decision-making by discovering patterns that may be hard to discern with the human eye. Based on our work, two potential applications are highlighted here.

The first application is the opportunistic diagnosis of small-cell lung cancer (SCLC) from CT scans. The key challenge is to distinguish SCLC nodules from benign and non-small-cell lung cancer (NSCLC) subtypes, such as adenocarcinoma and squamous cell carcinoma (SCC). SCLC, which represents only 13% of all lung cancers, has an aggressive doubling time as short as 25 [(Harris et al. 2012)](https://paperpile.com/c/33loEO/ag2uT) days and an abysmal 5-year survival rate of 2.9% in its extensive stage (ES) when the disease cannot be encompassed into a tolerable radiation field. A delay of 1–2 months between initial imaging and diagnosis could be within the doubling time of the tumor, and therefore, early detection of SCLC is critical to expedite diagnosis and treatment. In our work [(Shah et al. 2021)](https://paperpile.com/c/33loEO/lLG3X), the goal was to study the feasibility of distinguishing SCLC from benign and NSCLC (specifically, SCC and adenocarcinoma) nodules based on the CT scans only, which, if reliable, could lead to early treatment and better outcomes. To that end, CT scans were collected from the Palo Alto VA medical center and 103 patients were included based on the presence of a nodule > 1 cm in greatest diameter with confirmed cancer or benign diagnosis. The studies were then divided into group A (non-contrast scans) and group B (contrast-enhanced scans). For the analysis, a typical radiomics workflow was adopted (**Fig. 3**, left). A radiologist with more than 10 years of experience delineated every solid nodule > 1 cm in greatest diameter in every scan. A subset (18) of the cases were also annotated by a second radiologist with more than 5 years of experience, in order to assess inter-observer variability and robustness of the analysis. The Pyradiomics [(van Griethuysen et al. 2017)](https://paperpile.com/c/33loEO/wtj4v) feature pipeline was used to extract 105 features from the segmented 3D tumor volume. The extracted features that met the standards of the Image Biomarkers Standardization Initiative [(Zwanenburg et al. 2020)](https://paperpile.com/c/33loEO/YcEHc). Out of these, 59 quantitative first-order and texture characteristics were found to be robust between the two segmenters, and these were used with four ML classification models: support vector classifier, random forest (RF), XGBoost, and logistic regression, to develop radiomic models. The performance was evaluated using a receiver-operating-characteristic (ROC) curve. A final model was created using the RF classifier and aggregate minimum Redundancy Maximum Relevance to determine feature importance. The model achieved an area under the ROC (AUROC) curve of 0.81 and 0.88 on the non-contrast and contrast-enhanced groups, respectively. This suggests that a radiomics-based ML model can potentially be used for early prediction of SCLC on CT and allow for early treatment and better prognosis.

The second application is the prognostication of lung cancer patients based on their CT scans. In this study [(Mukherjee et al. 2020)](https://paperpile.com/c/33loEO/qYEKY), we focused on NSCLC patients and developed a deep-learning model based on convolutional neural networks (CNN) to predict their overall survival (**Fig. 3**, right). The goal was to stratify patients based on their risk of mortality so that the aggressiveness of their treatments can be tailored to the risk and lead to improved outcomes and quality of life. To that end, we collected four independent cohorts of patients with NSCLC from four medical centers: Stanford Hospital (cohort 1, n = 129), H. Lee Moffitt Cancer Center and Research Institute (cohort 2, n = 185), MAASTRO Clinic (cohort 3, n = 311), and Charité – Universitätsmedizin, Berlin (cohort 4, n = 84). For each patient, the nodules were delineated by radiologists at their respective institutions. We used a shallow 3D-CNN–based model architecture (LungNet) with only three convolution blocks with size 16 × 3 × 3 along with a 3D max-pooling layer with kernel size = 2, stride = 2. Three fully connected layers with decreasing sizes of feature vectors (i.e., 128, 64, and 64) were concatenated to reduce feature dimensions toward convergence of model training. As inputs to the model, we provided a 3D patch containing the nodule and another patch corresponding to its segmentation mask. Two versions of LungNet were developed: one with the imaging only and the other in which clinical features such as age, sex, histology, and stage, could be added as inputs. Both models were trained from scratch with a Cox regression loss function. Cohorts 1, 2, and 3 were used for model development, cohort 4 (from Berlin) was used solely as a test set. With the first three cohorts, we used a cross-validation training strategy to avoid overfitting and ensure a robust performance evaluation. Specifically, in a cyclic manner, we picked Cohort 1, Cohort 2, and Cohort 3 as the validation set and used the remaining two cohorts for training in the corresponding model. LungNet was implemented in TensorFlow (v 1.4). We showed that outcomes from LungNet are predictive of overall survival in all four independent survival cohorts as measured by concordance indices of 0.62, 0.62, 0.62, and 0.58 on cohorts 1, 2, 3, and 4, respectively. Further, the predicted outcomes show better risk-based stratification of patients than models trained with clinical variables such as age, sex, histology, and stage. Additionally, the risk stratification also worked well on early-stage cancers (stage 1 and stage 2); this is particularly promising since good risk stratification can facilitate tailored treatment from the early stages and lead to much better outcomes and quality of life later on. Overall, LungNet showed potential as a non-invasive predictor for prognosis in patients with NSCLC.

Despite such promising results, many challenges remain, including paucity of data, biases, and heterogeneity. Most studies, while promising, need to be validated on larger and more heterogeneous datasets to confirm their validity and usefulness in the clinic. This is particularly important because biases and heterogeneity in a dataset, from the demographic composition of patients to the machines and protocols used in CT acquisition can lead to models that do not generalize well on multi-institutional external cohorts. Of course, these challenges are nothing but new avenues of research, and we hope that MI models will facilitate routine decision-making in the clinic in the not-so-distant future.

**Summary**

MI can be applied to virtually any aspect of radiation research processing large datasets. Its application ranges from the establishment of ‘smart’ pipelines that automate data processing to truly novel data-driven discoveries that are not restricted by pre-defined analytical models. Machine-interpretable descriptions and a unifying ontology for the radiation sciences is an integral step towards meta-analysis through ML or other MI methods. Despite relevant data heterogeneity among patient demographics, some complex MI/ML pipelines can be successfully transferred across patient cohorts with appropriate integration of normalization techniques and personalized dose-effect estimates. In many cases, however, much work is still needed to address formidable barriers to data integration, like variations in radiation exposure reporting, pathology, and psychological testing practices over time. Following the prepared talks included in this symposium, lively discussion ensued among the panel of speakers regarding best practices and ways forward to harmonize and pool data, resources for overcoming obstacles in MI endeavors, and ideas for improving accessibility to large datasets and MI tools. Although still somewhat emergent, MI approaches to radiation sciences, such as those discussed in this Symposium, bode promising outcomes for big-data oriented discoveries, enabling researchers to interpret larger and more complex datasets to better understand radiation- and individual-specific parameters driving radiation-related outcomes. Such advancements can improve length and quality of life following planned radiation exposures.

**Acknowledgements**

LJW acknowledges the integral contributions of Andrew Green, PhD, Marcel van Herk, PhD, Eliana Vasquez Osorio, PhD, and Marianne Aznar, PhD, of the University of Manchester and Yimei Li, PhD, Thomas E Merchant, DO, PhD, and Austin M Faught, PhD, of St. Jude Children’s Research Hospital.

**Funding**

PNS acknowledges funding from the EURATOM research and training programme 2019-2020 under grant agreement No 900009, and the Alan Turing Institute. LJW acknowledges funding from the American Society for Radiation Oncology (ASTRO) and the American Association of Physicists in Medicine (AAPM).

**Declaration of Interest**

We have no conflicts of interest to disclose.

**Notes on contributors**

***Lydia J Wilson***, PhD, is a medical physics clinician scientist in the Department of Radiation Oncology at St. Jude Children’s Research Hospital where she divides her effort evenly between clinical and research endeavors. In this role, she works to improve the lives of tomorrow's cancer patients by engaging in research that pushes the boundaries of what is possible in radiation oncology while providing excellent clinical care to patients today. Her research aims to improve the long-term quality of life of childhood cancer survivors. Specifically, Dr. Wilson’s interests include better quantifying the delivered dose from fractionated and adaptive treatments and deepening our understanding of links between therapeutic radiation exposures and health effects.

***Frederico C. Kiffer***, PhD, is a postdoctoral research scientist in the Department of Anesthesia and Critical Care at the Children’s Hospital of Philadelphia. His work involves understanding the effects of charged-particle radiation on the brain and behavior to assess its risks and efficacy in space and in the clinic. To answer pressing questions related to radiation health, Dr. Kiffer uses novel deep learning approaches to estimate rodent behavior from video recordings.

***Daniel C. Berrios***, MD, is a physician scientist with additional training and experience in data science. Since 2014 he has been participating in the development and deployment of the NASA GeneLab database and NASA Life Sciences Data Archives. Prior to that, his experience at NASA included projects involving science data integration, semantic web technologies, spaceflight experiment operations and data systems design.

***Sylvain V. Costes***, PhD, obtained his doctorate from University of California at Berkeley in 1999, in the nuclear engineering department, where he worked on radiation biology and bioinformatics. He spent four years at the National Cancer Institute, 12 years at the Lawrence Berkeley National Laboratory working on NASA Human Research Program activities and the DOE low dose program. He joined NASA Ames Research Center in 2016 where he has been managing NASA Open Science efforts for Space Biology. He has been the Space Biosciences Research Branch Chief since 2020.

***Olivier Gevaert***, PhD, is an associate professor at Stanford University focusing on developing machine-learning methods for biomedical decision support from multi-scale data. He is an electrical engineer by training with additional training in MI, and a PhD in bioinformatics at the University of Leuven, Belgium. He continued his work as a postdoc in radiology at Stanford and then established his lab in the department of medicine in biomedical informatics. The Gevaert lab focuses on multi-scale biomedical data fusion primarily in oncology and neuroscience. The lab develops ML methods including Bayesian, kernel methods, regularized regression and deep learning to integrate molecular data or omics. The lab also investigates linking omics data with cellular and tissue data in the context of computational pathology, imaging genomics & radiogenomics.

***Bruno Matarèse***, PhD, is a physicist in biology and medicine with a particular interest in radiobiology and the use of combined electromagnetic and acoustic technology for early detection of cancer.

***Jack Miller***, PhD serves as a Subject Matter Expert in space radiation at NASA Ames Research Center, specializing in the physics of interactions of ionizing space radiation with materials and biological organisms.

***Pritam Mukherjee***, PhD, is a staff scientist at the National Institutes of Health Clinical Center. His primary research interests are ML applications for radiology. An electrical engineer by training, he obtained his PhD at the University of Maryland College Park. He went on to do a postdoctoral stint with Dr. Olivier Gevaert’s lab at Stanford University, where he worked on the application of ML to biomedical imaging and mining electronic medical records for the diagnosis and prognosis of diseases.

***Kristen Peach***, PhD is a plant biologist and data curator at NASA GeneLab. She holds a PhD in Ecology, Evolution and Marine Biology from the University of California Santa Barbara.

***Paul Schofield***, PhD, is Professor of Biomedical Informatics at the University of Cambridge and a Fellow of the Alan Turing Institute. His main interests are in the sharing and use of big biomedical data and particularly biosemantics in the discovery of pathological processes and the genetics of disease.

***Luke T. Slater*** achieved their PhD in 2020 at the University of Birmingham, where they continue to work as a postdoctoral fellow. Their research interests center around creating, questioning, and exploring computational representations of conceptual semantics across different domains and contexts; working to improve understanding and insight across scientific, public, and clinical frames of reference.

***Britta Langen***, PhD, is a radiobiological scientist in the Department of Radiation Oncology at UT Southwestern Medical Center. Her basic research integrates high-throughput gene expression analysis and statistical programming for biomarker discovery of treatment response, normal tissue risk, and radiation carcinogenesis. For this purpose, she established the RadioBioinformatics Initiative with Dr. Marcela Davila at University of Gothenburg, Sweden, and developed a ML framework for omics-based radiation research. Her pre-clinical cancer research focuses on superoxide dismutase mimetics as radiomodulators and enhancers of radioimmunotherapy.

**References**

[Althubaiti, Sara, Andreas Karwath, Ashraf Dallol, Adeeb Noor, Shadi Salem Alkhayyat, Rolina Alwassia, Katsuhiko Mineta, et al. 2019. “Ontology-Based Prediction of Cancer Driver Genes.” *Scientific Reports* 9 (1): 17405.](http://paperpile.com/b/33loEO/74zn8)

[Beasley, William, Maria Thor, Alan McWilliam, Andrew Green, Ranald Mackay, Nick Slevin, Caroline Olsson, et al. 2018. “Image-Based Data Mining to Probe Dosimetric Correlates of Radiation-Induced Trismus.” *International Journal of Radiation Oncology, Biology, Physics* 102 (4): 1330–38.](http://paperpile.com/b/33loEO/QQkR9)

[Berrios, Daniel C., Jonathan Galazka, Kirill Grigorev, Samrawit Gebre, and Sylvain V. Costes. 2021. “NASA GeneLab: Interfaces for the Exploration of Space Omics Data.” *Nucleic Acids Research* 49 (D1): D1515–22.](http://paperpile.com/b/33loEO/ee8e0)

[Birschwilks, M., M. Gruenberger, C. Adelmann, S. Tapio, G. Gerber, P. N. Schofield, and B. Grosche. 2011. “The European Radiobiological Archives: Online Access to Data from Radiobiological Experiments.” *Radiation Research* 175 (4): 526–31.](http://paperpile.com/b/33loEO/CWGvs)

[Chen, Jun, Azza Althagafi, and Robert Hoehndorf. 2021. “Predicting Candidate Genes from Phenotypes, Functions and Anatomical Site of Expression.” *Bioinformatics*  37 (6): 853–60.](http://paperpile.com/b/33loEO/YJsBl)

[Clarke, Daniel J. B., Lily Wang, Alex Jones, Megan L. Wojciechowicz, Denis Torre, Kathleen M. Jagodnik, Sherry L. Jenkins, et al. 2019. “FAIRshake: Toolkit to Evaluate the FAIRness of Research Digital Resources.” *Cell Systems* 9 (5): 417–21.](http://paperpile.com/b/33loEO/c6NGA)

[Clarke, M., R. Collins, S. Darby, C. Davies, P. Elphinstone, V. Evans, J. Godwin, et al. 2005. “Effects of Radiotherapy and of Differences in the Extent of Surgery for Early Breast Cancer on Local Recurrence and 15-Year Survival: An Overview of the Randomised Trials.” *The Lancet* 366 (9503): 2087–2106.](http://paperpile.com/b/33loEO/LoSRq)

[Cook, Daniel L., Jose L. V. Mejino, and Cornelius Rosse. 2004. “Evolution of a Foundational Model of Physiology: Symbolic Representation for Functional Bioinformatics.” *Studies in Health Technology and Informatics* 107 (Pt 1): 336–40.](http://paperpile.com/b/33loEO/eiWXw)

[Decherchi, Sergio, Elena Pedrini, Marina Mordenti, Andrea Cavalli, and Luca Sangiorgi. 2021. “Opportunities and Challenges for Machine Learning in Rare Diseases.” *Frontiers of Medicine* 8 (October): 747612.](http://paperpile.com/b/33loEO/g0D1Z)

[Griethuysen, Joost J. M. van, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G. H. Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo J. W. L. Aerts. 2017. “Computational Radiomics System to Decode the Radiographic Phenotype.” *Cancer Research* 77 (21): e104–7.](http://paperpile.com/b/33loEO/wtj4v)

[Harris, Kassem, Inga Khachaturova, Basem Azab, Theodore Maniatis, Srujitha Murukutla, Michel Chalhoub, Hassan Hatoum, et al. 2012. “Small Cell Lung Cancer Doubling Time and Its Effect on Clinical Presentation: A Concise Review.” *Clinical Medicine Insights. Oncology* 6 (May): 199–203.](http://paperpile.com/b/33loEO/ag2uT)

[~~Hastings, Janna, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, and Christoph Steinbeck. 2016. “ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites.”~~ *~~Nucleic Acids Research~~* ~~44 (D1): D1214–19.~~](http://paperpile.com/b/33loEO/9URYJ)

[Hayamizu, Terry F., Richard A. Baldock, and Martin Ringwald. 2015. “Mouse Anatomy Ontologies: Enhancements and Tools for Exploring and Integrating Biomedical Data.” *Mammalian Genome: Official Journal of the International Mammalian Genome Society* 26 (9-10): 422–30.](http://paperpile.com/b/33loEO/Di2GH)

[Heemsbergen, Wilma D., Abrahim Al-Mamgani, Marnix G. Witte, Marcel van Herk, Floris J. Pos, and Joos V. Lebesque. 2010. “Urinary Obstruction in Prostate Cancer Patients from the Dutch Trial (68 Gy vs. 78 Gy): Relationships with Local Dose, Acute Effects, and Baseline Characteristics.” *International Journal of Radiation Oncology, Biology, Physics* 78 (1): 19–25.](http://paperpile.com/b/33loEO/amf9a)

[Jackson, Rebecca, Nicolas Matentzoglu, James A. Overton, Randi Vita, James P. Balhoff, Pier Luigi Buttigieg, Seth Carbon, et al. 2021. “OBO Foundry in 2021: Operationalizing Open Data Principles to Evaluate Ontologies.” *Database: The Journal of Biological Databases and Curation* 2021 (October). https://doi.org/](http://paperpile.com/b/33loEO/SvHQb)[10.1093/database/baab069](http://dx.doi.org/10.1093/database/baab069)[.](http://paperpile.com/b/33loEO/SvHQb)

[Jebsen, Nina L., Clement S. Trovik, Henrik C. F. Bauer, Anders Rydholm, Odd R. Monge, Kirsten Sundby Hall, Thor Alvegård, and Øyvind S. Bruland. 2008. “Radiotherapy to Improve Local Control Regardless of Surgical Margin and Malignancy Grade in Extremity and Trunk Wall Soft Tissue Sarcoma: A Scandinavian Sarcoma Group Study.” *International Journal of Radiation Oncology\*Biology\*Physics* 71 (4): 1196–1203.](http://paperpile.com/b/33loEO/9NPiB)

[Kulmanov, Maxat, Fatima Zohra Smaili, Xin Gao, and Robert Hoehndorf. 2021. “Semantic Similarity and Machine Learning with Ontologies.” *Briefings in Bioinformatics* 22 (4). https://doi.org/](http://paperpile.com/b/33loEO/KuFzC)[10.1093/bib/bbaa199](http://dx.doi.org/10.1093/bib/bbaa199)[.](http://paperpile.com/b/33loEO/KuFzC)

[McWilliam, Alan, Jason Kennedy, Clare Hodgson, Eliana Vasquez Osorio, Corinne Faivre-Finn, and Marcel van Herk. 2017. “Radiation Dose to Heart Base Linked with Poorer Survival in Lung Cancer Patients.” *European Journal of Cancer*. https://doi.org/](http://paperpile.com/b/33loEO/zsotY)[10.1016/j.ejca.2017.07.053](http://dx.doi.org/10.1016/j.ejca.2017.07.053)[.](http://paperpile.com/b/33loEO/zsotY)

[Mukherjee, Pritam, Mu Zhou, Edward Lee, Anne Schicht, Yoganand Balagurunathan, Sandy Napel, Robert Gillies, et al. 2020. “A Shallow Convolutional Neural Network Predicts Prognosis of Lung Cancer Patients in Multi-Institutional CT-Image Data.” *Nature Machine Intelligence* 2 (5): 274–82.](http://paperpile.com/b/33loEO/qYEKY)

[Palma, Giuseppe, Serena Monti, Vittoria D’Avino, Manuel Conson, Raffaele Liuzzi, Maria Cristina Pressello, Vittorio Donato, et al. 2016. “A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage.” *International Journal of Radiation Oncology\*Biology\*Physics*. https://doi.org/](http://paperpile.com/b/33loEO/PnT56)[10.1016/j.ijrobp.2016.04.033](http://dx.doi.org/10.1016/j.ijrobp.2016.04.033)[.](http://paperpile.com/b/33loEO/PnT56)

[Robinson, Peter N., and Melissa A. Haendel. 2020. “Ontologies, Knowledge Representation, and Machine Learning for Translational Research: Recent Contributions.” *Yearbook of Medical Informatics* 29 (1): 159–62.](http://paperpile.com/b/33loEO/QkSo8)

[Schofield, Paul N., Ulrike Kulka, Soile Tapio, and Bernd Grosche. 2019. “Big Data in Radiation Biology and Epidemiology; an Overview of the Historical and Contemporary Landscape of Data and Biomaterial Archives.” *International Journal of Radiation Biology* 95 (7): 861–78.](http://paperpile.com/b/33loEO/u7M3h)

[Schofield, Paul N., John P. Sundberg, Beth A. Sundberg, Colin McKerlie, and Georgios V. Gkoutos. 2013. “The Mouse Pathology Ontology, MPATH; Structure and Applications.” *Journal of Biomedical Semantics* 4 (1): 18.](http://paperpile.com/b/33loEO/BydoO)

[Shah, Rajesh P., Heather M. Selby, Pritam Mukherjee, Shefali Verma, Peiyi Xie, Qinmei Xu, Millie Das, Sachin Malik, Olivier Gevaert, and Sandy Napel. 2021. “Machine Learning Radiomics Model for Early Identification of Small-Cell Lung Cancer on Computed Tomography Scans.” *JCO Clinical Cancer Informatics* 5 (June): 746–57.](http://paperpile.com/b/33loEO/lLG3X)

[Smaili, Fatima Zohra, Xin Gao, and Robert Hoehndorf. 2020. “Formal Axioms in Biomedical Ontologies Improve Analysis and Interpretation of Associated Data.” *Bioinformatics*. https://doi.org/](http://paperpile.com/b/33loEO/ASWXT)[10.1093/bioinformatics/btz920](http://dx.doi.org/10.1093/bioinformatics/btz920)[.](http://paperpile.com/b/33loEO/ASWXT)

[Suh, Eugene, Kayla L. Stratton, Wendy M. Leisenring, Paul C. Nathan, Jennifer S. Ford, David R. Freyer, Jennifer L. McNeer, et al. 2020. “Late Mortality and Chronic Health Conditions in Long-Term Survivors of Early-Adolescent and Young Adult Cancers: A Retrospective Cohort Analysis from the Childhood Cancer Survivor Study.” *The Lancet Oncology* 21 (3): 421–35.](http://paperpile.com/b/33loEO/LeQkI)

[Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.” *CA: A Cancer Journal for Clinicians* 71 (3): 209–49.](http://paperpile.com/b/33loEO/EPtdt)

The Gene Ontology Consortium. 2021. [“The Gene Ontology Resource: Enriching a GOld Mine.” *Nucleic Acids Research* 49 (D1): D325–34.](http://paperpile.com/b/33loEO/AIHtg)

[Torre, Siegel, Ward, and Jemal. 2016. “Global Cancer Incidence and Mortality Rates and Trends—An UpdateGlobal Cancer Rates and Trends—An Update.” *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*.](http://paperpile.com/b/33loEO/kQP6u) <https://aacrjournals.org/cebp/article-abstract/25/1/16/157144>[.](http://paperpile.com/b/33loEO/kQP6u)

[Veiga, Catarina, Pei Lim, Virginia Marin Anaya, Edward Chandy, Reem Ahmad, Derek D’Souza, Mark Gaze, Syed Moinuddin, and Jennifer Gains. 2021. “Atlas Construction and Spatial Normalisation to Facilitate Radiation-Induced Late Effects Research in Childhood Cancer.” *Physics in Medicine & Biology*. https://doi.org/](http://paperpile.com/b/33loEO/ijgaE)[10.1088/1361-6560/abf010](http://dx.doi.org/10.1088/1361-6560/abf010)[.](http://paperpile.com/b/33loEO/ijgaE)

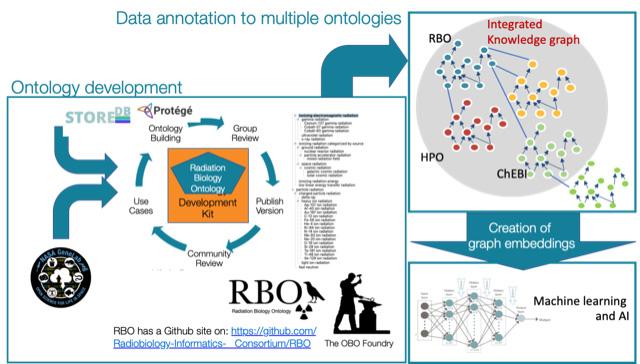
[Wilkinson, Mark D., Michel Dumontier, I. Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (March): 160018.](http://paperpile.com/b/33loEO/unN49)

Wilkinson MD, Dumontier M, Sansone S-A, da Silva Santos LOB, Prieto M, McQuilton P, Gautier J, Murphy D, Crosas M, Schultes E. 2018. Evaluating FAIR-Compliance Through an Objective, Automated, Community-Governed Framework. bioRxiv.418376.

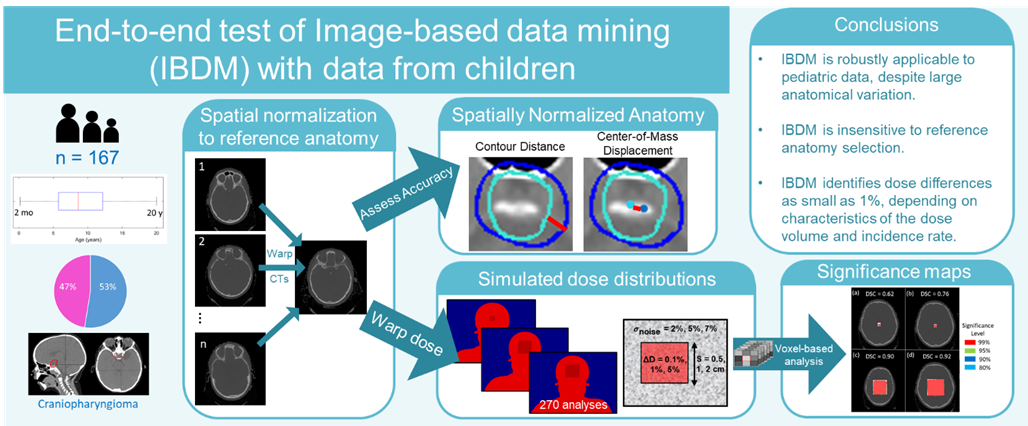
Wilson LJ, Bryce-Atkinson A, Green A, Li Y, Merchant TE, van Herk M, Vasquez Osorio E, Faught AM, Aznar MC. “Image-based data mining applies to data collected from children.” *Physica Medica*. 2022;99:31-43.

[Zwanenburg, Alex, Martin Vallières, Mahmoud A. Abdalah, Hugo J. W. L. Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, et al. 2020. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping.” *Radiology* 295 (2): 328–38.](http://paperpile.com/b/33loEO/YcEHc)

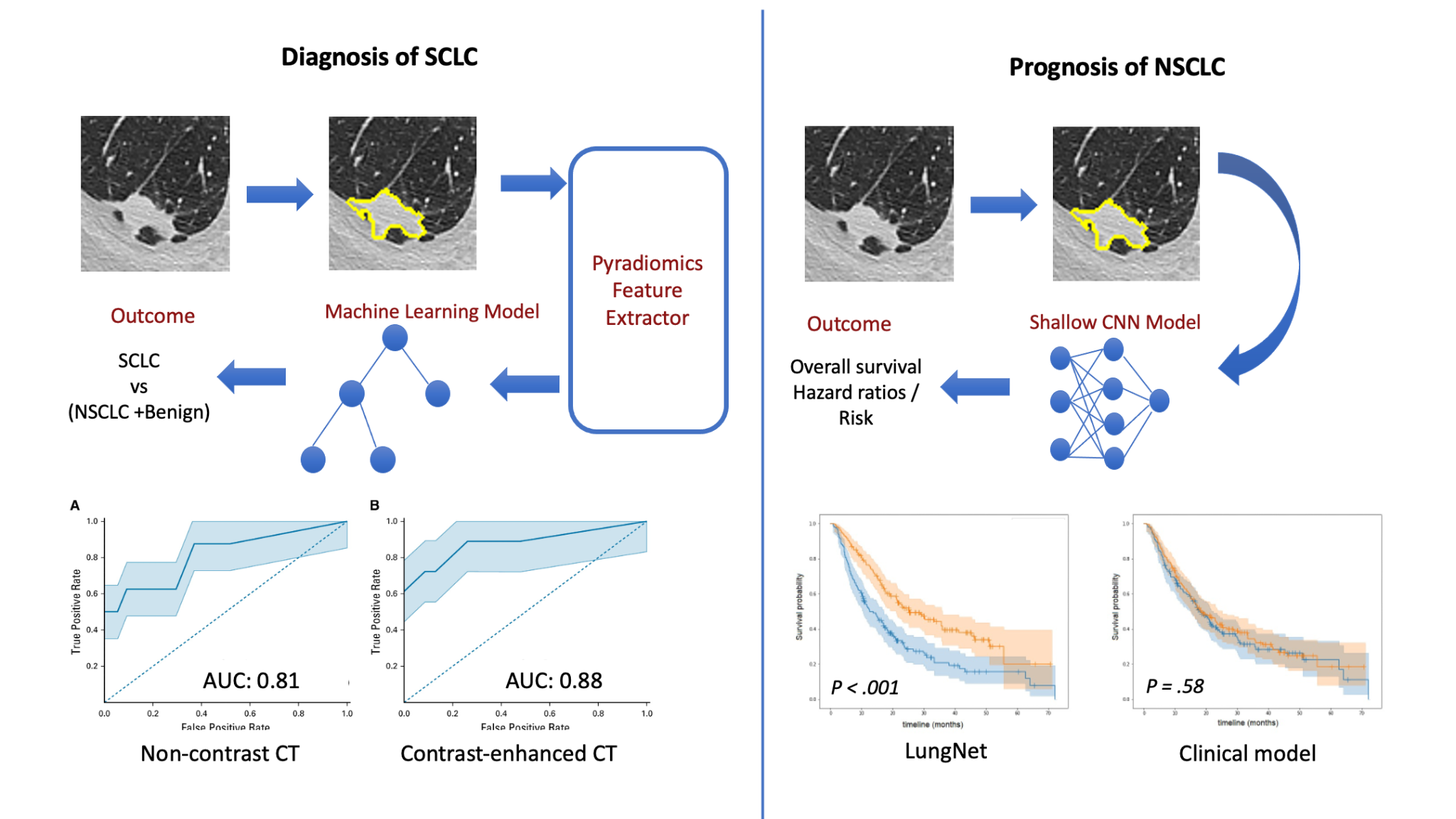
**Figures**



**Figure 1.** Graphical overview of the Radiation Biology Ontology including the integrated knowledge, plans for ongoing review and revision, and link for public access.



**Figure 2.** Graphical overview of the methods and major conclusions reported in the presentation. For the simulated dose distributions, σnoise is the magnitude of random noise, ΔD is the dose deviation, and S is the side length of the volume over which that dose difference exists. The significance maps show significant voxels overlaid on the test volume and DSC is the Dice Similarity Coefficient comparing the collection of significant voxels to the volume in which we introduced a dose deviation.



**Figure 3.** Graphical overview of the methods and results covered in the symposium. (left figure) Diagnosing small-cell lung cancer (SCLC) using a radiomics workflow – image features extracted from the segmented nodule using pyradiomics were used to train machine learning models for discriminating SCLC from non-small-cell lung cancer (NSCLC) and benign nodules. (right figure) Prognosticating overall survival for NSCLC – a deep learning model was developed to predict overall survival based on CT scans. CNN: convolutional neural network; CT: computed tomography.