Applications of Deep Learning to Differential Equation Models in Oncology

by

Cameron Meaney

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Applied Mathematics

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Lance Munn Professor, Dept. of Radiation Oncology, Harvard Medical School
Supervisor(s):	Mohammad Kohandel Professor, Dept. of Applied Math, Univ. of Waterloo
Internal Member(s):	Sander Rhebergen Associate Professor, Dept. of Applied Math, Univ. of Waterloo
	Sivabal Sivaloganathan Professor, Dept. of Applied Math, Univ. of Waterloo
Internal-External Member:	Ali Ghodsi Professor, Dept. of Statistics and Actuarial Science, Univ. of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The integration of quantitative tools in biology and medicine has led to many groundbreaking advances in recent history, with many more promising discoveries on the horizon. Conventional mathematical models, particularly differential equation-based models, have had great success in various biological applications, including modelling bacterial growth, disease propagation, and tumour spread. However, these approaches can be somewhat limited due to their reliance on known parameter values, initial conditions, and boundary conditions, which can dull their applicability. Furthermore, their forms are directly tied to mechanistic phenomena, making these models highly explainable, but also requiring a comprehensive understanding of the underlying dynamics before modelling the system. On the other hand, machine learning models typically require less prior knowledge of the system but require a significant amount of data for training. Although machine learning models can be more flexible, they tend to be black boxes, making them difficult to interpret.

Hybrid models, which combine conventional and machine learning approaches, have the potential to achieve the best of both worlds. These models can provide explainable outcomes while relying on minimal assumptions or data. An example of this is physics-informed neural networks, a novel deep learning approach that incorporates information from partial differential equations into the optimization of a neural network. This hybrid approach offers significant potential in various contexts where differential equation models are known, but data is scarce or challenging to work with. Precision oncology is one such field.

This thesis employs hybrid conventional/machine learning models to address problems in cancer medicine, specifically aiming to advance personalized medicine approaches. It contains three projects. In the first, a hybrid approach is used to make patient-specific characterizations of brain tumours using medical imaging data. In the second project, a hybrid approach is employed to create subject-specific projections of drug-carrying cancer nanoparticle accumulation and intratumoral interstitial fluid pressure. In the final project, a hybrid approach is utilized to optimize radiation therapy scheduling for tumours with heterogeneous cell populations and cancer stem cells.

Overall, this thesis showcases several examples of how quantitative tools, particularly those involving both conventional and machine learning approaches, can be employed to tackle challenges in oncology. It further supports the notion that the continued integration of quantitative tools in medicine is a key strategy in addressing problems and open questions in healthcare.

Acknowledgements

Modern scientific research is a highly collaborative effort. The days of the renaissance men who mastered and drove progress in multiple fields are long gone. And as someone who has dedicated nearly a decade of study to a single discipline, I still find it challenging to consider myself a true expert in even one field. This thesis, though written by me, is no exception: it would not have been possible without the direct and indirect assistance of many individuals.

I express my deepest gratitude to my direct collaborators on the projects included in this thesis and to any others who generously shared their time and expertise with me. Your contributions, big or small, have not gone unnoticed and are sincerely appreciated.

I would like to thank Dr. Matthew Scott, whose mentorship and guidance during my early research days ignited my passion for quantitative biology, which has since blossomed into a career in mathematical medicine. His expertise, enthusiasm, and encouragement early on have been invaluable throughout my academic journey.

I extend my thanks to the other current and former members of the mathematical medicine group at UW, who have provided me with a supportive and inspiring research environment. Their kindness, expertise, and collaborative spirit have been invaluable to me throughout the years. I hope that I have also been able to give back to them in some measure and express my sincerest gratitude for their friendship and guidance.

I would like to give special recognition to a few members of the research group who have played instrumental roles in my academic journey. Moriah, your help and friendship in those early days of my Master's was critical to my success. Michelle, I cannot thank you enough for your guidance and frank, honest feedback throughout which has helped me not only in research, but in shaping my career path as well. And Brydon, you have been an endless well of knowledge and an essential resource for me throughout my PhD. Having someone as bright, driven, and selfless as you one step ahead of me in our career journeys has provided me with immense motivation and inspiration. Your expertise and generosity have been incredibly valuable, and I am grateful to have you as a mentor and friend.

Lastly, and most importantly, I would like to express my sincere gratitude to my supervisor Dr. Mohammad Kohandel who has been a part of my academic journey from the very beginning of my undergraduate degree to the very end of my PhD. Together, we have collaborated on numerous projects, attended conferences, and communicated nearly every week for the past 10 years. You have provided me with countless opportunities to enrich my research and career, which I cannot thank you enough for. Although one day our collaboration will inevitably come to an end, it will be a sad day, and I will enjoy it until that day arrives.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA Anti-Angiogenic Agents 5

ADC Apparent Diffusion Coefficient 42–44, 60

ANN Artificial Neural Networks 20, 21, 24, 25, 27

BED Biologically Effective Dose 90, 91

CSC Cancer Stem Cell 3, 4, 7, 91–93, 109, 112, 114, 115, 117

CT Computed Tomography 4, 69, 70, 117

DWI Diffusion Weighted Imaging 4, 7, 39, 40, 42, 43, 58, 60

ECM Extracellular Matrix 3

EPR Enhanced Permeability and Retention 62

FEM Finite Element Method 13–16, 48, 81

FeTS Federated Tumor Segmentation 42–45, 50, 58

GBM Glioblastoma Multiforme 36, 37, 39, 40, 42, 44, 45, 48

HAP Hypoxia-Activated Prodrugs 5, 117

HIFU High Intensity Focused Ultrasound 118

IFP Interstitial Fluid Pressure 4, 7, 61–63, 65–69, 72–74, 76, 79, 81, 83, 85–88, 117

L-BFGS Limited-Memory Broyden-Fletcher-Goldfarb-Shanno 23, 70, 72

LQ Linear Quadratic 9, 90, 106, 114

MRI Magnetic Resonance Imaging 4, 7, 37–40, 58, 60, 62, 63, 116

MSE Mean-Squared Error 22, 26, 27, 29

OAR Organs at Risk 90, 91

ODE Ordinary Differential Equations 10, 11, 13, 90, 93, 106, 115

PDE Partial Differential Equations 10, 13–16, 24–27, 29, 31, 33–35, 39, 40, 45–50, 59, 62, 63, 65–68, 70, 72, 81, 82, 86, 90, 117

PI Proliferation-Invasion 13, 15, 16, 24, 37–40, 44–49, 56, 59, 60, 116

PINN Physics-Informed Neural Networks 7, 25–27, 29, 31, 33, 34, 45, 46, 48, 62, 63, 65, 67, 68, 70, 76, 81, 85, 86, 88, 116–118

SGD Stochastic Gradient Descent 23

SIR Susceptible-Infected-Recovered 10, 11

TCP Tumour Control Probability 90

TME Tumour Microenvironment 3, 4

Chapter 1

Introduction

In the 20th century, the introduction of quantitative methods revolutionized physics, enabling remarkable scientific progress. Today, the application of these quantitative tools has expanded, with increasing complexity and utilization in diverse fields. Notably, the last few decades have witnessed rapid progress in biology and medicine, largely attributable to the integration of quantitative methods, including groundbreaking technologies like medical imaging and genomics. Integrating quantitative tools with life sciences has opened up new avenues for understanding basic biological processes and solving biological problems, leading to valuable insights with the potential to transform medicine and improve global health outcomes.

This increase of quantitative tools in medicine has paved the way for new insights and breakthroughs in cancer research. Cancer remains one of the most pressing global health challenges of our time, affecting millions of people worldwide [107]. Despite the tremendous strides made in cancer research, understanding and treating the disease remains a daunting challenge. Every patient is unique, and the dynamic nature of tumours, which evolve over time and space, makes predicting cancer progression and developing an effective treatment plan even more complex. Medical imaging, which is often used to guide treatment decisions, can be both expensive and difficult to interpret accurately. In addition, cancer cells are adept at evolving to bypass even our best therapies, leading to relapse and treatment failure. All of these factors make it challenging to develop personalized treatment plans that are tailored to the unique traits and circumstances of a given patient, and underscore the critical need for innovative approaches that can account for the complex and dynamic nature of cancer. However, advances in mathematics and computation and their continued integration into modern cancer research has provided a suite of powerful tools for cancer researchers. Mathematical models, data analysis, and computational simulations allow for

uncovering the complex relationships underlying cancer data, enabling a new way to tackle many of the traditional challenges that have plagued scientific progression.

1.1 What is Cancer?

Cancer is not one single disease, but a category of diseases characterized by common traits, particularly the over-proliferation and spread of abnormal cells within the body. These cells can form tumours, invade surrounding tissue, and metastasize to other parts of the body through the bloodstream or lymphatic system. There are many different types of cancer, each with its own unique characteristics and treatment options. Despite this variability, a set of key traits, originally identified by Hanahan and Weinberg, known as the "hallmarks of cancer" [40] commonly occur throughout all types. These are:

1. **Sustaining Proliferative Signalling**: Cancer cells can divide, and signal other cells to divide, even when they shouldn't
2. **Evading Growth Suppressors**: Cancer cells can ignore signals from other cells to stop dividing
3. **Activating Invasion and Metastasis**: Cancer cells can relocate from their original position in the body, sometimes creating new tumours where they end up
4. **Enabling Replicative Immortality**: Cancer cells can divide indefinitely, unlike normal cells which have a limited number of divisions
5. **Inducing Angiogenesis**: Cancer cells can stimulate the growth of new blood vessels to supply them with nutrients and oxygen to help them grow
6. **Resisting Cell Death**: Cancer cells can resist the types of cell death which would normally eliminate abnormal or damaged cells
7. **Avoiding Immune Destruction**: Cancer cells can evade detection by the immune system and avoid destruction by immune cells
8. **Tumour-Promoting Inflammation**: Cancer cells can contribute to chronic inflammation which can promote the growth and survival of cancer cells
9. **Genome Instability and Mutation**: Cancer cells can mutate faster than normal cells and can accumulate genetic mutations that alter their behavior and contribute to cancer development

10. **Deregulating Cellular Energetics**: Cancer cells can alter their metabolism to obtain energy in different ways than normal cells, allowing them to survive and grow even in challenging conditions

These hallmarks of cancer are the result of complex interactions between the tumour cells and their microenvironment, and understanding them is crucial for the development of effective cancer treatments.

1.1.1 The Tumour Microenvironment

A wealth of research works has made it clear that tumours are not simply collections of abnormal, aggressive cells, but that the conditions surrounding a tumour are arguably equally as important as the traits of the cells themselves [40]. The Tumour Microenvironment (TME) is comprised of the components surrounding a tumour, both cellular and noncellular, including blood vessels, immune cells, and the Extracellular Matrix (ECM). The TME plays a critical role in cancer progression and response to therapy, as it can create a variety of pressures and stresses that influence tumour behavior. For example, the TME can be hypoxic (low in oxygen), which can promote genetic instability and alter cellular metabolism [84]. Additionally, high levels of interstitial fluid pressures are often observed in the TME, contributing to a wide range of pro-cancer behaviours and treatment resistance [78, 110, 111]. Investigations of the high fluid pressures in the TME is the focus of Chapter 4 of this thesis. Importantly, these factors in the TME can vary temporally and spatially within the tumour, giving it a degree of heterogeneity. Understanding the TME, tumour heterogeneity, and their roles in cancer progression is therefore crucial for the development of effective cancer treatments.

1.1.2 Cancer Stem Cells

A crucial source of intratumoral heterogeneity comes from the presence of different cell types within a tumour that have distinct characteristics. Until recently, cellular heterogeneity within a tumour was explained by the clonal evolution model [86]. According to this model, cellular heterogeneity arises within the tumour due to the accumulation of successive mutations which confer an advantageous adaptation in the TME. Furthermore, these mutations are often accelerated by further genomic instability, allowing for increased mutation rates and a higher likelihood of further mutations. However, the Cancer Stem Cell (CSC) model has emerged in recent years as a fuller explanation of intratumoral heterogeneity [127]. In this model, CSCs are a subpopulation of cells within tumours that

drive their growth, having the ability to both self-renew and differentiate into different cell types. CSCs have been shown to be resistant to conventional cancer therapies and are thought to play a key role in tumour progression, recurrence, and metastasis [127]. This is in part due to their ability to regenerate a tumour even after treatment has eliminated most of the bulk tumour cells. CSCs have been identified in a variety of cancers, and understanding the role of CSCs in tumorigenesis and their interaction with the TME is crucial for the development of effective cancer therapies. Modelling the role of CSCs in resistance to radiotherapy is the focus of Chapter 6 of this thesis.

1.2 Medical Imaging

Medical imaging has become an indispensable tool in modern cancer medicine, providing clinicians with a window into the internal workings of the body and enabling the detection, diagnosis, and monitoring of cancer. With its ability to capture images of the body at various levels of resolution, from the macroscopic to the molecular, medical imaging has transformed the way we approach cancer, allowing us to visualize the disease in real-time and at various stages of progression. As such, it is a key component of many cancer treatment plans, from initial diagnosis to monitoring the efficacy of therapies. Additionally, medical imaging provides a wealth of data that can be used as input to quantitative models. Given its importance in cancer medicine, medical imaging will play a crucial role in the research outlined in this thesis, as we seek to develop novel approaches to analyzing and interpreting complex imaging data to better understand cancer and improve treatment outcomes. In particular, Magnetic Resonance Imaging (MRI) and its subtypes, specifically Diffusion Weighted Imaging (DWI), are extensively used throughout Chapter 3 to examine brain tumours. Additionally, Chapter 4 employs Computed Tomography (CT) imaging of mice to evaluate liposome accumulation and Interstitial Fluid Pressure (IFP).

1.3 Cancer Therapies

Cancer therapies have advanced significantly from the crude surgeries and highly toxic compounds employed throughout history. Today, modern cancer treatment has four main pillars: surgery, radiation therapy, chemotherapy, and immunotherapy. Surgery, which involves the physical removal of tumours, is often used for localized cancers that have not yet spread to other parts of the body. Radiation therapy utilizes electromagnetic radiation to kill cancer cells and shrink tumours. It is commonly applied to tumours that

are difficult to remove surgically or following surgery in attempt to kill non-resected cells and prevent recurrence. Chemotherapy uses drugs to kill cancer cells and is commonly used for cancers that have spread beyond the primary tumour site. And the youngest pillar is immunotherapy, which involves the use of drugs or other substances to stimulate the own immune system of the patient to attack cancer cells. Immunotherapy is often used for cancers that have not responded to other treatments or have a high probability of responding to immunotherapy. Together, these four pillars form the basis of modern cancer treatment and are often used in combination to provide the best possible outcomes.

In this thesis, two types of cancer therapies are the focus: radiation therapy and nanoparticle therapy. Nanoparticle therapy involves the use of tiny particles, typically smaller than 100 nanometers in size, to selectively target cancer cells. These nanoparticles (or equivalently, drug-carrying liposomes) can be engineered to deliver drugs or other therapeutic agents directly to the tumour site, minimizing damage to healthy tissues. In addition, some nanoparticles can be designed to respond to specific stimuli, such as changes in pH or temperature, which can enhance their tumour-targeting properties. Nanoparticle therapy is a rapidly growing field and has shown promise in preclinical studies [133]. Some of my previous research projects (not included in this thesis) have focused on cancer nanotherapies. In one project, I examined the combination of Hypoxia-Activated Prodrugs (HAP) and Anti-Angiogenic Agents (AA) administered through cancer nanoparticles, showing a theoretical advantage over separate administration [75]. In another project, I contributed to a review article on the use of *in silico* methods to advance the clinical translation of cancer nanomedicine [81]. The work presented in Chapter 4 involves the use of cancer drug-carrying liposomes. Chapter 5 focuses on the other primary therapy in this thesis, optimizing the application of radiation therapy to heterogeneous tumour populations.

1.4 Mathematical Oncology

Mathematical oncology has emerged as a rapidly growing subfield within cancer research, integrating the power of quantitative tools and modelling approaches to tackle the complexities of cancer. This field has blossomed in recent years with the increase in data from medical imaging, advanced biomarking, and genetic sequencing providing a wealth of information to fuel mathematical models and simulations. Moreover, the greater acceptance of these quantitative tools by biologists and clinicians has led to a more fruitful and collaborative approach to cancer research, with interdisciplinary teams working together to uncover new insights into the disease. It has become an increasing focus for applied mathematicians, and many pharmaceutical and biotechnology companies has even begun

hiring modelling teams to supplement their in-house research. Mathematical oncology provides a powerful framework for the interpretation of complex data and the development of personalized treatment plans that can account for the dynamic nature of cancer. As this field continues to grow and evolve, it holds great promise for improving cancer outcomes and advancing our understanding of this complex disease.

1.4.1 Personalized Oncology

Medical science is often challenged by the unique nature of each individual patient, which makes predicting treatment effects extremely difficult. Consequently, researchers and clinicians often rely on probabilities and design treatments for the average patient. Personalized medicine, sometimes referred to as precision medicine, is an emerging field that aims to leverage the individual traits of patients to guide decisions on the prevention, diagnosis, and treatment of diseases. By taking additional patient-specific information into account, such as a patient genetic profile, environmental factors, or specific data obtained through imaging or biopsy, personalized medicine can improve patient outcomes by identifying the most effective therapies and preventive measures for each individual.

Personalized oncology, a subfield of personalized medicine, focuses on tailoring cancer treatments and prevention strategies based on unique patient tumour biology, genetic makeup, and other relevant factors. This approach enables clinicians to select the most effective treatments for each individual, potentially reducing side effects and improving overall survival rates. Well known examples are the development of cancer vaccines and immunotherapies designed to harness the immune system to fight cancer. For example, in breast cancers where the HER2 protein is overexpressed, the monoclonal antibody therapy Herceptin (Trastuzumab) has proven effective; however, it does not work as well in cases where the protein is not overexpressed [46]. Personalized oncology seeks to identify such patient-specific traits and incorporate them into treatment plans. This idea is the goal of much current research, and is slowly making its way into mainstream scientific thinking (see [1] for example).

Quantitative analysis plays a crucial role in advancing personalized oncology, and the central goal of this thesis is to demonstrate how mathematical and computational tools can be leveraged to identify patient-specific traits and utilize them in cancer medicine. With the power of personalized oncology, researchers and clinicians are paving the way for more precise and effective cancer treatments, ultimately improving patient care and outcomes.

[1] https://www.youtube.com/watch?v=9nmsSo2QbIs

1.5 Summary of Thesis

This thesis focuses on mathematical oncology and presents the findings of three research projects, each of which have been published in or submitted to scientific journals. As published works have already been subject to significant scrutiny through the peer review process, they have been, in general, minimally altered from their published or submitted forms. Alterations to them have been primarily for the purposes of adding relevant background information and increasing the clarity and flow of the full thesis. Though this does entail some redundancy in the explanation of the methods, it is worth keeping their original explanations to ensure proper context for the results of each project.

Chapter 2 provides an overview of the key quantitative methods used throughout the thesis, including differential equation modelling, numerical simulations of mathematical models, machine learning, artificial neural networks, and Physics-Informed Neural Networks (PINN). In the subsequent chapters, these methods are employed to tackle problems in cancer research.

Chapter 3 details an algorithm for characterizing brain tumours in a patient-specific manner, utilizing MRI, DWI, and PINNs. This work, entitled *Deep Learning Characterization of Brain Tumours with Diffusion Weighted Imaging* was published in the *Journal of Theoretical Biology* in 2022 [73].

Chapter 4 presents a novel method for predicting the accumulation of drug-carrying liposomes in tumour tissue and estimating IFP in a patient-specific manner, leveraging the power of mathematical modelling and PINNs. This work, entitled *Prediction of Intratumoral Fluid Pressures and Liposome Accumulation Using Deep Neural Networks* is currently in submission to *Scientific Reports*.

Chapter 5 introduces a new approach to optimizing radiation therapy in heterogeneous cell populations with CSCs, using mathematical modelling and machine learning to develop personalized treatment plans. This work was published in the *Journal of Mathematical Biology* in 2022 and was entitled *Temporal optimization of radiation therapy to heterogeneous tumour populations and cancer stem cells*.

Finally, Chapter 6 summarizes the key findings of this thesis and offers conclusions and future directions for research in this field.

Chapter 2

Quantitative Methods

2.1 Introduction

Quantitative methods, or in more sophisticated terms, *in silico* methods, offer a means to address problems in biology and medicine that would otherwise be intractable. Beyond addressing existing challenges, these methods can also lead to novel insights and approaches that can be subsequently validated through experimentation. Many (most, frankly) mathematical models used in practice however are simply too complex to be solved analytically by hand. As a result, researchers must employ computational and statistical techniques to solve these models and extract meaningful conclusions. This is especially true in the context of cancer, where the inherently dynamic and complex nature of tumours makes them exceedingly difficult to investigate. However, this complexity provides an ideal context for quantitative methods to demonstrate their effectiveness. This chapter provides an overview of the mathematical and computational methods used in the various projects within this thesis.

2.2 Mathematical Modelling

As a general term, a mathematical model simply refers to any quantitative, abstract representation of a system. Models in cancer medicine range in sophistication, from simple probabilistic models - *how likely is a patient to still be alive 5 years following a diagnosis of acute myeloid leukemia?* - to extremely complex spatiotemporal models - *given imaging of a brain tumour, what will be the voxel-by-voxel tumour cellularity one week later?*

Variability in modelling extends beyond just their complexity though: every model has its own strengths and weaknesses, capabilities and use cases, assumptions and omissions. No mathematical model is perfect, and indeed, no grand unifying theory for cancer modelling exists, but the selective application of mathematical modelling in medicine, when used for the right purpose, can have tremendous benefits. They can allow for the ability to estimate quantities that are difficult, costly, or ethically challenging to measure directly, provide a structural framework for analysis or inference, aid in building a deeper understanding of the mechanisms underlying complex medical phenomena, increase convenience and automation in the standard clinical pipeline, and produce novel, testable hypotheses which could be later confirmed by experimentation or clinical trials. Naturally, mathematical modelling is a key concept in this thesis.

2.2.1 Differential Equations

Differential equations serve as fundamental tools in mathematical modelling, and all projects described in this thesis employ them. The core principle behind differential equations is that the quantities of interest can be expressed as continuous functions of continuous variables, and that mathematical relationships can be established between these functions and their rates of change. Importantly, differential equations inherently assume that the quantities of interest are deterministic, meaning that the evolution of the system under investigation is reproducible and will produce the same outcome each time. In physics, most quantities of interest (e.g., position, velocity, acceleration, mass, momentum, force) are continuous, deterministic variables, exhibiting values that exist within a continuous range without unattainable intermediate values and without randomness in their temporal evolution. Many fundamental laws of physics, which establish relationships between these quantities, are predicated on these assumptions, such as Newton's second law, which can be succinctly expressed as a differential equation. In many applications in biology however, these assumptions can become somewhat tenuous.

The classical deterministic model of radiation-induced cell killing, known as the Linear Quadratic (LQ) model, exemplifies this issue. The LQ model is given by the equation:

$$S(D) = e^{-\alpha D - \beta D^2} \tag{2.1}$$

where S represents the proportion of surviving cells, D denotes the radiation dose applied to the cells, and α and β are tissue-specific radiobiological parameters. This model is widely used to model radiative cell killing [18, 49]. Although the LQ model is generally accurate in practice, it exhibits limitations in certain cases. For instance, according to this

equation, the proportion of surviving cells never reaches zero for any applied dose, implying that tumour regrowth is inevitable, which in reality, it is not. And furthermore, for small tumours consisting of only a few cells, one would expect a high likelihood that radiation would eradicate every cell in such a population. Zaider and Minerbo [135] reformulated this question using stochastic modelling (a birth-death process) rather than differential equation modelling and demonstrated that the classical understanding falters in cases where the cell population is small, though remains a reasonable approximation when the population is sufficiently large.

A similar limitation of differential equation modelling arises in the context of mutant fixation probabilities, which refers to the probability that a single mutant will become fixed in a population of a fixed size—in other words, that all individuals in the population will eventually be descendants of this mutant. In a differential equation model, the probability of mutant fixation would be either 0 or 1 depending on the mutant's fitness relative to the rest of the population. Though in reality of course, there is a nonzero chance that the mutant dies before having a chance to divide. Differential equation modelling is intrinsically ill-suited for tackling such problems. The key factor that renders differential equation modelling untenable in these scenarios is that the quantities of interest are discrete, not continuous. For instance, populations are discrete quantities (i.e., it is not possible to have 2.5 individuals), and thus, the quantities of interest do not adhere to continuous functions, leading to the breakdown of differential equation models. However, when the quantities of interest are sufficiently large (e.g., in a tumour with approximately 10^9 cells), approximating the population as a continuous function becomes more reasonable. In essence, a differential equation models the mean of the process, which is typically adequate for continuous or large enough discrete quantities, though it may become problematic for small discrete quantities.

Differential equations are broadly classified into two main categories: Ordinary Differential Equations (ODE) and Partial Differential Equations (PDE). ODEs model quantities as functions of a single independent variable which, in the context of medical modelling, is typically time. Throughout history, many ODE models have played crucial roles in the field of medicine. A prominent class of ODE models that has garnered significant attention in recent years, particularly due to the COVID-19 pandemic, is the Susceptible-Infected-Recovered (SIR) models, which are utilized to simulate the spread of infectious diseases through populations. Let $S(t)$, $I(t)$, and $R(t)$ represent the number of susceptible, infected, and recovered individuals, respectively, as functions of time. The most basic form of an

SIR model can be represented by the following system of equations:

$$\frac{dS}{dt} = -k_1 S \tag{2.2}$$

$$\frac{dI}{dt} = k_1 S - k_2 I \tag{2.3}$$

$$\frac{dR}{dt} = k_2 I \tag{2.4}$$

where k_1 and k_2 are parameters that depend on the characteristics of the virus and the population under study. When solving such a model, there is typically assumed to be a single infected individual in an otherwise susceptible population, and hence, the initial conditions of $S(0) = N - 1$, $I(0) = 1$, $R(0) = 0$ are selected for the total population N. The solution to this system of equations yields three time-dependent functions describing the sizes of the three compartments within the total population, with an example solution depicted in Figure 2.1. Models like this were instrumental in projecting the spread of COVID-19 and played a key role in informing public health strategies. Moreover, the SIR model is highly adaptable, with various studies incorporating additional factors such as vaccination, re-susceptibility, and preventive measures like social distancing. Notably, I contributed to a study early in the pandemic that employed an extended SIR model to investigate these very elements [29].

A cancer-specific application of ODEs is in modelling the proliferation of cancer cells. Two of the most widely used models for simulating cancer cell proliferation are the exponential growth model and the logistic growth model, represented by Equations (2.5) and (2.6), respectively:

$$\frac{dN}{dt} = rN \qquad \text{(Exponential Growth)} \tag{2.5}$$

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \qquad \text{(Logistic Growth)} \tag{2.6}$$

In these equations, r is the proliferation rate, K is the carrying capacity, and $N(t)$ is the population of cancer cells at time t. These models allow for the projection of cancer cell populations over time. Note the key difference between these models: the exponential growth model is characterized by unbounded growth - with a doubling time of $t_{1/2} = \ln(2)/r$ - whereas the logistic growth model features bounded growth - with an upper bound defined by the carrying capacity K. Unlike the SIR model, both of these models are easily solved by hand, as they are simple separable equations: a solution is given for initial condition

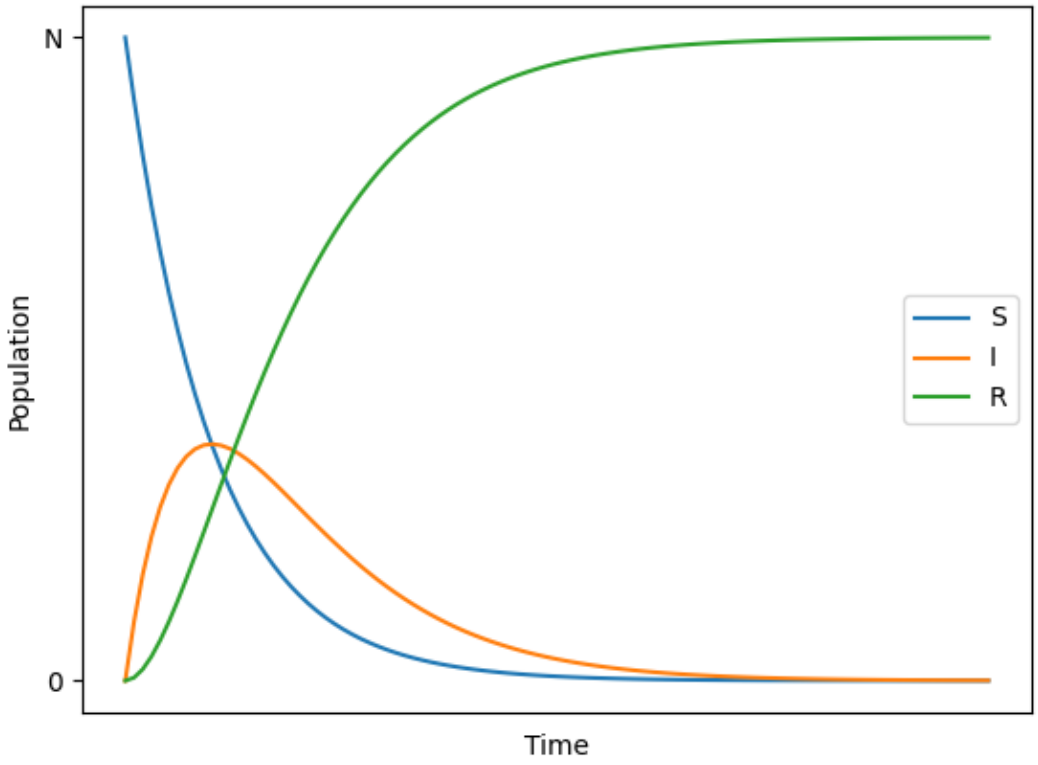

Figure 2.1: An example solution of the basic SIR ODE model using equations (2.2)-(2.4).

$N(0) = N_0$ in equations (2.7) and (2.8) below.

$$N(t) = N_0 e^{rt} \qquad \text{(Exponential Growth)} \tag{2.7}$$

$$N(t) = \frac{N_0 e^{rt}}{1 - \frac{N_0}{K}(1 - e^{rt})} \qquad \text{(Logistic Growth)} \tag{2.8}$$

These growth models are used extensively in chapters 3 and 5.

In contrast to ODEs, PDEs are differential equations involving functions of multiple variables. PDEs are often used to model quantities that depend on both time and space, with up to three spatial dimensions considered. We can derive a PDE from the logistic growth model given in equation (2.6) by introducing a term of the form $D\nabla^2 N$ (where ∇^2 denotes the Laplacian operator), resulting in the reaction-diffusion equation or *Fisher equation*, as follows:

$$\frac{\partial N}{\partial t} = D\nabla^2 N + rN\left(1 - \frac{N}{K}\right) \tag{2.9}$$

In this equation, $N(\vec{x}, t)$ represents the cell density at position $\vec{x}$ and time t, with $\vec{x} = (x, y, z)$ if the equation is in three dimensions. This equation was originally applied to model the spatiotemporal progression of brain tumours by James D. Murray [82, 83] and when used in this context, is referred to as the Proliferation-Invasion (PI) model. Most ODE models are relatively straightforward to solve, either analytically or computationally; PDE models however, often present significant challenges in solving. While analytical techniques can be useful, they are limited to cases with sufficiently simple equations, domains, and boundary conditions (namely, those with symmetry). In most cases, especially those encountered in practical medical applications, PDE models can be solved only through numerical, computational techniques.

2.2.2 Finite Element Methods and FEniCS Project

There are several different methods for numerically solving PDEs. The most sophisticated of these is a powerful approach called the Finite Element Method (FEM). The fundamental idea behind FEM is to discretize the spatial domain of interest into smaller, non-overlapping subdomains called 'elements.' These elements are typically geometrically simple shapes, such as intervals (1D), triangles (2D), or tetrahedra (3D) that collectively form a mesh over the entire spatial domain. The solution to the PDE is approximated within each element using basis functions, which are simple functions defined locally on each element.

Typically, the basis functions are chosen to be constants, linear functions, or quadratic functions. The main hallmark of FEM is the ability to solve PDEs on complicated meshes that other solution methods would have tremendous difficulty handling.

FEM is based on the variational formulation (or weak formulation) of PDEs. This formulation represents a generalization of the original PDE, which might not be as intuitive from an analytical perspective, but it facilitates numerical computations. The weak formulation extends the solution space to include functions that may not necessarily satisfy the original PDE pointwise, thereby enhancing its computational tractability. As an illustrative example, let us convert the Poisson equation into its variational form. In differential form, the Poisson equation is given by

$$-\nabla^2 u(\vec{x}) = f(\vec{x}) \quad , \quad \vec{x} \in \Omega \tag{2.10}$$

where Ω is the spatial domain over which we solve the equation, and $f(\vec{x})$ is a known function. For generality, let's consider the case that the boundary conditions involve both a Dirichlet and Neumann component; that is:

$$u = 0 \quad , \quad \vec{x} \in \Gamma \qquad \text{(Dirichlet)} \tag{2.11}$$

$$\frac{\partial u}{\partial n} = 0 \quad , \quad \vec{x} \in \partial\Omega - \Gamma \qquad \text{(Neumann)} \tag{2.12}$$

where $\Gamma \in \partial\Omega$ and the partial derivative with respect to n denotes the rate in the change in the direction of an outward normal vector to the domain boundary. We now define a function, $v(\vec{x})$ which is in the function space $V = \{v \in H^1(\Omega), v|_\Gamma = 0\}$. In this definition, $H^1(\Omega)$ is the Sobolev Space on Ω defined by $H^1(\Omega) = \{v \in L^2(\Omega), \nabla v \in L^2(\Omega)^d\}$ and $L^2(\Omega)$ is the space of all square-integrable functions on Ω. We refer to the function $v(\vec{x}) \in V$ as the test function.

To derive the variational form of the Poison equation, we multiply the differential form (equation (2.10)) by the test function, $v(\vec{x})$:

$$-v\nabla^2 u(\vec{x}) = v f(\vec{x}). \tag{2.13}$$

Next, the PDE is integrated over the domain of interest, Ω:

$$-\int_\Omega v\nabla^2 u(\vec{x})\ d\vec{x} = \int_\Omega v f(\vec{x})\ d\vec{x} \tag{2.14}$$

and integration by parts is performed on the 2nd derivative term as follows:

$$-\int_\Omega v\nabla^2 u(\vec{x})\ d\vec{x} = \int_\Omega \nabla u \cdot \nabla v\ d\vec{x} - \oint_{\partial\Omega} v\frac{\partial u}{\partial n}\ ds. \tag{2.15}$$

Now notice that the 2nd term of the integration by parts expansion is equal to zero since on the boundary, $v(\vec{x}) = 0$ on Γ or $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega - \Gamma$. Hence, we can then write the full variational form of the Poisson equation as

$$\int_\Omega \nabla u \cdot \nabla v \; d\vec{x} = \int_\Omega v f(\vec{x}) \; d\vec{x} \tag{2.16}$$

and the formulation of the problem is to find the function $u(\vec{x}) \in V$ such that the integral equation (2.16) is satisfied $\forall \; v \in V$. This variational form ensures that the solution satisfies the PDE in an average or integrated sense over the entire domain, rather than pointwise.

For many PDEs, including the ones focused on in this thesis, the dependent variables involve both space and time. Take, for example, the PI model explained above. Since our final variational formulation (2.16) involves integration over the spatial domain of interest, how is the temporal variable incorporated into this setup? The trick is to discretize our temporal domain into subintervals, then to solve the purely spatial variational formulation at each time step. For simplicity, consider the backward Euler time discretization of the PI model:

$$\frac{N_i - N_{i-1}}{\Delta t} = D\nabla^2 N_i + rN_i\left(1 - \frac{N_i}{K}\right) \tag{2.17}$$

where N_i represents the solution at the current time step and N_{i-1} represents the solution at the previous time step. Rearranging, and following the same procedure as above to derive the variational form, we arrive at

$$0 = \int_\Omega vN_i - vN_{i-1} + (\Delta t)D\nabla u \cdot \nabla v - (\Delta t)rvN_i\left(1 - \frac{N_i}{K}\right) \; d\vec{x} \tag{2.18}$$

This variational form can be solved iteratively, using the solution at the previous time step to inform the subsequent step. And furthermore, this same procedure can be performed for any chosen time discretization scheme. For the FEM applications presented in this thesis, all time discretizations are either backward Euler or Crank-Nicolson.

The basis functions used to approximate the solution are expressed as linear combinations of nodal values at specific points, called "nodes," within each element. The objective is to determine the values of these nodal coefficients that lead to the best approximation of the true solution, subject to boundary conditions. This results in a system of linear algebraic equations, which can be assembled and solved for the entire mesh to obtain the numerical solution. This process can be summarized by the following steps:

1. Discretize the domain into finite elements and create a mesh

2. Choose basis functions and express the approximate solution in terms of these basis functions
3. Formulate the weak form of the PDE by multiplying by a test function and integrating over the domain
4. Assemble the system of algebraic equations based on the weak form
5. Solve the system of equations to obtain the nodal coefficients of the solution
6. Use the nodal coefficients to reconstruct the approximate solution over the entire domain

This process, particularly the meshing of the domain into elements, can be quite complicated, especially for nonstandard domains. For this reason, software packages which automate this process have been developed which streamline and simplify the setup of FEM. One such package is the FEniCS Project, an open-source software platform for the numerical solution of PDEs using the FEM. With FEniCS, complicated meshes can be generated by combining simple geometries and complicated systems of equations can be implemented by simply converting them to their variational forms and inputting the results. In figure 2.2, a comparison of the 1D PI model solved using a finite difference scheme and with a FEM implemented in FEniCS is shown. Notice the simplicity in the FEniCS code. More details on the FEniCS project and FEM in general can be found in [4, 63, 104]

2.3 Machine Learning

Differential equations are considered mechanistic models in that they are based on specific mechanisms founded on fundamental laws of physics or biology and produce deterministic outcomes. Each term in the equations can be associated with a specific process or cause; for instance, the terms of the PI model (equation (3.1)) correspond to the processes of cellular diffusion and proliferation. However, in many situations we wish to examine, the underlying processes may not be clearly defined or may not lend themselves to closed-form mathematical representation. Or furthermore, the effect or form of these processes may be the very thing we are trying to discover (as is the case in emerging techniques such as symbolic regression [50]). In such cases, mechanistic modelling may be insufficient and alternative methods are necessary.

Statistical or data-driven models, on the other hand, are derived not from first principles, but rather from observed data. They aim to identify patterns, correlations, and

```python
#####################################################################
### Solution of Fisher Equation in 1D Manually
#####################################################################
import numpy as np
import matplotlib.pyplot as plt

# Model Parameters
D = 1.0  # Diffusivity
r = 1.0  # Proliferation rate
K = 1.0  # Carrying capacity

# Spatial domain parameters
L = 10.0    # Length of the spatial domain
n_x = 100   # Number of spatial grid points
dx = L/n_x  # Spatial step size

# Time-stepping parameters
T = 1.0     # Total simulation time
n_t = 1000  # Number of time steps
dt = T/n_t  # Time step size

# Create spatial grid points
x = np.linspace(-L/2, L/2, n_x)

# Define initial condition
u_n = np.exp(-x**2)

# Define time-stepping loop
for n in range(n_t):
    # Compute second spatial derivative using central differences
    laplacian_u = np.zeros((u_n.shape))
    laplacian_u[1:-1] = (u_n[2:] - 2*u_n[1:-1] + u_n[:-2]) / dx**2
    reaction = r * u_n * (1 - u_n / K)
    # Update solution using forward Euler method
    u_n = u_n + dt*(D*laplacian_u + reaction)
```

```python
#####################################################
### Solution o Fisher Equation in 1D Using FEniCS
#####################################################
from dolfin import *

# Define Problem Parameters
D = 1.0   # Diffusivity
r = 1.0   # Proliferation Rate
K = 1.0   # Carrying Capacity

# Create Mesh
mesh = IntervalMesh(100, -10.0, 10.0)
V = FunctionSpace(mesh, 'CG', 1)
u = Function(V)
v = TestFunction(V)

# Define Computational Parameters
T = 1.0      # Total simulation time
n_t = 1000   # Number of Time Steps
dt = T/n_t   # Time Step Length

# Define Initial condition
u_0 = Expression('exp(-pow(x[0],2))', degree=1)
u_n = interpolate(u_0, V)

# Define the PDE problem
F = u*v*dx - u_n*v*dx + dt*D*dot(grad(v),grad(u))*dx - dt*r*v*u*(1-u/n_max)*dx

# Define time-stepping loop
t = 0 # set initial time
for n in range(num_steps):
    t = t + dt # update Time
    solve(F == 0, u) # solve at time step
    u_n.assign(u) # update previous solution
```

Figure 2.2: A comparison of an implementation of the Fisher equation (equation (3.1) in one spatial dimension, solved with a manually coded finite difference method (left) vs. a finite element method implemented using FEniCS (right).

relationships within data without necessarily specifying fundamental dynamics. The parameters and forms of statistical models may similarly lack physical or biological interpretations. The primary focus of these models is simply to achieve high predictive accuracy irrespective of mechanisms of action, making them particularly useful in scenarios where these underlying mechanisms are complex or poorly understood. This category of models is commonly referred to as machine learning.

The are two general categories of machine learning: classification and regression. The key distinction between them lies in whether the output is discrete or continuous. In classification problems, inputs are transformed into discrete outputs; in regression problems, inputs are transformed into continuous outputs. The primary focus of this thesis is on regression problems, particularly function inference problems, in which data is used to tune a statistical model that approximates a given function. Linear regression is a simple example of a statistical regression model, wherein known data informs the slope and y-intercept of a linear equation. The resulting equation can be used for extrapolation, predicting an outcome y for a new input x.

The main goal of statistical regression models is to generate a numerical mapping between inputs and outputs based on a set of data, which can then be generalized and applied to future data on which the model was not trained. A key challenge therefore is to ensure that the model resulting from our training process truly uncovers the underlying relationships such that it is capable of extrapolation, and is not simply applicable exclusively to the input-output pairs on which it was trained. For instance, in theory, if we were to validate a statistical model on the same set of data on which that model was trained, then we should achieve 100% accuracy in our predictions, provided that the statistical model itself is capable of mimicking the complexity of the underlying function (i.e. a linear function will never achieve 100% accuracy on quadratically-distributed data). In such situations, the model is essentially just memorizing input-output pairs, rather than learning the true relationship between inputs and outputs more generally. Of course, this is obviously an undesirable quality of a statistical model, though it is not an uncommon one. In these cases where a model does not generalize well to new data, the model is considered to be 'overfit.' Many strategies have been employed to address this issue, the most common being to split the available data into training and testing sets. The machine learning model is then trained on the training sets, and evaluated on the testing set. In theory, this allows for an estimation of the true accuracy of the model, since it is evaluated on data that it has not yet seen. In many cases, the data allotted for training is further split into subcategories of training data and validation data. Then, the accuracy of the model is continuously evaluated on the validation set throughout training to identify when the model is beginning to overfit, and end training early. Validation datasets are also useful

for tuning model hyperparameters. Techniques such as randomizing input order, ensemble learning, cross-validation, and regularization are also employed to enhance generalization and accuracy when applying the model to new data.

Bias-Variance Trade-off

Another fundamental concept in the design of machine learning models is the bias-variance tradeoff, which describes the tradeoff between two sources of error that contribute to the overall prediction error of a model: bias and variance. Bias refers to the systematic error introduced by approximating a real-world problem with a simplified model. A model with high bias makes assumptions that do not accurately capture the underlying data distribution, leading to systematic underfitting and a reduced ability to generalize. Variance, on the other hand, measures the sensitivity of model predictions to changes in the training data. A model with high variance is overly sensitive to noise in the training data and tends to overfit, resulting in high variability in predictions for new data.

A crucial goal in designing a machine learning model then, is to achieve a balance between bias and variance to minimize the total prediction error. A model that is too simple may exhibit high bias and low variance, producing consistent but inaccurate predictions. Consider my dog (a yellow labrador named Owen) for example, whose favourite trick is to spin around in a circle at the input command of 'twirl.' However, as the canine brain is arguably a heavily biased model relative to what humans would like them achieve, often, his response to any of my input commands is simply to twirl around - especially frustrating when the intended command of 'leave it' results in him twirling around with my shoe in his mouth. Adorable, yet incorrect. Conversely, a model that is too complex may exhibit low bias and high variance, producing accurate predictions on the training data but performing poorly on unseen data. The bias-variance tradeoff reflects the inherent tension between the need for flexibility to capture complex patterns (low bias) and the need for stability to avoid overfitting (low variance).

Model selection and hyperparameter tuning are crucial processes in managing the bias-variance tradeoff. Techniques such as regularization, cross-validation, and ensemble learning can also help mitigate the tradeoff by reducing variance while preserving model expressiveness. Ultimately, finding the optimal balance between bias and variance is essential for building robust and accurate predictive models.

2.3.1 Artificial Neural Networks

In this thesis, the primary machine learning models employed are Artificial Neural Networks (ANN), a powerful technique inspired by the structure and function of the human brain. The human brain consists of neurons that accept a set of signals and output a combination signal. The first attempt to emulate this artificially was the perceptron, which accepted a vector of 0-1 signals and outputted a 0-1 signal from them. The perceptron included a set of weights, w_1, ..., w_n, and a bias, b, which transformed the input signals into an output signal using the following formula:

$$P(\vec{x}) = H\left(b + \sum_{i=1}^{N} w_i x_i\right) \tag{2.19}$$

where $H(v)$ is the Heaviside function defined by

$$H(v) = \begin{cases} 0, & v \leq 0 \\ 1, & v > 0 \end{cases} \tag{2.20}$$

In this context, the Heaviside function serves as what is called an activation function. Researchers began constructing networks of these perceptrons with layers that sequentially fed into each other. Eventually, the discontinuous output signals of the neural network were replaced with continuous outputs by substituting the Heaviside function, which produces a discontinuous signal, with various continuous functions. Several activation functions are popular in modern literature, including sigmoid functions such as $\varphi(v) = \left(\frac{1}{1+e^{-v}}\right)$, the hyperbolic tangent function $\varphi(v) = \tanh(v)$ and more recently, the rectified linear unit (ReLU) function given by

$$\varphi(v) = \begin{cases} 0, & v \leq 0 \\ v, & v > 0 \end{cases} \tag{2.21}$$

With this generalization from discontinuous to continuous outputs, the concept of a perceptron was generalized to the neuron. The output of a neuron, given an input signal $\vec{x}$, is then computed by

$$y(\vec{x}) = \varphi\left(b + \sum_{i=1}^{N} w_i x_i\right) \tag{2.22}$$

where $\varphi(v)$ is the chosen activation function and N is the total number of input signals feeding into the neuron. See figure 2.3 for a comparison of common activation functions.

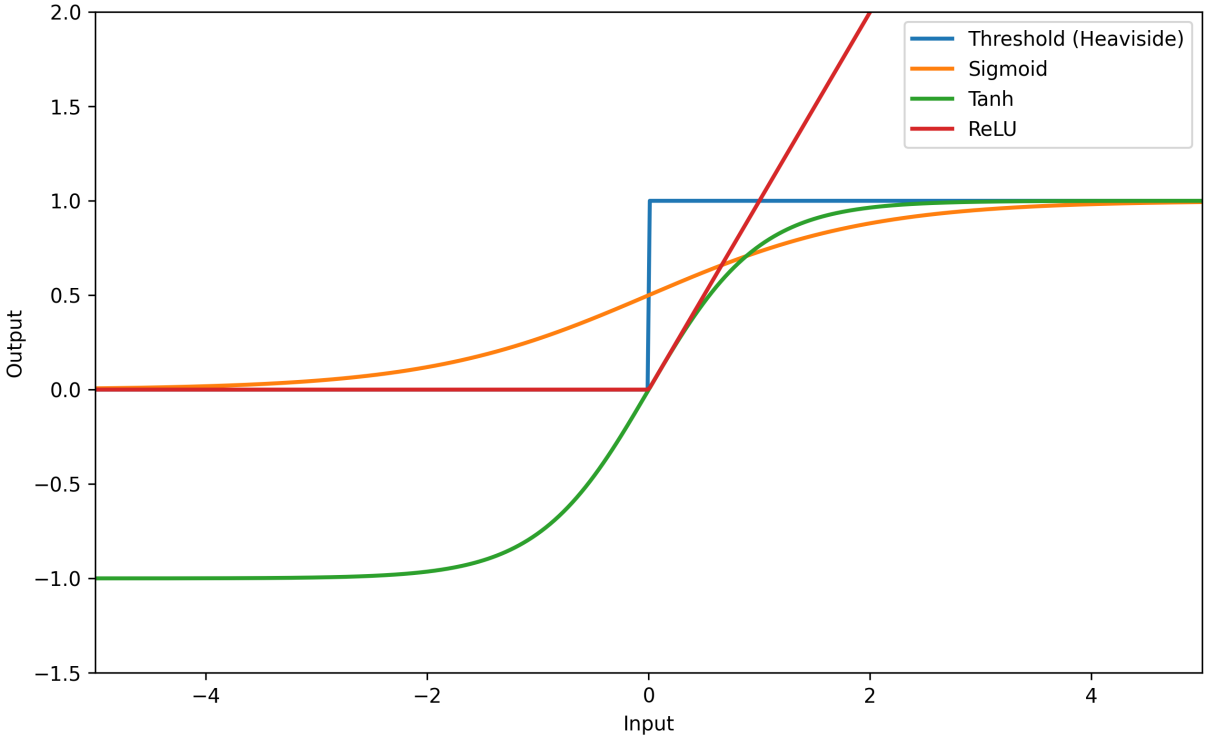

Figure 2.3: A comparison of common activation functions used in ANNs.

In typical modern applications, neural networks comprise not just a single neuron, but numerous layers of neurons with each layer containing multiple neurons. For these multilayered ANNs, the first layer is called the input layer and the final layer is called the output layer. The layers between the input and output layers are typically referred to as the hidden layers since their actions and outputs are not typically observed.

Given an input-output pair, denoted $(\vec{x}, \vec{y})$, a neural network accepts the input $\vec{x}$ and computes a prediction for the output, $\hat{y}$. When training a neural network, the goal is to minimize the error between the known data, $\vec{y}$, and the prediction of the network, $\hat{y}$. In order to calculate these network predictions, a process called forward propagation is used. We can write the output of layer l of the network as

$$\vec{y}^{(l)} = \varphi\left(\vec{v}^{(l)}\right) \tag{2.23}$$

where

$$\vec{v}^{(l)} = b^{(l)} + W^{(l)}\vec{y}^{(l-1)} \tag{2.24}$$

In this formula, φ is the chosen activation function, which can vary between layers in general, but for simplicity here, we will assume that the same activation function is used throughout the network. $W^{(l)}$ is the matrix of weights connecting layer l to layer $l+1$, where the ij entry corresponds to the weight connecting the ith neuron in layer l to the jth neuron in layer $l+1$. Similarly, $b^{(l)}$ is the added bias in layer l. Furthermore, the zeroth layer is simply the input to the network, $y^{(0)} = \vec{x}$ and the output of the final layer of the network (with a depth L) is simply the output of the network, $\hat{y} = \vec{y}^{(L)}$.

Using this output signal of the full network, the error can be calculated between the predicted and true outputs using a loss function, which we denote as $e(\hat{y})$. Different choices can be made for this loss function, but the most common one for regression problems is the Mean-Squared Error (MSE) loss function, given by

$$e = \frac{1}{I}\sum_{i}(y_i - \hat{y}i)^2 \tag{2.25}$$

where the sum is computed over the width of the output layer, and I is the length of the output vector. This loss can be used to update the weights and biases of the network through the process of backpropagation. Specifically, let us denote the current values of the weights and biases at layer l as $W_n^{(l)}$ and $b_n^{(l)}$, and the updated values of these after backpropagation by $W_{n+1}^{(l)}$ and $b_{n+1}^{(l)}$. Then

$$W_{n+1}^{(l)} = W_n^{(l)} - \eta\frac{\partial e}{\partial W_n^{(l)}} \tag{2.26}$$

$$b_{n+1}^{(l)} = b_n^{(l)} - \eta\frac{\partial e}{\partial b_n^{(l)}} \tag{2.27}$$

where η is a hyperparameter called the learning rate which controls the size of the gradient descent steps performed. The difficulty with this updating formula of course, is in efficient computation of the gradients. To do this, the chain rule is applied to compute the gradients recursively, beginning with the output layer and ending with the input layer. Let $\delta^{(l)}$ denote the local gradient at layer l. For the output layer (layer L), we can directly compute this using the partial derivatives of the loss function with respect to the output of the neurons:

$$\delta^{(L)} = \frac{\partial e}{\partial \vec{y}^{(L)}} \odot \varphi'\left(\vec{v}^{(L)}\right) \tag{2.28}$$

where $\odot$ denotes element-wise multiplication, and φ' is the derivative of the activation function. For the hidden layers, the error terms can be computed recursively using the error terms of the subsequent layer:

$$\delta^{(l)} = \left(W^{(l)}\right)^T \delta^{(l+1)} \odot \varphi'\left(\vec{v}^{(l)}\right) \tag{2.29}$$

With these local gradients for each layer, the full gradient of the loss function with respect to the weights and biases can be computed as follows:

$$\frac{\partial e}{\partial W_n^{(l)}} = \delta^{(l)} \cdot \left(\vec{y}^{(l-1)}\right)^T \tag{2.30}$$

$$\frac{\partial e}{\partial b_n^{(l)}} = \delta^{(l)} \tag{2.31}$$

Once the error terms for all layers have been computed, the gradients can be used to update the weights and biases of the network according to equations (2.26)-(2.27). Importantly, within this training procedure, it is most useful to think of the loss function as a function not of the network outputs, but rather as a function of the weights and biases. The process of forward propagation and backpropagation is iteratively performed for each input-output pair in the training dataset, typically for multiple epochs until a predefined stopping criterion is satisfied.

The optimization procedure described above is known as gradient descent, which is a common optimizer used in regression problems. In practice, a variant of gradient descent called Stochastic Gradient Descent (SGD) is more frequently employed. In SGD, rather than training using the entire dataset at each iteration, a random subset of the data, called a mini-batch, is used for training. This variation of the gradient descent method often leads to faster convergence and better generalization due to the inherent noise introduced by selecting random subsets of the data. Another variant of gradient descent is Adam (Adaptive Moment Estimation), which is used in this thesis. Adam optimization is similar to SGD but differs in that it does not use a single global learning rate; instead, it maintains separate learning rates for each parameter and adaptively adjusts them throughout training based on the first and second moments of the gradients. This adaptive approach typically results in faster convergence than standard SGD.

Another optimization algorithm used in this thesis is the Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimizer. L-BFGS is a quasi-Newton method that approximates the second-order Hessian matrix using a limited amount of computer memory. Unlike gradient-based optimizers such as SGD and Adam, L-BFGS relies on curvature information to find the optimal solution more efficiently. It maintains a low-rank approximation of the inverse Hessian matrix, which is updated iteratively using a history of past gradients and parameter updates. This low-memory approach allows L-BFGS to be suitable for large-scale optimization problems where the exact computation or storage of the Hessian matrix is computationally prohibitive. L-BFGS often exhibits faster convergence and better performance on certain types of problems, particularly those with smooth and continuous objective functions, compared to first-order gradient-based methods.

Regularization of Artificial Neural Networks

A crucial concept in machine learning is regularization, which is used to prevent overfitting and improve the generalization of models to unseen data. For ANNs, the standard approach is to incorporate a term into the loss function which penalizes large weights, discouraging the model from relying too heavily on any single feature. In artificial neural networks, common regularization techniques include L1 and L2 regularization, which add the sum of absolute or squared weights, respectively, to the loss function. Another popular regularization method is dropout, which involves randomly dropping neurons from the network during training, forcing the model to learn redundant representations and rely on a more diverse set of features. These regularization strategies ultimately improve the robustness and predictive performance of the neural network on previously unseen data. One can also think of regularization as a method for increasing the bias of a model (as in the bias-variance trade-off). The primary ANN technique used in this thesis (physics-informed neural networks, explained below) is based upon a particular method of regularization which allows the model to be highly generalizable to unknown data.

2.3.2 Physics-Informed Neural Networks

Incorporating prior knowledge into the optimization process can significantly enhance the performance and robustness of machine learning models. Consider a scenario with N known data points, where the objective is to identify the underlying relationship between these points and extrapolate to new data. In the absence of any additional information, perfect accuracy could be achieved on the N training points by using an N-degree polynomial. However, in most cases, this choice will not represent the optimal model selection. For instance, when an additional data point is added to the training set (making it $N+1$ data points), an $(N+1)$-degree polynomial is unlikely to become the new optimal model. Fortunately, in many scientific applications, we possess more information than just the raw data points, and this information can be leveraged to improve model robustness. In this example with N data points, if it was also known that the underlying dynamics were linear, the model could be simplified, reducing the number of parameters to fit from N to 2. Imposing such system information into the model reduces its variance, thereby improving its extrapolation capabilities for future data.

Integrating known linear underlying dynamics into a model is relatively straightforward. However, some problems may involve prior knowledge that is less easily incorporated. For instance, in cancer modelling, system information is often captured in the form of a PDE or perhaps even a system of PDEs. The PI model (equation (3.1)) is a classic

example of this. Models such as this are considerably more complex, posing challenges when integrating them into machine learning models. Moreover, the data itself might be difficult to align with the desired outcome. In particular, much medical data comes from medical imaging, which is typically collected at discrete time points, while continuous-time predictions are often desired. PINNs provide a method for incorporating complex prior knowledge, specifically PDEs, into a machine learning model, resulting in more robust and accurate representations of the underlying system dynamics. PINNs were first introduced in a series of works and patent by Raissi et al [96, 97, 98].

An essential advantage of incorporating existing knowledge into optimizations is the increased interpretability of models that are typically considered black boxes. There have been numerous calls for enhancing the interpretability of machine learning models, and integrating prior knowledge in this way is a promising approach to achieving this goal. Additionally, the incorporation of a governing mathematical model into the optimization of the machine learning model allows for it to be trained using significantly less data. This is particularly relevant in the context of medicine, where data scarcity is a common issue. Furthermore, for applications in personalized medicine, data is even more scarce since a sufficient amount of patient-specific data is required to draw relevant conclusions.

Given that standard neural network models are typically trained solely using data, one might ask whether, given a sufficient amount of data, a data-only neural network approach could yield comparable performance to PINNs. However, the amount of data necessary to achieve comparable results is extremely large, and often disqualifying in the context of medical applications to do so. Similarly, training solely using the PDE is also not a viable option since without specifying initial and boundary conditions (which are often unknown or extremely complex is this context), the PDE cannot be directly solved. It is only through hybrid methods like PINNs, which combine both data and the governing model, that accurate predictions can be achieved using limited data. In continuous PINNs, the importance of the data vs. the PDE model can be separated and individually weighted during training, though in discrete PINNs, this is not possible.

The goal of a PINN is to train a deep ANN to mimic a function with potentially multidimensional input and output from a small number of data points. This may initially seem naive and infeasible, since in situations where the amount of data is small enough, there may not even be guaranteed convergence of the training algorithm. However, with the incorporation of additional information in the form of a PDE model, it is possible to achieve this objective.

Consider that N measured input-output pairs from a target function $n(\vec{x}, t)$ are randomly selected from the input domain. We can train a neural network to approximate

$n(\vec{x}, t)$ by minimizing the MSE between the known outputs at the N points and the predictions of the network at those N points. We set up a neural network that accepts the inputs $\vec{x}$ and t and outputs the value of the solution function, $n(\vec{x}, t)$. Let us denote this network as the function

$$F : (\vec{x}, t) \rightarrow n \tag{2.32}$$

The error between the network predictions and the true known data points can then be quantified using

$$MSE_n = \frac{1}{N} \sum_{i=1}^{N} [F(\vec{x}_i, t_i) - n_i]^2. \tag{2.33}$$

The goal now is to include the governing PDE model in the deep learning optimization. Consider that the PDE is of the general form

$$0 = f(n; \vec{x}, t) \tag{2.34}$$

where $n(\vec{x}, t)$ is the desired solution of the problem, $\vec{x}$ and t are the spatial and temporal input variables respectively, and the function $f(\cdot, \cdot)$ is a function of the input variables and the solution function itself, which can involve nonlinear and differential operators. This setup can accommodate many common model types in physical and biological modelling, including conservation laws, diffusion processes, advection-reaction-diffusion equations, and kinetic equations. We can then evaluate how well our function approximation network satisfies the PDE using the formula

$$MSE_f = \frac{1}{N} \sum_{i=1}^{N} f\left(F(\vec{x}_i, t_i); \vec{x}_i, t_i\right)^2. \tag{2.35}$$

By adding both of these two MSE losses in a joint loss function, we create a loss that incorporates information from both the known data and the PDE:

$$MSE = MSE_n + MSE_f \tag{2.36}$$

If we were training only with the data points, the loss function would include only the first term. By adding the second term, which quantifies the satisfaction of the PDE, we regularize the deep learning model by constraining the solution space to those that sufficiently satisfy the PDE model.

Through several examples in the original papers outlining PINNs [96, 97], it is demonstrated that this formulation of the loss function results in the network converging more

accurately and more efficiently than using the data alone. This is expected, as altering the loss function in this manner incorporates more information from which the neural network can learn. We will showcase this by applying a PINN model to a simple example. Consider the problem of solving the heat equation in one spatial dimension with given initial condition and boundary conditions:

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T, \quad x \in [0,1],\ t \in [0,1] \tag{2.37}$$

$$T(x,0) = 0 \tag{2.38}$$

$$T(0,t) = 1,\ n(1,t) = 0 \tag{2.39}$$

This model setup could be used to model a metal bar which starts at temperature $T = 0$ and is heated on its left side with an element of temperature $T = 1$, for example. This equation can be solved using a finite difference numerical scheme. In solving, we divide the input domain into 200 spatial steps and 100 time steps, employing a backward Euler time discretization scheme. The thermal diffusivity parameter is set to a value of $\alpha = 0.5$. The solution can be observed at the top of Figure 2.4.

Using 200 data points (input-output pairs) randomly selected from the full solution of the heat equation problem above (which represents only 1% of the total data), we set up a PINN that accepts x and t as input, and outputs the value of the temperature T at position x and time t. The network structure consists of 8 hidden layers, each with 20 neurons. Activation functions are chosen to be tanh throughout the entire network except for the output layer, which simply outputs the raw value. Input data is normalized to the interval [-1, 1] before training, and output data is normalized to the interval [0, 1]. We add a custom loss function to the network, which implements equation (2.36). The network is trained using these data points with an Adam optimizer over 10,000 iterations. The ANN is implemented using the TensorFlow framework. An example result from this optimization can be seen in Figure 2.4, where the known and predicted solutions of the temperature are compared. The MSE between the exact and predicted full solution (at all inputs, not just the training points) is 7.91e-4. It is worth noting how remarkable this result is - extremely high accuracy can be achieved in a function approximation network while training on just 1% of the available data. Often, the cost or difficulties associated with acquiring enough data to train a deep learning model for many real-world purposes makes doing so impossible. But here, we show that by incorporating a governing PDE model into the optimization, convergence and high accuracy in the prediction of a solution can be achieved with what would normally be considered a prohibitively small amount of data.

It is also worth noting that the machine learning model was able to make these pre-

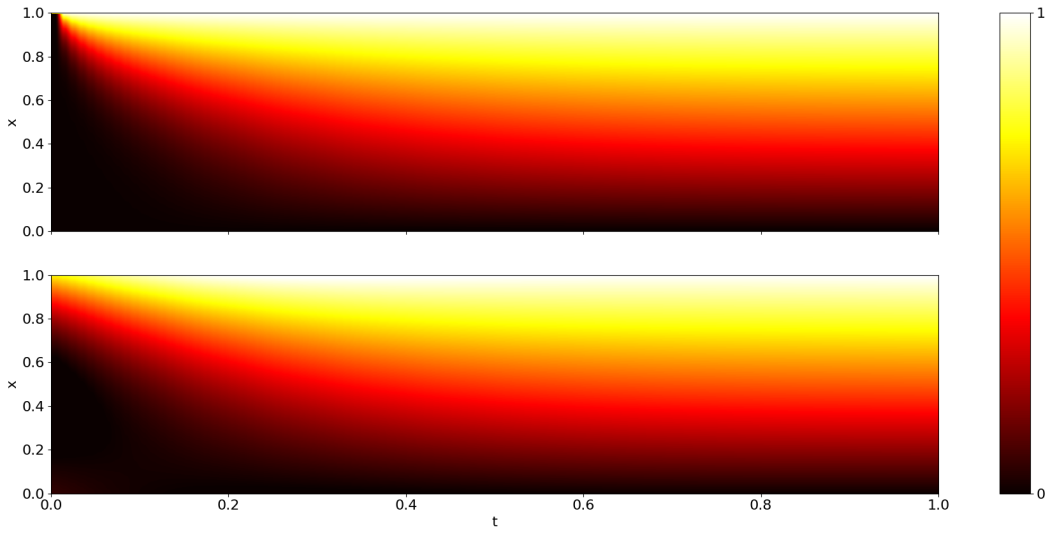

Figure 2.4: **Top**: The solution of the heat equation initial/boundary value problem in equations (2.37)-(2.39) solved using a finite difference method with backward Euler time-stepping on a $N_x = 200, N_t = 100$ grid. **Bottom**: the solution of the problem as predicted by the PINN model after training on just 200 input-output pairs from the exact solution.

dictions while being ignorant of the initial and boundary conditions of the problem. In theory, we could have specified these in our optimization by further adding terms to insist that our predicted solution adhered to these conditions. However, the network is capable of producing accurate estimates without these being specified. The reason why this is possible is that information from the initial and boundary conditions is contained within the data points themselves. In other words, the network infers the initial and boundary conditions from the data without them needing to be specified explicitly. This is an important benefit because in many applications in medicine, the initial and boundary conditions are unknown or too complicated to reasonably implement. Take the example of a growing tumour (which is the subject of Chapter 3). The initial condition of a growing tumour is presumably a single cancerous cell, but when and where that initial cancer cell emerged is impossible to know. Furthermore, boundary conditions on tumour cells as they proliferate will depend on the anatomy surrounding the tumour location, which can be extremely difficult to model. Hence, PINNs bypassing explicit specification of these conditions is a significant advantage.

In the example presented in Figure 2.4, 1% of the known data is used to train the network and achieve remarkably accurate solutions. However, it is interesting to explore the number of data points necessary for training in this context. To investigate this, we conduct the same experiment on the heat equation as described above, but with a varying number of known data points. This is accomplished by choosing a random number of data points up to 10% of the known data (2000 input-output pairs), training the network, comparing the full predicted solution to the full exact solution, and calculating the error. We perform this process for 100 different data set sizes and showcase the results in Figure 2.5. Though clearly randomness is apparent in these results, notice how the error in the network predictions decreases as more data is used to train the model. However, observe that relatively high accuracy is still attained even when small amounts of data are used for training. In the case with the least data, using just 21 points to train the model results in a MSE of 0.0009, which is remarkably accurate given the amount of data. A similar analysis was performed on a different PDE problem in [96, 97] which arrived at similar results.

PINNs are Robust to Measurement Noise

The original PINN papers also demonstrate that these networks are remarkably robust to noise in the system. We showcase this robustness in our example as well. After selecting our training data points (200 for each of the runs in this example), we introduce random

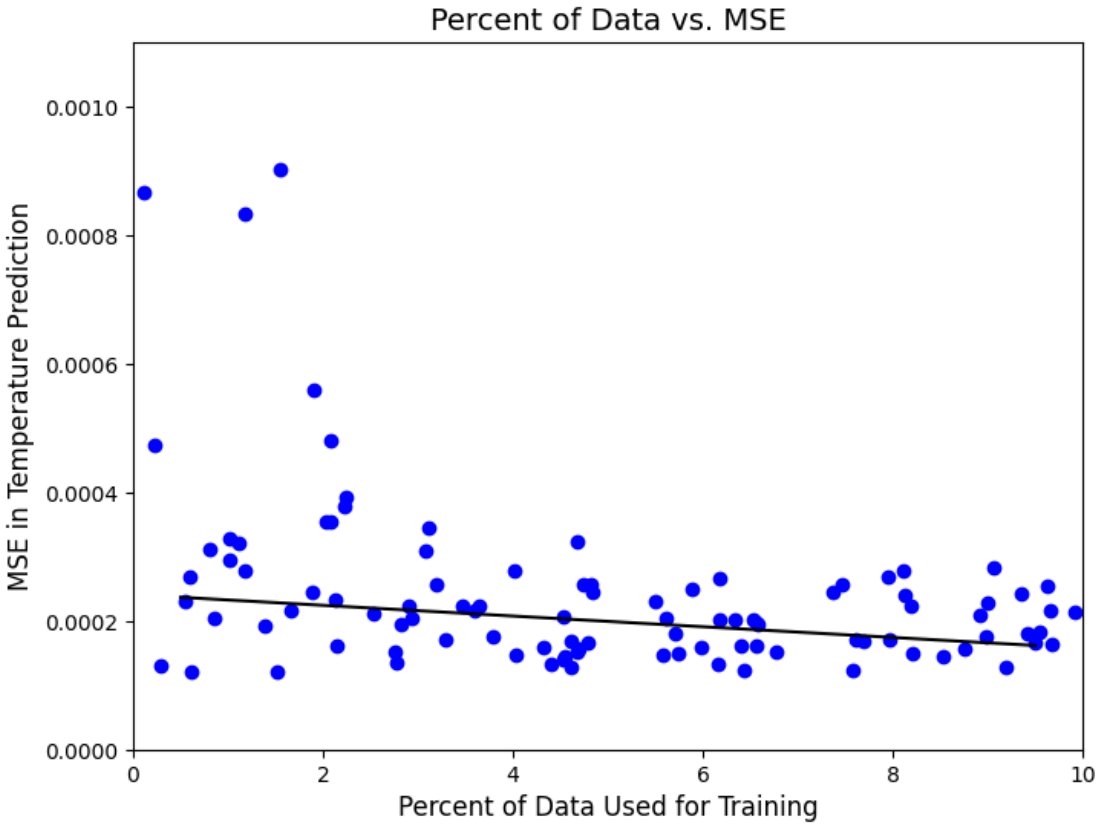

Figure 2.5: Error in network predictions as a function of the amount of training data provided with the line of best fit. The x-axis gives the percentage of the full spatiotemporal domain that is used for training, which had a total of 200,000 points (meaning 10% corresponds to 2000 input-output pairs). The y-axis gives the MSE between the predicted and known temperature over the entire spatiotemporal domain. Notice that relatively high accuracy is still achieved even when an extremely small amount of data is used for training.

noise into each data point measurement using the following formula:

$$y_{noisy} = y_{exact}(1 + \texttt{Noise} * (2\epsilon - 1)) \tag{2.40}$$

where `Noise` is the maximum possible noise that can be applied to each output, ϵ is a random number in the interval [0,1], and y_{exact} is the data output free from noise. In other words, each output is multiplied by a random number in the interval [1-`Noise`, 1+`Noise`]. This process allows us to observe how the network predictions vary with different degrees of noise added to the system. This is an important consideration, as real-world observations often come with associated measurement noise, especially in the case of medical data.

We examine the network performance in the presence of noise in the same way as we did above for varying amounts of known data. Specifically, we choose a series of values for `Noise`, run the PINN, and assess the error in the predictions. The results can be seen in Figure 2.6. As expected, the error rises with increased noise; however, it is again noteworthy that the increase is slight, and that high accuracy is still achieved even at high noise levels.

PINNs can Estimate PDE Model Parameters and Fields

In this thesis, we focus on applications to personalized medicine. Ideally, we would like to develop a PDE model that is specific to each patient. However, a key challenge in creating an equation for a particular patient is that the parameters appearing in that PDE model may depend on characteristics specific to that patient; in other words, the PDE model parameters may be patient-specific. In the context of training PINNs, since the loss function depends upon the PDE to calculate the loss based on the network predictions, if the parameters in the PDE model are unknown, then the loss cannot be calculated.

A key benefit of PINNs, however, is that unknown parameters do not prevent the application of the deep learning model. This is because the unknown parameters can be estimated during the process of network optimization. This is done in the same way as the network weights themselves are optimized. In the TensorFlow framework, the PDE model parameters are defined as trainable parameters, and their values are updated iteratively throughout training. Once the model has converged, the values of these parameters will have also converged. Examples of this are shown in the original papers outlining PINNs [96, 97] where the authors demonstrate that a parameter appearing in Burgers equation can be accurately estimated using this method.

We showcase this here by again applying the model to the problem outlined in equations (2.37)-(2.39), except we now assume that the value of the thermal diffusivity, α, is unknown.

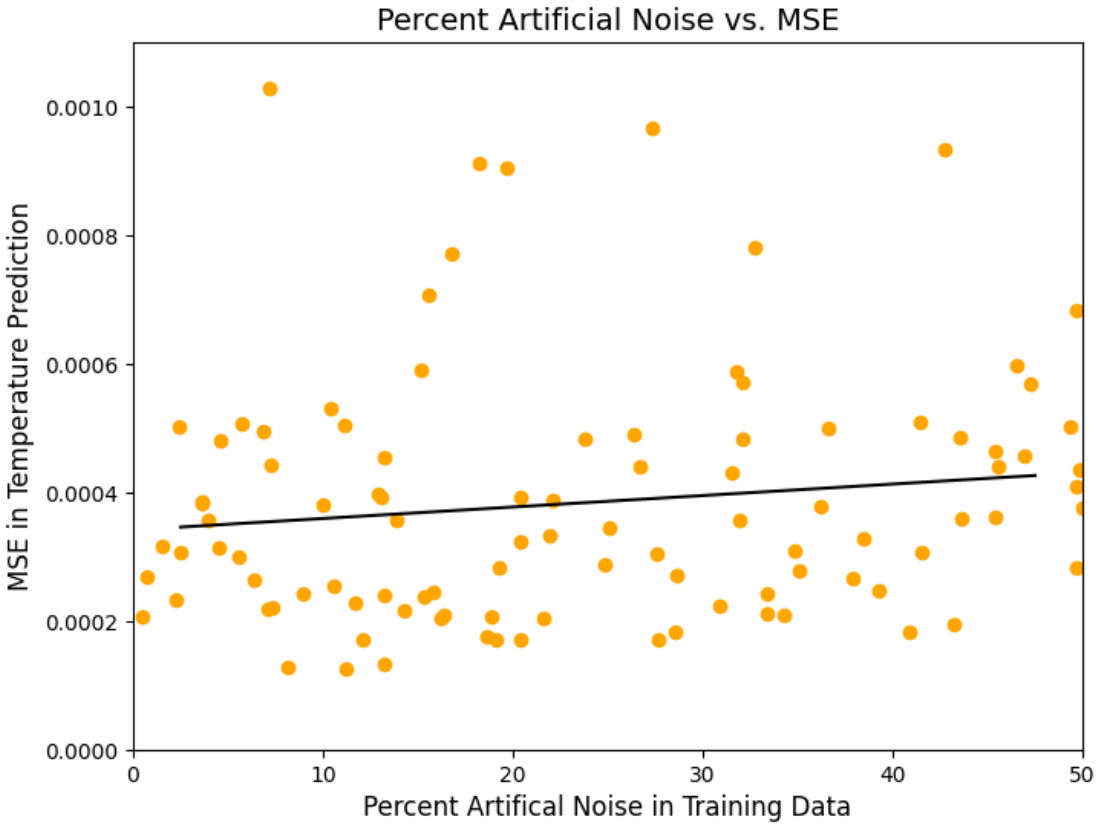

Figure 2.6: Error in network predictions as a function of the level of artificial noise added to the training data with line of best fit. The x-axis gives the percentage noise applied to the data in training, as calculated using equation (2.40). The y-axis gives the MSE between the predicted and known temperature over the entire spatiotemporal domain. Notice that relatively high accuracy is still achieved even when high error is added to the data.

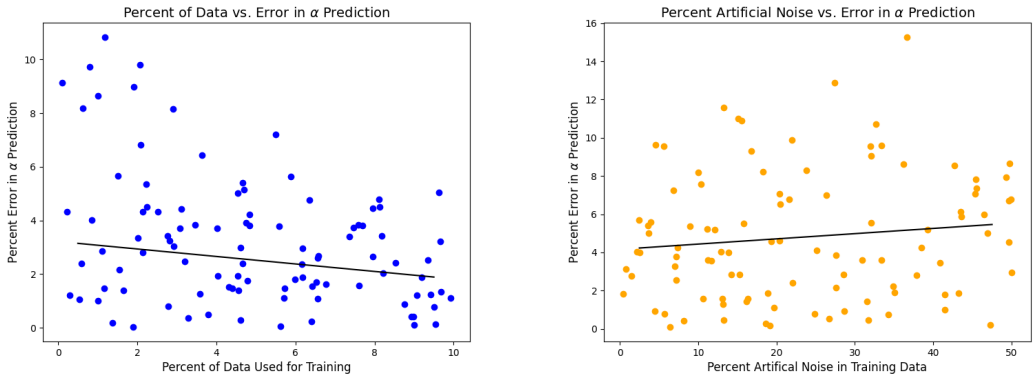

Figure 2.7: Error in the estimation of the thermal diffusivity parameter α for different amounts of known data (left) and different amounts of noise (right) with lines of best fit.

In Figure 2.7 below, we show the error in the estimation of α for different amounts of known data and different amounts of noise. Again observe that the error decreases with increased data and increases with increased noise, as expected, though the predictions remain remarkable robust throughout.

Oftentimes, however, PDE models involve more than just unknown parameters. They can also include entire unknown fields appearing in the equations. This is addressed in a slightly different way than it is for parameters. Each field appearing in the PDE model is made an output of the PINN, like the solution itself. Then throughout training, the relationship between the input parameters and these other fields is approximated by minimizing the loss involving the PDE. In the original papers outlining PINNs, the authors showcase this by applying PINNs to the two-dimensional Navier-Stokes equations. They demonstrate that the pressure appearing in those equations can be estimated to an extremely high degree of accuracy by applying a PINN to the problem. This approach can be applied to other scenarios as well, such as deriving intratumoral pressure after the administration of drug-carrying liposomes, which is the subject of Chapter 4.

PINNs with Discrete Data

An additional important capability of PINNs is that they can operate on both continuous data (occurring over a continuous time range) and discrete data (occurring only at initial and final times). Consider that rather than knowing N randomly selected input-output pairs from the entire domain, we instead know the full spatial solution at two separate

times, $n(\vec{x}, t_1)$ and $n(\vec{x}, t_2)$. This is of particular importance to this thesis since the primary data used originates from medical imaging, which falls into the discrete data category. When dealing with discrete data, the core idea of PINNs remains the same, though the math and network setup need to be augmented slightly. First, the PDE must be discretized using a Runge-Kutta time stepping algorithm with q stages (where q is chosen to be the number of intermediate points at which the solution is predicted). We can write this Runge-Kutta scheme as

$$k_i(\vec{x}) = n_1(\vec{x}) + \Delta t \sum_{j=1}^{q} a_{ij} g\left(k_j(\vec{x}), t_1 + c_j \Delta t\right) \tag{2.41}$$

$$n_2(\vec{x}) = n_1(\vec{x}) + \Delta t \sum_{j=1}^{q} b_j g\left(k_j(\vec{x}), t_1 + c_j \Delta t\right) \tag{2.42}$$

where k_i is the ith Runge-Kutta stage ($i \in [0, q]$), a_{ij}, b_i, and c_i are the Runge-Kutta scheme parameters (commonly expressed in a Butcher Tableau), $\Delta t = t_2 - t_1$ is the time between the initial and final times, and g is the right hand side of the PDE rewritten as $\frac{\partial n}{\partial t} = g(n; \vec{x}, t)$. Note that each $k_i(\vec{x})$ is a prediction of $n(\vec{x}, t)$ at an intermediate time between t_1 and t_2 corresponding to the Runge-Kutta parameter c_i. When typically using a Runge-Kutta scheme, an implicit matrix equation is solved for all q of the $k_i(\vec{x})$ profiles, which can then be used to calculate the $n_2(\vec{x})$ from $n_1(\vec{x})$. This process would be performed iteratively over many time steps to estimate the solution at later times. In this case, however, both $n_1(\vec{x})$ and $n_2(\vec{x})$ are known, but the function $g(\cdot, \cdot)$ cannot be computed since boundary conditions, parameters, and/or fields appearing in the right hand side may be unknown. Instead, this numerical scheme can be inverted to write the known initial and final solutions as

$$n_1(\vec{x}) = k_i - h \sum_{j=1}^{q} a_{ij} g(k_j, t_0 + h c_j) \quad , \quad n_2(\vec{x}) = k_i - h \sum_{j=1}^{q} (b_j - a_{ij}) g(k_j, t_0 + h c_j) \tag{2.43}$$

The PINN is now set up to receive $\vec{x}$ as input and output the Runge-Kutta stages $k_1, k_2, ..., k_q$. Evaluating the prediction of the initial and final data snapshots using these stages and the Runge-Kutta scheme, the loss function for the discrete case can be written as

$$MSE = MSE_0 + MSE_1 \tag{2.44}$$

where

$$MSE_0 = \frac{1}{N} \sum_{i=1}^{N} (n_1 - n_{1,pred})^2 \quad , \quad MSE_1 = \frac{1}{N} \sum_{i=1}^{N} (n_2 - n_{2,pred})^2. \tag{2.45}$$

In the above, N is the number of data points at each of the two times, n_1 and n_2 are the known solutions at the initial and final times, and $n_{1,pred}$ and $n_{2,pred}$ are the solutions at the initial and final times predicted by the neural network. Note that this loss function still incorporates information from both the known data and the PDE like in the continuous case; however, these two sources of information can no longer be separated term-by-term as they were before.

Similar experiments can be performed in this discrete case as in the continuous case, showcasing its insensitivity to noise (and some are performed in [96, 97]). Unknown parameters and fields in the PDE can also be incorporated in the same way, either defining unknown PDE model parameters as trainable parameters or adding the unknown fields as an additional outputs of the network.

Chapter 3

Deep Learning Characterization of Brain Tumours with Diffusion Weighted Imaging

This chapter contains the first project of this thesis which was published in the *Journal of Theoretical Biology* in 2022 [73]. The data used in the study was obtained from Unity Health Toronto, and ethics approval was obtained for the study from Unity Health Toronto and the University of Waterloo. The author list on the publication is below.

Cameron Meaney[1], Sunit Das[2,3], Errol Colak[3,4], Mohammad Kohandel[1]

[1]Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
[2]Division of Neurosurgery, St. Michael's Hospital, Toronto, Canada
[3]Faculty of Medicine, University of Toronto, Toronto, Canada
[4]Department of Medical Imaging, St. Michael's Hospital, Toronto, Canada

3.1 Introduction

The most prevalent form of malignant brain tumour among adults is Glioblastoma Multiforme (GBM). Modern treatments for GBMs involve a combination of surgery, chemotherapy, and radiotherapy, yet despite the most aggressive therapies, mean survival after diagnosis is only 15 months [6, 56, 113]. GBMs are commonly distinguished from other brain tumours by their dense necrotic core and surrounding shell of peritumoral edema, both of

which pose challenges for effective therapy [87, 132]. In spite of continuous advancements in our understanding of the biology of GBMs and in the imaging techniques used to observe them, their full extent is often inaccurately assessed by state of the art imaging: many studies have noted that their growth and invasion tend to be underestimated [65, 87, 132]. As treatments are typically designed based on this knowledge, there is potential for therapies to be misapplied, possibly contributing to recurrence or treatment resistance. Methods and technologies for more accurately assessing GBM growth and invasion are sought by clinicians and researchers, with the hope of using this knowledge to make more accurate growth predictions and design more effective treatments, especially in a patient-specific manner.

Much of the information available when examining GBMs comes from medical imaging, specifically multi sequence MRI [65, 126]. Qualitative conclusions can be drawn from MRIs which can be used by clinicians to inform recommendations for observation and treatment. The explosion of big data in recent years however, has also allowed for analysis of medical imaging data quantitatively, which has proven tremendously fruitful. Unfortunately, the barrier to entry in quantitative analysis of medical imaging is high, requiring substantial mathematical and computational tools. Accordingly, mathematical modelling and machine learning have become crucial tools with the capability to make quantitative analysis of imaging data accessible enough for clinical use.

Many works have used mathematical models to predict the extent, future growth, and response to treatment of GBMs [3, 35, 41, 43, 44, 55, 60, 72, 76, 82, 83, 119]. In particular, the most common model used to simulate GBM progression is the PI model. The PI model is a relatively simple model which describes GBM progression being governed by two distinct processes: diffusion and proliferation. Each of these processes is characterised by a model parameter dictating the importance of that process to the overall progression. With estimates of these two parameters, along with an initial map of cellularity, the PI model can be used to predict tumour progression and response to treatment. However, making accurate estimates of these two parameters is quite challenging. Given the amount of data that standard numerical tools require, the costs, difficulties, and potential risks associated with obtaining the necessary imaging data are often disqualifying. Researchers must then resort to techniques capable of making predictions based on smaller data sets which, while possible, often introduces a new set of challenges and inaccuracies.

The first published technique that attempted to estimate patient-specific parameters based on imaging data was developed by Swanson et al. in a paper [118] and subsequent patent [117]. Their method used summary measurements taken from imaging combined with several mathematical assumptions on the governing model, which led to a set of equations which could be solved to yield estimates for the model parameters. For many

years, this method served as the only one for estimating tumour progression parameters in a patient-specific manner. A key advantage of this method is that it is mathematically simple, requiring no more than a few measurements and basic algebra to derive estimates. However, this simplicity arises from the fact that the data available from tumour imaging is used only in summary measurements (such as radius) while more granular measurements (such as voxelwise signal) are unused. Other recent works have attempted to create more sophisticated methods to accomplish patient-specific brain tumour characterization, though they each tackle slightly different problems and utilize a range of different techniques. Konukoglu et al. [55], for example, estimated both the diffusivity and proliferation parameters of the PI model for brain tumour growth using a time series of multimodal MRIs. Interestingly, their method did not rely on estimated tumour cellularities as ours does (explained below), but rather, relied on tumour delineation on imaging formulated using Eikonal equations. Schuefele et al. [103] similarly attempt to estimate the parameters of the PI model, but use only one patient MRI, generating a pre-image using an atlas of healthy scans. From the observed image and the generated image, they were able to use a modified Picard iteration algorithm to estimate the patient parameters. Theoretically, one important advantage of this method is that it relies on only one MRI, rather than two or more, as most longitudinal studies do. However, their method also requires a large existing imaging dataset before making predictions on a particular patient - an important disadvantage, given that data can often be scarce depending on the specific application of interest. Subramanian et al. [114] take a similar approach, utilizing a single brain image and generating a second using brain atlases. In their work, the PI model is extended to a multi-equation system including the effect of advection as well as brain tissue elasticity. They were able to derive estimates for parameters governing diffusion, proliferation, and mass-effect for both synthetically generated tumours as well as a patient dataset. Gooya et al. [37] also estimated the parameters in a reaction-diffusion-advection model with brain tissue elasticity, though their predictions came simultaneous with a segmentation and registration process. Tunc et al. [123] compared parameter estimates for different models of brain tumour growth including simple reaction-diffusion, reaction-diffusion-advection, and reaction-diffusion-mass-effect, and did so using a covariance matrix evolution method. In another variation of the standard PI model, Meghdadi et al. [76] estimate the parameters of a reaction-diffusion-convection model using multiparameteric MRIs with particle swarm optimization and genetic algorithms. Pati et al. [89] utilize multiparametric MRI to estimate diffusivity, a mass effect parameter, and the number of days since tumour inception. Interestingly, their method simply used labelled images from The Cancer Imaging Archive (TCIA) and is largely independent of the choice of model. Finally, Hormuth et al. [44] take a similar approach to the one we outline in this work by using multiparametric MRI to match parameter values using solutions of a reaction-diffusion model. Like us, they utilize

DWI to estimate tumour cellularity, which they use as input to PDE solvers. Parameters are then matched to the patient case through comparing to solutions of the PI model.

All of the above works have advantages and disadvantages, and examine related problems which vary in slight, though crucial ways, making direct comparisons challenging; so, no method has emerged as a consensus favourite for addressing the general question. Ideally, a tumour progression parameter estimation algorithm would have several advantageous characteristics. First, it should be able to make accurate estimations based on minimal information, as data relating to a particular patient and their tumour can be hard to come by, leaving researchers and clinicians relying on general data or existing databases. Second, the estimates produced by the algorithm should be insensitive to error, an asset in this setting as standardization across image acquisition protocols - to ensure appropriate comparisons between images obtained at different facilities and with different machine operators - is still not common [22, 26, 30, 45, 79]. Third, the presence of various alterations of the PI model being used in the literature suggests that the algorithm should be easily generalized to include other model forms. And lastly, the algorithm should be capable of easily incorporating advances in imaging, registration, and segmentation, as advances in these fields have proven to be frequent and significant of late.

In this paper, we develop a specialized deep learning model capable of characterizing brain tumours which has the desired properties discussed above. Our method uses multi sequence MRI at two time points and produces estimates of the PI model parameters along with a full prediction of the tumour progression between the two imaging times. Involved in our pipeline are several important steps including segmentation, conversion to tumour cellularity, and network training; each of which are described in the methods section below. We explain in detail the two distinct scenarios in which we apply our model: synthetic sensitivity and patient data. The outcomes of the tumour segmentations, conversions to cellularity, and model application to both the synthetic and patient data cases are detailed in the Results section. In the conclusion, the work is summarized, limitations are discussed, and directions for future research are noted.

3.2 Methods

This paper outlines a deep learning model capable of producing accurate, patient-specific estimates of the PI model parameters and simultaneously giving a full forecasting of the intermediate GBM progression curve. As known data, the neural network requires the voxel-by-voxel tumour cellularity throughout the relevant brain volume at each of the two imaging times. The following sections detail the key steps and ideas involved in deriving

these cellularity profiles from MRI data, including data the PI model, data acquisition, image segmentation, and conversion to cell density. We also explain our deep learning model and detail the two cases in which we apply it: synthetic sensitivity and patient data. Our procedure pipeline can be seen in the top part of figure 3.1.

3.2.1 Proliferation Invasion Model

The mathematical model most commonly used to describe the progression of GBMs is a PDE called the proliferation-invasion model which has been used and discussed in many previous works [3, 35, 41, 72, 76, 82, 83, 119]. The PI model has the form

$$\frac{\partial n}{\partial t} = D\nabla^2 n + rn\left(1 - \frac{n}{n_{max}}\right) \tag{3.1}$$

where $n = n(\vec{x}, t)$ represents the tumour cell density at position $\vec{x}$ and time t. The key characterizing parameters of the PI model are D and r, which represent the tumour cell diffusivity and proliferation rate respectively. With estimates of these two parameters, the PI model can be solved and the tumour fully characterised. Also present in the model is the cell carrying capacity n_{max}, which quantifies the maximum biologically feasible tumour cellularity at a location. Typically, the value of n_{max} is straightforward to estimate as it is simply assumed to be the maximum observed cell density. Furthermore, when using the PI model in our pipeline, we work with a nondimensionalised form of the equation which considers only the ratio $\frac{n}{n_{max}}$. For this reason, we artificially set $n_{max} = 1$ for all calculations without loss of generality. We chose the PI model for our study because it is relatively simple, well understood, and widely used in the literature.

3.2.2 Data Acquisition

The imaging data used in our study was collected as part of previous studies and was obtained for use in this work through a data sharing agreement between the University of Waterloo and Unity Health Toronto. Research ethics board approval was obtained from both the University of Waterloo and Unity Health Toronto and all ethics guidelines were followed in the study. The dataset consists of MRI data stored in DICOM format for five patients, each of whom was diagnosed with GBM. For each patient, MRI was performed in two instances with the T1, T1-GAD, T2, T2-FLAIR, and DWI sequences collected, giving 10 sequences per patient, each of which is 3D and voxelized. As explained below, the first four of these sequences are used exclusively in a segmentation algorithm to separate the

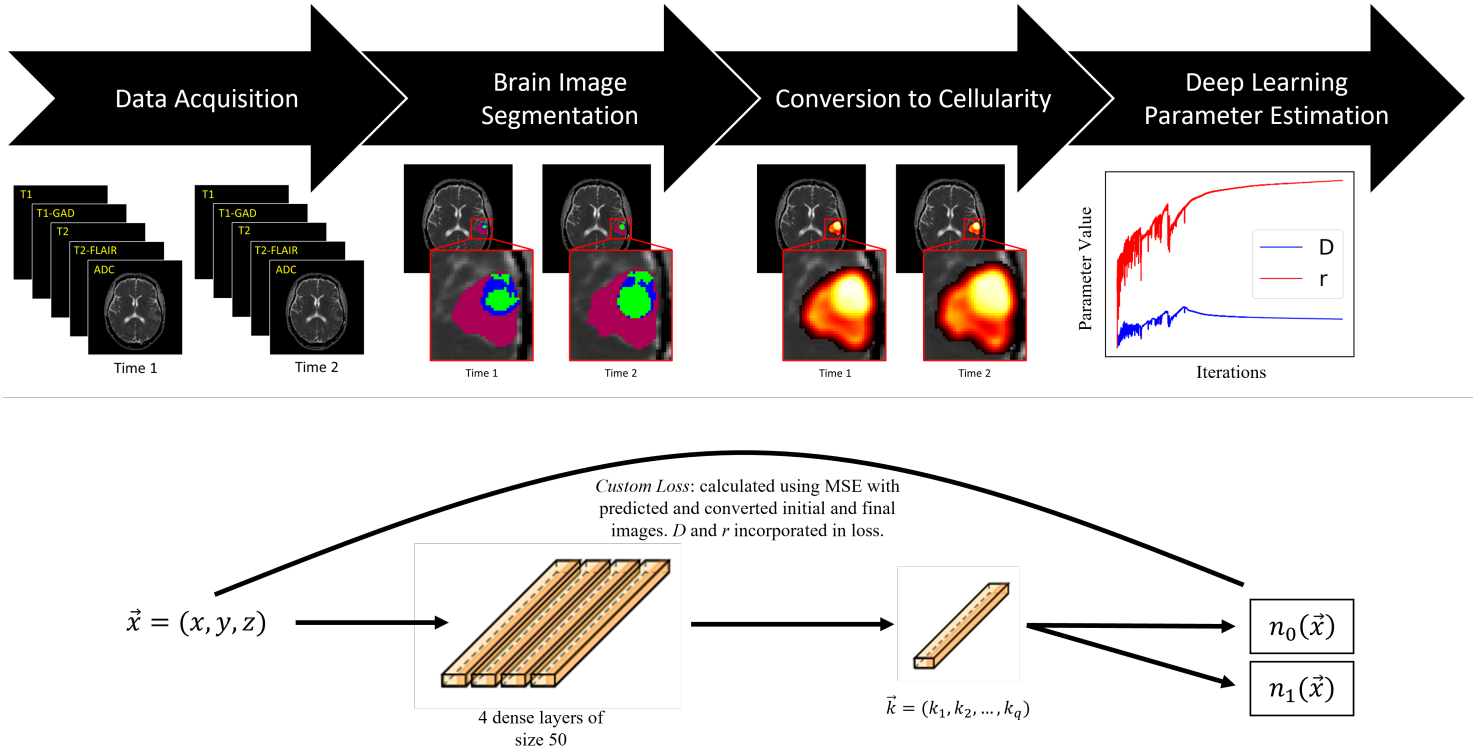

Figure 3.1: **Top**: Pipeline for our parameter estimation model. Following data acquisition, the T1, T1-GAD, T2, and T2-FLAIR images are used for segmenting the tumour voxel-by-voxel. The segmentation results are used to convert the ADC to a map of tumour cellularity. The resulting tumour cellularities at the two imaging times are the known data required to train our deep learning model to derive PI model parameter estimates and intermediate progression curves. **Bottom**: Schematic of our deep learning model. The voxelized spatial coordinates are used as input to four dense layers, each of size 50. The output of these layers is used to predict the q intermediate Runge-Kutta stages which are then used to predict the initial and final cellularity profiles. These predicted profiles are compared with the cellularity profiles derived from imaging to generate a loss which is backpropogated through the network to update the weights and model parameters.

tumour tissue from normal, healthy brain tissue. The DWI however, is used exclusively to derive a map of the tumour cellularity. Importantly, what is actually required for this derivation of cellularity is the map of Apparent Diffusion Coefficient (ADC) which is often obtained explicitly but can also be derived from a DWI and the T2 (null, $b = 0$) image [128]. In this study, the ADC was precalculated from the DWI and given to us directly. A summary of our dataset can be seen in table 3.1.

As our study aims to characterize the natural progression of the tumour, a requirement for our data was that no anticancer treatment was performed between the imaging times for each patient. This requirement severely limits the amount of data suitable for inclusion in our study, since treatment for GBM is usually initiated shortly after detection. This limitation is further discussed in the conclusion. Since the only required data for our study are the images and time between them; the dataset was de-identified with respect to all patient identifiers.

Patient ID	Days Between Imaging	Diagnosis
001	121	GBM grade IV
002	26	GBM grade IV
003	51	GBM grade IV
004	111	GBM grade IV
005	76	Anaplastic astrocytoma, progressed to GBM grade IV

Table 3.1: Summary of the patient data used in this study.

3.2.3 Image Segmentation

Brain tumour segmentation from imaging is a well-studied problem in the literature - even a cursory search will return myriad results with countless segmentation approaches to consider [37, 38, 48, 66, 136]. In this work, we rely on the Federated Tumor Segmentation (FeTS) initiative software which is an open source toolkit which utilizes machine learning techniques to perform tumour boundary detection [88, 105, 106] (`https://www.fets.ai/`). It is developed and maintained by the Centre for Biomedical Image Computing and Analytics at the University of Pennsylvania. The software utilizes a fusion of numerous deep learning models to complete its segmentation. More specifically, the FeTS segmentation algorithm uses T1, T1-GAD, T2, and T2-FLAIR sequences and classifies each voxel into one of four tissue categories: peritumoral edema, enhancing proliferative, necrotic, or non-tumoral. As is explained below, this classification into four tissue types rather than the

simple binary classification of tumoral vs. non-tumoral enables a more sophisticated conversion to cellularity.

Our original DICOM images were first converted to NIfTI files as required by FeTS. Then for each time and patient respectively, the T1, T2, T2-FLAIR, and DWI images were rigidly co-registered to the T1-GAD image using the Greedy diffeomorphic registration algorithm [90]. As part of this co-registration process, the 3D voxel size on all images was standardized to (1mm, 1mm, 1mm). Skull stripping of the images, if necessary, is also required as part of the FeTS segmentation algorithm. Note that in our dataset, patient 004's images were given to us pre skull stripped, and therefore this step was skipped from the segmentation algorithm for patient 004. The results of the segmentation algorithm applied to our patient dataset are given in section 3.1 below.

3.2.4 Calculation of Tumour Cell Density

The known data required to train our deep learning model for a patient is the tumour cellularity profile at two instances and the time between them. Accordingly, deriving the tumour cellularity profile from imaging is an important step in our method. Several works in the literature exist which explore methods of obtaining tumour cell density from imaging, the most common method relying on the ADC image. In particular, a correlation between ADC and tumour cell density has been identified, specifically that

$$\text{Low ADC Signal} \longrightarrow \text{Restricted Fluid Motion} \longrightarrow \text{High Cellularity.}$$

The biological logic behind this line of thinking is that the largest barrier to the Brownian motion of water molecules present in brain tissue is cell membranes, and that an increase in cellularity prevents the free diffusion of water molecules which is observed on an ADC image. Furthermore, this thinking is supported by more than simply a biological rationale, with many experimental works reporting strong correlation between ADC and tumour cellularity, particularly in gliomas [27, 34, 39, 64, 115, 124, 134]. Most convincingly, Surov et al. [116] provided a meta-analysis of the correlations between ADC and cellularity for various types of tumours, noting that gliomas exhibited one of the highest correlations over many cancer types. However, it is worth noting that despite the literature supporting its validity, this correlation is not perfect, specifically being shown to lose accuracy in certain b-value ranges of the original DWI image used to derive the ADC [130]. Nevertheless, this correlation has led to the development of a mathematical expression relating the local

ADC and tumour cellularity, which is given by

$$\frac{n(\vec{x})}{n_{max}} = \frac{ADC_{water} - ADC(\vec{x})}{ADC_{water} - ADC_{min}} \tag{3.2}$$

where ADC_{water} is the ADC in free water, ADC_{min} is the minimum ADC observed in the tumour, and n_{max} is the tumour cell carrying capacity also found in the PI model (equation (3.1)). Note that ADC_{min} is the minimum of the ADC values observed in the region identified as tumour by the segmentation algorithm, and therefore that ADC_{min} changes from patient to patient. This expression assumes a simple negative linear relationship between ADC and cellularity with the maximum cellularity coinciding with the minimum observed ADC and the minimum cellularity coinciding with the maximum observed ADC: that of water. Many previous mathematical works have used this relationship to derive tumour cell density profiles [8, 9, 77, 121].

While this relationship has the benefits of being simple and experimentally validated, in certain situations, it can lead to tumour cellularity profiles that are not biologically reasonable. Specifically, in a recent review of diffusion imaging of brain tumours, Maier et al. [65] noted that ADC values in brain tumour tissue occasionally clearly exceed ADC values of normal gray and white matter, and furthermore that ADC values tend to be highest in the centre of the tumour and decrease toward the tumour boundary. Indeed, this was observed in some of the patient images obtained for this study (see figure 3.2). In such cases, using the above conversion relationship between ADC and cellularity would therefore result in cellularity profiles with higher densities at the tumour boundary and lower densities at the tumour core - the opposite of what is typically observed for GBM. Maier et al. explain that this apparent contradiction between ADC and cellularity is due to the presence of tissue necrosis which causes degradation of cell membranes, reducing the restricted diffusion of water molecules, and hence increasing the observed ADC [65]. Importantly however, these necrotic cells in high ADC regions are still meaningful for tracking the tumour progression and need to be counted in the conversion even if these cells are not proliferating. The PI model accounts for these non-proliferating cells by using a logistic growth term which attenuates the proliferation rate as cellularity increases. Another potential issue is that due to the sharp nature of tumour segmentation maps, the cell density profile resulting from this conversion will have a sharp drop-off at the tumour boundary. Again, this is not the biology of GBMs, which tend to invade local tissues with a smaller number of cells than is present in the tumour core. Furthermore, both of these drawbacks lead to results that are opposite of the results typical of the PI model, which produces profiles with a dense core and decreasing cellularity toward the tumour boundary.

However, the additional knowledge obtained from using FeTS for segmentation allows for the creation of a more sophisticated conversion relationship. Specifically, consider a

voxelwise conversion function as follows.

$$\frac{n(\vec{x})}{n_{max}} = \begin{cases} 0, & \text{if } \vec{x} \in \text{healthy voxels} \\ \frac{ADC_{max} - ADC(\vec{x})}{ADC_{max} - ADC_{min}}, & \text{if } \vec{x} \in \text{enhancing proliferative or peritumoral edema voxels} \\ 1, & \text{if } \vec{x} \in \text{necrotic voxels} \end{cases} \tag{3.3}$$

In this relationship, the additional knowledge from classes of tissue identified by FeTS is leveraged to ensure that the resulting cellularity profiles agree with GBM biology and the PI model. It is assumed that in necrotic tissue, the cells have ceased proliferating and have therefore reached their carrying capacity, n_{max} (normalized to 1 here). In proliferative or edematous voxels, the established relationship is used, and in healthy tissues the cellularity is assumed to be zero. While this form alleviates the problem of low cellularity in the tumour core, it does not address the sharp discontinuities occurring at the boundary between tissue types. To address this, a mean filter with kernel size 3x3x3 (corresponding to a 3mm cube) is applied to the resulting cellularity array using the `scipy` function `ndimage.convolve` to achieve the desired smoothness around the tumour boundaries. We stress that this step is numerically, not biologically, motivated and is necessary to reduce numerical errors related to solving the PDE. It is also worth noting that applying a mean filter in this way slightly enlarges the tumour volume (by 2 voxels, in this case). Though this new conversion model is not perfect, it is able to remove many of the problems with the existing conversion by incorporating information from a more sophisticated segmentation algorithm.

3.2.5 Deep Learning Parameter Estimation

Neural networks can be used for function inference; in other words, to copy the action of a target function on given inputs. If given N input-output pairs from a target function, $n(\vec{x}, t)$, a neural network can be trained to approximate the function by minimizing the mean squared error between the expected and predicted value of that function. In this case, the function we use a neural network to approximate is the tumour cell density as a function of space and time. Since this function is quite complicated and the data we use to approximate it is scarce and noisy, the data points alone are not enough to accurately infer the function. Luckily though, we have more information available to us than simply the data points; specifically, we assume that the tumour cell density obeys the PI model and can use this information to supplement the training of our function inference network.

Our method for incorporating the PI model into our neural network optimization is motivated by the concept of physics-informed neural networks (PINNs) developed in a

series of papers and a patent by Raissi et al. [96, 97, 98]. The key idea of PINNs is to improve function inference networks by augmenting the loss function used in training to include a quantification of the satisfaction of the governing PDE model by the network predictions. Through several examples in the original papers explaining PINNs, it is shown that this formulation results in the network converging more accurately and more efficiently than by using the data alone. This is not surprising, since by altering the loss function in this way, more information has been added from which the neural network can learn. Of particular importance, it is also shown that the optimization remains remarkably robust when noise is incorporated into the known data. In our case, we augment the loss function in such a way that the network is penalized when the PI model is not well satisfied. To do this, the PI model is discretized using a Runge-Kutta time stepping algorithm with q stages. Then, the Runge-Kutta scheme is rearranged with the initial and final solutions as functions of the q stages,

$$\tilde{n}_0 = k_i - \Delta t \sum_{j=1}^{q} a_{ij} g(k_j, t_0 + c_j \Delta t) \quad , \quad \tilde{n}_1 = k_i + \Delta t \sum_{j=1}^{q} (b_j - a_{ij}) g(k_j, t_0 + c_j \Delta t) \quad (3.4)$$

where k_i is the ith Runge-Kutta stage ($i \in [1, q]$), a_{ij}, b_i, and c_i are the Runge-Kutta scheme parameters selected using a Gaussian quadrature rule, Δt is the time between the initial and final cellularity snapshots, and g is the right hand side of the PI model (equation (3.1)) rewritten as $\frac{\partial n}{\partial t} = g(n; \vec{x}, t)$. In this formulation of the Runge-Kutta scheme, each intermediate stage, k_i, can be used to produce an estimate of the initial and final cell density profiles, $\tilde{n}_0$ and $\tilde{n}_1$. Hence the q estimates for the initial and final cellularities can be compared to the known cell densities obtained through imaging, n_0 and n_1, as explained above and fed to the loss function as a quantification of the satisfaction of the PDE.

Inputted to the network are the spatial coordinates of the voxels $\vec{x}$, and outputted are the Runge-Kutta stages $k_1, k_2, ..., k_q$ which are used to quantify the satisfaction of the PDE. Evaluating the prediction of the initial and final data snapshots using these stages and the Runge-Kutta scheme, the loss function can be written as:

$$MSE = MSE_0 + MSE_1 \quad (3.5)$$

where

$$MSE_0 = \frac{1}{N} \sum_{i=1}^{N} (n_0 - \tilde{n}_0)^2 \quad , \quad MSE_1 = \frac{1}{N} \sum_{i=1}^{N} (n_1 - \tilde{n}_1)^2. \quad (3.6)$$

In the above, N is the total number of voxels in each image, n_0 and n_1 are the known solutions at the initial and final times, and $\tilde{n}_0$ and $\tilde{n}_1$ are the solutions at the initial and

final times predicted by the neural network. In other words, the loss function is the sum of the mean squared errors at the initial and final cellularity snapshots over all of the voxels in the image. As the network is trained, MSE approaches zero and the network weights converge to the values which best mimic the action of the cell density function. Note that the intermediate Runge-Kutta stages (k_i) are estimates of the cellularity profile between the two known images at times determined by the c_j Runge-Kutta parameters, hence giving estimates of the intermediate cellularity along the way. A visual inspection of these profiles allows for a check of network overfitting, which was done for each case in our work. The network input has a size equal to the spatial dimension of the cellularity and output equal to the total number of voxels. This structure was the same as was used in [97].

The key capability of this style of network optimization is that parameters appearing in the PDE can be estimated during the training of the network. In the same way that the network optimizes the value of its weights, it can iteratively optimize the PDE parameters to find the values which best fit the data. This allows the network optimization to act as not only a method for solving the PI model, but also as a parameter estimation technique. To do this, a final custom layer is added to the network which calculates the loss using the PDE and unknown parameters which can be backpropogated through to update these unknown parameters. Another important capability of the model is to make predictions without the need to specify boundary conditions acting on model (3.1). As explained in [96, 97, 98], the boundary conditions are assumed to be applied to the system which generated the data and are therefore incorporated into the network training through the data itself, rather than through explicit specification. This means that the network is able to infer the impact of the boundary conditions and take this into account when making predictions.

3.2.6 Sensitivity Analysis on Synthetic Tumours

To showcase the model capabilities and assess the places where the error is most likely to be high, we first apply our deep learning model to a set of synthetically generated tumours. Each synthetic tumour is a computationally generated tumour cellularity profile for which the full cellularity progression and PI model parameter values are known. In particular, we perform analysis on the model by applying it to a series of tumours generated using random selections of parameters, all originating from a Gaussian initial condition. Additionally, although the model operates in three spatial dimensions when applied to the patient data (as described above), here we perform the synthetic tests in one spatial dimension rather than three. This decision was made simply to reduce the computational

expense in performing the analysis: our testing indicated that each three-dimensional case requires roughly 20 times the computational run time as the one-dimensional case. Importantly, this reduction in spatial dimension affects the code only in the encoding of equation (3.1) (partial derivatives only need to be calculated in one dimension instead of three), and in the number of outputs of the neural network (which effectively reduces by the cube root of the number of voxels in each dimension). It would have also been possible to treat the 3D tumour as a 1D mathematical problem through the assumption of spherical symmetry, with the tumour radius, r, acting as the lone spatial parameter. This would preserve the 3D nature of the problem while saving significant computational expense. We did not do this however, because switching between coordinate systems when interfacing the FEM and PINN codes is quite cumbersome. The deep learning model is easily adaptable to different dimensions of the spatial input. In the development of the code, isolated 3D cases were also tested, with results similar to those in the 1D case.

In order to generate tumour cellularity profiles, the nondimensionalized PI model was solved with zero-flux boundary conditions using the FEniCS Project finite element PDE solving software [4, 63]. This solving was performed on a spatial interval of length 1 nondimensional unit with 100 spatial points, and over a time interval of 10 nondimensional days. To obtain a fully dimensional solution, random values for D and r were selected and used to scale the spatial axis and total time. These values were randomly selected within the ranges $D \in [0, 0.1]$ (mm^2/day) and $r \in [0, 1.0]$ (1/day), which were intentionally selected to exceed the ranges in [41], who investigated the ranges typical for patient GBMs. To dimensionalize these tests, the results simply must be scaled by the appropriate scaling factors. Specifically, the number of days between images can be found by multiplying the nondimensional time by $(1/r)$ and the tumour size can be found by multiplying the nondimensional length by $\sqrt{D/r}$. For example, if the initial and final snapshots have a nondimensional time between them of $\tilde{t} = 5$, then a proliferation rate of $r = 0.1$ per day would correspond to the images being taken 50 days apart. In other words, the identification of dimensional parameters using the model is equivalent to the identification of nondimensional parameters with appropriate scaling. After dimensionalizing, 2 random time steps are selected (from the 500 time steps in the full solution) to be the initial and final images given to the parameter estimation network. Additionally, in the original papers describing PINNs [96, 97, 98], random noise was added to the known data in testing to display the model's robustness. To assess this robustness in our model, we perform similar tests by adding noise to the generated solutions of the PI model and analyzing our model accuracy. To do this, a random maximum noise, ϵ_{max}, is selected from the interval $[0, 0.05]$. Then, each voxel in the initial and final cellularity profiles is multiplied by a random value in the interval $[1-\epsilon, 1+\epsilon]$ where $\epsilon \in [-\epsilon_{max}, \epsilon_{max}]$. In other words, up to 5% random noise

is added to each voxel of the known data individually. As the PI model is numerically solved with known parameter values, the predicted values can be compared to the exact values and the prediction error quantified. Examining this error over a range of times between images and data noises can give an idea of when the network is more likely to have an error in its predictions. This was done for 2500 synthetically-generated tumours.

3.2.7 Implementation of Application to Patient Data

Using data obtained from patient imaging, as described in sections 2.2-2.4, the deep learning model can be used to estimate parameters for patients' tumours. This is done for the five patients outlined in section 2.2. For each patient, the 3D grid of spatial inputs and cell density outputs were flattened into 1D arrays. The initial and final cellularity profiles are used to estimate the values of D and r specific to that tumour. To calculate these, each patient's images are given to the deep learning model 10 times to generate 5 estimates for D and r. A bootstrapping algorithm is then used to derive the parameter estimate and 95% confidence interval for each patient's parameters. To do this, 1000 sets of size 10 are generated by randomly sampling with replacement from the original 10 network approximations. The estimate for each parameter is then calculated by taking the mean of the means of each set, and the 95% confidence interval is calculated by finding the minimum width around the mean such that 95% of samples are contained within that interval.

3.2.8 Neural Network Implementation

For all examples and simulations in this paper, the neural network was implemented in python using TensorFlow 2.3.1. The number of intermediate Runge-Kutta stages was always chosen to be $q = 100$. The network structure in all cases consisted of 4 fully connected layers, with 50 nodes per layer, and all nodes utilized the built-in hyperbolic tangent activation function, and Adam optimizer. The Adam optimizer is an extension of the stochastic gradient descent algorithm, more details on which can be found in [53]. Network inputs and outputs were normalized to the interval [-1,1] prior to training, which requires rewriting the PDE in the loss function layer using the new scaled variables. Network structure and relevant training hyperparameters were either taken to be the default, as optimized by Raissi *et al.* [96, 97, 98] or were chosen heuristically based on testing the convergence and run time. The code was run on an AMD EPYC 7542 2.9 GHz CPU and an NVIDIA Tesla A100 GPU. In the 1D synthetic case, the network was trained for 10000 iterations and assumed a constant learning rate of $\eta = 0.001$. Each synthetic tumour required approximately 30-45 seconds of run time. In the synthetic case, the underlying data used

to train the network matched the PDE exactly (with some added noise) and hence it was amenable to training. In the patient data case however, the PDE model is a far rougher approximation of the data resulting from our conversion method, so the training algorithm takes far longer to converge than was required in the synthetic case. For each patient, the optimization was run for 1.5 million Adam iterations, which was chosen based on a visual inspection of the graph of the loss functions. The learning rate was chosen to begin at $\eta = 0.01$ and decay by 2% every 5000 iterations. This meant that in the final iteration, the learning rate was $\eta = 2.3e-05$. Each tumour requiring approximately 5-10 hours of run time, depending on its size.

3.3 Results

3.3.1 Image Segmentation and Data Preprocessing Results

Brain tumour segmentation was performed on our 5 patient dataset using FeTS, which categorized each voxel in the brain tumour image into one of four tissue types: peritumoral edema, enhancing proliferative, necrotic, or non-tumoral. Total final tumour volume for each of the five patients (from 001 to 005) are 89.1cm^3, 12.1cm^3, 30.1cm^3, 67.8cm^3, 53.7cm^3 and the proportion of the tumour volumes segmented into each of the 4 tissue types can be seen in table 3.2. Representative cross sections of this conversion can be seen in figure 3.2, and figure 3.3 contains an additional figure with axial, sagittal, and coronal cross sections of the tumour. These segmentations are used to calculate the cellularity profiles using the procedure outlined above; figure 3.4 shows representative cross sections of these cellularities for each patient. Of note, it is interesting that patients 001 and 005 have image segmentation results which found no proliferative region within the tumour - only necrotic and edematous regions were present. Other segmentation algorithms may give slightly different results for these tissues, and it is important to note that replacing the segmentation method used in our code is simple. More discussion on this is provided in the conclusion section.

3.3.2 Sensitivity Analysis on Synthetic Tumours

The results of the sensitivity analysis performed on synthetically generated tumours can be seen in figure 3.5. In these plots, the average percent error in the prediction of the values of the two parameters is shown as a function of Δt and ϵ_{max}. As can be seen, parameters

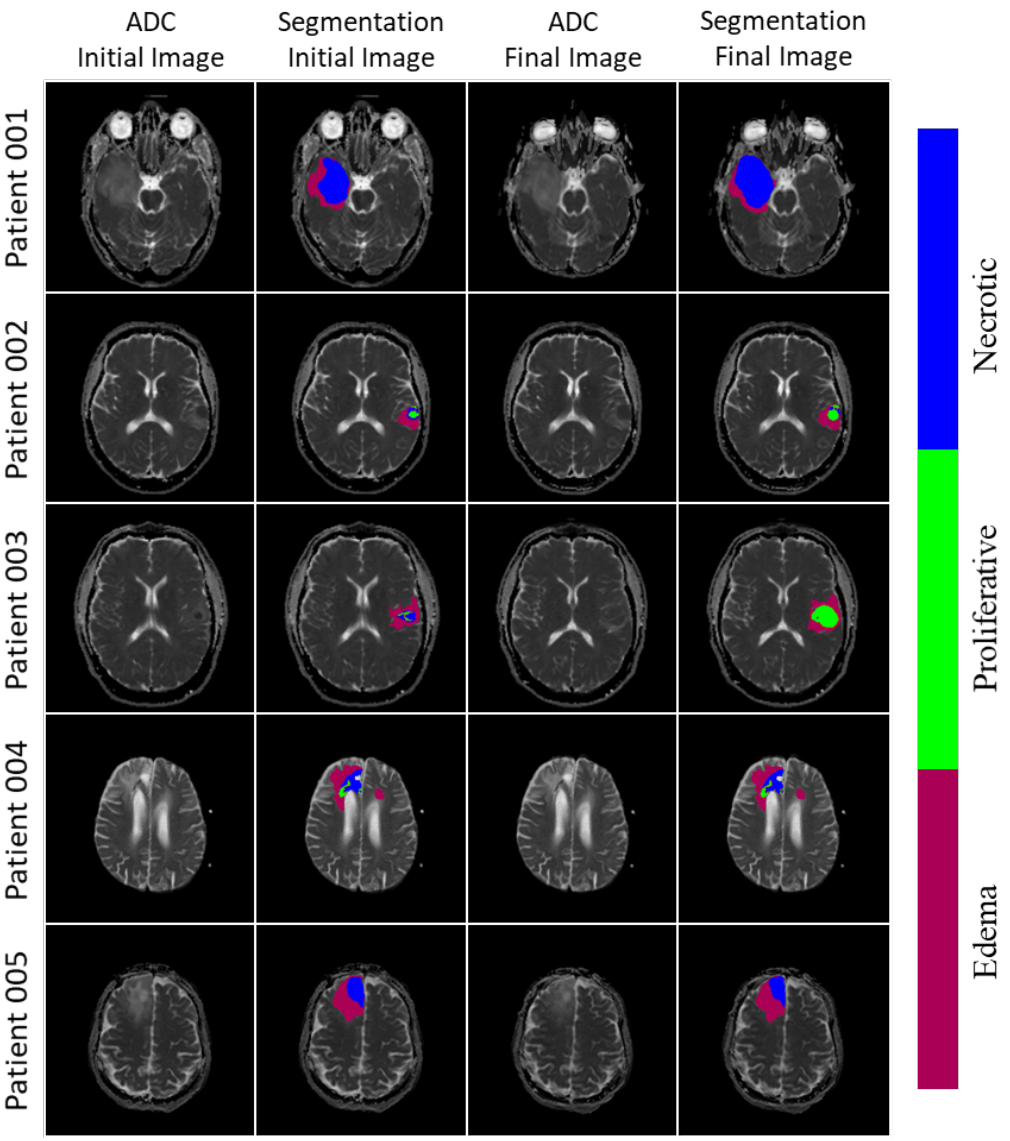

Figure 3.2: Representative cross sections of segmentation results for our patient dataset. Segmentations were performed by the FeTS software which uses the T1, T1-GAD, T2, and T2-FLAIR images to classify the image voxel-by-voxel into the tissue categories peritumoral edema, enhancing proliferative, necrotic, or non-tumoral. The first and third columns show the ADC at the initial and final image times respectively. The second and fourth columns show the ADC with the segmentation results superimposed at the initial and final image times respectively.

Patient ID	Necrotic Proportion		Proliferative Proportion		Edema Proportion	
	Initial	Final	Initial	Final	Initial	Final
001	0.61	0.67	0.00	0.01	0.39	0.33
002	0.11	0.05	0.11	0.19	0.77	0.76
003	0.13	0.02	0.03	0.41	0.83	0.57
004	0.14	0.14	0.07	0.07	0.79	0.79
005	0.19	0.20	0.00	0.00	0.81	0.80

Table 3.2: Proportions of each patient's initial and final tumour segmented into the tissue types: peritumoral edema, enhancing proliferative, necrotic, or non-tumoral.

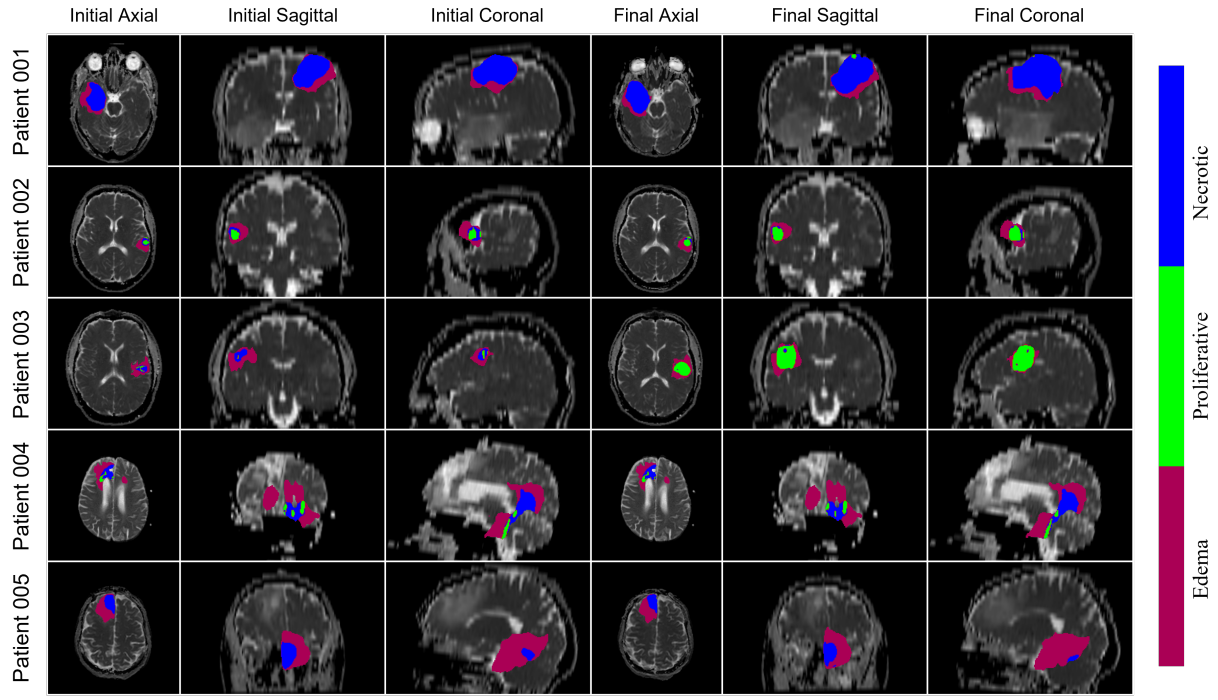

Figure 3.3: Segmentation results of the three spatial axes for the initial and final time images of each patient. Segmentations were performed by the FeTS software which uses the T1, T1-GAD, T2, and T2-FLAIR images to classify the image voxel-by-voxel into the tissue categories peritumoral edema, enhancing proliferative, necrotic, or non-tumoral.

are able to be estimated with reasonably high accuracy in most cases, with 89.7% of runs having an error of less than 5% in both parameters and only 1.1% having an error higher than 10% for both parameters. Our model shows the highest error in its predictions in

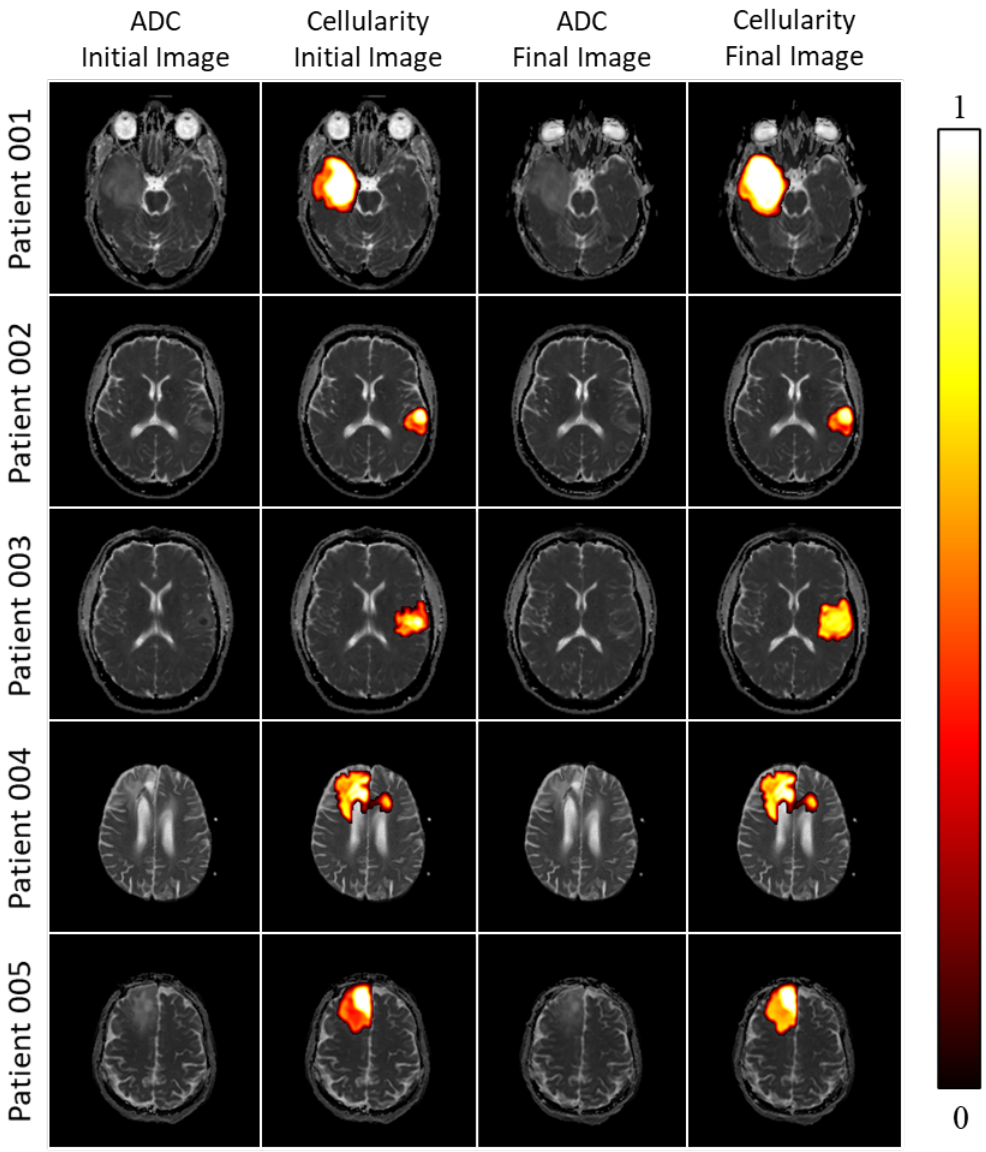

Figure 3.4: Representative cross sections of cellularity conversion from ADC and tumour segmentations for our patient dataset. Conversions were performed using equation (3.3). The first and third columns show the ADC at the initial and final image times respectively. The second and fourth columns show the ADC with the calculated cellularity superimposed at the initial and final image times respectively. The scale maximum is the tumour cellularity normalized by n_{max}.

cases where the time between the images is small: less than 1 nondimensional day (or $1/r$ dimensional days). In these cases, and particularly when noise is also high, higher errors are present in the predictions of the two parameters, especially for the diffusivity D. The reasoning for the higher error in these cases is that when Δt is small, the difference between the cellularity profiles at the initial and final times is also small. Therefore, the small absolute differences in cellularities resulting from image noise lead to high relative differences between the images, which can lead to larger differences in the predictions of the parameters. Similarly, the model is able to produce highly accurate approximations of the parameters even in cases when noise is high as long as the time between the images is not small. Importantly, this analysis suggests that the cases which are most likely to contain an error in the predictions of the parameters are the ones whose initial and final cellularity profiles are very similar. Intuitively, it is more challenging to extract precise parameter estimates from cellularity profiles that show little progression between them than from ones which show noticeable progression.

An important advantage of our model is that it trains on one patient's data alone and requires no other data or existing imaging database in order to make predictions. This does however mean that there is no meaningful split of our data into training and testing sets as is done in many machine learning applications. This means that it is difficult to assess our model's performance prior to application since, by its nature, all relevant data must be incorporated into the training. It is for this reason that we chose to apply the model to synthetic tumours since it allows for an analysis of the parameter approximations in cases where the true values are known. Specifically, it sheds light on cases where we should have higher or lower confidence in model predictions, which can inform the interpretation of the model results when applied to patient data.

3.3.3 Application to Patient Data

The parameter estimates made by our deep learning model on our patient dataset can be seen in table 3.3. For each patient, the parameter estimate and 95% confidence interval are given for both the diffusivity, D, and proliferation rate, r. Also shown are the average of the final loss values over all of the bootstrap samples. Observe that the range of predicted proliferation rates over the 10 patient set is large, ranging over approximately 7 orders of magnitude. Furthermore, notice how these estimates coincide with the imaging and cellularity profiles seen in figure 3.4. For example, there is a large amount of visible growth during the 51 days between images for patient 003, leading the model to predict a high proliferation rate. In contrast, the difference between the cellularity profiles for patient 004 appears small and the time between these images is 111 days, leading the model to

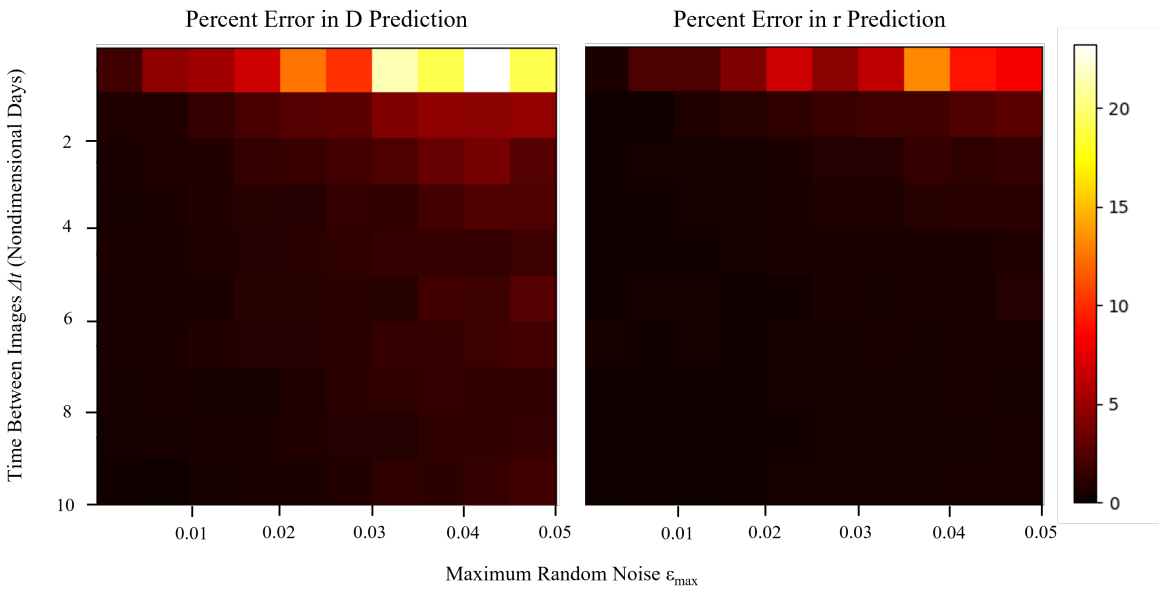

Figure 3.5: Percent errors for the application of the deep learning model on synthetic tumours for D (left) and r (right). Maximum noise is selected by uniformly sampling over the interval $[0, 0.05]$ and nondimensional time between images is uniformly sampled in $[0, 10]$ nondimensional days. Results are averaged into 10 equal width bins along each axis and averaged.

converge to a small value of r. For the most part, our parameter estimates listed in table 3.3 fall in the lower range of other estimates found in the literature [41, 60, 101, 102, 122]. This is most likely because of the bias in data inclusion in our work, since patients who had multiple images without treatment between them were likely those whose tumours were growing slowly enough that they did not require immediate treatment. It could also be due to the difference between considering different growth models, for example, exponential vs. logistic growth in equation (3.1). Some previously cited works consider an exponential growth model, which may result in a different prediction for proliferation rates, especially for larger tumours. Indeed, the five patients considered in our study may fall into this category. In figure 3.6, representative cross sections of the initial and final cellularity profiles as well as profiles at a middle time (the middle Runge-Kutta k_i) are shown.

After obtaining the patient-specific estimates for the PI model parameters, they can be used to predict future tumour progression. To do this, the final cellularity profile for each patient, as derived from imaging, is used as an initial condition to simulate the solution of the PI model with the estimated parameter values. Like for the synthetic tumour case, solving the PI model is done using the FEniCS project solver [4, 63]. The results of this tumour progression prediction can also be seen in figure 3.6.

Patient ID	Estimated D (mm^2/day)	Estimated r (1/day)	Average Loss
001	$7.55\text{e}{-4} \pm 1.54\text{e}{-4}$	$2.25\text{e}{-2} \pm 7.95\text{e}{-4}$	1.32e−3
002	$2.60\text{e}{-2} \pm 3.53\text{e}{-3}$	$9.27\text{e}{-3} \pm 5.99\text{e}{-4}$	4.99e−4
003	$1.18\text{e}{-2} \pm 1.36\text{e}{-3}$	$7.14\text{e}{-2} \pm 6.46\text{e}{-3}$	4.40e−4
004	$1.38\text{e}{-9} \pm 3.22\text{e}{-11}$	$1.11\text{e}{-3} \pm 4.22\text{e}{-6}$	1.40e−4
005	$9.25\text{e}{-4} \pm 4.52\text{e}{-5}$	$5.87\text{e}{-3} \pm 1.76\text{e}{-4}$	5.35e−4

Table 3.3: The mean and 95% confidence interval estimates for the PI model parameters for each of the patient data cases as predicted by our deep learning model using 10 runs of the model and 1000 bootstrap samples. The final column of the table is the average terminal loss after training the 10 bootstrap samples for each patient, shown to indicate the goodness of fit.

3.4 Conclusion

In this work, we develop a pipeline that utilizes a deep learning model to make patient-specific estimates of the parameters of the well-known PI model for brain tumour progres-

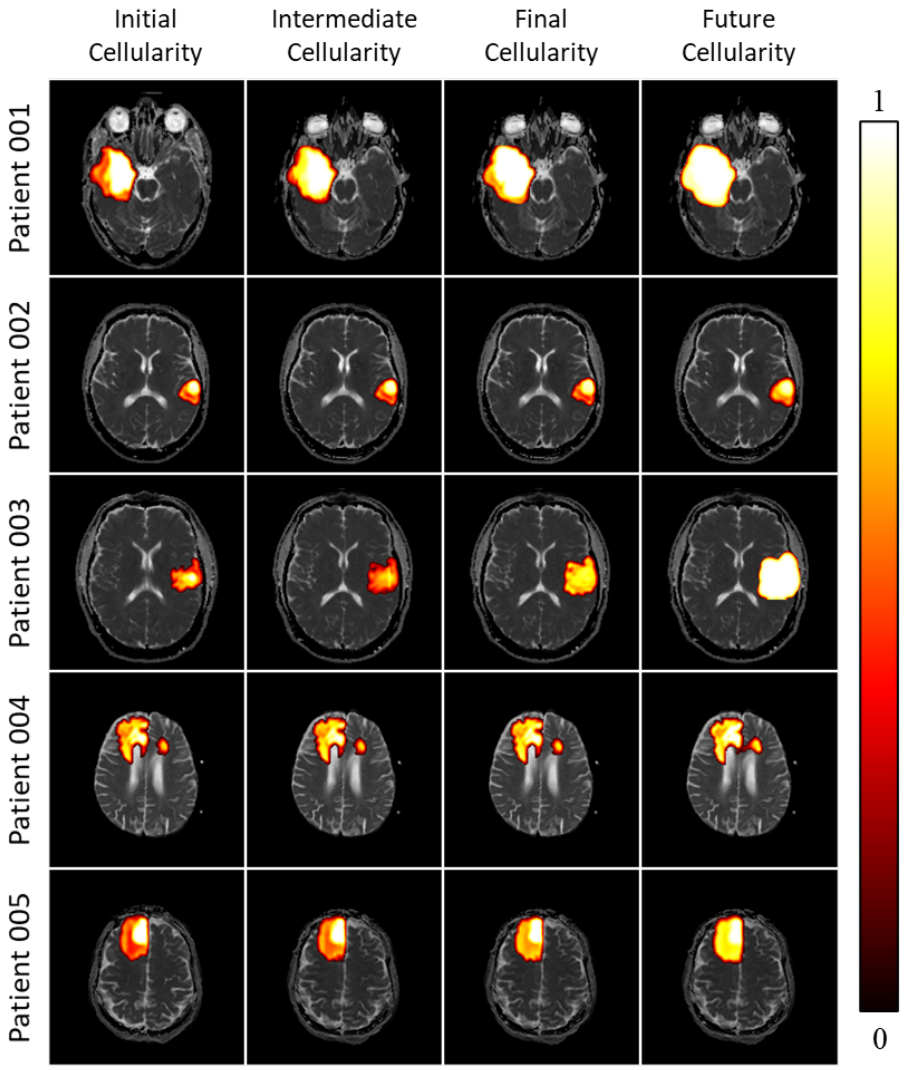

Figure 3.6: The results of the deep learning optimization model applied to the five patients in our dataset. The cross sections shown are representative slices of the 3D tumours. The cellularity is superimposed on the ADC at four times (from left to right): the initial imaging time, halfway between the two imaging times, the final imaging time, and 90 days after the final imaging time. The initial and final time images are converted from ADC data, while the intermediate time in estimated using the neural network, and the future time cellularity is calculated using the PI model. The scale maximum is the tumour cellularity normalized by n_{max}.

sion and use these estimates to make personalized estimations of tumour growth parameters. The pipeline relies on knowledge of five MRI sequences, including T1, T1-GAD, T2, T2-FLAIR, and DWI - the first four of which are used in image segmentation and the final of which is used to derive a map of tumour cellularity. The personalized parameter estimates produced by our model can be used in predictions of tumour growth and treatment response all while relying on minimal data, which is obtained during standard tumour imaging. This clearly has both theoretical and clinical utility. One particular advantage of this method is that it requires data only from the patient about whom predictions are being made. This sidesteps the major hurdle of requiring a large existing dataset which commonly plagues the application of machine learning models to problems in medicine.

While our method was capable of deriving reasonable predictions, there are several parts of our work that could be improved. First, no segmentation algorithm is perfect, and improvements to segmentation performance would directly lead to more accurate deep learning predictions. One may question whether our choice of segmentation algorithm is optimal, or whether the accuracy of the tumour segmentation results obtained herein are sufficient. Luckily, including different segmentation algorithms into our pipeline is a simple process - meaning that as benchmark performance of brain tumour segmentation algorithms continues to improve, these gains can be seamlessly incorporated into our optimization. Additionally, we should also stress that brain tumour segmentation is not the focus of this paper, nor do we claim that our work contributes to state-of-the-art brain tumour segmentation. For this reason, we do not investigate other brain tumour segmentation algorithms. Our choice of FeTS as segmentation algorithm has the advantage of classifying tissue into one of several categories, which allows for a more detailed conversion to cell density. Since our conversion method is able to utilize the voxel-by-voxel tissue classification, we believe that the resulting cellularity profiles are likely more realistic than the standard method. However, more in depth studies in the literature leading to more accurate rules for cellularity derivation would be of tremendous use to improving the accuracy of our work, and clearly there is room in the literature for more investigation here. Furthermore, changes in this regard could similarly be seamlessly incorporated into our work. Additionally, our conversion method does not account for the presence of tumour cells beyond the identified tumour boundary despite the critical importance of these cells when considering tumour recurrence. Identifying and quantifying these cells which extend beyond this boundary is obviously a crucial clinical goal and challenge, and future methods capable of obtaining this information from imaging can easily be added into our pipeline by altering the conversion formula (3.3) to include them. The largest source of error in our model however is likely due to differences in MRI machine calibration and operator protocols. As has been noted elsewhere [22, 26, 30, 45, 79], standardization across image

acquisition protocols is clearly necessary in order to advance AI in medicine. When considering any of the above potential sources of error however, it is important to recall that our deep learning optimization is very robust with respect to noise within the data (as is explored in depth in [97, 96, 98]). Given this, even with changes to any of these parts of our procedure, changes to the parameter estimates are likely to be small. This lack of sensitivity to data noise is a key advantage of our method.

Though the PI model is a well-established tool for modelling the progression of brain tumours, other more sophisticated models could provide benefits. Specifically, the PI model is a one-population model where all cells are considered to diffuse, proliferate, and otherwise act identically. It does not take into account other factors such as tumour necrosis, vascularization, pressures, or advection. Additionally, note that the choice of setting necrotic voxels to a normalized cellularity of 1 may be somewhat misleading as cells in necrotic tissues are not expected to proliferate and diffuse as they would be considered to by equation (3.1), representing a limitation. For example, this setup would theoretically allow for cells in necrotic voxels (and therefore saturated voxels) to diffuse into non-saturated voxels and continue proliferating. This is particularly problematic given the infinite diffusion speed imposed by Laplacian operators in reaction-diffusion systems. Though clearly this effect is small in our application. However, it is important to note that this is not a limitation of this assumption, but again a limitation of the PI model and one-population cell models in general. Our deep learning model could easily be applied to a different model equation (or equations) to estimate the parameters appearing in that model. Switching the PDE model used in the neural network requires only augmenting the loss function which quantifies the agreement of the data to the PDE. Additionally, works using the PI model [82, 83, 118] often consider a spatially-dependent diffusivity,

$$D(\vec{x}) = \begin{cases} D_G, & \text{in gray matter} \\ D_W, & \text{in white matter} \end{cases}$$

where differences between gray and white matter are considered. This could also be incorporated into our model by obtaining a map classifying brain voxels into gray and white matter, then adding a third parameter for the deep learning model to optimize. This idea could be further generalized to include DTI-derived diffusion tensor fields or spatially-dependent proliferation rates. There is also the possibility of pretraining the deep learning model on a larger set of patients prior to patient-specific training. As currently, this method requires a new network to be trained from scratch for each patient, pretraining could greatly reduce the computational expense and potentially lead to more accurate predictions. This would however negate one key advantage of our method, being that our algorithm relies solely on the patient of interest's images and requires no additional data. The most notable

shortcoming of our work though, is that we are unable to compare the growth predictions that result from our deep learning model to a third image time to assess their reasonability. The reason we are unable to do this is simply a lack of data availability due to the specifics of our model setup. In particular, a patient's data was only admissible to this work if it contained all five relevant MRI sequences (T1, T1-GAD, T2, T2-FLAIR, DWI/ADC) at two times with no anticancer treatment administered between them. Typically, it is standard in clinical procedure to perform one set of imaging for diagnostic purposes and then a second set for treatment guidance, and often only one set of imaging is performed to serve both purposes simultaneously. This means that finding cases where three sets of imaging were performed without treatment between them is challenging, and is the reason why our work does not contain any. However, this restriction could be bypassed if the requirement of no treatment could be ignored. This could be possible by adding the effect of treatment into the PI model. For example, the PI model could be augmented to include the effect of treatment by adding terms to the right hand side of equation (3.1) as has been done in many other works [70, 71, 95, 100, 101]. This would allow for the analysis of patients who have treatment between their images as well as the potential to make patient-specific estimates of mathematical parameters related to treatment such as the radiobiological parameters in the linear-quadratic model. Finally, recent works [69, 131] have noted differences in gliomas based on patient sex and have called for the inclusion of this information in the analysis. For our work, it would be interesting to observe the differences in parameter predictions for male and female patients, though this would require a much larger dataset than we have used in order to be meaningful. This will be considered in future studies.

We hope that this work serves as an example of the benefits of deep learning applied to the analysis of medical imaging. Techniques such as this can act in addition to the standard clinical workflow, providing clinicians with evidence-based predictions which incorporate patient data that can be used to aid in disease management.

Chapter 4

Prediction of Intratumoral Fluid Pressures and Liposome Accumulation Using Deep Neural Networks

This chapter contains the second project of this thesis which is currently in submission to *Scientific Reports*. The data was obtained from previous studies on which Shawn Stapleton was the first author. The author list on the submitted manuscript is below.

Cameron Meaney[1], Shawn Stapleton[2,3], Mohammad Kohandel[1]

[1]Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
[2]MD Anderson Cancer Center, Houston, Texas, USA
[3]Department of Radiology, University of Washington, Seattle, Washington, USA

4.1 Introduction

A key obstacle in the efficacy of anticancer chemotherapeutics is the chaotic, inefficient vasculature which commonly plagues solid tumours [78]. This irregular tumour vasculature causes decreased nutrient delivery, increased hypoxia, impaired drainage, and notably, high interstitial fluid pressure (IFP) [21]. High IFP in tumours has been associated with cancer progression and resistance to both chemo- and radio-therapy [78]. Accordingly, the level of

IFP present in tumour tissues is a relevant clinical factor for forecasting tumour progression and designing optimal treatments. Unfortunately, measuring the level of IFP *in vivo* directly is challenging, relying on invasive procedures which would make it disqualifying in a clinical setting. Motivation therefore exists for the development of noninvasive methods capable of predicting high IFP, especially quantitatively, in a patient-specific manner.

Fortunately, recent advancements in medical imaging and liposome technology have provided an avenue to accomplish this. Specifically, a phenomenon termed the Enhanced Permeability and Retention (EPR)) effect - similarly caused by the aforementioned irregular tumour vasculature - results in a selective accumulation of liposomes within tumour tissue over healthy tissue [32, 91]. While this preferential accumulation hasn't yet been leveraged into a significant increase in drug efficacy over standard of care, the phenomenon is well-documented and able to be observed through imaging [28, 32, 42]. This imaged accumulation can then be combined with sophisticated quantitative methods to derive estimates for intratumoral IFP.

An established mathematical model which has been used in many studies ([13, 14, 12, 15, 62, 109, 110, 111, 120], for example) quantitatively links interstitial liposome accumulation to intratumoral IFP. Given the IFP, in addition to the tissue- and liposome-specific model parameters, a map of liposome accumulation could be predicted pre-administration by solving the mathematical model. However, given that IFP is difficult to accurately measure *in vivo*, it is more useful to consider the inverse of this problem: namely, given the intratumoral accumulation of liposomes post-administration, it should be theoretically possible to derive the underlying IFP which resulted in that spatial distribution of liposome accumulation. Using standard mathematical techniques, this inverse problem is quite challenging to solve. However, a novel deep learning technique called physics-informed neural networks (PINNs) is well-suited to handle problems of this nature. PINNs are a type of neural network which incorporates data from a mathematical model into network optimization [96, 97]. In this case, data from liposome accumulation imaging can be used in combination with the established mathematical model to estimate the IFP in that tissue. Spatially varying predictions of IFP for an individual patient could then theoretically be used in predicting disease progression, optimizing disease therapy, or evaluating treatment response.

Estimation of IFP using noninvasive methodologies has been the aim of several previous studies. Bhandari et al conducted a series of investigations that leveraged MRI data from brain tumours. In their initial study [15], they employed PDEs to solve for pressure distributions directly, utilizing parameter values for normal and tumour tissues from the literature. These pressure distributions were then used to predict the transport of cancer drug-carrying liposomes, deploying equations similar to those applied in this study. In a

subsequent study [12], Bhandari et al incorporated variable factors into their modelling, such as heterogeneous vasculature, theoretically enabling patient-specific estimations of IFP. In another work from the same group [14], the authors extended their approach by solving for pressure using PDEs, and using the derived pressure values to predict the transport of various cancer chemotherapeutics. This was done by selecting tissue-specific parameters for normal and tumour tissues, substituting values for drug-specific parameters specific for each chemotherapeutic agent, and solving the equations to obtain concentration maps. Further advancements were made in [13], where dynamic contrast-enhanced imaging was utilized to estimate patient-specific parameter values, enabling predictions of IFP and drug distribution for individual patients. Soltani et al [109] introduced the concept of angiogenesis into the model and, similar to previous studies, and solved for pressure using PDEs with fixed parameter values. Liu et al. [62] also applied MRI data to predict IFP by extracting key tumour measurements and using them to derive parameter values that informed the PDE for pressure. In a larger study, Swinburne et al. [120] employed the PDE model for pressure and transport, applying it to a large dataset of 41 brain tumour patients. The study generated voxel-by-voxel estimates of pressure as well as mean pressure values for each tumour. Importantly however, each of these works solved the forward problem, first selecting or estimating parameter values, then solving for pressure and subsequently drug distribution. None of them solved the inverse problem, beginning with drug distribution and using it to derive pressure, as we do here, which could lead to more specific and accurate IFP predictions.

In this paper, we develop a specialized deep learning model, based on the concept of PINNs, which is capable of predicting the distribution of intratumoral pressure *in vivo* from liposome accumulation data derived from imaging. In order to make predictions, our model requires the spatial map of liposome accumulation within the tumour at some time post-administration as well as estimates of the parameters present in the mathematical model, which could be estimated in a patient-specific manner or taken from the literature as done in the previous studies mentioned above. We apply our method to an animal dataset from a previously published liposome accumulation study [110]. In the materials and methods section below, we explain the data acquisition and preprocessing, as well as describe the key concepts and steps in our method including the PDE model and deep learning model used to make predictions of pressure. In the results section, we show the IFP predictions made by our model on the animal dataset and compare to the measured IFP to assess our model accuracy. We also conduct a sensitivity analysis on our model by applying it to a series of synthetically generated tumours and liposome accumulation maps. In the conclusion, we summarize the paper and discuss limitations and directions for future research.

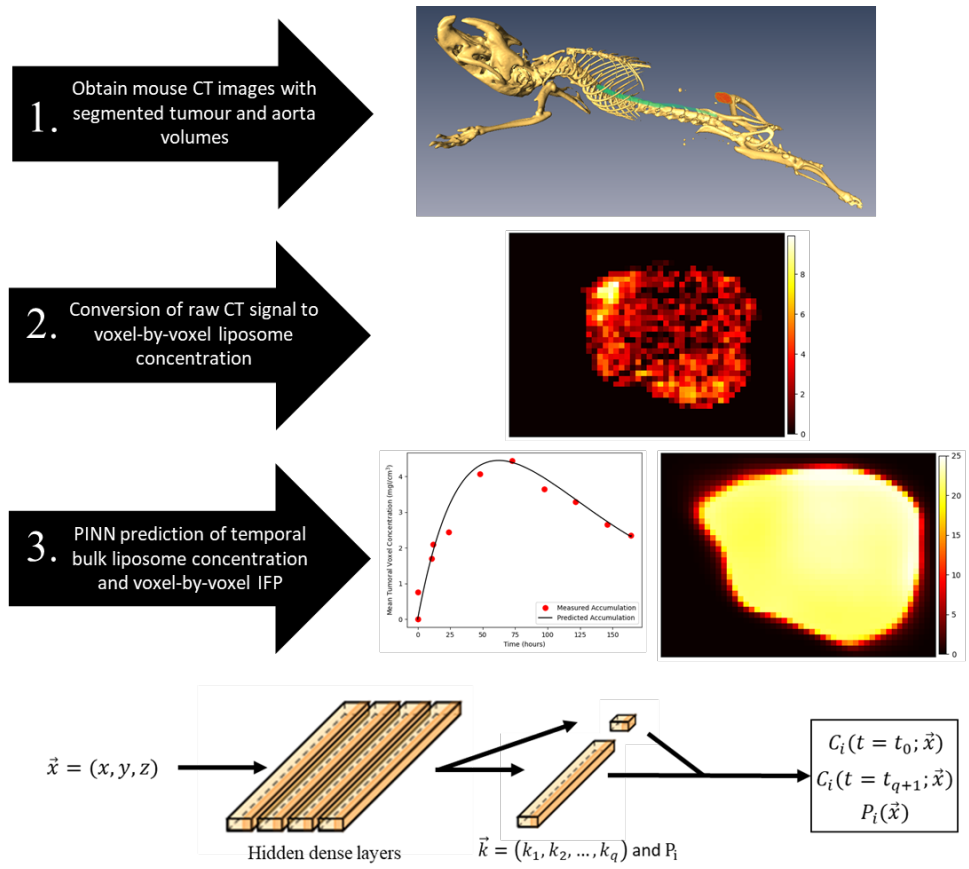

Figure 4.1: Diagram illustrating the workflow of our deep learning model. Following acquisition of CT imaging of segmented xenograft mouse tumours, the data is first cropped and converted to voxelwise liposome concentration at the final imaging time point. Then, our deep learning algorithm uses this data to predict the IFP and liposome accumulation curve. For our dataset, the predictions for liposome accumulation can be compared to additional measured liposome accumulation datapoints obtained through imaging, and the prediction for IFP can be compared to measured IFP using a wick-in-needle measurement. On the bottom, a visual representation of our deep learning model is shown. The spatial coordinates of each voxel are input to the network which consists of 4 dense layers each with 50 nodes, followed by a final dense layer on $q + 1$ nodes. The outputs of this final layer are the predictions of intermediate liposome concentration at each time and pressure at the inputted voxel. These values are passed to the PDE model to compute the initial and final measured liposome accumulation maps derived from imaging, which are used to compute the loss and update the network.

4.2 Materials and Methods

4.2.1 Mathematical Model

The key concept of PINNs, which we employ in the design of our deep learning model, is to incorporate information from a governing mathematical model into the training of the neural network through the loss function. In our work, we rely on a PDE model to relate liposome transport and IFP within tumour tissue, which is given by equation (4.1). This PDE has been used to model the transport of nanoparticles in tumour tissue in several previous works ([12, 13, 14, 15, 62, 109, 110, 111, 120], for example), many of which provide a detailed derivation of the equation itself.

$$\frac{\partial C_i(\vec{x},t)}{\partial t} = \frac{L_p S}{V}\left(P_v - P_i(\vec{x})\right)(1-\sigma)\,C_p(t) + \nabla\cdot\left(fKC_i(\vec{x},t)\nabla P_i(\vec{x})\right) - k_d C_i \tag{4.1}$$

In this model, $C_i(\vec{x},t)$ and $C_p(t)$ are the concentrations of liposomes in the interstitium and plasma respectively. Note that the interstitial concentration is allowed to vary throughout the tumour both spatially and temporally, whereas the plasma concentration is assumed to vary only temporally. In other words, the plasma concentration of liposomes is assumed to be equal in each voxel in the tumour image, and only change its concentration over time. In healthy tissues, pressures are tightly regulated to ensure homeostasis, though in tumours, these pressures can become unregulated. The first term of the right hand side of equation (4.1) describes fluid flow from the plasma into the interstitium, which is governed primarily by the difference between the IFP, $P_i(\vec{x})$, and the microvascular pressure, P_v, which is assumed to be constant in space and time. This fluid motion is also dictated by the vascular hydraulic conductivity, L_p, the vessel surface area per unit tissue volume, S/V, and the reflection coefficient, σ. Since $\frac{L_p S}{V}$ always appears as a grouping, we consider it as a single parameter throughout our work and refer to it as the vessel escape rate. Within the interstitial space, nanoparticles undergo both a convective and diffusive process. The second term on the right hand side of equation (4.1) describes the convection of nanoparticles from areas of high IFP to areas of low IFP. This process also depends on two parameters: the retardation coefficient, f, and the interstitial hydraulic conductivity, K. While in the interstitium, nanotherapeutics also undergo elimination, either through natural degredation, fluid phase uptake, or cellular update. We term this grouped effect the elimination rate, and denote it k_d. In the above, ∇ is the gradient operator which operates in as many spatial dimensions as the problem contains. Additionally, note that $\frac{L_p S}{V}$, σ, f, K, and k_d are all allowed to vary spatially throughout the domain of interest. In particular, it is often assumed that each parameter takes on two values: one within tumour tissue and

one within healthy tissue. See table 4.1 below for a summary of the parameters appearing in the PDE model. Note that some studies which use this PDE model for liposome transport also assume liposome diffusion in the interstitium, which is incorporated into the model by adding a term of the form $D\nabla^2 C_i$ where D is the liposome diffusivity. In our case, since the study from which our data comes did not include diffusivity, we omit it when operating on the mouse data as well. We have included the diffusivity in the sensitivity analysis section below however, for completeness, because many other studies do include it, and since it is a simple addition.

Symbol	Name	Units	Tissue- vs. Particle- Specific
$\frac{L_p S}{V}$	Vessel Escape Rate	1/mmHg*s	Particle
P_v	Microvascular Fluid Pressure	mmHg	Tissue
σ	Reflection Coefficient	1	Tissue
K	Interstitial Hydraulic Conductivity	cm^2/mmHg*s	Particle
f	Retardation Coefficient	1	Tissue
k_d	Elimination Rate	1/s	Tissue
D	Diffusivity	cm^2/s	Particle

Table 4.1: Summary of parameters included in the PDE model (4.1), which has been used in many previous studies. Note that the diffusivity is not included in our model when operating on our mouse dataset since the work this data originates from [110], and to which we compare our predictions to, did not include it. The diffusivity has been included in the sensitivity analysis however because it is included in many other studies.

4.2.2 Deep Learning Model for Estimation of Interstitial Fluid Pressure

From a purely mathematical perspective, given the function $C_i(\vec{x}, t)$, it is theoretically possible to derive the underlying IFP, $P_i(\vec{x})$, using the liposome transport PDE model (equation (4.1)). Doing so however, is no simple task since the IFP can vary spatially throughout the volume of the tumour. From a practical perspective, it is even more challenging since data on $C_i(\vec{x}, t)$ is obtained via imaging, meaning that continuous measurements are not known, but rather, data is typically only available at a number of discrete times; in this case, two of them. The problem of pressure identification is then to find the best fit approximation of $P_i(\vec{x})$ such that equation (4.1) is satisfied and $C_i(\vec{x}, t_1) = u_1(\vec{x})$ and $C_i(\vec{x}, t_2) = u_2(\vec{x})$, where $t_1 < t_2$ and the profiles $u_1(\vec{x})$ and $u_2(\vec{x})$ are the known liposome accumulation maps obtained through imaging. This can be mathematically framed as a function inference problem in which we seek to minimize the error in the known data

points as well as satisfy the governing PDE. Hence, it is well-suited for the application of PINNs.

Consider the time discretization of the liposome transport model according to an implicit Runge-Kutta scheme with q stages over the interval $[t_1, t_2]$. We denote the Butcher tableau parameters a_{nm}, b_n, and c_n, and write the intermediate Runge-Kutta stages as $k_n(\vec{x})$ where $n \in [1, q]$. Note that each $k_n(\vec{x})$ is a prediction of $C_i(\vec{x}, t)$ at an intermediate time between t_1 and t_2 corresponding to the Runge-Kutta parameter c_n. We can write this numerical scheme as

$$k_n(\vec{x}) = u_1(\vec{x}) + \Delta t \sum_{m=1}^{q} a_{nm} g\left(k_m(\vec{x}), t_1 + c_m \Delta t\right) \tag{4.2}$$

$$u_2(\vec{x}) = u_1(\vec{x}) + \Delta t \sum_{m=1}^{q} b_m g\left(k_m(\vec{x}), t_1 + c_m \Delta t\right) \tag{4.3}$$

where $u_1(\vec{x})$ and $u_2(\vec{x})$ are the measured liposome accumulation maps at t_1 and t_2 respectively, $\Delta t = t_2 - t_1$, and $g(\cdot, \cdot)$ is the right hand side of equation (4.1) written as $\frac{\partial C_i}{\partial t} = g(C_i, t)$. When typically using a Runge-Kutta scheme, an implicit matrix equation is solved for all q of the $k_n(\vec{x})$ profiles, which can then be used to calculate the $u_2(\vec{x})$ from $u_1(\vec{x})$. This process would be performed iteratively over many time steps to estimate the solution at later times. In this case however, both $u_1(\vec{x})$ and $u_2(\vec{x})$ are known, but the function $g(\cdot, \cdot)$ cannot be computed since the IFP, $P_i(\vec{x})$, is unknown. Instead, we can invert this numerical scheme to write the known initial and final liposome accumulation maps as

$$u_1(\vec{x}) = k_n(\vec{x}) - \Delta t \sum_{m=1}^{q} a_{nm} g\left(k_m(\vec{x}), t_1 + c_m \Delta t\right) \tag{4.4}$$

$$u_2(\vec{x}) = k_n(\vec{x}) + \Delta t \sum_{m=1}^{q} (b_m - a_{nm}) g\left(k_m(\vec{x}), t_1 + c_m \Delta t\right). \tag{4.5}$$

This formulation of the Runge-Kutta scheme allows the known data (liposome accumulation snapshots) and the governing model (equation (4.1)) to be included in the loss function of our network training, as explained below.

We create a neural network which takes the spatial position of a voxel $\vec{x}$ as input, and outputs the liposome accumulation in that voxel at each intermediate time in the Runge-Kutta scheme, $(k_1(\vec{x}), ...k_q(\vec{x}))$ (hence, 3 inputs and q outputs). These q intermediate stages can then be used to compute estimations for the initial and final liposome accumulation

snapshots, $C_i(\vec{x}, t_1)$ and $C_i(\vec{x}, t_2)$, using equations (4.4) and (4.5) which can be compared to the known liposome accumulation maps derived from imaging, $u_1(\vec{x})$ and $u_2(\vec{x})$. We choose a mean squared error metric to quantify the differences, and hence, the loss function used in network training is written as

$$MSE = \frac{1}{N} \sum_{\vec{x}} [C_i(\vec{x}, t_1) - u_1(\vec{x})]^2 + [C_i(\vec{x}, t_2) - u_2(\vec{x})]^2 \tag{4.6}$$

where N is the total number of voxels in the liposome accumulation image and the sum is computed over all voxels in the image.

While the above formulation allows the inclusion of all of the available information into the optimization, it does not address the problem of unknown pressure. Fortunately, a key capability of this style of network optimization is that quantities appearing in the PDE model can be estimated during the training of the network. In the same way that the network optimizes the values of its weights, it can iteratively optimize the parameters or functions appearing in the model to find which values best fit the known data. In the original papers outlining PINNs, Raissi et al [96, 97] showed how PINNs could accurately estimate the values of unknown parameters in Burgers' equation and identify pressure fields appearing in the Navier-Stokes equations. More recently, we used this technique in combination with brain tumour segmentation algorithms to characterize human glioblastoma multiforme by estimating patient-specific parameters appearing in a common model of brain tumour progression [73]. To make these estimations, a final custom layer is added to the neural network which calculates the loss using the PDE and unknown IFP. This custom loss layer can be backpropogated through to update the values of the unknown pressure in each voxel. In other words, this layer implements the PDE and loss function (equation (4.6)) by including an array for IFP, each voxel of which is an independent, trainable parameter which is optimized to the data and PDE throughout the network training. As the network trains, $u_1(\vec{x})$ and $u_2(\vec{x})$ converge to $C_i(\vec{x}, t_1)$ and $C_i(\vec{x}, t_2)$, and $P_i(\vec{x})$ should converge to the true pressure underlying the system.

One important note is that boundary conditions are not needed to train the network using the data and PDE. As explained in [96, 97], the boundary conditions are assumed to be present in the system which generated the data, and therefore, information on the boundary conditions is assumed to be present in the data itself. Therefore, there is no need to directly specify them in the optimization. It would be possible to add further terms to the loss function to quantify the satisfaction of particular boundary conditions to the network predictions - in much the same way that this is done for the PDE - but this would be challenging since defining the boundary on which these boundary conditions act is nontrivial in the context of complicated tumour geometries. Additionally, the boundary

conditions that would typically be applied to the liposome transport model are continuity conditions at the tumour boundary and a zero-at-infinity condition, both of which are not straightforward additions to this form of discretized model.

4.2.3 Mouse Data Collection and Preprocessing

The mouse liposome accumulation maps used in this study were obtained as part of a previous investigation conducted by Stapleton et al [110]. In that study, 15 female SCID mice were injected with the MDA-231 human breast adenocarcinoma tumour cell line and allowed to grow until reaching a volume of 140 mm^3. A liposome-based CT contrast agent was then prepared and administered to all mice considered in the study. Liposome accumulation CT imaging was performed at various time points using a micro-CT system. To ensure consistency, a laser positioning system was used to place the mice in approximately the same orientation for successive scans, and manual rigid registration was performed after imaging. The tumour volume and descending aorta were then manually contoured on each CT data set.

A second set of mice were also used to measure intratumoral IFP. As the wick-in-needle pressure measurement method can disrupt the transport of liposomes, a separate set of mice was required in order to measure IFP. The intratumoral IFP was calculated by taking 3 or 4 measurements of IFP and finding the mean values. Importantly, the mice that were imaged and used to measure liposome accumulation were a different set of mice that were used for measuring IFP.

For each mouse, the average signal in the aorta volume was used to calculate the plasma concentration after scaling with a species-specific factor (50.1 HU/mgI·cm^3) and hematocrit factor (0.5 unitless), as described in [111]. The resulting plasma concentration at each imaging time was used to fit a continuous plasma concentration function of the form $C_p(t) = ae^{-bt}$, which was then used in equation (4.1). Calculating the interstitial concentration of liposomes is not as simple as scaling the CT signal. Specifically, each tissue voxel contains a vascular compartment with concentration C_p and volume fraction ϵ_p, a cellular compartment with concentration C_c and volume fraction ϵ_c, and an interstitial compartment with concentration C_i and volume fraction ϵ_i. The total concentration can then be written as

$$C_{total} = \epsilon_p C_p + \epsilon_c C_c + \epsilon_i C_i \tag{4.7}$$

Based on the low fraction of endocytic cells observed experimentally by Stapleton et al [110], we assume that the cellular concentration of liposome was negligible, and therefore that

$C_c \approx 0$. This assumption was similarly made by Stapleton et al [111] in their modelling. In addition, at early times, close to 100% of the injected liposomes remain in the plasma compartment, allowing us to make the assumption that $C_i \approx 0$. Using the measured value of C_p from the aorta and the average voxel tissue concentration at an early time (10 minutes in this case), the interstitial volume fraction can be calculated. At future times, the interstitial concentration of liposomes can be solved for by using the same equation, but by measuring the full tumour concentration C_{total} and the plasma concentration using the aorta signal C_p, and by using the previously calculated value for the interstitial volume fraction ϵ_i. This approach yields a voxel-by-voxel interstitial concentration of liposomes derived directly from CT imaging. Herein, we assume the values of $\epsilon_i = 0.30$ and $\epsilon_p = 0.03$ as calculated by Stapleton et al [110].

After conversion to concentration, the tumour contour was used to crop the image to the tumour area with a small buffer and remove liposome signal outside of the tumour area. While it is not a perfect assumption since liposomes are expected to move slightly outside of the tumour area, we are focused on the magnitude and shape of the pressure profile inside the tumour, and thus ignore liposomes outside of its volume. Figure 4.2 shows representative cross sections of the resulting conversions for each mouse.

4.2.4 Neural Network Implementation

All simulations were implemented in python using TensorFlow 2.3.1. The number of intermediate Runge-Kutta stages was always chosen to be $q = 100$. The network structure in all cases consisted of 4 fully connected layers with 50 nodes per layer, and all nodes utilized the built-in hyperbolic tangent activation function. The training uses a combination of an Adam optimizer [52] (an extension of the stochastic gradient descent algorithm) and an L-BFGS optimizer [61]. The PDE and data were nondimensionalized using the variable scaling factors of $\tilde{t} = t * (P_v \mu_T)$, $\tilde{x} = x * \left(\sqrt{\frac{\mu_T}{\eta_T}}\right)$, and $\tilde{P} = P/P_v$. Additionally, prior to training, the network inputs were normalized to the interval [-1,1] and the network outputs were normalized to the interval [0,1]. Note that scaling the inputs and outputs in this way requires a transformation of the PDE to the space of these new variables. Choices for network structure and training, unless otherwise noted, were chosen based on the results of Raissi et al. [96, 97] who published the original works outlinng PINNs. All code was run on an AMD EPYC 7542 2.9 GHz CPU and an NVIDIA Tesla A100 GPU. In the synthetic cases, each sample was optimized first over 30,000 Adam iterations, then the optimizer was switched to L-BFGS, and the training resumed until convergence (defined using a tolerance of 1.0e-8) or a maximum of 10,000 L-BFGS iterations was reached. In

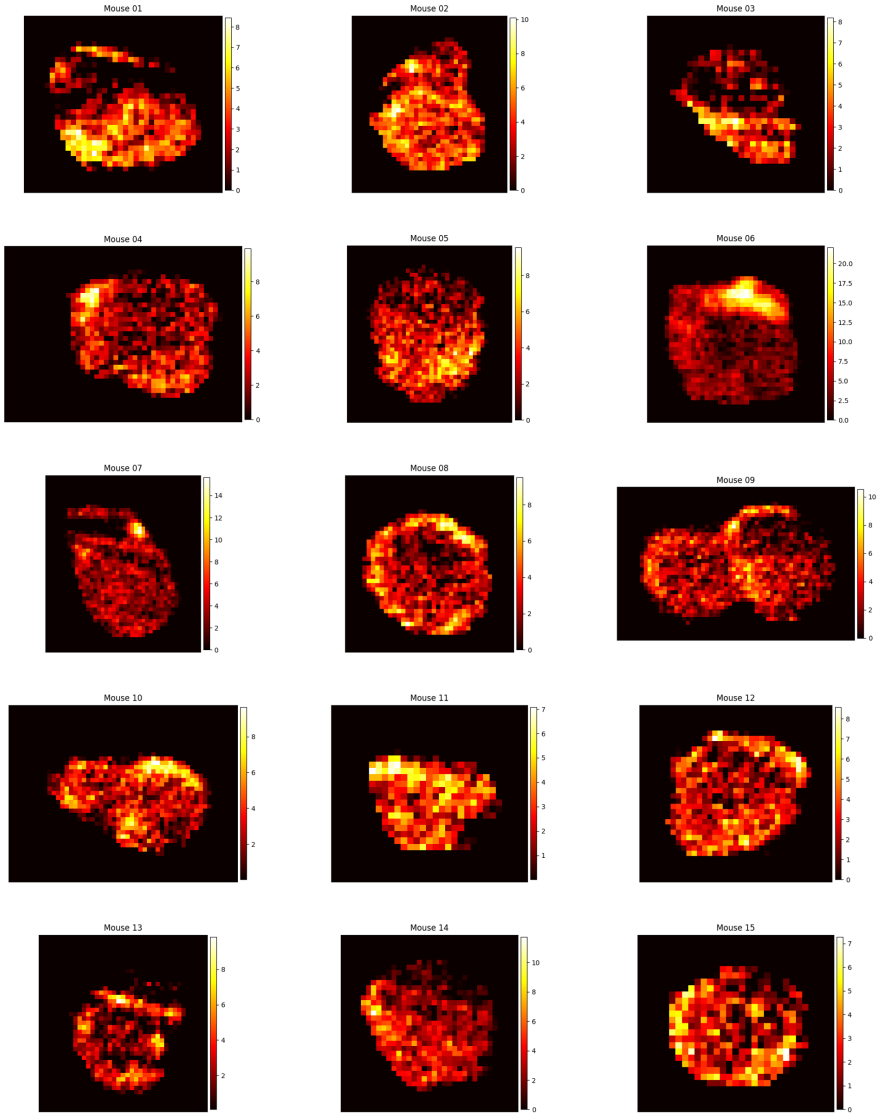

Figure 4.2: Final time spatial liposome accumulation maps for each of the 15 mice considered in our study. Raw CT signal in HU is converted to liposome accumulation using the above preprocessing procedure. Colour bar values are in units of mgI/cm^3. Though the full tumour concentration image is in 3D, a representative vertical slice approximately half way through the tumour core is shown here. These distributions are used as known data for the deep learning model and are compared to the network predictions to calculate the loss value during training.

the animal data cases, the dual optimizer was again used, however training took place over 300,000 Adam iterations and a maximum of 30,000 L-BFGS iterations. Each synthetic case required approximately 10 minutes of run time, whereas each animal case required anywhere from 2-12 hours of run time, depending on its size. The drastic difference in training time between the synthetic and animal data cases is a result of the adherence of the data to the underlying PDE model. Specifically, in the synthetic case, the underlying data used to train the network matches the PDE closely (since the PDE was used to generate the data), and hence the PDE model is not an additional imposed assumption, making it amenable to training. Whereas in the case of the mouse data, the PDE model is a far rougher approximation of the experimental data, since it was generated *in vivo* (which in reality does not perfectly adhere to the PDE), meaning that the PDE is an additional imposed assumption, making it less amenable to training and take far longer to converge. This, in addition to the difference in total voxels due to the spatial dimension, results in significantly larger run times for the mouse data case compared to the synthetic data case.

4.3 Results

4.3.1 Predictions on Mouse Tumours

Using liposome accumulation maps derived from the imaging data of Stapleton et al [110], we implemented our deep learning model to perform voxel-by-voxel estimations of both liposome accumulation and IFP for each of the 15 mice included in our study. The model uses the final liposome accumulation map from imaging as known data and generates estimations of the temporal progression of liposome concentration as well as the spatial distribution of IFP within the tumour, both voxel-by-voxel. We consider the initial imaging to take place immediately prior to liposome administration, and therefore assume that the initial image is identically of zero magnitude. The parameters incorporated in equation (4.1) were assumed to correspond to the estimates provided in the original study by Stapleton et al [110], which are detailed in Table 4.2.

In the study conducted by Stapleton et al [110], each mouse was imaged at multiple time points. Using these time series data, we computed the corresponding liposome concentration maps for each time point. The neural network then utilized the final time point of each of these datasets to generate voxel-by-voxel prediction of liposome concentration at all times and IFP. Figure 4.3 presents a comparison of the measured and predicted liposome accumulations for each mouse, achieved in our predictions by calculating the mean

Parameter	Units	Value
$\left(\frac{L_P S}{V}\right)_N$	1/mmHg·s	2.52e-6
$\left(\frac{L_P S}{V}\right)_T$	1/mmHg·s	2.6e-6
σ_N	1	1.0
σ_T	1	0.2
f_N	1	1.0
f_T	1	0.5
K_N	cm^2/mmHg·s	8.53e-8
K_T	cm^2/mmHg·s	4.13e-8
P_v	mmHg	25
k_{d_N}	1/s	1.65e-6
k_{d_T}	1/s	5.10e-6

Table 4.2: Parameter values used for the mouse simulations in our study. All parameter values were taken from Stapleton et al [110]. Note that each parameter other than P_v is given by a step function, with one value in the normal tissue and another value in the tumour tissue. The subscript N denotes the value in the normal region and the subscript T denotes to the value in the tumour region.

voxel concentration within the tumour region. Note that the mice were not all imaged at the same number of time points.

Observe that for most of the mice considered in our study, the prediction of our deep learning model showed agreement with the observed results obtained through imaging. Some cases however, did not show good agreement. For example, the liposome accumulation comparison for mouse 13 shows a predicted peak concentration much higher than the observed peak concentration. Several factors could account for this discrepancy. First, since the network only uses data from a single time in order to make predictions, the time chosen to be used as known data can impact the results. In mouse 1 for example, the predicted liposome accumulation peak is higher than the observed peak; however, looking at the data points themselves, it is reasonable to question whether the final data point is an outlier as it appears out of step with the rest of the data. If a different time point had been selected to be used as known data, perhaps the error between the measured and predicted accumulation would have been less. Of course in practice, it would not be standard to conduct imaging at numerous times in this way, so one would not know whether an imaged time represented an outlier. Such is the nature of working with scarce, noisy data. Below, we compare our predictions for liposome accumulation and IFP generated using imaging from different times.

Another potential source of error comes from the selection parameters used in equation (4.1). In these simulations, the parameter values used are those in table 4.2. In reality however, it would be reasonable to assume - and indeed, has been observed - that the tissue-specific parameters in the model can vary subject to subject. Changing the values of these parameters would impact the resulting predicted liposome accumulation curve and the prediction for IFP. Of course, obtaining estimates for the values of these parameters on a subject-by-subject basis is challenging, especially if pressure is also unknown. More discussion on this is provided in the conclusion section below.

In addition to the spatiotemporal accumulation of liposomes, our network also predicts the voxel-by-voxel IFP. Figure 4.4 shows a representative cross section of the intratumoral pressure for each of the mice in our study. Stapleton et al [110] performed wick-in-needle pressure experiments to measure intratumoral IFP in 15 mice, though importantly, these 15 mice were a different 15 mice than were administered liposomes and imaged. This is because the wick-in-needle measurement method disrupts the liposome transport process, and therefore wouldn't produce an accurate match of pressure and liposome distribution. However, the mice used for pressure measurements were otherwise identical to those used for imaging, so we can compare the sets of pressure measurements. In figure 4.5, a comparison of the measured and predicted average intratumoral IFP for the mice is shown. Each dot represents an individual mouse, and the dotted line represents the average of each case. Note that the interstitial pressure, P_i must be lower than the microvascular pressure, $P_v = 25$ mmHg, and therefore P_v is the upper bound of the average intratumoral IFP. For the measured average IFP values, the minimum, maximum, and mean values were 7.8 mmHg, 24.18 mmHg, and 17.84 mmHg whereas for the predicted IFP average IFP values, they were 12.57 mmHg, 22.22 mmHg, and 19.77 mmHg.

While comparing pressure predictions mouse-to-mouse would be preferable, our dataset simply does not allow it. We therefore resort to comparing the distributions of measured and predicted pressure. Visually, and based on their minima, maxima, and means, the predictions are relatively similar. The most notable discrepancies are in the predicted mean pressure being higher and range of pressures being narrower. Likely, the discrepancy is a function of two factors. First, a sample size of 15 is relatively small, and therefore randomness may be enough to explain the differences. Additionally however, in Stapleton et al, the mean intratumoral pressure was calculated by performing 3 or 4 wick-in-needle measurements in the tumour volume, then calculating their mean, whereas here, the mean IFP was obtained by calculating the mean over every tumoral voxel. The averaging of a smaller number of measurements used to calculate the observed mean in [110] could add additional error to the result, and given the small size of the tumour, may result in measurements made closer to the tumour boundary or slightly outside the tumour region,

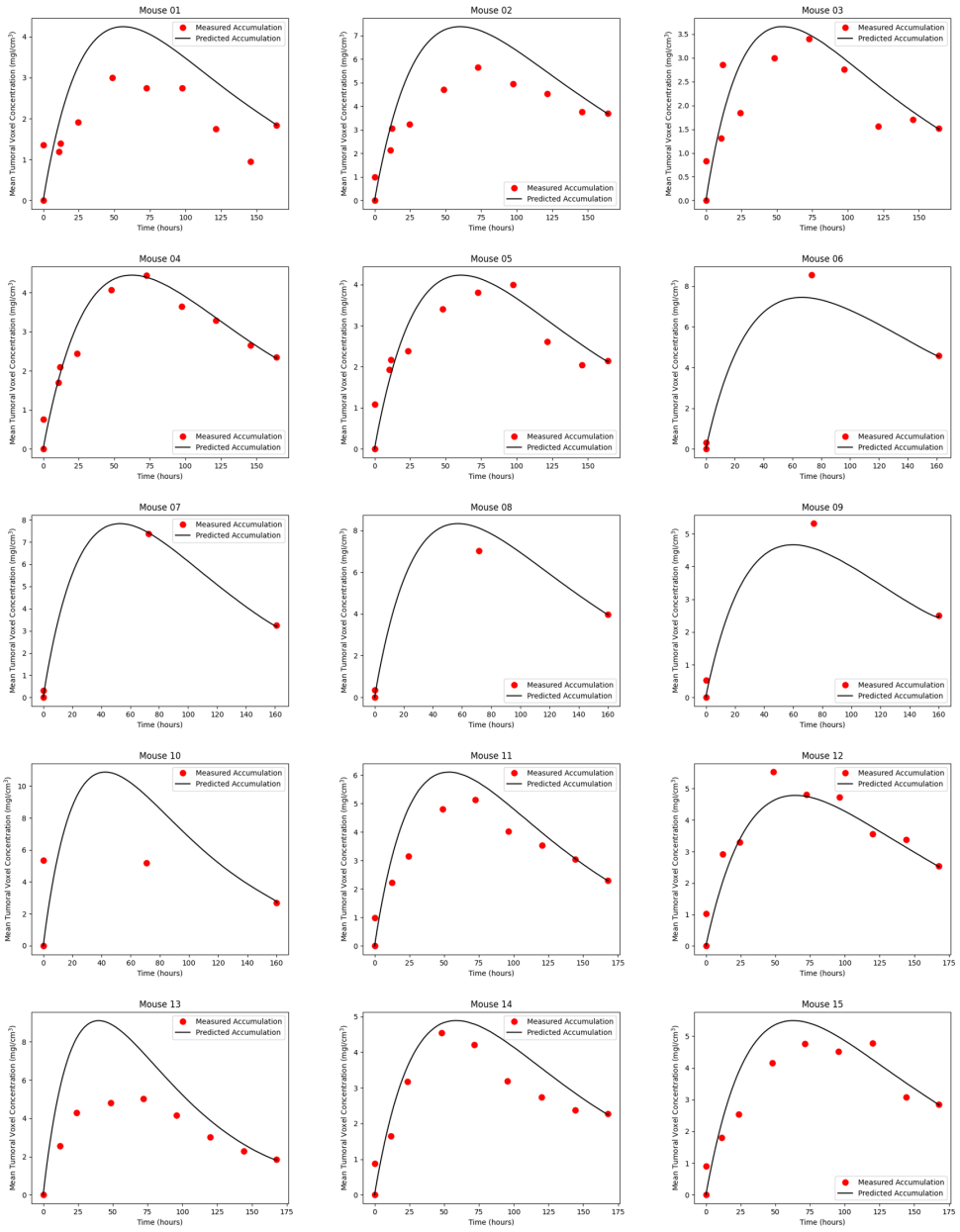

Figure 4.3: Comparison of the measured (red points) and predicted (black line) accumulation of liposomes in the tumour region for each of the 15 mice included in our study. Mean tumoral voxel concentration is calculated by taking the average concentration of all voxels in the labelled tumour volume at each time step.

where the pressure tends to be lower. Furthermore, in [110], it is noted that the pressure measurements themselves can cause a decrease in the tumour IFP, suggesting that perhaps multiple measurements could bias toward a slightly smaller result.

4.3.2 Predictions Using Different Time Points

One key question is how the time of the chosen liposome accumulation image affects the predictions of the network. In other words, how long after injection of liposomes is required in order to derive an accurate estimation of IFP? Specifically, our algorithm is reliant on the final time image, though in our dataset, we have imaging at several times for each mouse, meaning that we could choose any of them to be used as known data in our network optimization. In this section, we do just this, assuming our final image to be each of those available from our dataset and comparing the network predictions. In each case, we derive the IFP and the liposome accumulation up to that time point, then use the PINN to project the concentration forward to the final available time, using the derived pressure to do so. This procedure allows us to analyze the discrepancies in predicted liposome accumulation and IFP when employing different time points as the basis for predictions. Due to the significant computational runtime associated with this process, we restricted our analysis to mouse 01 as a representative illustration. Figure 4.6 presents the results of this analysis.

On the left side of Figure 4.6, the predicted liposome accumulation curves based on the different chosen imaging time as known data is shown. Each of the curves passes through exactly two data points, the first (at time and concentration 0) and the point whose corresponding distribution is used as the final image. Several aspects should be noted here. First, note that the prediction resulting from using the first imaging time (10 minutes post injection) has been omitted from the graph since it's height is significantly higher than the others, making the graph difficult to interpret for the remaining curves. Next, notice that the qualitative behaviour of the curve remains constant irrespective of the imaging time used; specifically, the time at which the maximal average concentration is reached remains constant. Additionally, as noted previously, notice that the final measured time point in the measured accumulation appears to be somewhat of an outlier compared to the general trend, which results in a predicted accumulation curve that does not align well with the rest of the data. In contrast, utilizing one of the intermediate time points as known data yields a curve that more closely matches the remaining data points. Quantitatively, the error between the measured and predicted accumulation is shown in Table 4.3. Based on these residuals, the optimal time to image in this case is 72 hours post injection. Though any time between 11 and 120 hours appear to give reasonable fits to the remaining datapoints. Furthermore, the poor fit of the 144 and 164 hour predictions compared to the other

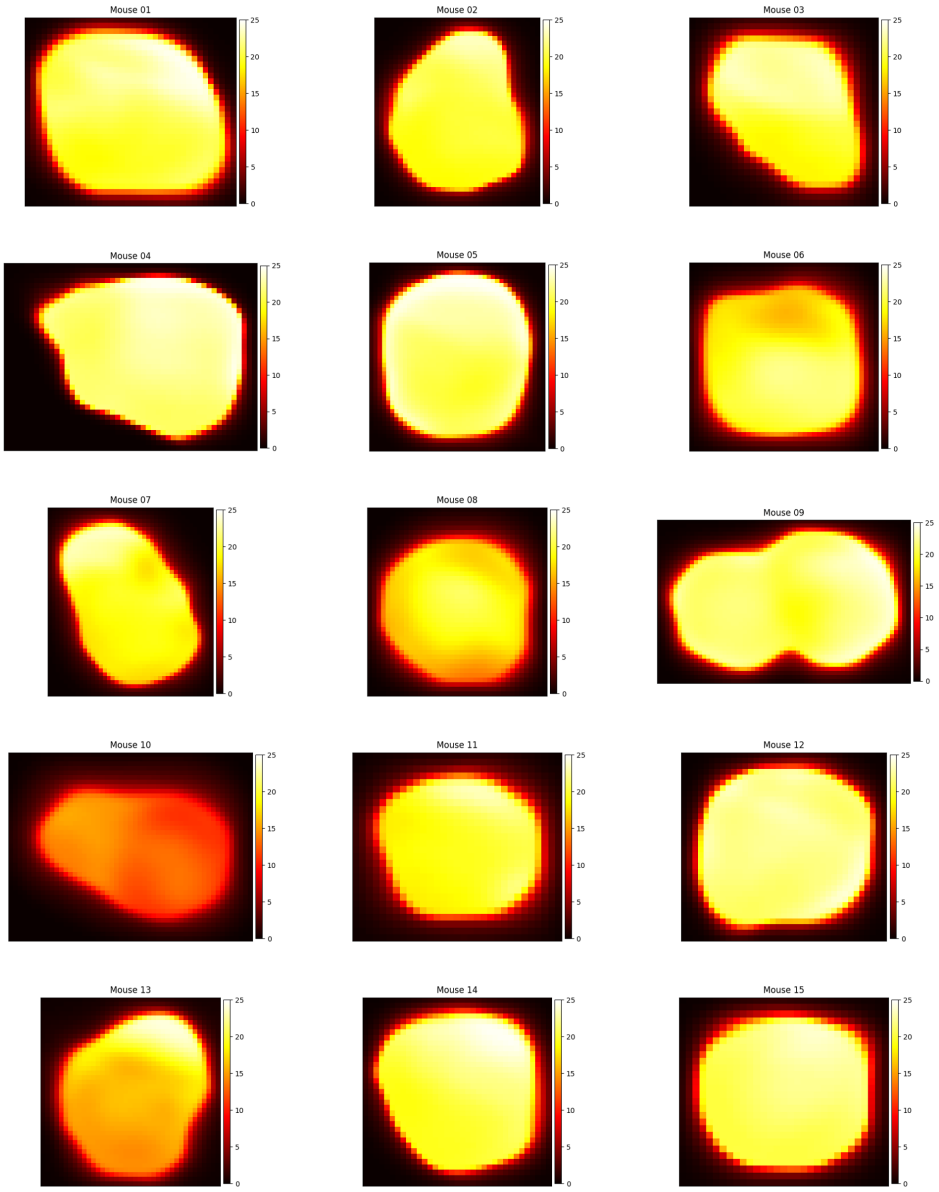

Figure 4.4: Spatial distributions of intratumoral IFP for each of the 15 mice considered in our study. Colour bar values are in units of mmHg - note that $P_v = 25$ mmHg is an upper bound for IFP (P_i. Though the full tumour pressure image is in 3D, a representative vertical slice approximately half way through the tumour core is shown here. These distributions are used in equation 4.1 to calculate the loss value during training.

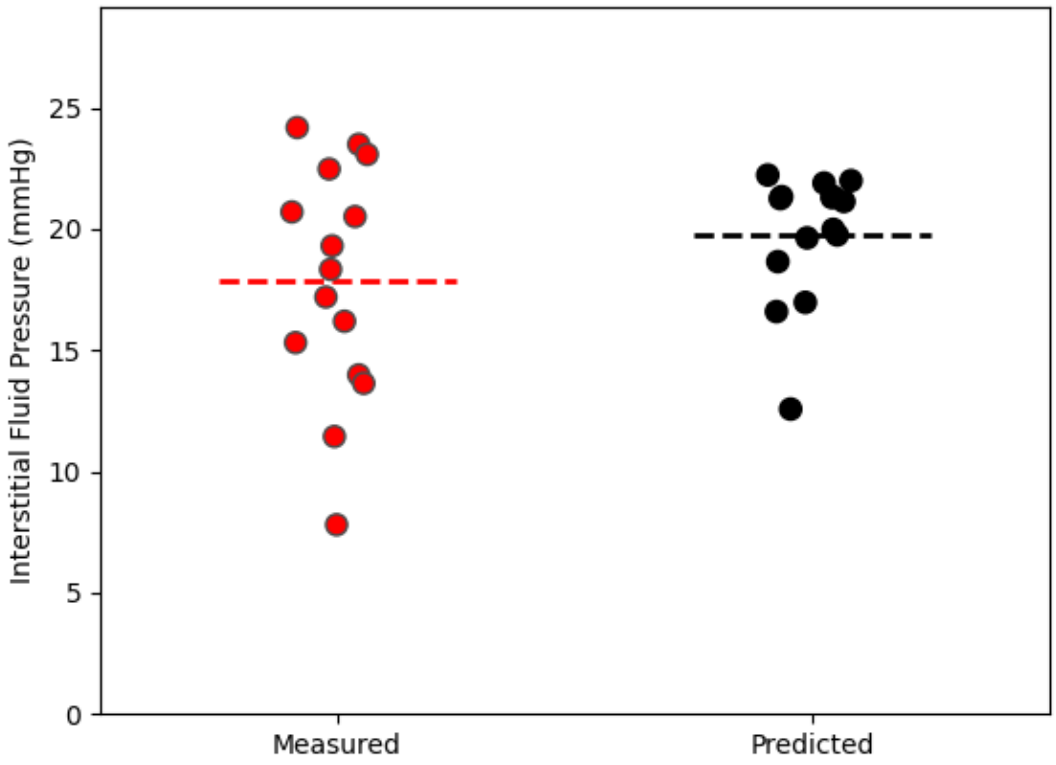

Figure 4.5: Comparison of the measured (red points) and predicted (black points) intratumoral IFP. Measured IFP was obtained in [110] who took an average over 3 or 4 wick-in-needle pressure measurements in the tumour volume. Predicted IFP was obtained by taking the mean over all voxels in the tumour region. Horizontal dotted lines are placed at the mean of each case. For the measured average IFP values, the minimum, maximum, and mean values were 7.8 mmHg, 24.18 mmHg, and 17.84 mmHg whereas for the predicted IFP average IFP values, they were 12.57 mmHg, 22.22 mmHg, and 19.77 mmHg.

datapoints suggests that these data points are outliers, matching our previous intuition. In practical scenarios, it is uncommon to conduct imaging at multiple time points like this, rendering it challenging to discern whether a given time point is likely to be an outlier. This challenge underscores the inherent complexity of working with noisy data, which may be difficult or costly to acquire.

Imaging Time (hours)	Prediction MSE
0.17	3024.31
11	2.85
12	3.11
24	3.30
48	2.91
72	2.77
98	3.82
120	3.18
144	8.18
164	12.01

Table 4.3: Quantification of the error between the measured and predicted total accumulation from using different imaging times as known data. MSE is calculated by finding the sum of the squared errors between each measured accumulation and the predicted accumulation at that time.

On the right side of Figure 4.6, the prediction of average intratumoral IFP is shown. Importantly, the predictions are relatively stable with respect to the final imaging time. Excluding the initial measurement time of 10 minutes, all predicted IFP values fall within the range of [21.20 mmHg, 23.43 mmHg], which is easily explained by the choice of imaging time as known data. Based on this observation, we suggest that a single image, obtained one day post-administration, could suffice to generate reasonable predictions for both the complete liposome accumulation curve and the intratumoral IFP.

4.3.3 Sensitivity Analysis on Synthetic Tumours

In addition to the application of the model on liposome accumulation maps derived from mouse experiments, we also apply the model to a series of synthetically generated liposome accumulation maps. By synthetically generated, we mean that the liposome accumulation maps are derived computationally through solving of the liposome transport model directly.

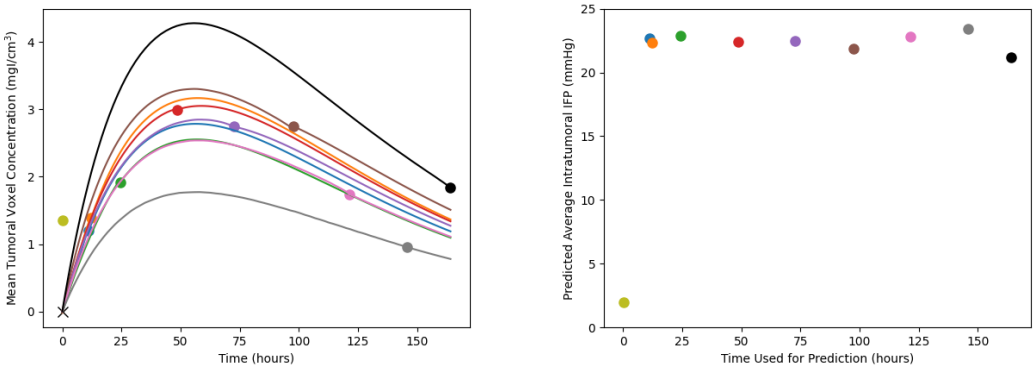

Figure 4.6: Differences in deep learning model predictions based on which final time point is used as input. **Left**: the different mean liposome concentrations throughout the tumour. The red points are the measured accumulations calculated from imaging and the black lines are the predictions of the network. Each line goes through exactly two known points (the point marked with an 'x' at time and concentration zero, and the dot with matching colour to the line). Note that the prediction using the first time point (at 10 minutes) is omitted from the graph as it has a high maximum which makes the rest of the graph difficult to interpret. **Right**: the mean intratumoral IFP as a function of the time point used for predictions (with colour matched to the colour on the left). Notice that, with the exception of the first time point at 10 minutes, the mean intratumoral pressure remains relatively steady while changing the known data.

In particular, an IFP profile, initial condition, tumour geometry, and set of parameters can be selected and used to solve equation (4.1) directly; then, the pressure identification deep learning model can be applied to the result. This serves two purposes. First, since the underlying IFP is known, we can compare the network predictions to the true pressure to assess the error, allowing us to showcase the model capabilities. Second, through generating many of these profiles using various PDE parameter values, we can assess areas where we expect the network to perform well and poorly. Though the mouse experiments above are performed in three spatial dimensions, we choose to perform the following sensitivity analysis in just one spatial dimension. This choice is made purely to reduce the computational expense involved: each increase in the spatial dimension of the problem increases the computation run time by a factor of the number of voxels in the dimension, which typically needs to be in the range of approximately 30 in order to obtain meaningful results. Importantly, this reduction in spatial dimension affects the code only in the encoding of equation (4.1) (partial derivatives only need to be calculated in one spatial dimension) and in the amount of data on which the network trains. All other parts of the network, including its architecture and training, are unchanged based on the spatial dimension of the problem. It would have also been possible to generate 3D synthetic data and treat it as a 1D mathematical problem through the assumption of spherical symmetry, with the radius, r, acting as the lone spatial parameter. This would preserve the 3D nature of the problem while saving significant computational expense. We did not do this however, because switching between coordinate systems when interfacing the FEM and PINN codes is quite cumbersome.

To perform this sensitivity analysis, we assume a baseline of parameter values as given in table 4.4. Then, using these parameters, the IFP is obtained from the PDE

$$\nabla^2 P_i(x) = -\alpha(x)^2 (P_e(x) - P_i(x)) \tag{4.8}$$

which was originally derived by Baxter and Jain [11]. This equation is solved over an interval of length 10cm, equally partitioned into 200 units. The solving is accomplished with a finite element method with a zero-at-infinity boundary condition. To implement this computationally, a zero Dirichlet boundary condition is used and the computational domain is extended farther than necessary, which adds extra computational expense to the code solving, but is a better approximation of the infinite boundary condition. In the above equation, $\alpha(x)$ is given by

$$\alpha(x) = \begin{cases} \sqrt{\frac{1}{K_T}\left(\frac{L_p S}{V}\right)_T}, & \text{if } x \in \text{tumour} \\ \sqrt{\frac{1}{K_N}\left(\frac{L_p S}{V}\right)_N}, & \text{if } x \notin \text{tumour} \end{cases} \tag{4.9}$$

where the individual parameters are given in table 4.4. Similarly, we can write

$$P_e(x) = \begin{cases} P_v, & \text{if } x \in \text{tumour} \\ 0, & \text{if } x \notin \text{tumour} \end{cases} \tag{4.10}$$

with P_v also given in table 4.4. In the above, we define the tumour area as an interval of diameter 2cm in the centre of the computational domain. Using this pressure, equation (4.1) is solved over the same spatial domain using a finite element method in space and a Crank-Nicolson time stepping algorithm. The solution is obtained over the time interval of $[0, 200]$ hours using 150 equidistant time points. A zero flux boundary condition is applied to the PDE at the edges of the computational domain. This yields the function $C_i(x, t)$, from which the initial and final times can be selected as $C_i(x, t_1)$ and $C_i(x, t_2)$ and used as data for training of our deep learning model.

Parameter	Units	Baseline Value	Min Value	Max Value
D	cm^2/s	7.5e-7	0	7.5e-6
f_T	1	0.5	0.4	0.6
K_T	cm^2/mmHg*s	4.13e-9	1.0e-9	1.0e-6
$\left(\frac{L_P S}{V}\right)_T$	1/mmHg*s	2.6e-6	1.0e-7	1.0e-4
Noise	1	0.025	0	0.05
P_v	mmHg	25	5	50
σ_T	1	0.19	0	0.5
k_d	1/s	5.1e-6	1.0e-7	1.0e-5

Table 4.4: Parameter values used in the sensitivity analysis.

The deep learning model is used to predict pressure for the base parameter set as well as the base parameter set with one parameter altered to be the minimum or maximum of a reasonable range. The baseline values are chosen to be the same as those estimated by Stapleton et al [110] as explained above (apart from the noise, which is obviously not added to the mouse data), and the ranges are chosen to be purposefully broad to span a large subset of the parameter space. Note that we have included D as a parameter here even though we did not include it when making predictions on our mouse dataset. We include it here since many works which examine this or similar equations have included diffusion, and it is a simple addition. For each of these cases, the deep learning model is used with the synthetically generated liposome accumulation maps to derive an estimation for P_i, which can be compared to the P_i obtained from solving equation (4.8). This error can then be quantified. In summary, the procedure for our sensitivity analysis is:

1. Select parameters by taking the baseline parameter set, and possibly altering one parameter
2. Solve equation (4.8) with the chosen parameters to obtain the IFP, $P_i(x)$.
3. Use the chosen parameters and calculated P_i to solve equation (4.1) and obtain the liposome accumulation over time, $C_i(x,t)$.
4. Use the initial and final time distributions of $C_i(x,t)$ as known data in the deep learning model to estimate the best fit approximation of $P_i(x)$.
5. Compare the calculated and predicted $P_i(x)$ and report the relative mean squared error between the profiles.

For each parameter set, the deep learning network is run 5 times, and a bootstrapping algorithm with 1000 bootstrap samples is used to obtain the mean and 95% confidence interval estimate for the IFP prediction error. These can be seen in figure 4.7. Note that the model is able to approximate the error to within 3% error in all cases other than the maximum noise case. The highest error in the prediction of IFP (of about 6.5%) occurs in this high noise case, which is unsurprising since adding noise to the voxelwise concentration is expected to lead to errors in the model predictions. Similarly, the lowest error occurs when the noise is minimal. Of the remaining parameters, the predictions appear to be most sensitive to $\frac{L_p S}{V}$ and P_v, with larger values leading to larger errors for both. A relatively large error is also observed in the minimum k_d case.

Note that the sensitivity analysis above considers onyl a local sensitivity analysis; in other words, scenarios where only one variable is altered at a time. However, it is important to acknowledge that the model may have lower accuracy in other regions of the parameter space that require simultaneous alterations of multiple parameters, which is only able to be observed through a global sensitivity analysis. Due to the large number of variables involved in the model and the associated computational costs, conducting a comprehensive exploration of the entire parameter space poses a significant challenge. Ideally, a grid search across the complete parameter space could be performed by discretizing each parameter range into subintervals and systematically selecting parameters from within each subinterval. This would enable the observation of areas with high and low prediction errors, providing a more comprehensive understanding of potential sources of prediction inaccuracies. Nevertheless, the computational burden associated with exploring the entirety of the parameter space renders this approach infeasible.

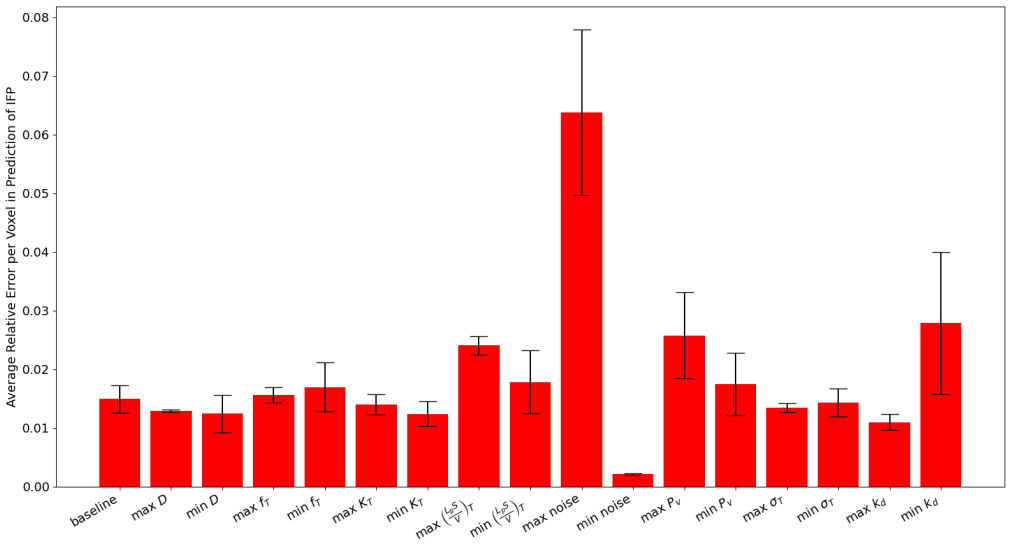

Figure 4.7: Average relative mean squared error in the prediction of IFP (P_i) for synthetic tumours generated using particular parameter sets. The columns denote either the set of baseline parameters in table 4.4 or the set of baseline parameters with one of them made either the maximum or minimum of a chosen range, also given in table 4.4. The height of each bar represents the average error over the 1000 bootstrap samples and the error bounds represent the 95% confidence intervals.

4.4 Conclusion

In this work, we have developed a deep learning model for predicting intratumoral liposome accumulation and intratumoral IFP. The model utilizes the spatial distribution of liposome concentration throughout a tumour volume as input, which can be obtained through imaging. This approach has the potential to enable personalized predictions of the spatiotemporal distribution of cancer nanoparticles or other chemotherapeutic agents for individual patients. Additionally, IFP is a critical measure for tumours, with implications for disease progression and treatment response, knowledge of which could provide valuable information to clinicians. Conceptually, patients could receive an injectable dye, undergo imaging, and have the resulting data used in our model to predict key tumour characteristics. We applied our model to data derived from mouse xenograft tumour imaging and synthetically generated tumours. The inclusion of synthetic cases demonstrates the high accuracy of our model and robustness to measurement noise. Furthermore, the agreement between our model's predictions and measured liposome accumulation in mouse experiments reinforces confidence in the model accuracy.

Our model offers several key advantages. Firstly, it requires minimal input data to make predictions, specifically a single image from a single patient. This contrasts with many machine learning models that require extensive image databases. Accordingly however, this necessitates the limitation of needing to retrain the model for each patient individually, which may incur significant computational costs. Another advantage is the model's capability to handle measurement noise, a crucial attribute given the inherent errors present in medical imaging data. We demonstrated this by applying the model to synthetic tumours with added noise and observing that a 5% noise added voxelwise to our known data resulted in an average error in the prediction of pressure of only approximately 6.5%. Similar results have been reported in previous studies using PINNs [73, 96, 97]. The voxel-by-voxel prediction capability also allows for spatial predictions of drug distribution and improved understanding of individual tumours. Though it is important to acknowledge that this application to synthetic tumours constitutes a local sensitivity analysis rather than a comprehensive global sensitivity analysis since we have not investigated scenarios where multiple parameters are simultaneously perturbed from their baseline values. Conducting a thorough exploration of the entire parameter space is computationally expensive, but its inclusion could provide valuable insights and a more comprehensive understanding of the model behavior.

Despite these advantages, there are areas for improvement. Direct comparisons of predicted to measured IFP is not possible in our work, as pressure measurements using a wick-and-needle technique can affect liposome transport itself, compromising the accuracy

of liposome accumulation data for network training. Theoretically, liposome imaging could precede pressure measurements or additional interventions, but the nature of our data did not support this. Instead, the data in our study used two separate sets of mice, by which we can compare the two groups to see the differences in IFPs, but are unable to compare for particular mice. Additionally, our model does not account for liposomes outside the tumour volume, which are expected to be present in reality. Though since we are primarily interested in intratumoral IFP, we don't expect the small amount of particles outside the tumour volume to have a large effect on the derivation of intratumoral IFP.

Our largest limitation to clinical translation however, is the difficulty of measuring the parameters of the liposome transport model (equation 4.1). Ideally, we would like to estimate both the parameter values and IFP simultaneously. While PINNs can theoretically achieve this, the challenge lies in uniquely identifying these values given available information. In particular, in equation 4.1, $\frac{L_p S}{V}$ and $(P_v - P_i)$ are multiplied together, allowing only their product to be uniquely identified. Accordingly, we are forced to either fix the parameter values and derive the IFP, or fix the IFP and derive the parameter values. In our study, we chose to fix the parameter values using the estimates made by Stapleton et al [110], and derive the IFP. This data offered an opportunity since the parameters were estimated for the specific set of mice and contained measurements of intratumoral IFP to which we could compare our predictions. In the absence of these parameter estimates, representative values from the literature could be used, or a range of parameter values could be sampled and used in the optimization.

Alternatively, there may be a path toward in estimating both the parameter values and pressure simultaneously; however, more information than we considered would need to be included in order to do this. For example, the optimization method could use all of the imaging time points together in order to make a prediction, rather than just an initial and final image. Though inherent difficulties exist in this too. First, in practice a series of images is unlikely to be performed, so this method would not scale well. And second, using many images would require that these images were accurately coregistered together, which is theoretically possible in cases where sophisticated registration algorithms exist (like in the context of brain tumours), though is very challenging in situations where they do not. Another potential avenue to estimate both the parameters and pressure simultaneously is through the incorporation of additional mathematics into the optimization. Specifically, consider the idea that the PDE for pressure, equation (4.8), is included in the optimization. Then, the deep learning model could produce estimates for the unknown parameters of interest, use these values to predict IFP using equation (4.8), then use the parameters and IFP to calculate the loss. This idea has potential downsides as well however. For one, doing so requires that an additional PDE is solved at each optimization iteration,

adding a significant amount of computational expense. Furthermore, this optimization is conceptually different since the network would no longer be predicting the pressure directly, but rather, predicting the parameters, then using them to derive the pressure. Since the parameters are assumed to be constant within the volume of the tumour, this would force the network to lose some of the voxel-by-voxel specificity in predictions of intratumoral IFP.

Similarly, our model also relies on the assumption of a spatially-uniform plasma concentration throughout the tumour volume, which aligns with the approach employed by Stapleton et al. [110] in their analysis. Essentially, this assumption implies that variations in the imaged distribution of liposomes within the tumour can be solely attributed to underlying spatial variations in IFP, rather than the plasma concentration per voxel. In reality however, both the IFP and plasma concentration per voxel are expected to vary throughout the tumour volume, and therefore impact the resulting spatially-varying accumulation. This is especially true considering that both high IFP and impaired plasma concentration are effects of an underlying dysfunctional tumour vasculature. More accurately, we expect the plasma concentration of liposomes to be approximately uniform throughout the vessels, but for the vessel density itself to vary throughout the tumour. Similar to the challenge posed by unknown parameter values described above, without additional data or mathematical assumptions, it is necessary to fix one variable to uniquely predict the other. Given the experimental data available to us, we have opted to fix the plasma concentration and predict the pressure.

Though in theory, it is possible to separate the effects of IFP and plasma concentration for a given subject. However, accomplishing this would require either 1) obtaining more data or 2) incorporating additional mathematical assumptions. For more data, if vessel perfusion imaging were available alongside liposome accumulation imaging for our subjects, it could be used to derive a spatial distribution of plasma concentration. This information could then be incorporated into our model, allowing for the decoupling of the effects of IFP and plasma concentration. Other types of data could also allow for a similar separation of these effects. For instance, if liposome accumulation data were available for multiple dyes within the same subject, it could be possible to mathematically decouple the two fields. The different transport parameters (such as hydraulic conductivity or diffusivity) associated with each dye could be utilized to mathematically separate the effects of these processes. Similarly, the application of antiangiogenic agents before liposome administration could allow for the observation of changes in liposome accumulation under different tumour vasculatures induced by the antiangiogenics, thus similarly enabling the separation of IFP and vasculature effects. For incorporating more mathematics, a different set of model equations could be assumed, coupling both the IFP and the plasma concentration to an

underlying vascular density. Then, in theory, the PINN could be used to predict this underlying vasculature and the IFP and plasma concentration could be derived from it. However, preliminary tests have shown that this process incurs substantial computational expense, making it infeasible at present. Nonetheless, it represents a potential path forward for simultaneous prediction, provided computational challenges are overcome.

In summary, this exploratory study presents a novel machine learning approach to predict liposome accumulation and IFP from imaging data, advancing personalized medicine. We hope this work showcases the potential benefits of deep learning in oncology and medical imaging, offering clinicians a powerful tool for forecasting tumour progression, designing effective treatments, and predicting treatment efficacy.

Chapter 5

Temporal Optimization of Radiation Therapy to Heterogeneous Tumour Populations and Cancer Stem Cells

This chapter contains the third project of this thesis which was published in the *Journal of Mathematical Biology* in 2022 [74]. The author list on the publication is below.

Cameron Meaney[1], Mohammad Kohandel[1] Arian Novruzi[2]
[1]Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
[2]Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada

5.1 Introduction

As one of the pillars of modern cancer treatment, radiation therapy has been the subject of countless research works. Despite its origins in cancer medicine tracing back to over a century ago, many questions and challenges still remain. Of particular interest in recent years, improving our ability to accurately predict the efficacy of a particular course of radiation prior to treatment remains an important clinical goal. In theory, if the efficacy of a given treatment prescription can be predicted prior to administration, then many treatments can be considered and the optimal selected. Furthermore, if patient-specific knowledge can be incorporated into these predictions, then patient-specific optimization of radiation becomes a reality. Typically, researchers rely on mathematical and computational models to

evaluate such predictions, and many promising advancements have been made. In practice however, this has proven far easier said than done, since patient-specific measurements are often difficult and costly to obtain and can come with significant uncertainty. Furthermore, the sheer number of physical and biological factors at play in predicting treatment outcomes makes accurate mathematical predictions challenging.

Many previous works have examined different ways to optimize radiation therapy including temporally [1, 5, 16, 31, 33, 59, 129], spatially [2, 51, 72, 112], and patient scheduling [25]. Generally, these studies perform mathematical optimizations, occurring in four steps: first, select metrics; second, determine constraints; third, write governing model of treatment response; and fourth, perform mathematical optimization. When discussing radiation, several different metrics can be considered. The most common is to maximize cell kill, but others such as maximizing Tumour Control Probability (TCP), minimizing administered dose, and minimizing tissue complication probability are also commonly considered. Some studies consider multiple metrics, typically by forming a combined metric with weight coefficients. Constraints for the mathematical problem are chosen based on the physical and biological limitations of the situation. In some cases, this means limiting the total administered dose to a prescribed amount over the course of a day or a full treatment period. In other cases, constraints include timing considerations such as prohibiting radiation given on weekends. Models of treatment response typically involve ODEs or PDEs and use established models of radiation effect such as the LQ model. These models determine the effect of a chosen candidate treatment with respect to the chosen metric. Once the metric, constraints, and model have been fixed, a mathematical optimization can be performed where an optimal is selected from the many candidate treatments.

In this paper, we consider the problem of finding the temporal distribution of radiation dose which minimizes the total remaining number of cells at the end of treatment under a set of clinically-relevant constraints. Several previous works have examined similar questions. In Wein et al [129], the authors considered a spatially-varying population of homogeneous tumour cells and used dynamic reprogrammming to find the radiation schedule that optimizes the TCP under constraints to the Biologically Effective Dose (BED). They found that allowing the fractional dose amount to change over the course of treatment allowed for a higher TCP than the clinically-standard constant dose rate. Altman et al [5] obtained a similar result, showing that when maximizing cell kill over a homogeneous population of tumour cells subject to constrained total dose, the optimal distribution of radiation was again when the dose rate was allowed to vary throughout treatment. In Kim et al [51], the authors considered maximizing the BED while constraining the dose for nearby Organs at Risk (OAR)s. They argued that due to recent technological advancements allowing more precise localization of radiation dose on a tumour core, treatment

optimizations can focus primarily on maximizing anticancer effect. Even when including spatial effects in their optimization, the authors found that varying the dose per fraction could lead to an improvement over uniform schedules. Some analytical results were derived in Mizuta et al [80] related to optimal temporal dose distributions where cell kill was to be optimized and BED to an OAR minimized. They provided a theorem stating that in cases without tumour repolulation, if the OAR is sufficiently sensitive, then the optimal dose distribution is uniform over the treatment length, but if the OAR is not sufficiently sensitive, then the optimal distribution is to focus the allowable radiation into the least number of fractions possible. Bortfeld et al [16] built off of the work of Mizuta et al by including the effect of tumour repopulation. They similarly showed that in cases where the OAR is less sensitive, the optimal distribution was to focus radiation into the smallest number of fractions possible. Through their analysis, they were further able to conclude that for tumours with faster growth rates, the optimal distribution of radiation tended to be spread over a shorter time.

Despite the important differences between the above works with respect their particular problem formulation, they all arrive at similar conclusions. Specifically, that a nonuniform fractional dose distribution can outperform the clinically-standard uniform distribution, and that the optimal case tends to involve focusing allowable dose over shorter periods of time at higher dose rates. Clearly, such results are of tremendous clinical importance. However, the applicability of these works is somewhat lessened by the omission of a key feature of tumours: cellular heterogeneity. Each of the above works considers homogeneous tumours in which all cells proliferate, invade, and respond to treatment in the same way. In reality, tumours are heterogeneous and contain many different types of cells which can act in different ways. Of particular importance, CSCs are a type of tumour cell capable of self-renewal and differentiation. They are thought to represent a small subpopulation of the full tumour, yet nevertheless be the primary drivers of tumour growth and repopulation. Countless recent studies have noted their importance in understanding cancer progression and treatment ([7, 10], for example). Studies have also noted experimental techniques which could be used to identify stem-like subpopulations, relying typically on protein biomarkers [85]. Despite the imperfect natures of these biomarker methods, differences between biomarker positive and biomarker negative cells can be observed. With respect to radiation therapy, studies have examined the difference in response to radiation by CSCs and non-CSCs [57, 92] (or biomarker $+/-$ cells), finding that CSCs are commonly less sensitive to radiotherapy than their non-CSC counterparts. Furthermore, they observe that post-radiation tumours tend to have become less heterogeneous as sensitive cells are removed by radiation and all that remains are the resistant stem cells. Such observations suggest that considering a heterogeneous tumour population may be a crucial feature in

developing accurate models of radiation optimization. Indeed, many recent works have done just this.

Leder et al [59] considered a mathematical model with two distinct cell types: a radioresistant CSC-like cell and a radiosensitive non-CSC-like cell. Using a two compartment mathematical model along with a Monte Carlo optimization method, they found that mathematically optimized radiation schedules led to a longer survival time and greater reduction in tumour volume than the standard schedule in glioblastoma-bearing mice. Notably, their optimal schedule no longer had the qualitative behaviour observed in the one cell type cases described above. In fact, the hypofractionated (less fractions at higher dose rates) case performed the worst out of the four tested. In Forouzannia et al [31], the authors also considered a two cell type model and included plasticity rates between their CSC and non-CSC compartments. They considered five different schedules of radiation, and sought to find which schedules led to the minimum remaining cells and to find the resulting fraction of CSCs with respect to the full tumour population. Their results were similar to [59] in that neither the standard of care nor hypofractionation was optimal. In contrast however, Galochkina et al [33] also considered a two cell type model but arrived at the opposite result of [31] and [59]. They found that the uniform distribution of dose was typically optimal, and that cases with different optima generally only provided insignificant benefits. However, in [33]'s two cell model, the only difference between the cell types was their replenishment rates - importantly, they were assumed to be affected by radiotherapy in the same way. Given the differing conclusions of these models, it would be reasonable to suspect that a key force in determining the qualitative nature of an optimal radiation schedule is the difference in radiation responses between the cell types as opposed to differences in repopulation or plasticity rates. In this work, we set out to answer this question and determine which situations lead to qualitatively different optimal radiation schedules.

We begin the methods section by detailing the two cell type model that we use to determine the effect of radiation therapy. We then explain our optimization problem, including choice of metric and solution constraints. We provide some important analytical analysis of our model and its possible solutions, specifically proving that solutions to our model is typically 'bang-bang' in nature. Finally, we explain the projected gradient descent optimization technique that we employ to find the optimal solution. In the results section, we provide some example solutions of the optimization problem and distinguish between the two different qualitative categories into which solutions fall. We then provide analysis using tree classifiers of the key parameters which determine the behaviour of results, connect this analysis to previous literature, and discuss the clinical relevance. In the conclusion, the work is summarized, and directions for future research are noted.

5.2 Methods

5.2.1 Model

In this work, we consider a tumour consisting of two distinct populations of cells, denoted by n_1 and n_2. We use a system of two ODEs, shown in equations (5.1) and (5.2) and in Figure 5.1, to govern the time evolution of the cell populations.

$$\frac{dn_1}{dt} = r_1 n_1 \left(1 - \frac{n_1 + n_2}{n_{max}}\right) - n_1 f_1(t) - (\gamma_1 n_1 - \gamma_2 n_2), \tag{5.1}$$

$$\frac{dn_2}{dt} = r_2 n_2 \left(1 - \frac{n_1 + n_2}{n_{max}}\right) - n_2 f_2(t) + (\gamma_1 n_1 - \gamma_2 n_2). \tag{5.2}$$

Throughout this paper it is understood that this system is equipped with initial conditions at $t = 0$, so $n_1(0)$ and $n_2(0)$ are known.

Each cell type is assumed to proliferate according to a combined logistic growth law which is shown in the first term of each equation. The parameter n_{max} is the cell carrying capacity: the maximum number of cells that the tumour can sustain. As the total number of cells, $n_1 + n_2$, approaches n_{max}, the rate of increase of the cell population decreases. The proliferation rate parameters r_1 and r_2 differentiate the rates of growth between the cell types, with higher values corresponding to faster growth. When the total cell population is low compared to n_{max}, r_1 and r_2 are equivalent to exponential growth rates. The final term in each equation describes the plasticity between the cell types: cells of type 1 turning into cells of type 2 and the converse. Interestingly, recent works have show the bidirectional nature of plasticity between CSCs and non-CSCs [19, 68, 125]. For this reason, we include bidirectional plasticity in our model with respective plasticity rates of γ_1 (type 1 to type 2) and γ_2 (type 2 to type 1). The middle term of each equation describes the effect of radiotherapy on the cells, which is assumed to be governed by the well-known LQ model and is included in the functions $f_1(t)$ and $f_2(t)$ in its differential form. In our model, we assume that radiation is applied in a finite number of fractions and denote this number as N. We denote the time midpoint of fraction number i as τ_i, its length by $2\omega_i$, and it dose rate (in Gy/day) by a_i. With this formulation, we can write the functions $f_1(t)$ and $f_2(t)$

as shown in equations (5.3) and (5.4).

$$f_1(t) = \sum_{i=1}^{N} \left(\alpha_1 a_i + 2\beta_1 a_i^2 (t - \tau_i + \omega_i)\right) \varphi\left(\frac{t-\tau_i}{\omega_i}\right), \tag{5.3}$$

$$f_2(t) = \sum_{i=1}^{N} \left(\alpha_2 a_i + 2\beta_2 a_i^2 (t - \tau_i + \omega_i)\right) \varphi\left(\frac{t-\tau_i}{\omega_i}\right). \tag{5.4}$$

Note that $\varphi\left(\frac{t-\tau_i}{\omega_i}\right)$ is a rectangular window function centred at time τ_i with width $2\omega_i$ given by

$$\varphi\left(\frac{t-\tau_i}{\omega_i}\right) = \begin{cases} 1, & \text{if } |t-\tau_i| < \omega_i, \\ 0, & \text{otherwise.} \end{cases}$$

This defines the dose rate as an on-again-off-again function: a_i when radiation is being applied and zero when it's not. Also note the different values of the radiobiological parameters α_k and β_k for the different cell types. Higher values of these parameters mean that the cell type has a higher sensitivity to radiation.

A radiation treatment schedule is therefore fully characterized by the three N-dimensional vectors $a = (a_1, \ldots, a_N)$ (fraction dose rates), $\tau = (\tau_1, \ldots, \tau_N)$ (fraction time midpoints), and $\omega = (\omega_1, \ldots, \omega_N)$ (fraction half lengths). In this work, we are primarily interested in optimizing the distribution of dose over a treatment period, so we make two key assumptions on candidate radiation schedules. First, we assume that the length of each fraction is the same, meaning $\omega_i = \omega_j, \ \forall\ i, j$. Second, we assume that the fraction midpoints are fixed in time, once per day at the same time each day, meaning $\tau_{i+1} = \tau_i + 1$ day, $\forall\ i$. Note that this does not mean that radiation is administered every day: if $a_i = 0$, then no radiation will be given on day i. We seek to find the distribution of dose over the possible times, determined by the selection of a alone, that leads to the maximum cell kill. For any choice of a, the model can be solved to yield a population vs. time plot for each cell type and the total, incorporating the effect of radiation.

As one might expect, without constraints on the applied radiation, the optimal cell kill would result from simply increasing the dose rate of each fraction to its maximal level. But of course, in clinical practice, there are various constraints that prevent this from being a reasonable solution. In our optimization, we consider two constraints on our fractional dose rate schedule: total dose and maximal dose. We first insist that the total dose administered over all fractions is less than or equal to the total allowable amount, and for maximal dose,

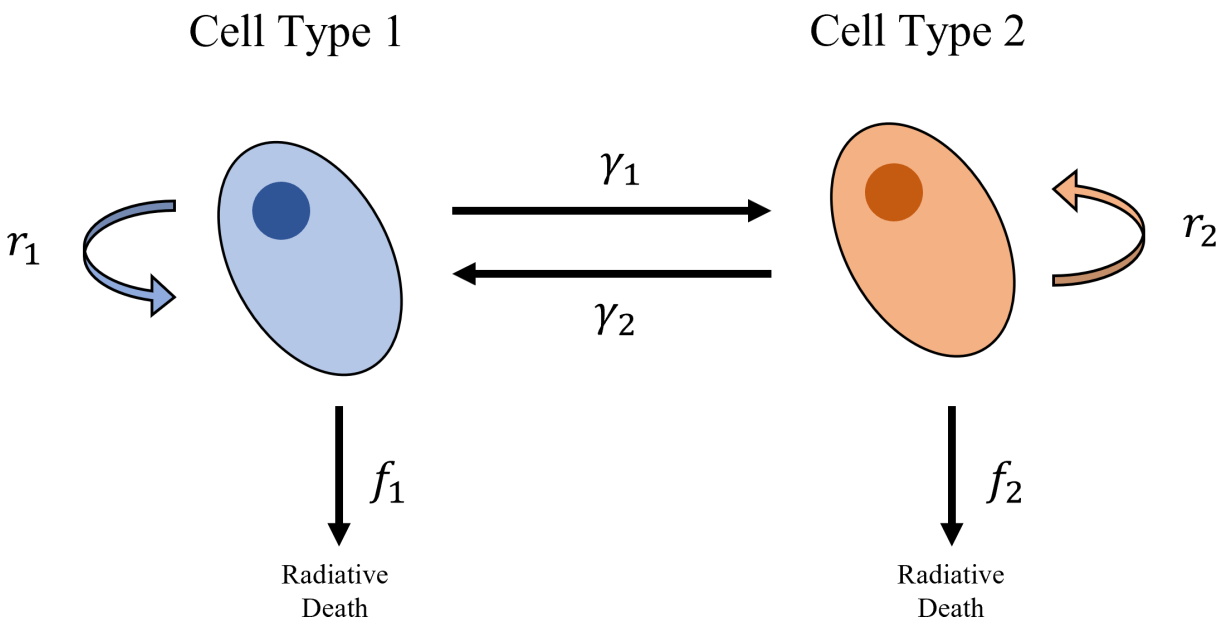

Figure 5.1: Schematic of the two cell type compartmental ODE model in equations (5.1) and (5.2). Each cell type is assumed to proliferate according to a logistic growth law with proliferation rate r_i. The plasticity rates between the cells types are given by γ_i. Cells are killed by radiation through the functions f_i which depend on the radiobiological parameters α_k and β_k.

we simply impose an upper bound on the dose rate given over each fraction, which read as

$$\sum_{i=1}^{N} 2a_i\omega =: \sum_{i=1}^{N} D_i \quad \leq \quad D_{tot}, \tag{5.5}$$

$$0 \leq 2a_i\omega =: D_i \quad \leq \quad D_{max}, \quad \forall i = 1, \ldots, N. \tag{5.6}$$

Clinically, the constraint (5.5) is applied to prevent the patient from receiving dangerous levels of radiation beyond what is necessary for treatment. We expect the inequality in (5.5) to be saturated in the optimal case. The clinical rationale for (5.6) is because of the different response times of healthy and cancerous tissues. As healthy tissues generally have a higher capacity to repair DNA damage, dose fractionation can limit the damage to healthy tissues while still incurring high damage to diseased tissues. Without this constraint, the model would be free to converge to a single hit of radiation at a dangerously high dose rate, removing the fractionation altogether. Here, we will exclusively consider cases where $ND_{max} > D_{tot}$, meaning that applying the maximal dose per fraction will result in surpassing the limit in total dose as these are the clinically relevant cases and the ones which produce interesting mathematical results. In cases where $ND_{max} \leq D_{tot}$, the optimal result is trivially to apply D_{max} at every fraction.

We therefore seek to answer the following question. For a tumour consisting of two distinct subpopulations of cancer cells and a fractionated dose schedule constrained by total dose and maximal dose, what choice of radiation dose rates, a, result in the minimal population of total cells remaining after treatment. In our analysis, we generalize to minimizing a weighted sum of cells, $e = p_1 n_1(T) + p_2 n_2(T)$ for the end of treatment time, T. Mathematically, we seek to find

$$a^* = \arg\min \{e(a),\ a \in \mathcal{A}\}, \tag{5.7}$$

where $\mathcal{A}$ is the vector space of fractional dose rates $a = (a_1, ..., a_N) \in \mathbb{R}^N$ ($a_i \geq 0\ \forall i$) satisfying (5.5) and (5.6).

5.2.2 Bang-Bang Structure of the Optimal Radiation Treatment

It is known, see for example [20], that the optimal distribution of resources in population models is of bang-bang type, i.e. the most favorable distribution of resources for a species to survive is when the resources are distributed spatially with patches of maximal or minimal values. If we consider our variable a as similar to the growth rate variable m in [20], our result represents a time extension of the bang-bang type result of the optimal solution.

In this section, we will show that modulo a subset where the gradient is constant, the solution to (5.7) is of bang-bang type, i.e. the components of optimal radiation $a^* = (a_1^*, \ldots, a_N^*)$ are of three types: i) they are maximal equal to D_{max}, ii) they are minimal equal to 0, or iii) they are such that derivative of the energy e on the direction of those components is constant. In our numerical experiments, we always found that the optimal solution a^* is purely of bang-bang type, i.e. the set of components of a^* where the energy is constant is empty.

Note that the solution of (5.1)-(5.2) is understood in the following sense. We say (n_1, n_2) solves (5.1)-(5.2) if $n_1, n_2 \in C^0([0,T]) \cap W^{1,\infty}(0,T)$ and

$$n_1(t) = n_1(0) + \int_0^t \left(r_1 n_1 \left(1 - \frac{n_1 + n_2}{n_{max}} \right) - n_1 f_1(t) - (\gamma_1 n_1 - \gamma_2 n_2) \right) dt, \tag{5.8}$$

$$n_2(t) = n_2(0) + \int_0^t \left(r_2 n_2 \left(1 - \frac{n_1 + n_2}{n_{max}} \right) - n_2 f_2(t) + (\gamma_1 n_1 - \gamma_2 n_2) \right) dt, \quad t \in (0,T). \tag{5.9}$$

We start with an a priori estimate for the solution (n_1, n_2) of (5.1)-(5.2).

Proposition 5.2.1. *Let $F(t, n_1, n_2) = (F_1(t, n_1, n_2), F_2(t, n_1, n_2))$ with*

$$F_1(t, n_1, n_2) = r_1 n_1 \left(1 - \frac{n_1 + n_2}{n_{max}} \right) - n_1 f_1(t) - (\gamma_1 n_1 - \gamma_2 n_2), \tag{5.10}$$

$$F_2(t, n_1, n_2) = r_2 n_2 \left(1 - \frac{n_1 + n_2}{n_{max}} \right) - n_2 f_2(t) + (\gamma_1 n_1 - \gamma_2 n_2). \tag{5.11}$$

Assume (5.1)-(5.2) has a solution $(n_1, n_2) \in W^{1,\infty}((0,T); \mathbb{R}^2)$. If $n_1(0) \geq 0$, $n_2(0) \geq 0$ then

$$0 \leq n_1 \leq \min \left\{ 1, \max \left\{ 1 - \gamma_1, \frac{\gamma_2}{r_1} \right\} \right\} n_{max} := n_{1,max}, \tag{5.12}$$

$$0 \leq n_2 \leq \min \left\{ 1, \max \left\{ 1 - \gamma_2, \frac{\gamma_1}{r_2} \right\} \right\} n_{max} := n_{2,max}, \tag{5.13}$$

$$0 \leq n_1 + n_2 \leq n_{max}. \tag{5.14}$$

Proof. Note that $F_1(t, 0, n_2) \geq 0$, $F_2(t, n_1, 0) \geq 0$ for all $n_1 \geq 0$, $n_2 \geq 0$. Based on, for example, [93, Lemma 1.1], non negativity of the initial conditions implies $n_1 \geq 0$, $n_2 \geq 0$

at all times given the non-negativity of the initial conditions. Furthermore,

$$\begin{aligned} F_1(t,n_1,n_2) &\leq r_1\left(1-\gamma_1-\frac{n_1}{n_{max}}\right)n_1+n_2\left(\gamma_2-r_1\frac{n_1}{n_{max}}\right) \\ &\leq 0, \qquad \forall\ n_1 \geq \max\left\{1-\gamma_1,\frac{\gamma_2}{r_1}\right\}n_{max}, \\ F_2(t,n_1,n_2) &\leq r_2\left(1-\gamma_2-\frac{n_2}{n_{max}}\right)n_2+n_1\left(\gamma_1-r_2\frac{n_2}{n_{max}}\right) \\ &\leq 0, \qquad \forall\ n_2 \geq \max\left\{1-\gamma_2,\frac{\gamma_1}{r_2}\right\}n_{max}. \end{aligned}$$

This implies $n_1 \leq \max\left\{1-\gamma_1,\frac{\gamma_2}{r_1}\right\}n_{max}$, $n_2 \leq \max\left\{1-\gamma_2,\frac{\gamma_1}{r_2}\right\}n_{max}$. Finally, we have

$$(n_1+n_2)' = (r_1n_1+r_2n_2)\left(1-\frac{n_1+n_2}{n_{max}}\right)-(f_1n_1+f_2n_2).$$

Considering the right hand side of the previous equation as a function of n_1+n_2, we see that it is negative as soon as $n_1+n_2=n_{max}$, which implies $n_1+n_2 \leq n_{max}$, whence in particular $n_1 \leq n_{max}$ and $n_2 \leq n_{max}$. □

Theorem 5.2.2. *For every $a \in \mathbb{R}^N$ the problem (5.1)-(5.2) has a unique solution $(n_1,n_2) \in W^{1,\infty}(\mathbb{R};\mathbb{R}^2)$, where in general $W^{1,\infty}((\alpha,\beta);\mathbb{R}^2)$ is the Sobolev space of continuous functions in $[\alpha,\beta]$ having almost everywhere bounded derivatives in (α,β).*

Proof. As f_1 and f_2 are discontinuous and bounded, the right space to look for the solution (n_1,n_2) is $W^{1,\infty}$ rather than commonly used space C^1.

Note that F is a Carathéodory function and from [23, Theorem 1.1, Chapter 2], there exists a local in time absolutely continuous solution in a certain interval $(-\alpha_0,\alpha_0)$. Taking into account the form of F, which is C^∞ in (n_1,n_2) and L^∞ in t, it implies that the local solution is in $W^{1,\infty}((-\alpha_0,\alpha_0);\mathbb{R}^2)$. As the solution remains bounded, see Proposition 5.2.1, it extends to $W^{1,\infty}(\mathbb{R};\mathbb{R}^2)$.

For the uniqueness we note that for every two solutions $n=(n_1,n_2)$, $\tilde{n}=(\tilde{n}_1,\tilde{n}_2)$ the function F satisfies $|F(t,n_1,n_2)-F(t,\tilde{n}_1,\tilde{n}_1)| \leq C(|n_1-\tilde{n}_1|+|n_2-\tilde{n}_2|)$, because n and $\tilde{n}$ are bounded. Then the uniqueness follows from Gronwall inequality.

□

Proposition 5.2.3. *The maps $N_1 : a \in \mathbb{R}^N \mapsto N_1(a) \in W^{1,\infty}(0,T)$, with $N_1(a) = n_1(\cdot;a)$, and $N_2 : a \in \mathbb{R}^N \mapsto N_2(a) \in W^{1,\infty}(0,T)$, with $N_2(a) = n_2(\cdot;a)$, are C^∞ from $\mathbb{R}^N$ into $W^{1,\infty}(0,T)$. Their first derivatives wrt a_i, denoted by $\partial_{a_i} n_1'$ and $\partial_{a_i} n_2'$ respectively, satisfy*

$$\begin{aligned}
\partial_{a_i} n_1' &= \left(r_1\left(1 - \frac{2n_1+n_2}{n_{max}}\right) - f_1 - \gamma_1\right)\partial_{a_i} n_1 + \left(\gamma_2 - r_1\frac{n_1}{n_{max}}\right)\partial_{a_i} n_2 - n_1 \partial_{a_i} f_1 \\
&=: G_1(t; \partial_{a_i} n_1, \partial_{a_i} n_2), \qquad (5.15)\\
\partial_{a_i} n_1(0) &= 0, \qquad (5.16)\\
\partial_{a_i} n_2' &= \left(\gamma_1 - r_2\frac{n_2}{n_{max}}\right)\partial_{a_i} n_1 + \left(r_2\left(1 - \frac{n_1+2n_2}{n_{max}}\right) - f_2 - \gamma_2\right)\partial_{a_i} n_2 - n_2 \partial_{a_i} f_2 \\
&=: G_2(t; \partial_{a_i} n_1, \partial_{a_i} n_2), \qquad (5.17)\\
\partial_{a_i} n_2(0) &= 0. \qquad (5.18)
\end{aligned}$$

where the prime notation refers to differentiation with respect to time. As a consequence, $n_1(T;a)$, $n_2(T;a)$ and $e(a)$ are C^∞ with respect to a and

$$\partial_{a_i} e(a) = p_1 \partial_{a_i} n_1(T;a) + p_2 \partial_{a_i} n_2(T;a). \qquad (5.19)$$

Proof. First we show that n_1 and n_2 are C^∞ with respect to a in $W^{1,\infty}(0,T)$. This can be done easily by using implicit function theorem as follows. Let $H = (H_1, H_2) : \mathbb{R}^N \times W^{1,\infty}((0,T);\mathbb{R}^2) \mapsto W^{1,\infty}((0,T),\mathbb{R}^2)$ be given as follows: for $a \in \mathbb{R}^N$ and $v = (v_1, v_2) \in W^{1,\infty}((0,T);\mathbb{R}^2)$ we define

$$H_1(a,v)(t) = v_1(t) - n_1(0) - \int_0^t \left(r_1 v_1\left(1 - \frac{v_1+v_2}{n_{max}}\right) - v_1 f_1(s;a) - (\gamma_1 v_1 - \gamma_2 v_2)\right) ds, \qquad (5.20)$$

$$H_2(a,v)(t) = v_2(t) - n_2(0) - \int_0^t \left(r_2 v_2\left(1 - \frac{v_1+v_2}{n_{max}}\right) - v_2 f_2(s;a) + (\gamma_1 v_1 - \gamma_2 v_2)\right) ds, \quad t \in (0,T). \qquad (5.21)$$

Now let $a^0 \in \mathbb{R}^N$ be fixed. We have $H(a^0, n(\cdot;a^0)) = 0$ and H is C^∞ wrt (a, v_1, v_2). The derivative of H wrt to v at $(a^0; n(\cdot;a^0))$ in direction w is given by

$$\begin{aligned}
\partial_v H_1(a^0, v; w)(t) &= w_1(t) - \int_0^t \left(\left[r_1\left(1 - \frac{2v_1+v_2}{n_{max}}\right) - f_1(t) - \gamma_1\right] w_1 + \left[\gamma_2 - r_1\frac{v_1}{n_{max}}\right] w_2\right) dt,\\
\partial_v H_2(a^0, v; w)(t) &= w_2(t) - \int_0^t \left(\left[\gamma_1 - r_2\frac{v_2}{n_{max}}\right] w_1 + \left[r_2\left(1 - \frac{v_1+2v_2}{n_{max}}\right) - f_2(t) - \gamma_2\right] w_2\right) dt.
\end{aligned}$$

Like for the existence of (n_1, n_2) in Theorem 5.2.2, it is easy to show that $\partial_v H(a^0; n(\cdot; a^0))$ defines an isomorphism from $W^{1,\infty}((0,T);\mathbb{R}^2)$ to itself. The implicit function theorem implies the existence of a C^∞ map $z = (z_1, z_2)$ from an open neighbourhood $A_0 \subset \mathbb{R}^N$ of a^0 into $W^{1,\infty}((0,T);\mathbb{R}^2)$ such that $H(a, z(\cdot; a)) = 0$ for all a in A_0. The uniqueness of solution for problem (5.1)-(5.2) implies $N_i(a) = n_i(\cdot; a) = z_i(\cdot; a)$, $i = 1, 2$, which proves the C^∞ differentiability of $n_i(\cdot; a)$.

The C^∞ differentiability wrt a of n in $W^{1,\infty}((0,T);\mathbb{R}^2)$ and the continuous embedding of $W^{1,\infty}((0,T);\mathbb{R}^2)$ in $C^0([0,T];\mathbb{R}^2)$ imply also the C^∞ differentiability wrt a of $n(T; a)$ and so of $e(a)$. Finally, the formulas (5.15)-(5.18) and (5.19) are proved easily from direct calculus. □

Theorem 5.2.4. *The minimization problem (5.7) has a solution $a^* \in \mathcal{A}$.*

Proof. We note first that (5.5) and (5.6) imply that $\mathcal{A}$ is a compact set. Indeed, clearly $\mathcal{A}$ is bounded. As the inequalities in (5.5) and (5.6) are not strict, we get that every limit point a^* of $\mathcal{A}$ satisfies (5.5) and (5.6), so $a^* \in \mathcal{A}$.

Now let (a^k), $a^k \in \mathcal{A}$, be a minimization sequence of (5.7). As $\mathcal{A}$ is bounded, necessarily (a^k) is a bounded sequence in $\mathbb{R}^N$ and then from the compactness of $\mathcal{A}$ there exists a subsequence of (a^k), still denoted (a^k), converging to a certain $a^* \in \mathcal{A}$.

From Proposition 5.2.1, with (a^k) instead of a, we have that (n_i^k), $i = 1, 2$, where $n_i^k = n_i(t; a^k)$, are bounded in $W^{1,\infty}(0,T)$. From classical Sobolev embeddings, we have that $W^{1,\infty}(0,T)$ is compactly embedded in $C^0([0,T])$, and therefore there exists a subsequence of $(n_i^k)_k$, $i = 1, 2$, still denoted (n_i^k), converging to n_i^* in $C^0([0,T])$.

As a result we get a sequence (a^k) in $\mathcal{A}$ converging to $a^* \in \mathcal{A}$, and the corresponding solutions n_i^k, $i = 1, 2$, to (5.8)-(5.9) converging in $C^0([0,T])$ to a certain function $n_i^* \in C^0([0,T])$, $i = 1, 2$

It implies that $n_i^* = n_i(t; a^*)$, $i = 1, 2$, because by replacing a^k, n_i^k, $i = 1, 2$, in (5.8)-(5.9) and passing in limit shows that n_i^*, $i = 1, 2$, solve (5.8)-(5.9) with $a = a^*$, which from the uniqueness of the solution to (5.8)-(5.9) implies $n_i^* = n_i(t; a^*)$, $i = 1, 2$. As $e(a)$ is continuous wrt a it follows that a^* solves (5.7). □

Lemma 5.2.5. *Let $a^* = (a_1^*, .., a_N^*)$ be a solution of (5.7). For all $i = 1, \ldots, N$ and all $t \in [0,T]$ we have $\partial_{a_i} n_1(t; a^*) \leq 0$ and $\partial_{a_i} n_2(t; a^*) \leq 0$, and therefore $\partial_{a_i} e(a^*) \leq 0$.*

Proof. Note that as n_1 and n_2 are differentiable in $W^{1,\infty}(0,T)$ implies that they are differentiable in $C^0([0,T])$. Using the boundedness of n_1 and n_2, see Proposition 5.2.1, we get

$$G_1(t; 0, y) \leq 0, \quad \forall y \leq 0, \qquad G_2(t, x, 0) \leq 0, \quad \forall x \leq 0,$$

which in combination with (5.15)-(5.18) implies $\partial_{a_i} n_1(t;a^*) \le 0$ and $\partial_{a_i} n_2(t;a^*) \le 0$, and therefore $\partial_{a_i} e(a^*) \le 0$, which proves the lemma. □

Remark 5.2.6. *This lemma suggests that the solution a^* should be on the boundary of $\mathcal{A}$. The following theorem shows that indeed $a^* \in \partial\mathcal{A}$ and typically a^* is of bang-bang type.*

Theorem 5.2.7. *Assume $m := \dfrac{D_{tot}}{D_{max}} \in \mathbb{N}$. Let $a^* = [a_1^*, \dots, a_N^*]$ be a solution of (5.7) and*

$$\partial_{a_{i_1}} e(a^*) \le \cdots \le \partial_{a_{i_m}} e(a^*) \le \partial_{a_{i_{m+1}}} e(a^*) \le \cdots \le \partial_{a_{i_N}} e(a^*) \le 0$$

be an increasing ordering of components of $(\partial_{a_i} e(a^*))$. We have

i) If $\partial_{a_{i_m}} e(a^*) < \partial_{a_{i_{m+1}}} e(a^*)$ then the solution a^* is of bang-bang type. More precisely

$$a_{i_1}^* = \cdots = a_{i_m}^* = \frac{D_{max}}{2\omega}, \qquad a_{i_{m+1}}^* = \cdots = a_{i_N}^* e(a^*) = 0.$$

ii) Otherwise

$$\partial_{a_{i_1}} e(a^*) \le \cdots \le \partial_{a_{k-1}} e(a^*) < \partial_{a_k} e(a^*) = \cdots = \partial_{a_{i_\ell}} e(a^*) < \partial_{a_{i_{\ell+1}}} e(a^*) \le \cdots \le \partial_{a_{i_N}} e(a^*)$$

for certain k, ℓ integers such that $1 \le k \le m < \ell \le N$, and

$$a_1^* = \cdots = a_{k-1}^* = \frac{D_{max}}{2\omega}, \qquad \partial_{a_k} e(a^*) = \cdots = \partial_{a_\ell} e(a^*), \qquad a_{\ell+1}^* = \cdots = a_N^* = 0.$$

Proof. We note that from Lemma 5.2.5, $\partial_{a_i} e(a^*) < 0$ for all $1 \le i \le m$ in the case i) and $\partial_{a_i} e(a^*) < 0$ for all $1 \le i \le \ell$ in the case ii).

Let us first prove i). We assume that the claim does not hold. Then necessarily there exists: i.1) an integer i with $1 \le i \le m$ and $a_i^* < \frac{D_{max}}{2\omega}$, or i.2) an integer j with $m < j \le N$ such that $a_j^* > 0$.

In the case i.1), if all $a_j^* = 0$, $m < j \le N$, we consider $z = (0, \dots, 0, 1, 0, \dots, 0)$, where 1 is on $i$the place and $a = a^* + hz$, $h > 0$ small. Then $a \in \mathcal{A}$ and

$$e(a^* + hz) - e(a^*) \approx h\partial_{a_i} e(a^*) < 0,$$

which is a contradiction and proves that this case cannot happen. Otherwise, if $a_j^* > 0$ for any integer j, $m < j \le N$, we consider $z = (0, \dots, 0, 1, 0, \dots, 0, -1, 0, \dots, 0)$, where 1 is on $i$the place and -1 in jth place. Then $a = a^* + hz \in \mathcal{A}$ for $h > 0$ small and

$$e(a^* + hz) - e(a^*) \approx h(\partial_{a_i} e(a^*) - \partial_{a_j} e(a^*)) < 0,$$

which is again a contradiction and shows that this case does not happens.

In the case i.2), from the constraint (5.5) necessarily there exists an integer i, $1 \leq i \leq m$ with $a_i^* < \frac{D_{max}}{2\omega}$. Then we consider $z = (0, \ldots, 0, 1, 0, \ldots, 0, -1, 0, \ldots, 0)$, where 1 is on $i$the place and -1 in jth place and conclude as in the case i.2) above.

For the proof of ii), if $k = 1$ and $\ell = N$, there is nothing to prove - the vector $(\partial_{a_i} e(a^*))$ is constant. The cases $k > 1$ or $\ell < N$ are proved similarly as i) above. Indeed, if we assume that the claim does not hold then there exits: ii.1) an integer i, $1 \leq i \leq k$, with $a_i^* < \frac{D_{max}}{2\omega}$, or ii.2) an integer j, $\ell + 1 \leq j \leq N$, with $a_j^* > 0$.

In the case ii.1), resp. ii.2), we proceed as in i.1), resp. i.2), above and conclude that this case cannot happen. □

Remark 5.2.8. *Proposition 5.2.3 shows that e is a C^∞ function and therefore a^* is a critical point of e only if $\partial_{a_i} e(a^*) = 0$ for all i. Theorem 5.2.7 shows that a^* is a critical point of a only if case ii) of Theorem 5.2.7 occurs with $k = 1$ and $l = N$. In all our numerical experiments we have found optimal solution a^* only of the type i) of Theorem 5.2.7. We conjecture that the case ii) of Theorem 5.2.7 does not happen and therefore a^* is not a critical point of e.*

5.2.3 Numerical optimization

To solve (5.7), we can use a gradient descent method. Note that our minimization problem is constrained and the set $\mathcal{A}$ is convex. However, we cannot prove nor disprove whether the objective function $e(a)$ is convex (we suspect that it is not). We therefore address the issue of convergence by utilizing the multiple start strategy. Many initial conditions are used to begin the algorithm, and it is verified that each of these converges to the same optimal value. To satisfy the constraints for a, the constrained optimization problem (5.7) is solved with the projected gradient descent method (see [47]) as given by Algorithm 1. We note that a^{k+1} of the problem (5.22) below could also be solved by using a quadratic programming algorithm,

$$\min \left\{ (a - \alpha^{k+1})^t \cdot (a - \alpha^{k+1}), \quad a \in \mathcal{A} \right\}.$$

Computing (5.19) requires one solution of (5.8)-(5.9) for n_1 and n_2, and the solution of (5.15)-(5.18) for $\partial_{a_i} n_1$ and $\partial_{a_i} n_2$ for all $i = 1, \ldots, N$. In terms of computational cost this is equivalent to $N + 1$ solutions of (5.8)-(5.9) at each step of the gradient descent. We can

Algorithm 1: Gradient descent method

Input:

$\epsilon > 0$ — tolerance

$\delta > 0$ — step size (sufficiently small and possibly dependent on iteration number);

$a^0 \in \mathbb{R}^N_+$ — initial guess

Output:

a^k — approximation to the solution a^* of problem (5.7)

1: choose $\epsilon > 0$ small

2: set $k = 0$ and choose $a^k = (a_i^k)$; set $(\partial_{a_i} e(a^k)) = (0)$

3: **repeat**

4: update the parameters

$$b_i^{k+1} = a_i^k - \delta \partial_{a_i} e(a^k), \quad i = 1, \dots, N,$$

with $\delta > 0$ a small "learning" parameter

5: satisfy the constraints by projecting b^{k+1} in $\mathcal{A}$ and set a^{k+1} to its projection:

$$a^{k+1} = \text{argmin} \left\{ \frac{1}{2} |a - b^{k+1}\|^2, \ a \in \mathcal{A} \right\} \tag{5.22}$$

6: set $k = k + 1$

7: compute $n_1(t; a^k)$, $n_2(t; a^k)$ solution of (5.1)-(5.2)

8: compute $\partial_{a_i} n_1(T; a^k)$, $\partial_{a_i} n_2^k(T; a^k)$ by solving (5.15)-(5.18), and $\partial_{a_i} e(a^k)$ by using (5.19), for $i = 1, \dots, N$

9: **until** $|(\partial_{a_i} e(a^k))| < \epsilon$

reduce this computational cost by using appropriate adjoint functions as follows. This will reduce the computation cost per iteration to one solution of (5.8)-(5.9) for n_1 and n_2, and one solution of (5.23)-(5.26) for h_1 and h_2, see (5.27) (the computing cost of $\partial_{a_i} f_j$ is zero because it is computed explicitly). In terms of computational cost this is equivalent to two solutions of (5.8)-(5.9) per iteration.

Proposition 5.2.9. *Let $a \in \mathbb{R}^N$. Then the system of ODEs*

$$\begin{aligned}
h_1' &= -\left[r_1\left(1-\frac{2n_1+n_2}{n_{max}}\right)-f_1-\gamma_1\right]h_1-\left[\gamma_1-r_2\frac{n_2}{n_{max}}\right]h_2, && (5.23)\\
h_2' &= -\left[\gamma_2-r_1\frac{n_1}{n_{max}}\right]h_1-\left[r_2\left(1-\frac{n_1+2n_2}{n_{max}}\right)-f_2-\gamma_1\right]h_2, && (5.24)\\
h_1(T) &= p_1, && (5.25)\\
h_2(T) &= p_2, && (5.26)
\end{aligned}$$

has a unique solution $(h_1, h_2) \in W^{1,\infty}(((0,T);\mathbb{R}^2)$. Furthermore,

$$\partial_{a_i} e(a) = -\int_0^T (n_1 h_1 \partial_{a_i} f_1 + n_2 h_2 \partial_{a_i} f_2)dt. \tag{5.27}$$

Proof. By changing the variable $s = T - s$, $s \in (0,T)$ and setting $g_i(s) = h_i(T-s)$, $i = 1,2$, we find that (g_1, g_2) solves a system of (linear) ODEs similar to (5.23)-(5.26) but with initial conditions at $s = 0$. Then we prove the existence and solution of (g_1, g_2), and so of (h_1, h_2), by proceeding as in Theorem 5.2.2 for (g_1, g_2) instead of (n_1, n_2).

Now we prove formula (5.27). Note that $\partial_{a_i} n_1$ and $\partial_{a_i} n_2$ solve (5.15)-(5.18). Then

$$\begin{aligned}
\partial_{a_i} e(a) &= p_1 \partial_{a_i} n_1(a;T) + p_2 \partial_{a_i} n_2(T;a)\\
&= h_1(T)\partial_{a_i} n_1(a;T) + h_2(T)\partial_{a_i} n_2(T;a)\\
&= \int_0^T (h_1 \partial_{a_i} n_1 + h_2 p_2 \partial_{a_i} n_2)' dt\\
&= \int_0^T (u_{1,i} h_1' + u_{2,i} h_2') + (u_{1,i}' h_1 + u_{2,i}' h_2))dt, && (5.28)
\end{aligned}$$

where for simplicity we wrote $u_{1,i} = \partial a_i n_1$, $u_{2,i} = \partial_{a_i} n_2$. Using (5.15)-(5.18) we can evaluate

$u'_{1,i}h_1 + u'_{2,i}h_2$ as follows

$$\begin{aligned}
u'_{1,i}h_1 + u'_{2,i}h_2 &= \left(r_1\left(1 - \frac{2n_1+n_2}{n_{max}}\right) - f_1 - \gamma_1\right)u_{1,i}h_1 + \left(\gamma_2 - r_1\frac{n_1}{n_{max}}\right)u_{2,i}h_1 \\
&- n_1h_1\partial_{a_i}f_1 \\
&+ \left(r_2\left(1 - \frac{n_1+2n_2}{n_{max}}\right) - f_2 - \gamma_2\right)u_{2,i}h_2 + \left(\gamma_1 - r_2\frac{n_2}{n_{max}}\right)u_{1,i}h_2 \\
&- n_2h_2\partial_{a_i}f_2) \\
&= \left(\left[r_1\left(1 - \frac{2n_1+n_2}{n_{max}}\right) - f_1 - \gamma_1\right]h_1 + \left[\gamma_1 - r_2\frac{n_2}{n_{max}}\right]h_2\right)u_{1,i} \\
&+ \left(\left[\gamma_2 - r_1\frac{n_1}{n_{max}}\right]h_1 + \left[r_2\left(1 - \frac{n_1+2n_2}{n_{max}}\right) - f_2 - \gamma_1\right]h_2\right)u_{2,i} \\
&- (n_1h_1\partial_{a_i}f_1 + n_2h_2\partial_{a_i}f_2) \\
&= -(u_{1,i}h'_1 + u_{2,i}h'_2) - (n_1h_1\partial_{a_i}f_1 + n_2h_2\partial_{a_i}f_2),
\end{aligned}$$

which in combination with (5.28) proves (5.27). □

Based on (5.27) we amend the Algorithm 1 as follows:

Algorithm 2: Gradient descent method amended

Input:

the same as in Algorithm 1

Output:

the same as in Algorithm 1

1: steps 1-2 the same as in Algorithm 1
2: **repeat**
3: steps 1-7 the same as in Algorithm 1
4: step 8 replaced by: compute h_1, h_2 by using (5.23)-(5.26) and $\partial_{a_i}e(a^k)$ by using (5.27), for $i = 1, \ldots, N$
5: **until** $|(\partial_{a_i}e(a^k))| < \epsilon$

This optimization procedure is implemented in python. To solve the system of ODEs, the `lsoda` method is used, and to compute the projected gradient, the `solve_ls` function from `qpsolvers` is used.

5.3 Results

As the set of variables to consider is large, we make several assumptions on the radiation schedule before attempting optimization. We choose $N = 30$, $D_{max} = 3\text{Gy}$, and $D_{tot} = 60\text{Gy}$ and consider the end of the treatment period to be the end of the 31^{st} day. This choice is easily changeable, but fixing it allows us to perform analysis. Since $ND_{max} > D_{tot}$, the optimal solution cannot simply be to apply maximum radiation at every fraction. Additionally, since a nondimensionalization of cell number would eliminate the parameter n_{max}, we set $n_{max} = 1$ for convenience - the results of the model are independent of the choice of n_{max}. Additionally, tumour cells typically have a known value for their ratio of radiobiological parameters, the α/β ratio. This ratio is frequently cited as being in the range 3-10 Gy, and fixing it allows us to greatly speed up numerical calculations. Furthermore, the effects of this ratio have been extensively studied in other works, and our primary purpose is to investigate the effect of radiation altogether, rather than the effect of each part of the LQ model. For these reasons, we fix $\alpha/\beta = 3$ for all simulations. We also assume that prior to radiation, the nondimensionalized cell fraction has reached its steady state. This steady state is easily obtained from the governing ODEs by setting the derivatives and the terms with f_1, f_2 to zero and solving for the cell numbers as $n_1(0) = 1/(1 + \gamma_1/\gamma_2)$ and $n_2(0) = 1/(1 + \gamma_2/\gamma_1)$, which we use as our initial condition. Using the optimization procedure outlined above, the optimal distribution of radiation can be identified for any given parameter set which consists of the eight parameters found in equations (5.1), (5.2), (5.3), and (5.4) (not including n_{max}). In order to complete the optimization, we must also specify the metric function weight coefficients, p_1 and p_2.

When using the optimization procedure on various parameter sets, a dominant qualitative behaviour quickly emerges; specifically, that the optimal radiation schedule is when the final 20 days receive the maximum allowable radiation and the first 10 days receive none. We refer to this case as the 'trivial' result since it is both the least interesting and most common among reasonable parameter sets. Intuitively, the reason for this result is that since we compute the metric at the end of treatment, the optimization tends to find that focusing radiation at the end of treatment does not allow time for regrowth. Importantly though, since the tumour cell population begins at its steady-state value in our optimization, the clinical interpretation of the trivial result is not that treatment onset should be delayed, but rather simply that the radiation should be focused over the smallest amount of time possible. More discussion on this point is included in the conclusion. Many parameter sets however do not fall into this pattern and show more varied optimal dose distributions. We refer to cases which do not optimize to the trivial result as 'nontrivial' results. For example, consider the optimization results shown in Figure 5.2. In this figure, we seek to

minimize the total cell number and therefore choose $p_1 = p_2 = 1$. We generate the parameter set of $r_1 = 0.4, r_2 = 0, \alpha_1 = 0.2, \beta_1 = \alpha_1/3, \alpha_2 = 0.0005, \beta_2 = \alpha_2/3, \gamma_1 = 0.1, \gamma_2 = 0.1$ and use our procedure to find the radiation distribution that minimizes our metric. Notice that the optimal distribution of radiation is to apply the maximal amount of radiation on all days other than days 3, 4, 6, 9, 10, 13, 16, 20, 24, and 27. In fact, this schedule results in an approximate 22% decrease in final cell number compared to the clinical case and an approximate 40% decrease in final cell number compared to the trivial case.

Cases such as the one displayed in Figure 5.2 are the outliers however, with the trivial distribution being the optimal for most parameter sets. This then begs the question of which parameter sets result in nontrivial optima and what are the key parameters which determine such behaviour. Answering these questions using analytical methods is quite challenging due to the number of variables included in the model and the complexity of the optimization process. We therefore result to numerical and data analysis techniques to understand the relationship between the parameters and the optimal radiation distribution.

5.3.1 Minimization of total cells

We seek to identify the relationship between the model parameters and the optimal radiation schedule for the case that the clinical goal is to kill the maximal number of cells. For this, we set $p_1 = p_2 = 1$ for our optimization and attempt to identify the subset of the parameter space which produces nontrivial optimization results.

We begin by generating an input-output data set where each sample consists of the values for the six free parameters along with a label of trivial or nontrivial for its optimization result. To create this dataset, we select clinically-reasonable ranges for each of the parameters and sample values from within those ranges. The ranges used for sampling from the parameter space are shown in Table 5.1. Note that for the proliferation and plasticity rates, parameters are uniformly sampled, whereas for the radiobiological parameters, they are sampled on a logarithmic scale. As a full optimization typically takes several minutes to find the optimum, in order to generate a large enough dataset for analysis, we slightly alter the optimization. Since we are only interested in labelling the sample as trivial or nontrivial, we do not need to complete the full optimization process. Instead, we can label the sample by setting our initial optimization guess as the trivial result, performing a single optimization iteration, and determining whether or not the new optimal guess has changed: if it has, then the trivial case must not be the optimal solution and there must therefore be a nontrivial result. This method allows for the generation of a much larger dataset than would otherwise be possible given computational limitations. With this method, a data set of 19,338 labelled parameter samples are generated and used for analysis

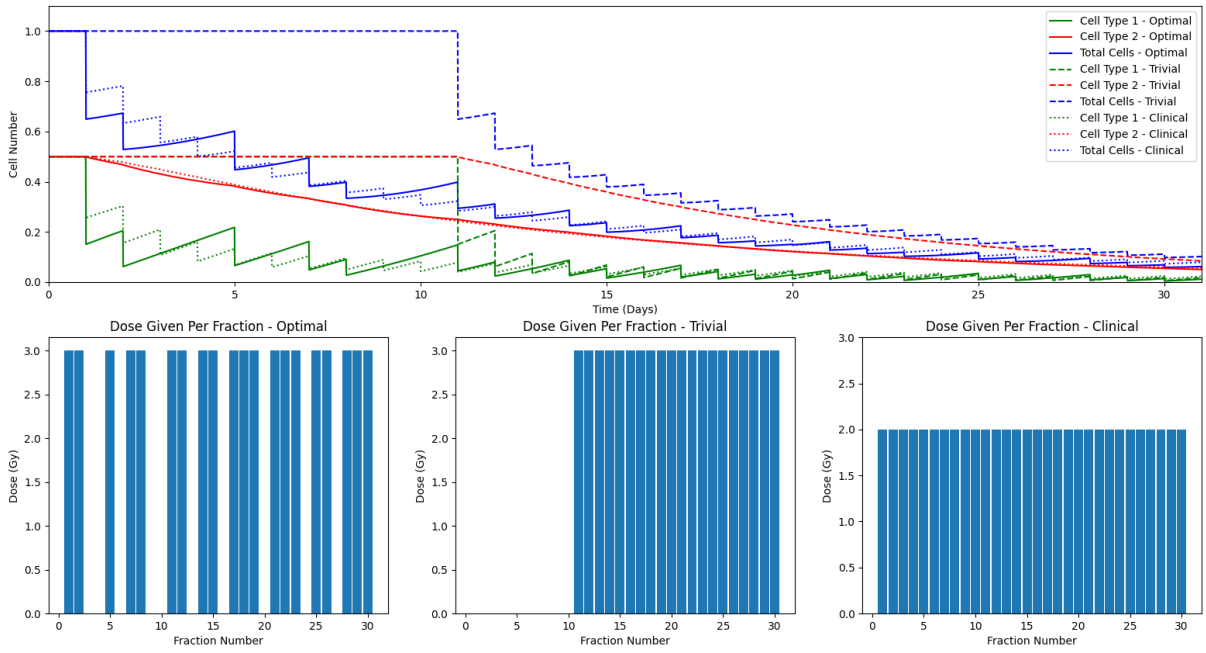

Figure 5.2: A comparison of model results with different radiation schedules. Cell type 1 is shown in green, cell type 2 in red, and total cells in blue. The schedule derived using the optimization procedure is plotted with the full line, the trivial case with the dashed line, and the clinical standard case with the dotted line. The parameters used for this case are: $r_1 = 0.4$ (1/day), $r_2 = 0$ (1/day), $\alpha_1 = 0.2$ (1/Gy), $\beta_1 = \alpha_1/3$ (1/Gy), $\alpha_2 = 0.0005$ (1/Gy), $\beta_2 = \alpha_2/3$ (1/Gy), $\gamma_1 = 0.1$ (1/day), $\gamma_2 = 0.1$ (1/day). Notice that the optimal case results in an approximate 22% decrease in final cell number compared to the clinical case and an approximate 40% decrease in final cell number compared to the trivial case.

Parameter	Symbol	Range Min (Unit)	Range Max (Unit)	Sample Type
Cell Proliferation Rate	r_i	0.3466 (1/day)	0.0347 (1/day)	Uniform
Linear LQ Parameter	α_i	e^{-10} (1/Gy)	1 (1/Gy)	Logarithmic
Cell Plasticity Rate	γ_i	0 (1/day)	0.3 (1/day)	Uniform

Table 5.1: Parameter sampling ranges and method for the minimization of total cells case.

To identify which parameters are the most important in determining the optimization result, we create a 2D histogram matrix as can be seen in Figure 5.3. In the upper triangular portion of this figure, a 2D histogram of each pair of model parameters is plotted with blue representing trivial results and orange representing nontrivial results. As can be clearly seen from these histograms, a separation between the optimization result is most apparent when considering the effect of radiation through the parameters α_1 and α_2 (plotted in this Figure on a log scale). Interestingly, the α_1 vs. α_2 histogram shows that nontrivial results are grouped along the axes, where one of the values is relatively large and the other relatively small. This motivates the idea that a key determining factor in the result is the ratio of the radiation effect parameters between the cell types. In the bottom part of Figure 5.3, the 2D histogram of the logarithm of the larger α_i is plotted against the logarithm of the larger α_i/α_j. A clear separation between the results is observed in this plot, showing that the most important factor is the difference in radiation effect between the cell types. To further illustrate this, a tree classifier (Figure 5.4) was created and trained using only α_1 and the ratio α_1/α_2. Limited to only this knowledge and a maximum decision depth of three, the tree classifier was able to obtain 88.6% accuracy. The tree classifier was created using the `DecisionTreeClassifier()` function from `sklearn`

In light of these results, it is worth noting that the situation where two distinct populations of cells have differing levels of radiation response is precisely what has been observed in previous studies examining CSC subpopulations. Specifically, [57] and [92] both observed that CSCs exhibit less sensitivity to radiotherapy than non-CSCs. Furthermore, revisiting the disagreement between the results of Leder et al [59] and Forouzannia et al [31] compared with Galochkina et al [33], we can more concretely evaluate our hypothesis that the reason for the disagreement was due to the different assumptions about the effect of radiation on the cell types. We find that in cases where the effect of radiation is constant between the cell types, the optimal schedule of radiation is trivial. With the increasing ability of researchers to identify CSC subpopulations within tumours, results such as these are clearly of importance for treatment planning.

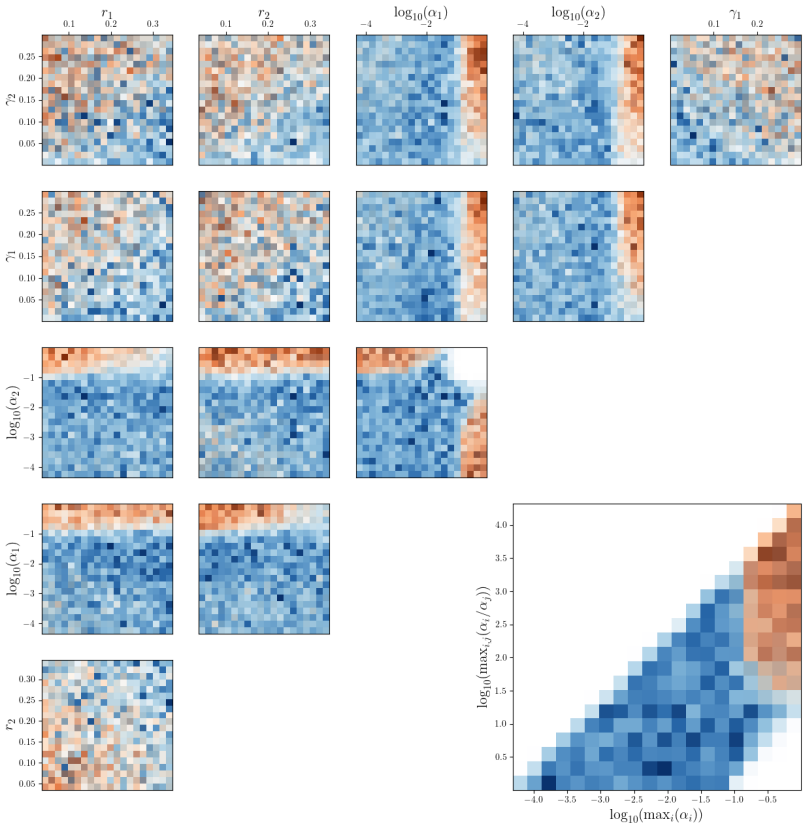

Figure 5.3: A 2D histogram matrix of each pair of model parameters for the case of minimization of total cells. Areas that are more solid blue represent cases where the optimal result was the trivial case and areas of more solid orange represent cases where the optimal result was nontrivial. Note that the parameters α_1 and α_2 are plotted on a base 10 log scale. Observe the clear separation in cases involving α_1 and α_2 and the lack of separation in cases that don't. In the bottom portion of the figure, a 2D histogram with the same colour scheme is included for the maximum α_i vs. the maximum α_i/α_j, with both axes on a base 10 log scale. Notice the clear separation between the optimization cases based on the ratio of the radiation effect parameters. The parameters r_1, r_2, γ_1, and γ_2 are in units of (1/day) while α_1 and α_2 are in units of (1/Gy).

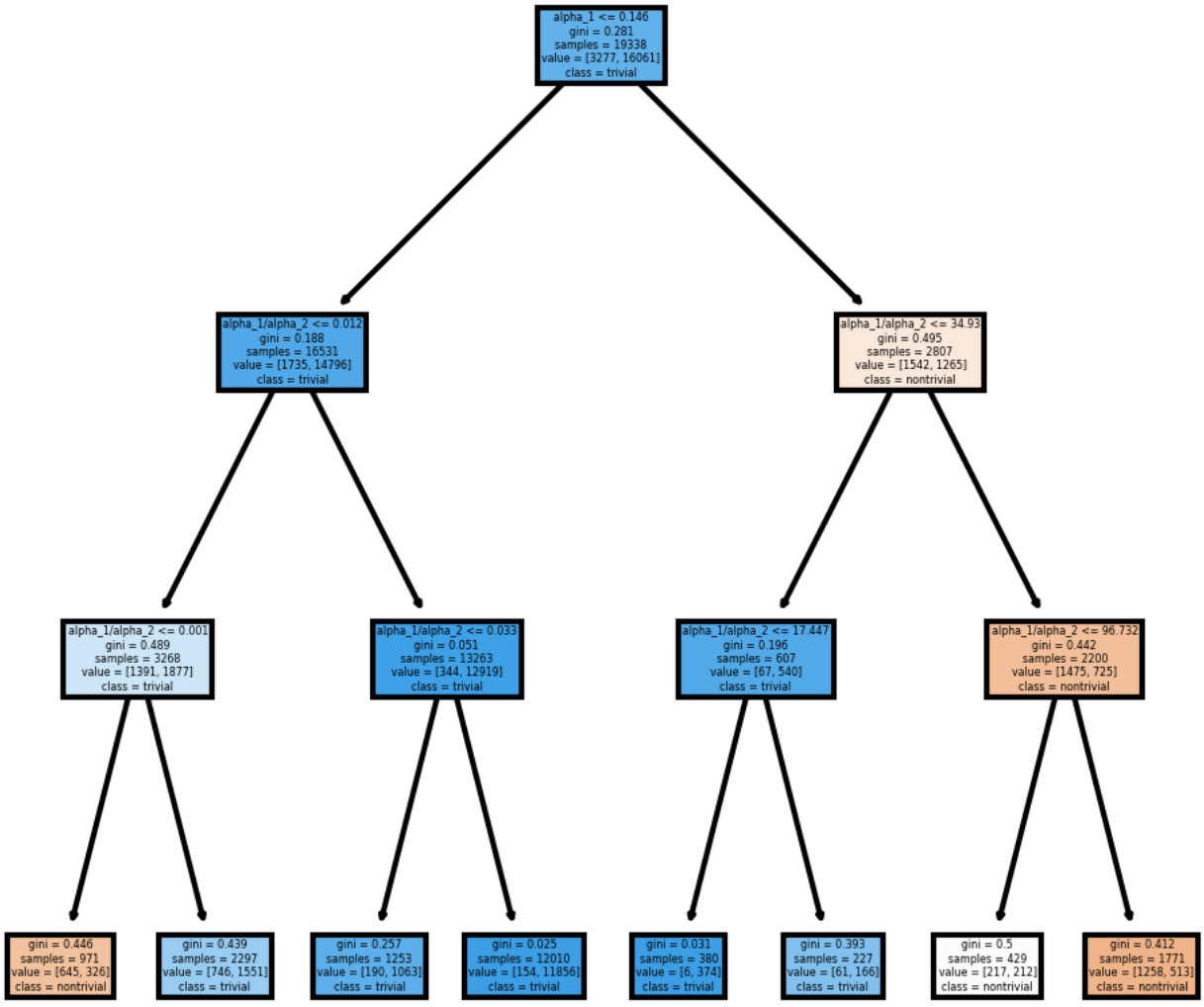

Figure 5.4: A tree classifier for the minimization of total cells case. The tree uses only α_1 and the ratio α_1/α_2 to make predictions and is limited to a maximum decision depth of three. Even with these limitations, the tree is nonetheless able to obtain 88.6% accuracy, showing that the qualitative nature of the optimization result is largely determined by the radiation effect parameters.

5.3.2 Minimization of cancer stem cells

Given the importance of CSCs to the progression of cancer, it has been proposed that treatments should prioritize the elimination of CSCs rather than the minimization of total cells. Accordingly, we consider the case where $p_1 = 0$ and $p_2 = 1$ to see the different strategies that evolve when seeking to minimize the size of a specific subpopulation. To do so, we designate n_2 as the CSCs and n_1 as the non-CSCs and restrict our parameter generation to align with this distinction. Specifically, given the results of the previous case, we generate samples where $0.1 \leq \alpha_1 \leq 1$ and $20 < \alpha_1/\alpha_2 < 100$ and create another histogram matrix. The results of this are shown in Figure 5.5 for a set of 17853 samples. In this figure, notice that there is a clear separation between the trivial and nontrivial cases on the plot of r_2 vs. γ_2. Importantly, the case of a tumour subpopulation which exhibits radioresistance, a small growth rate, but a high replenishment rate of normal cancer cells is precisely the behaviour expected of CSCs. These results buttress our qualitative conclusion from the last section that non-uniform radiation schedules can be preferable in cases where there is a known CSC population.

5.4 Conclusion

In this paper, we added a crucial element to the classic problem of temporal optimization of radiation – cellular heterogeneity. We created a differential equation model and an optimization procedure to identify which distribution of radiation dose resulted in the maximal cell kill for a given set of parameters governing a tumour's growth and response to treatment. We found that most simulated sets of governing parameters resulted in the optimal distribution being when the allowable dose was focused as much as possible at the end of treatment, a case which we termed the trivial result. Clinically however, this result does not suggest that the start of treatment should be delayed since, of course, the choice of treatment end time is somewhat arbitrary. Instead, the trivial result shows that optimal cell kill is achieved when radiation is grouped over the smallest amount of time possible. Since the tumour population begins at its steady-state value, the trivial result would actually suggest front-loading the radiation, rather than back-loading it. But some parameter sets exhibited nontrivial results where the available dose was spread across the treatment time. Of note, we showed that the optimal distribution is typically bang-bang, with each day's radiation either being the maximum allowable or zero. We find that the main factor in determining whether the optimal radiation dose will be nontrivial is the difference in the effect of radiation between the cell types. We specifically showed that

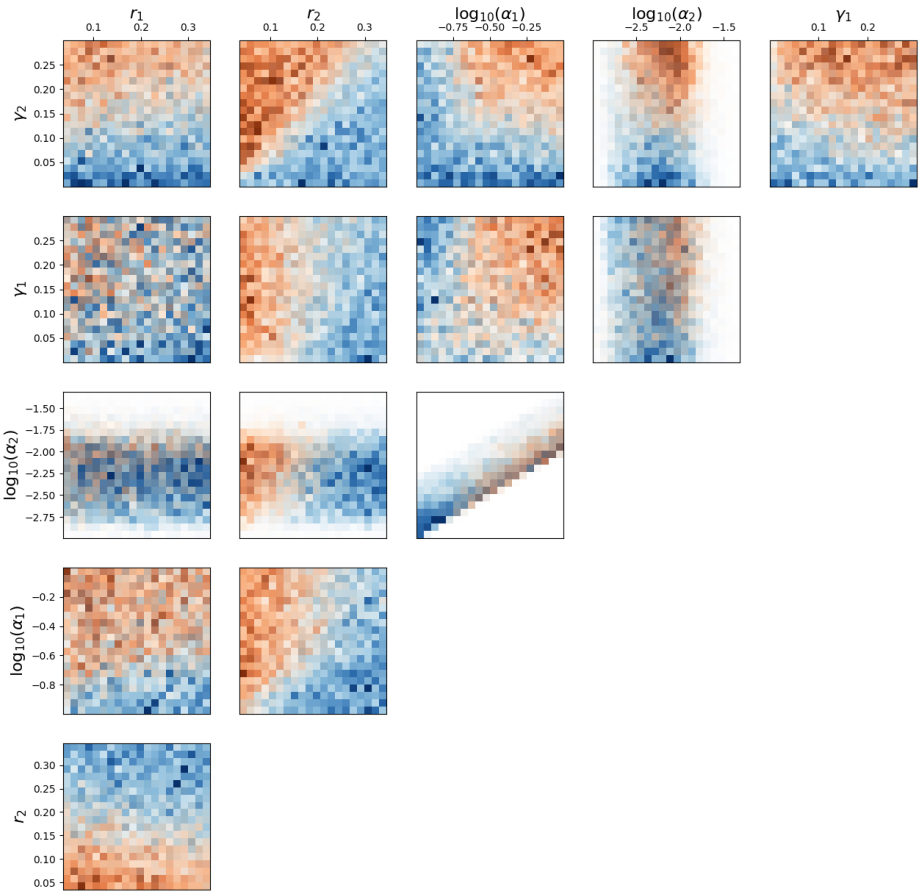

Figure 5.5: A 2D histogram matrix of each pair of model parameters for the case of minimization of CSCs. Areas that are more solid blue represent cases where the optimal result was the trivial case and areas of more solid orange represent cases where the optimal result was nontrivial. Note that the parameters α_1 and α_2 are plotted on a base 10 log scale. Observe that for minimization of CSCs with a restricted parameter space, the clearest separation is now seen in the plot of r_2 vs. γ_2. The parameters r_1, r_2, γ_1, and γ_2 are in units of (1/day) while α_1 and α_2 are in units of (1/Gy).

knowledge of the linear LQ parameters for the cell types alone was enough to predict the behaviour with over 88% accuracy.

There are several parts of our work that could be further generalized and limitations of our model which provide opportunities for future research directions. The assumptions made in creating our model can easily be changed: one could consider a Gompertzian growth, or a nonlinear plasticity, for example. The radiation schedule could also be generalized to include fractions of different lengths and start times. Continuous dose profiles could also be worked into the optimization which may be more appropriate in cases such as when the method of administration is brachytherapy. The assumptions made on our choices of parameters could also be changed or expanded. Importantly, in our model, the tumour was always assumed to be saturated prior to beginning treatment. However, diagnosing tumours before they reach this state is obviously a crucial clinical goal, making the case of non-saturated initial conditions important to consider. Due to this assumption, our model is biased to focusing radiation at the end of the schedule to prevent cells from simply regrowing. Based on our computational algorithm however, different initial conditions are challenging to include in our analysis. If extensions to our work could include varying initial conditions and oberve how this alters the predictions, it would be a valuable addition. Different α/β ratios are not included in this study since we are mainly interested in overall radiation effect, though it could be included to provide more general results or fixed to a value for a particular patient for more specific recommendations. These results could also be taken further, and one could attempt to find the precise relationship between the parameters and the qualitative behaviour beyond just the difference in radiation effect, though this would likely require a more efficient optimization procedure or significantly more computing power. Furthermore, actually predicting which days would receive radiation rather than just the qualitative behaviour would be interesting, though again, this would certainly require more data or computing power.

In modern clinical practice, determining the precise number and spatial distribution of CSCs within the tumour of a particular subject presents a considerable challenge. Consequently, it is reasonable to question the potential clinical utility of modelling approaches like the one presented here. Current clinical methods do not provide a means to accurately assess the proportion of CSCs within a tumour, nor their spatial distribution. However, there are biomarkers available that can be used to label cells with stem-like characteristics (such as CD44+/CD24- in breast cancer), which can be identified through single-cell imaging techniques like flow cytometry. Additionally, cells obtained from biopsies can be cultured in a laboratory or implanted in a mouse model to observe the proportion that form colonies, indicating stem-like properties. Nevertheless, performing these techniques for a specific patient prior to treatment is not a straightforward process. However, even

with the existing techniques, it is possible to make predictions regarding the proportion of stem-like cells within a tumour, given its type and other factors. Since our model is based on ODEs, we only require proportional information to initiate predictions. Although this information may not be readily available for a particular patient prior to treatment, it could enable predictions about the proportion of stem-like cells within a tumour.

Our results shed important light on the question of how differences in tumour subpopulations can lead to changes in optimal treatment protocol. This is particularly relevant when discussing CSCs, which are known to exist in tumours and display radioresistance compared to non-CSCs. We hope that this work acts as support for the idea that understanding the full tumour biology is important for designing effective treatments and the idea that mathematical tools can play a vital role in identifying cases where treatments can be improved. We hope that the results presented here can be used as hypotheses for clinical and experimental researchers to test our findings. Clearly, the recommendations which could arise from our results are of importance to clinical practice.

Chapter 6

Conclusion and Future Directions

This thesis focused on the complementary use of conventional and machine learning modelling to advance personalized oncology. The increasing importance of incorporating quantitative analysis in biology and medicine, alongside the growing availability of data, creates ample opportunities for these methods to make significant contributions. Researchers and pharmaceutical companies have made it a key part of modern research, and even regulatory agencies have started to accept results from quantitative analysis more readily. For example, the FDA has increasingly considered mathematical modelling in its drug approval process [24].

The three projects in this thesis each utilized a combination of conventional and machine learning models to address a problem in oncology. Two of these specifically used PINNs, a novel deep learning technique with tremendous potential for further application in this context.

In Chapter 3, PINNs were combined with the PI model and applied to characterize brain tumours in a patient-specific manner. Accurately estimating tumour cell diffusivity and proliferation rates can be challenging to do for a particular tumour, forcing researchers and clinicians to rely on averages or crude approximation techniques. To address this issue, we developed a deep learning model capable of accurately estimating these key parameters and predicting tumour progression. Our method used two sets of multi-sequence MRI data and relied on a preprocessing pipeline that included brain tumour segmentation and conversion to tumour cellularity. By applying our model to synthetic tumours and a clinical dataset of five GBM patients, we demonstrated the potential for estimating tumour parameters and predicting tumour progression in a patient-specific way. This work laid the foundation for further research aimed at aiding physicians in personalized diagnoses and treatments.

In Chapter 4, we estimated tumour IFP using CT imaging of drug-carrying liposomes. We developed a model capable of predicting voxel-by-voxel liposome accumulation and IFP in tumours using an approach based on an established PDE model of liposome transport and PINNs. We validated our model using data from animal experiments and computationally-generated tumours, demonstrating its potential for predicting intratumoral pressure maps that could be used for enhanced precision in forecasting tumour progression and designing treatments. This work further contributes to the idea of quantitative methods providing an opportunity for patient selection prior to treatment.

In Chapter 5, we focused on external beam radiation therapy, a crucial component of modern cancer treatments. Previous mathematical studies of radiation effect primarily assume intratumoral homogeneity, overlooking the potentially important heterogeneous nature of tumours and the presence of various subpopulations, such as CSCs, which exhibit higher radioresistance compared to non-CSCs. In this work, we generalized the problem of temporal optimization of dose distribution in radiation therapy to a two-cell-type model, accounting for tumour heterogeneity. We developed a mathematical model and a numerical optimization algorithm to find the optimal dose distribution for maximum cell kill. By creating a dataset of optimization solutions, we used machine learning tools to understand the relationships between model parameters and the qualitative behavior of optimization results. Our findings suggest key factors in predicting optimal radiation distribution which can guide treatment strategies in cases where clinicians have knowledge of tumour heterogeneity and CSC abundance, ultimately leading to more effective cancer therapies.

There are many problems to which the methods I have developed within this thesis can be applied. For instance, in some of my previous works, I designed mathematical models to analyze HAPs, a class of bioreductive compound that becomes activated in the presence of tumour hypoxia. Though these drugs have strong theoretical and preclinical experimental backing, they have struggled in clinical trials, with few success stories making it to the clinic. This discrepancy between theory and reality has largely been attributed to improper patient selection. Namely, not all tumours are hypoxic, and HAPs are unlikely to provide significant benefit to those patients who do not have tumour sufficient hypoxia. Methods similar to those used in Chapters 3 and 4 could be used to identify these patients. For example, if imaging could be done to observe a biomarker mathematically related to tumour hypoxia, then PINNs could derive key tissue traits and tissue oxygen distributions, which could be incorporated into patient selection or even treatment design.

Another potential avenue for further work is in the design of cancer nanoparticles. Based on the work in Chapter 4, we demonstrated that given an imaged distribution of injected dye into a tumour, the underlying IFP could be predicted. Then, this information could be used with the specific parameters for a cancer nanoparticle to predict drug distribution

and treatment efficacy. Consider however, the inverse problem: given a desired drug pharmacokinetic profile within a tumour (or pharmacodynamic effect on the tumour), find the nanoparticle-specific parameters which achieve this. PINNs could be used in a similar way to attempt to do this. We have already begun working on this project with promising initial results.

Additionally, consider High Intensity Focused Ultrasound (HIFU) therapy, in which high temperatures are used to kill cancer cells. HIFU therapy is a non-invasive technique that can target specific regions within a tumour, minimizing damage to surrounding healthy tissue. We could use PINNs to find the temperature profile and the correct parameters or methods necessary to achieve this profile and anticancer effect. By doing so, it may be possible to further optimize HIFU therapy, leading to more effective and precise cancer treatments with fewer side effects for patients.

Overall, this thesis aims to showcase examples of how mathematical modelling and machine learning can be used in combination to address problems in oncology. As is widely accepted within these fields, the integration of quantitative analysis in cancer is the future, and the next great leaps forward will stem from further interdisciplinary work. Therefore, this work is well poised to have an impact in the years ahead.

References

[1] J C L Alfonso, G Buttazzo, B García-Archilla, M A Herrero, and L Núñez. Selecting Radiotherapy Dose Distributions by Means of Constrained Optimization Problems. *Bulletin of Mathematical Biology*, 76(5):1017–1044, 2014.

[2] J C L Alfonso, G Buttazzo, Bosco García-Archilla, M A Herrero, and L Núñez. A class of optimization problems in radiotherapy dosimetry planning. *Discrete and Continuous Dynamical Systems - Series B*, 17(6):1651–1672, 2012.

[3] J. C. L. Alfonso, K. Talkenberger, M. Seifert, B. Klink, A. Hawkins-Daarud, K. R. Swanson, H. Hatzikirou, and A. Deutsch. The biology and mathematical modelling of glioma invasion: A review. *Journal of the Royal Society Interface*, 14(136), nov 2017.

[4] M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The fenics project version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[5] M B Altman, S J Chmura, J O Deasy, and J C Roeske. Optimization of the temporal pattern of radiation: An IMRT based study. *International Journal of Radiation Oncology Biology Physics*, 66(3):898–905, 2006.

[6] A. Amarouch and J. J. Mazeron. Radiotherapy plus concomitant and adjuvant Temozolomide for glioblastoma. *Cancer/Radiotherapie*, 9(3):196–197, 2005.

[7] M R Atashzar, R Baharlou, J Karami, H Abdollahi, R Rezaei, F Pourramezan, and Seyed H Zoljalali Moghaddam. Cancer stem cells: A review from origin to therapeutic implications. *Journal of Cellular Physiology*, 235(2):790–803, 2020.

[8] N. C. Atuegwu, L. R. Arlinghaus, X. Li, A. B. Chakravarthy, V. G. Abramson, M. E. Sanders, and T. E. Yankeelov. Parameterizing the logistic model of tumor

growth by DW-MRI and DCE-MRI data to predict treatment response and changes in breast cancer cellularity during neoadjuvant chemotherapy. *Translational Oncology*, 6(3):256–264, 2013.

[9] N. C. Atuegwu, D. C. Colvin, M. E. Loveless, L. Xu, J. C. Gore, and T. E. Yankeelov. Incorporation of diffusion-weighted magnetic resonance imaging data into a simple mathematical model of tumor growth. *Physics in Medicine and Biology*, 57(1):225–240, jan 2012.

[10] E Batlle and H Clevers. Cancer stem cells revisited. *Nature Medicine*, 23(10):1124–1134, 2017.

[11] Laurence T Baxter and Rakesh K Jain'. Transport of Fluid and Macromolecules in Tumors I. Role of Interstitial Pressure and Convection. *Microvascular Research*, 37:77–104, 1988.

[12] A. Bhandari, A. Bansal, A. Singh, and N. Sinha. Perfusion kinetics in human brain tumor with DCE-MRI derived model and CFD analysis. *Journal of Biomechanics*, 59:80–89, July 2017.

[13] Ajay Bhandari, Ankit Bansal, Rishav Jain, Anup Singh, and Niraj Sinha. Effect of Tumor Volume on Drug Delivery in Heterogeneous Vasculature of Human Brain Tumors. *Journal of Engineering and Science in Medical Diagnostics and Therapy*, 2(2):021004, May 2019.

[14] Ajay Bhandari, Ankit Bansal, Anup Singh, Rakesh Kumar Gupta, and Niraj Sinha. Comparison of transport of chemotherapeutic drugs in voxelized heterogeneous model of human brain tumor. *Microvascular Research*, 124:76–90, July 2019.

[15] Ajay Bhandari, Ankit Bansal, Anup Singh, and Niraj Sinha. Transport of Liposome Encapsulated Drugs in Voxelized Computational Model of Human Brain Tumors. *IEEE Transactions on NanoBioscience*, 16(7):634–644, October 2017.

[16] T Bortfeld, J Ramakrishnan, J N. Tsitsiklis, and J Unkelbach. Optimization of radiation therapy fractionation schedules in the presence of tumor repopulation. *INFORMS Journal on Computing*, 27(4):788–803, sep 2015.

[17] A Brahme. Development of radiation therapy optimization. *Acta Oncologica*, 39(5):579–595, 2000.

[18] David J. Brenner. The linear-quadratic model is an appropriate methodology for determining isoeffective doses at large doses per fraction. *Seminars in Radiation Oncology*, 18(4):234–239, October 2008.

[19] M C Cabrera. Cancer stem cell plasticity and tumor hierarchy. *World Journal of Stem Cells*, 7(1):27, 2015.

[20] R S Cantrell and C Cosner. Diffusive logistic equations with indefinite weights: population models in disrupted environments. *Proceedings of the Royal Society of Edinburgh*, 112A, 293-318, 1989.

[21] Peter Carmeliet and Rakesh K. Jain. Principles and mechanisms of vessel normalization for cancer and other angiogenic diseases. *Nature Reviews Drug Discovery*, 10(6):417–427, June 2011.

[22] I. S. Chua, M. Gaziel-Yablowitz, Z. T. Korach, K. L. Kehl, N. A. Levitan, Y. E. Arriaga, G. P. Jackson, D. W. Bates, and M. Hassett. Artificial intelligence in oncology: Path to implementation. *Cancer Medicine*, 10(12):4138–4149, 2021.

[23] E Coddington and N Leninson. *Theory of ordinary differential equations*. Tata Mc-Grow Hill Publishing Company Limited, TMH edition, 1972, 9th reprint 1987.

[24] Office of the Commissioner. Modeling & Simulation at FDA. *FDA*, November 2022. https://www.fda.gov/science-research/about-science-research-fda/modeling-simulation-fda.

[25] D Conforti, F Guerriero, and R Guido. Optimization models for radiotherapy patient scheduling. *4or*, 6(3):263–278, 2008.

[26] T. F. Dagi, F. G. Barker, and J. Glass. Machine Learning and Artificial Intelligence in Neurosurgery: Status, Prospects, and Challenges. *Neurosurgery*, 89(2):133–142, 2021.

[27] A. Darbar, M. Waqas, S. F. Enam, and S. D. Mahmood. Use of Preoperative Apparent Diffusion Coefficients to Predict Brain Tumor Grade. *Cureus*, mar 2018.

[28] D. C. Drummond, O. Meyer, K. Hong, D. B. Kirpotin, and D. Papahadjopoulos. Optimizing liposomes for delivery of chemotherapeutic agents to solid tumors. *Pharmacological Reviews*, 51(4):691–743, December 1999.

[29] Brydon Eastman, Cameron Meaney, Michelle Przedborski, and Mohammad Kohandel. Modeling the impact of public response on the COVID-19 pandemic in Ontario. *PLoS ONE*, 16(4 April), April 2021. Publisher: Public Library of Science.

[30] J. Elkhader and O. Elemento. Artificial intelligence in oncology: From bench to clinic. *Seminars in Cancer Biology*, 2021.

[31] F Forouzannia, H Enderling, and M Kohandel. Mathematical Modeling of the Effects of Tumor Heterogeneity on the Efficiency of Radiation Treatment Schedule. *Bulletin of Mathematical Biology*, 80(2):283–293, feb 2018.

[32] A. A. Gabizon. Selective tumor localization and improved therapeutic index of anthracyclines encapsulated in long-circulating liposomes. *Cancer Research*, 52(4):891–896, February 1992.

[33] T Galochkina, A Bratus, and Víctor M. Pérez-García. Optimal radiation fractionation for low-grade gliomas: Insights from a mathematical model. *Mathematical Biosciences*, 267:1–9, 2015.

[34] K. M. Gauvain, R. C. McKinstry, P. Mukherjee, A. Perry, J. J. Neil, B. A. Kaufman, and R. J. Hayashi. Evaluating pediatric brain tumor cellularity with diffusion-tensor imaging. *American Journal of Roentgenology*, 177(2):449–454, 2001.

[35] N. Gaw, A. Hawkins-Daarud, L. S. Hu, H. Yoon, L. Wang, Y. Xu, P. R. Jackson, K.e W. Singleton, L. C. Baxter, J. Eschbacher, A. Gonzales, A. Nespodzany, K. Smith, P. Nakaji, J. R. Mitchell, T. Wu, K. R. Swanson, and J. Li. Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Scientific Reports*, 9(1), dec 2019.

[36] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1994.

[37] Ali Gooya, Kilian M. Pohl, Michel Bilello, Luigi Cirillo, George Biros, Elias R. Melhem, and Christos Davatzikos. Glistr: Glioma image segmentation and registration. *IEEE Transactions on Medical Imaging*, 31:1941–1954, 2012.

[38] Stefan Grivalsky, Martin Tamajka, and Wanda Benesova. Segmentation of gliomas in magnetic resonance images using recurrent neural networks. *42nd International Conference on Telecommunications and Signal Processing (TSP)*, page 763, 2019.

[39] R. K. Gupta, T. F. Cloughesy, U. Sinha, J. Garakian, J. Lazareff, G. Rubino, L. Rubino, D. P. Becker, H. V. Vinters, and J. R. Alger. Relationships between choline magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma. *Journal of Neuro-Oncology*, 50(3):215–226, 2000.

[40] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, March 2011. Publisher: Elsevier.

[41] H. L. P. Harpold, E. C. Alvord, and K. R. Swanson. The Evolution of Mathematical Modeling of Glioma Proliferation and Invasion. *J Neuropathol Exp Neurol*, 66(1), 2007.

[42] K. J. Harrington, S. Mohammadtaghi, P. S. Uster, D. Glass, A. M. Peters, R. G. Vile, and J. S. Stewart. Effective targeting of solid tumors in patients with locally advanced cancers by radiolabeled pegylated liposomes. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 7(2):243–254, February 2001.

[43] Cosmina Hogea, Christos Davatzikos, and George Biros. An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. *Journal of Mathematical Biology*, 56:793–825, 6 2008.

[44] David A. Hormuth, Karine A. Al Feghali, Andrew M. Elliott, Thomas E. Yankeelov, and Caroline Chung. Image-based personalization of computational models for predicting response of high-grade glioma to chemoradiation. *Scientific Reports*, 11, 12 2021.

[45] L. S. Hu, L. Wang, A. Hawkins-Daarud, J. M. Eschbacher, K. W. Singleton, Pa. R. Jackson, K. Clark-Swanson, C. P. Sereduk, S. Peng, P. Wang, J. Wang, L. C. Baxter, K. A. Smith, G. L. Mazza, A. M. Stokes, B. R. Bendok, R. S. Zimmerman, C. Krishna, A. B. Porter, M. M. Mrugala, J. M. Hoxworth, T. Wu, N. L. Tran, K. R. Swanson, and J. Li. Uncertainty quantification in the radiogenomics modeling of EGFR amplification in glioblastoma. *Scientific Reports*, 11(1):1–14, 2021.

[46] Clifford A. Hudis. Trastuzumab — Mechanism of Action and Use in Clinical Practice. *New England Journal of Medicine*, 357(1):39–51, July 2007. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMra043186.

[47] A N Iusem. On the convergence properties of the projected gradient method for convex optimization. *Computational and Applied Mathematics*, 22(1):37–52, 2003.

[48] Ali Işin, Cem Direkoğlu, and Melike Şah. Review of mri-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102:317–324, 2016.

[49] L. Jones, P. Hoban, and P. Metcalfe. The use of the linear quadratic model in radiotherapy: a review. *Australasian Physics & Engineering Sciences in Medicine*, 24(3):132–146, September 2001.

[50] Liron Simon Keren, Alex Liberzon, and Teddy Lazebnik. A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge. *Scientific Reports*, 13(1):1249, January 2023. Number: 1 Publisher: Nature Publishing Group.

[51] M Kim, R D Stewart, and M H Phillips. A feasibility study: Selection of a personalized radiotherapy fractionation schedule using spatiotemporal optimization. *Medical Physics*, 42(11):6671–6678, 2015.

[52] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. arXiv:1412.6980 [cs].

[53] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *ICLR*, 12 2015.

[54] Donald Knuth. *The TEXbook*. Addison-Wesley, Reading, Massachusetts, 1986.

[55] Ender Konukoglu, Olivier Clatz, Bjoern H. Menze, Bram Stieltjes, Marc André Weber, Emmanuel Mandonnet, Hervé Delingette, and Nicholas Ayache. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. *IEEE Transactions on Medical Imaging*, 29:77–95, 1 2010.

[56] Matthew Koshy, John L. Villano, Therese A. Dolecek, Andrew Howard, Usama Mahmood, Steven J. Chmura, Ralph R. Weichselbaum, and Bridget J. McCarthy. Improved survival time trends for glioblastoma using the SEER 17 population-based registries. *Journal of Neuro-Oncology*, 107(1):207–212, March 2012.

[57] C Lagadec, E Vlashi, L Della Donna, Y H Meng, C Dekmezian, K Kim, and F Pajonk. Survival and self-renewing capacity of breast cancer initiating cells during fractionated radiation treatment. *Breast Cancer Research*, 12(1):1–13, 2010.

[58] Leslie Lamport. *LaTeX — A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, second edition, 1994.

[59] K Leder, K Pitter, Q Laplant, D Hambardzumyan, B D Ross, T A Chan, E C Holland, and F Michor. Mathematical modeling of pdgf-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, 156(3):603–616, 2014.

[60] Jana Lipkova, Panagiotis Angelikopoulos, Stephen Wu, Esther Alberts, Benedikt Wiestler, Christian Diehl, Christine Preibisch, Thomas Pyka, Stephanie E. Combs, Panagiotis Hadjidoukas, Koen Van Leemput, Petros Koumoutsakos, John Lowengrub, and Bjoern Menze. Personalized radiotherapy design for glioblastoma: Integrating mathematical tumor models, multimodal scans, and bayesian inference. *IEEE Transactions on Medical Imaging*, 38:1875–1884, 8 2019.

[61] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, August 1989.

[62] Long Jian Liu, Stephen L. Brown, James R. Ewing, Brigitte D. Ala, Kenneth M. Schneider, and Mordechay Schlesinger. Estimation of Tumor Interstitial Fluid Pressure (TIFP) Noninvasively. *PLOS ONE*, 11(7):e0140892, July 2016.

[63] A. Logg, G. N. Wells, and J. Hake. *DOLFIN: a C++/Python Finite Element Library*, chapter 10. Springer, 2012.

[64] H. Lyng, O. Haraldseth, and E. K. Rofstad. Measurement of cell density and necrotic fraction in human melanoma xenografts by diffusion weighted magnetic resonance imaging. *Magnetic Resonance in Medicine*, 43(6):828–836, 2000.

[65] S. E. Maier, Y. Sun, and R. V. Mulkern. Diffusion imaging of brain tumors. *NMR in Biomedicine*, 23(7):849–864, aug 2010.

[66] M. Malathi and P. Sinthia. Brain tumour segmentation using convolutional neural network with tensor flow. *Asian Pacific Journal of Cancer Prevention*, 20:2095–2101, 2019.

[67] Andreas Mang, Spyridon Bakas, Shashank Subramanuan, Christos Davatzikos, and George Biros. Integrated biophysical modeling and image analysis application to neuro-oncology. *Annual Reviews in Biomedical Engineering*, 22:309–341, 2020.

[68] N D Marjanovic, R A Weinberg, and C L Chaffer. Cell plasticity and heterogeneity in cancer. *Clinical Chemistry*, 59(1):168–179, 2013.

[69] Susan Christine Massey, Paula Whitmire, Tatum E. Doyle, Joseph E. Ippolito, Maciej M. Mrugala, Leland S. Hu, Peter Canoll, Alexander R.A. Anderson, Melissa A. Wilson, Susan M. Fitzpatrick, Margaret M. McCarthy, Joshua B. Rubin, and Kristin R. Swanson. Sex differences in health and disease: A review of biological sex differences relevant to cancer with a spotlight on glioma. *Cancer Letters*, 1(498):178–187, 2021.

[70] C. Meaney, G. G. Powathil, A. Yaromina, L. J. Dubois, P. Lambin, and M. Kohandel. Role of hypoxia-activated prodrugs in combination with radiation therapy: An in silico approach. *Mathematical Biosciences and Engineering*, 16(6):6257–6273, 2019.

[71] C. Meaney, S. Rhebergen, and M. Kohandel. In silico analysis of hypoxia activated prodrugs in combination with anti angiogenic therapy through nanocell delivery. *PLoS Computational Biology*, 16(5), may 2020.

[72] C Meaney, M Stastna, M Kardar, and M Kohandel. Spatial optimization for radiation therapy of brain tumours. *PLoS ONE*, 14(6), 2019.

[73] Cameron Meaney, Sunit Das, Errol Colak, and Mohammad Kohandel. Deep learning characterization of brain tumours with diffusion weighted imaging. *Journal of Theoretical Biology*, 557:111342, January 2023.

[74] Cameron Meaney, Mohammad Kohandel, and Arian Novruzi. Temporal optimization of radiation therapy to heterogeneous tumour populations and cancer stem cells. *Journal of Mathematical Biology*, 85(5):51, October 2022.

[75] Cameron Meaney, Sander Rhebergen, and Mohammad Kohandel. In silico analysis of hypoxia activated prodrugs in combination with anti angiogenic therapy through nanocell delivery. *PLOS Computational Biology*, 16(5):e1007926, May 2020. Publisher: Public Library of Science.

[76] N. Meghdadi, M. Soltani, H. Niroomand-Oscuii, and N. Yamani. Personalized image-based tumor growth prediction in a convection–diffusion–reaction model. *Acta Neurologica Belgica*, 120(1):49–57, feb 2020.

[77] H. J. Meyer, N. Garnov, and A. Surov. Comparison of Two Mathematical Models of Cellularity Calculation. *Translational Oncology*, 11(2):307–310, apr 2018.

[78] Michael Milosevic, Anthony Fyles, David Hedley, Melania Pintilie, Wilfred Levin, Lee Manchul, and Richard Hill. Interstitial Fluid Pressure Predicts Survival in Patients

with Cervix Cancer Independent of Clinical Prognostic Factors and Tumor Oxygen Measurements. *Cancer Research*, 61, 2001.

[79] S. Mitra. Deep Learning with Radiogenomics towards Personalized Management of Gliomas. *IEEE Reviews in Biomedical Engineering*, 3333(c):1–17, 2021.

[80] M Mizuta, S Takao, H Date, N Kishimoto, K L Sutherland, R Onimaru, and H Shirato. A mathematical study to select fractionation regimen based on physical dose distribution and the linear-quadratic model. *International Journal of Radiation Oncology Biology Physics*, 84(3):829–833, 2012.

[81] Farshad Moradi Kashkooli, M. Soltani, Mohammad Souri, Cameron Meaney, and Mohammad Kohandel. Nexus between in silico and in vivo models to enhance clinical translation of nanomedicine. *Nano Today*, 36:101057, February 2021.

[82] J. D. Murray. *Mathematical Biology: I. An Introduction, Third Edition*. Springer, 3rd edition, 2002.

[83] J D Murray. *Mathematical Biology II: Spatial Models and Biomedical Applications, Third Edition*. Springer, 3rd edition, 2002.

[84] Barbara Muz, Pilar de la Puente, Feda Azab, and Abdel Kareem Azab. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. *Hypoxia*, 3:83–92, December 2015.

[85] Long V. Nguyen, Robert Vanner, Peter Dirks, and Connie J. Eaves. Cancer stem cells: An evolving concept. *Nature Reviews Cancer*, 12(2):133–143, 2012.

[86] P. C. Nowell. The clonal evolution of tumor cell populations. *Science (New York, N.Y.)*, 194(4260):23–28, October 1976.

[87] Uresh Patel, Ankush Patel, Charles Cobb, Tara Benkers, and Sandra Vermeulen. The management of brain necrosis as a result of srs treatment for intra-cranial tumor. *Translational Cancer Research*, 3:373–382, 8 2014.

[88] S. Pati, U. Baid, M. Zenk, B. Edwards, M. Sheller, G. A. Reina, P. Foley, A. Gruzdev, J. Martin, S. Albarqouni, Y. Chen, R. T. Shinohara, A. Reinke, D. Zimmerer, J. B. Freymann, J. S. Kirby, C. Davatzikos, R. R. Colen, A. Kotrotsou, D. Marcus, M. Milchenko, A. Nazeri, H. Fathallah-Shaykh, R. Wiest, A. Jakab, M. Weber, A. Mahajan, L. Maier-Hein, J. Kleesiek, B. Menze, K. Maier-Hein, and S. Bakas. The Federated Tumor Segmentation (FeTS) Challenge. *ArXiv*, pages 1–15, 2021.

[89] Sarthak Pati, Vaibhav Sharma, Heena Aslam, Siddhesh P. Thakur, Hamed Akbari, Andreas Mang, Shashank Subramanian, George Biros, Christos Davatzikos, and Spyridon Bakas. Estimating glioblastoma biophysical growth parameters using deep learning regression. *Lecture Notes in Computer Science*, 12658 LNCS:157–167, 2021.

[90] Sarthak Pati, Vaibhav Sharma, Heena Aslam, Siddhesh P. Thakur, Hamed Akbari, Andreas Mang, Shashank Subramanian, George Biros, Christos Davatzikos, and Spyridon Bakas. Estimating glioblastoma biophysical growth parameters using deep learning regression. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12658 LNCS:157–167, 2021.

[91] Dan Peer, Jeffrey M. Karp, Seungpyo Hong, Omid C. Farokhzad, Rimona Margalit, and Robert Langer. Nanocarriers as an emerging platform for cancer therapy. *Nature Nanotechnology*, 2(12):751–760, December 2007. Number: 12 Publisher: Nature Publishing Group.

[92] T M Phillips, W H McBride, and F Pajonk. The response of CD24-/low/CD44+ breast cancer-initiating cells to radiation. *Journal of the National Cancer Institute*, 98(24):1777–1785, 2006.

[93] M Pierre. Global existence in reaction-diffusion systems with control of mass: a survey. *Milan J. Math.*, 78:417–455, 2010.

[94] F Pizarro and A Hernández. Optimization of radiotherapy fractionation schedules based on radiobiological functions. *British Journal of Radiology*, 90(1079):1–5, 2017.

[95] G. Powathil, M. Kohandel, S. Sivaloganathan, A. Oza, and M. Milosevic. Mathematical modeling of brain tumors: Effects of radiotherapy and chemotherapy. *Physics in Medicine and Biology*, 52(11):3291–3306, jun 2007.

[96] M. Raissi and G. E. Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, mar 2018.

[97] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, feb 2019.

[98] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics Informed Learning Machine, 2020.

[99] J Ramakrishnan. *Dynamic Optimization of Fractionation Schedules in Radiation Therapy*. PhD thesis, Massachusetts Institute of Technology, 2013.

[100] R. Rockne, E. C. Alvord, J. K. Rockhill, and K. R. Swanson. A mathematical model for brain tumor response to radiation therapy. *Journal of Mathematical Biology*, 58(4-5):561–578, apr 2009.

[101] R. Rockne, J. K. Rockhill, M. Mrugala, A. M. Spence, I. Kalet, K. Hendrickson, A. Lai, T. Cloughesy, E. C. Alvord, and K. R. Swanson. Predicting the efficacy of radiotherapy in individual glioblastoma patients in vivo: A mathematical modeling approach. *Physics in Medicine and Biology*, 55(12):3271–3285, 2010.

[102] Russell C. Rockne, Andrew D. Trister, Joshua Jacobs, Andrea J. Hawkins-Daarud, Maxwell L. Neal, Kristi Hendrickson, Maciej M. Mrugala, Jason K. Rockhill, Paul Kinahan, Kenneth A. Krohn, and Kristin R. Swanson. A patient-specific computational model of hypoxia-modulated radiation resistance in glioblastoma using 18f-fmiso-pet. *Journal of the Royal Society Interface*, 12, 2 2015.

[103] Klaudius Scheufele, Shashank Subramanian, Andreas Mang, George Biros, and Miriam Mehl. Image-driven biophysical tumor growth model calibration. *SIAM Journal on Scientific Computing*, 42:B549–B580, 2020.

[104] L Ridgway Scott. *Introduction to Automated Modeling using FEniCS*. Computational Modeling Initiative LLC, 2018.

[105] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports*, 10(1):1–12, 2020.

[106] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S Bakas. Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation. *Brainlesion*, 1(11383):92–104, 2019.

[107] Rebecca L. Siegel, Kimberly D. Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1):17–48, 2023. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21763.

[108] Joshua H. Smith and Joseph A.C. Humphrey. Interstitial transport and transvascular fluid exchange during infusion into brain and tumor tissue. *Microvascular Research*, 73(1):58–73, January 2007.

[109] M. Soltani and P. Chen. Numerical Modeling of Interstitial Fluid Flow Coupled with Blood Flow through a Remodeled Solid Tumor Microvascular Network. *PLoS ONE*, 8(6):e67025, June 2013.

[110] Shawn Stapleton, Michael Dunne, Michael Milosevic, Charles W. Tran, Matthew J. Gold, Ali Vedadi, Trevor D. McKee, Pamela S. Ohashi, Christine Allen, and David A. Jaffray. Radiation and heat improve the delivery and efficacy of nanotherapeutics by modulating intratumoral fluid dynamics. *ACS Nano*, 12(8):7583–7600, August 2018. Publisher: American Chemical Society.

[111] Shawn Stapleton, Michael Milosevic, Christine Allen, Jinzi Zheng, Michael Dunne, Ivan Yeung, and David A. Jaffray. A Mathematical Model of the Enhanced Permeability and Retention Effect for Liposome Transport in Solid Tumors. *PLOS ONE*, 8(12):e81157, December 2013. Publisher: Public Library of Science.

[112] N A Stavreva, P V Stavrev, B Warkentin, and B G Fallone. Investigating the effect of cell repopulation on the tumor response to fractionated external radiotherapy. *Medical Physics*, 30(5):735–742, may 2003.

[113] R. Stupp, S. Taillibert, A. Kanner, W. Read, D. M. Steinberg, B. Lhermitte, S. Toms, A. Idbaih, M. S. Ahluwalia, K. Fink, F. Di Meco, F. Lieberman, J. J. Zhu, G. Stragliotto, D. D. Tran, S. Brem, A. F. Hottinger, E. D. Kirson, G. Lavy-Shahaf, U. Weinberg, C. Y. Kim, S. H. Paek, G. Nicholas, J. Burna, H. Hirte, M. Weller, Y. Palti, M. E. Hegi, and Z. Ram. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma a randomized clinical trial. *JAMA - Journal of the American Medical Association*, 318(23):2306–2316, 2017.

[114] Shashank Subramanian, Klaudius Scheufele, Naveen Himthani, and George Biros. Multiatlas calibration of biophysical brain tumor growth models with mass effect. *Lecture Notes in Computer Science*, 12262 LNCS:551–560, 2020.

[115] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, T. Okuda, L. Liang, Y. Ge, Y. Komohara, Y. Ushio, and M. Takahashi. Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas. *Journal of Magnetic Resonance Imaging*, 9(1):53–60, 1999.

[116] A. Surov, H. J. Meyer, and A. Wienke. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis. *Oncotarget*, 8(35):59492–59499, 2017.

[117] K. R. Swanson, E. C. Alvord, N. D. Alvord, J. D. Murray, and R. Rockne. Method and System for Characterizing Tumors, mar 2013.

[118] K. R. Swanson, E. C Alvord, and J. D. Murray. A quantitative model for differential motility of gliomas in grey and white matter. *Cell Proliferation*, 33(5):317–329, 2000.

[119] K. R. Swanson, C. Bridge, J. D. Murray, and E. C. Alvord. Virtual and real brain tumors: Using mathematical modeling to quantify glioma growth and invasion. *Journal of the Neurological Sciences*, 216(1):1–10, dec 2003.

[120] Nathaniel Swinburne, Eve LoCastro, Ramesh Paudyal, Jung Hun Oh, Neil K. Taunk, Akash Shah, Kathryn Beal, Behroze Vachha, Robert J. Young, Andrei I. Holodny, Amita Shukla-Dave, and Vaios Hatzoglou. Computational Modeling of Interstitial Fluid Pressure and Velocity in Non-small Cell Lung Cancer Brain Metastases Treated With Stereotactic Radiosurgery. *Frontiers in Neurology*, 11:402, May 2020.

[121] F. Szczepankiewicz, D. van Westen, E. Englund, C. F. Westin, F. Ståhlberg, J. Lätt, P. C. Sundgren, and M. Nilsson. The link between diffusion MRI and tumor heterogeneity: Mapping cell eccentricity and density by diffusional variance decomposition (DIVIDE). *NeuroImage*, 142:522–532, nov 2016.

[122] Mindy D. Szeto, Gargi Chakraborty, Jennifer Hadley, Russ Rockne, Mark Muzi, Ellsworth C. Alvord, Kenneth A. Krohn, Alexander M. Spence, and Kristin R. Swanson. Quantitative metrics of net proliferation and invasion link biological aggressiveness assessed by mri with hypoxia assessed by fmiso-pet in newly diagnosed glioblastomas. *Cancer Research*, 69:4502–4509, 5 2009.

[123] Birkan Tunc, David Hormuth, George Biros, and Thomas E. Yankeelov. Modeling of glioma growth with mass effect by longitudinal magnetic resonance imaging. *IEEE Transactions on Biomedical Engineering*, 68:3713–3724, 12 2021.

[124] M. Uhl, C. Altehoefer, U. Kontny, K. Il'yasov, M. Büchert, and M. Langer. MRI-diffusion imaging of neuroblastomas: First results and correlation to histology. *European Radiology*, 12(9):2335–2338, sep 2002.

[125] S M van Neerven, M Tieken, L Vermeulen, and M F Bijlsma. Bidirectional interconversion of stem and non-stem cancer cell populations: A reassessment of theoretical models for tumor heterogeneity. *Molecular and Cellular Oncology*, 3(2):1–9, 2016.

[126] Javier E. Villanueva-Meyer, Marc C. Mabray, and Soonmee Cha. Current clinical brain tumor imaging. *Clinical Neurosurgery*, 81:397–415, 9 2017.

[127] Jane E. Visvader and Geoffrey J. Lindeman. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature Reviews. Cancer*, 8(10):755–768, October 2008.

[128] Huo Vrenken, Petra Pouwels, Jeroen J G Geurts, Dirk L Knol, Chris H Polman, Frederik Barkhof, and Jonas A Castelijns. Altered diffusion tensor in multiple sclerosis normal-appearing brain tissue: Cortical diffusion changes seem related to clinical deterioration. *Journal of Magnetic Resonance Imaging*, 22:628–636, 2006.

[129] L M Wein, J E Cohen, and J T Wu. Dynamic optimization of a linear-quadratic model with incomplete repair and volume-dependent sensitivity and repopulation. *International Journal of Radiation Oncology Biology Physics*, 47(4):1073–1083, 2000.

[130] Nathan S. White and Anders M. Dale. Distinct effects of nuclear volume fraction and cell diameter on high b-value diffusion mri contrast in tumors. *Magnetic Resonance in Medicine*, 72:1435–1443, 11 2014.

[131] Paula Whitmire, Cassandra R. Rickertsen, Andrea Hawkins-Daarud, Eduardo Carrasco, Julia Lorence, Gustavo De Leon, Lee Curtin, Spencer Bayless, Kamala Clark-Swanson, Noah C. Peeri, Christina Corpuz, Christine Paula Lewis-De Los Angeles, Bernard R. Bendok, Luis Gonzalez-Cuyar, Sujay Vora, Maciej M. Mrugala, Leland S. Hu, Lei Wang, Alyx Porter, Priya Kumthekar, Sandra K. Johnston, Kathleen M. Egan, Robert Gatenby, Peter Canoll, Joshua B. Rubin, and Kristin R. Swanson. Sex-specific impact of patterns of imageable tumor growth on survival of primary glioblastoma patients. *BMC Cancer*, 20(1):1–10, 2020.

[132] Graeme F. Woodworth, Gavin P. Dunn, Elizabeth A. Nance, Justin Hanes, and Henry Brem. Emerging insights into barriers to effective brain tumor therapeutics. *Frontiers in Oncology*, 4 JUL, 2014.

[133] Yihan Yao, Yunxiang Zhou, Lihong Liu, Yanyan Xu, Qiang Chen, Yali Wang, Shijie Wu, Yongchuan Deng, Jianmin Zhang, and Anwen Shao. Nanoparticle-Based Drug Delivery in Cancer Therapy and Its Role in Overcoming Drug Resistance. *Frontiers in Molecular Biosciences*, 7, 2020.

[134] Y. Yin, O. Sedlaczek, B. Müller, A. Warth, M. González-Vallinas, B. Lahrmann, N. Grabe, H. U. Kauczor, K. Breuhahn, I. E. Vignon-Clementel, and D. Drasdo. Tumor Cell Load and Heterogeneity Estimation from Diffusion-Weighted MRI Calibrated with Histological Data: An Example from Lung Cancer. *IEEE Transactions on Medical Imaging*, 37(1):35–46, jan 2018.

[135] M. Zaider and G. N. Minerbo. Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Physics in Medicine & Biology*, 45(2):279, February 2000.

[136] Ke Zeng, Spyridon Bakas, Aristeidis Sotiras, Hamed Akbari, Martin Rozycki, Saima Rathore, Sarthak Pati, and Christos Davatzikos. Segmentation of gliomas in pre-operative and post-operative multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. *Brainlesion*, 10154:184–194, 2016.