12-2023

Deep Learning and Generative AI Approaches for Automated Diagnosis and Personalized Treatment: Bridging Machine Learning, Medicine, and Biomechanics in Predicting Tissue Mechanics and Biomaterial Properties.

Yasin Shokrollahi
Florida Institute of Technology, yshokrollahi2020@my.fit.edu

Deep Learning and Generative AI Approaches for Automated Diagnosis and Personalized Treatment: Bridging Machine Learning, Medicine, and Biomechanics in Predicting Tissue Mechanics and Biomaterial Properties.

by

Yasin Shokrollahi

A dissertation submitted to the College of Engineering and Science of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Melbourne, Florida
December 2023

We the undersigned committee hereby approve the attached dissertation,
"Deep Learning and Generative AI Approaches for Automated Diagnosis and Personalized Treatment: Bridging Machine Learning, Medicine, and Biomechanics in Predicting Tissue Mechanics and Biomaterial Properties."
by
Yasin Shokrollahi

_____

Linxia Gu, Ph.D.
Professor
Mechanical and Civil Engineering
Major Advisor

_____

Darshan G. Pahinkar, Ph.D.
Assistant Professor
Mechanical and Civil Engineering

_____

Ryan T. White, Ph.D.
Assistant Professor
Mathematics and Systems Engineering

_____

Pengfei Dong, Ph.D.
Assistant Professor
Biomedical Engineering and Science

_____

Xianqi Li, Ph.D.
Assistant Professor
Mathematics and Systems Engineering

_____

Ashok Pandit, Ph.D., P.E.
Professor and Department Head
Mechanical and Civil Engineering

Abstract

Deep Learning and Generative AI Approaches for Automated Diagnosis and Personalized Treatment: Bridging Machine Learning, Medicine, and Biomechanics in Predicting Tissue Mechanics and Biomaterial Properties.

Author: Yasin Shokrollahi

Advisor: Linxia Gu, Ph.D.

Machine learning, particularly deep neural networks, has demonstrated significant potential in predicting high-dimensional tasks across various domains. This work encompasses a detailed review of Generative AI in healthcare and three studies integrating machine learning with finite element analysis for predicting biomechanical behaviors and properties. Initially, we provide a comprehensive overview of Generative AI applications in healthcare, focusing on Transformers and Denoising Diffusion models and suggesting potential research avenues to address existing challenges.

Subsequently, we addressed soccer-related ocular injuries by combining finite element analysis and machine learning to predict retinal mechanics following a soccer ball hit rapidly. The prediction errors are between 3.03% and 16.40% for peak von Mises stress and maximum principal strain, respectively.

Furthermore, we developed end-to-end deep learning tools as an efficient alternative to finite element methods for predicting stress-strain fields within arterial walls, which is crucial for understanding atherosclerosis. Utilizing U-Net-based convolutional neural networks and conditional generative adversarial networks, combined with ensemble and transfer learning techniques, we developed models that accurately predicted von Mises stress and strain fields in the artery.

Lastly, we developed a finite element-based machine learning model for predicting the mechanical properties of bioglass-collagen composite hydrogels. This model effectively predicted Young's modulus and Poisson's ratio with 95% and 83% R-squared values.

This work aims to advance the intersection of machine learning, biomechanics, and healthcare. By developing accurate models for predicting biomechanical behavior and mechanical properties and providing a comprehensive review of Generative AI applications, we hope to contribute to addressing pervasive public health issues.

Table of Contents

List of Figures

List of Tables

Acknowledgement

First and foremost, I extend my deepest gratitude to my research advisor, Professor Linxia Gu. Her unwavering support, insightful guidance, and expert training were invaluable to my journey. It has been a privilege to be her graduate student. I also wish to express my heartfelt appreciation to my qualification and dissertation committee members, Professor Ryan T. White, Professor Darshan G. Pahinkar, Professor Pengfei Dong and Professor Xianqi Li. Their wisdom and constructive feedback have been instrumental to this work. Additionally, I acknowledge the members of our lab, from past and present Ph.D. students to master's students and the incredibly dedicated undergraduates. Their discussions and support have enriched this experience immensely, and I have learned from each of them. I am thankful for the friendships I forged within the lab and across various campus departments.

I extend my thanks to the mechanical engineering and biomedical engineering departments, graduate advisors, and student advisors for their continuous support. Special thanks go to my loving and supportive family. My parents' endless love, support, and guidance have been my foundation since day one, and I owe them everything – quite literally and figuratively. My wonderful sisters, who I hold dear, have always been my strongest supporters and advocates. In the grand tapestry of life, we are threads intertwined with others, and it is through these connections that we find meaning and purpose. This journey has been a tapestry woven with the threads of many, and for that, I am profoundly grateful.

In closing, I am grateful for everyone who has been part of this journey. Your collective support has made all the difference. The pursuit of knowledge is a collaborative endeavor, and I am humbled by the contributions of all who have been part of this journey.

Thank you all!

Dedication

In the Name of God, the most Gracious, the most Merciful

... Dedicated to all those that needed a second chance ..

Chapter 1
Introduction

Advanced computational models, imaging methods, and machine learning approaches, including deep learning methods, have shown great promise in forecasting stress-strain field distributions, assessing rupture risks, and predicting the mechanical properties of tissues and composite materials. Machine learning methods, particularly deep learning methods, have been extensively utilized to solve engineering problems, such as predicting atomic and molecular properties, crystal structures and stability, stent expansion, and retinal mechanics (Guo et al. 2021; Shokrollahi, Dong, Kaya, et al. 2022; Dong et al. 2020; Garg et al. 2022). Specifically, deep neural network (DNN) approaches have proven efficient in uncovering unique structures. Given sufficient training data, DNNs can use high-dimensional feature vectors from the original data and learn the nonlinear relationships between the feature vectors and expected outputs (Ziletti et al. 2018; Salmenjoki, Alava, and Laurson 2018) and (Ye et al. 2019; Ye et al. 2021) demonstrated that DNNs could accurately map the effective mechanical properties (Young's modulus and Poisson's ratio) and microstructures of composites. (Ford et al. 2021) used supervised machine learning to predict the homogenized elastic properties of two-phase materials. (C. Yang et al. 2020) predicted binary composites' stress–strain curve using principal component analysis and convolutional neural networks (CNN). This thesis focuses on the application of Deep Learning and Generative AI across three distinct but interconnected scales: micro-level, meso-level, and macro-level. Integrated machine learning and biomechanics were applied to predict arterial mechanics, ocular injuries induced by soccer ball impact, and bioglass-collagen composite hydrogel properties.

The first case study regards Atherosclerosis, a multifaceted cardiovascular ailment characterized by plaque accumulation in the arterial walls, leading to arterial constriction and stiffening. This condition is a major cause of heart disease, stroke, and other cardiovascular disorders, making it a leading cause of global mortality. It is influenced by various genetic, environmental, and lifestyle factors, such as elevated cholesterol levels, hypertension, tobacco use, obesity, and diabetes (Weber and Noels 2011; Libby, Ridker, and

Maseri 2002). The mechanical properties of the plaque, including its rigidity and resilience, are critical in determining its stability and potential for rupture. Forecasting the stress-strain field distributions of plaque is key to identifying areas susceptible to rupture. Various studies have highlighted the potential of computational models and imaging methods in predicting the stress-strain field distributions of atherosclerotic plaques and assessing their rupture risk (Church and Miller 2016; Krams et al. 1997).

Additionally, ocular injuries are frequently reported in robust sports, predominantly affecting individuals under 25. Soccer is the most common cause of sports-related ocular trauma (Yan 2020) due to the high speed of the soccer ball and lack of protection at all playing levels. While the orbit, the bone structure around the eyeball, can absorb the kinetic energy of a fast-moving soccer ball, experiments have shown that deformable soccer balls can directly contact the eyeball (Vinger and Filipe 2004), leading to various levels of ocular injury such as hyphemia, corneal abrasions, traumatic retinal edema, retinal breaks and detachment, macular hole, and choroidal hemorrhages (Ghosh and Bauer 1995; MacEwen 1989; Reed et al. 2002; J.A.C. Filipe, Fernandes, et al. 2003; J.C. Filipe, Rocha-Sousa, et al. 2003; Vinger and Filipe 2004; Yan 2020; Kent et al. 2007; Gökçe et al. 2013; Horn et al. 2000; Hassan et al. 2018), potentially resulting in vision loss. Notably, hyphemia and retinal hemorrhage are observed in over 30% of soccer ball trauma cases (J.A.C. Filipe, Fernandes, et al. 2003). They are caused by the rupture of blood vessels in the eye, leading to blood leakage into the anterior chamber or vitreous, partially or entirely obscuring vision and potentially leading to permanent vision problems (Jalali 2003; Saxena et al. 2003; Wollensak and Spoerl 2004). (Horn et al. 2000) reported 13 retinal injury cases caused by soccer balls, half of which required surgical intervention. (J.A.C. Filipe, Fernandes, et al. 2003) noted that severe ocular lesions could occur in soccer players without any symptoms and at all skill levels. (Leshno et al. 2021) analyzed posterior eye injuries in children and found that soccer-ball blunt trauma could cause major posterior eye injuries, even without external injury or ocular complaints, suggesting that all soccer players should consider eye protection, regardless of age.

In tissue engineering, biocompatible materials, or biomaterials, are used to create hydrogels or scaffolds for repairing or replacing damaged or diseased tissues. To minimize

scar tissue formation at the interface with host tissues, biomimetic materials that mimic the physicochemical properties of native tissue are preferred. Collagen (COL), the most abundant protein in mammals (Eglin et al. 2006), can be reinforced with bioglass (BG) to create composite scaffolds with enhanced mechanical properties for bone tissue engineering applications (Sarker et al. 2015). (Kajave et al. 2021) incorporating BG into COL reduced swelling and improved COL hydrogels' stability and rheological properties (i.e., yield stress). (Gurumurthy and Janorkar 2021) reported in a review paper that COL scaffold stiffness significantly increased upon adding BGs. The mechanical properties of BG-COL composite scaffolds can be estimated using mathematical models or finite element (FE) methods (Sousa et al. 2021). Homogenization methods, such as the double inclusion method, Mori–Tanaka mean field method, and self-consistent approaches, are generally practical for simple microstructures (Das et al. 2015; Hori and Nemat-Nasser 1993; C.-C. Yang and Huang 1996; Hua and Gu 2013). The representative volume element (RVE) technique in the FE method is widely used to estimate the effective properties of composite materials (Zhou et al. 2016; G.-D. Cheng, Cai, and Xu 2013; Omairey, Dunning, and Sriramula 2019). Previous studies have shown that the mechanical properties of BG-COL scaffolds depend on BG concentration, spatial distribution, particle size, and fabrication process (Swaminathan, Ghosh, and Pagano 2006).

In summary, predicting the arterial mechanics, ocular injuries, and biomaterial properties in tissue engineering is crucial for improving treatment strategies and preventing adverse outcomes. This thesis aims to delve deeper into the potential of machine learning methods, specifically deep learning, in predicting tissue material properties, assessing the risk of cardiovascular and ocular injuries, and ultimately contributing to developing more effective preventative and treatment strategies.

Dissertation Structure

This dissertation is dedicated to employing deep learning for the diagnosis, biomechanical analysis of cardiovascular diseases, prediction of ocular injuries, and material properties of biological tissues. We initiate our exploration with a crucial application of

generative AI in healthcare: a review, then diagnosis through medical imaging—then progress to predicting biomechanics of cardiovascular ailments, analyzing ocular injuries incurred during robust sports activities, and finally, estimating the material properties of biological tissues. Each of these pivotal areas is elaborated upon in the subsequent chapters.

Chapter 2: A Comprehensive Review of Generative AI in Healthcare The rapid progression of Artificial Intelligence (AI), specifically Generative AI models like Transformers and Denoising Diffusion models, has revolutionized the healthcare sector. These models have been employed for various applications, including, but not limited to, medical imaging, protein structure prediction, clinical documentation, radiology interpretation, medical coding and billing, and drug design. Their application has significantly improved clinical diagnosis, data reconstruction, and drug synthesis processes. This review paper aims to provide a comprehensive overview of the applications of Generative AI in healthcare, focusing on Transformers and Denoising Diffusion models. It suggests potential avenues for future research to address existing challenges and meet the changing needs of the healthcare industry. It is intended as a detailed guide for researchers and practitioners interested in healthcare applications of Generative AI, shedding light on the current state of the art, ongoing challenges, and potential future directions.

Chapter 3: Rapid Prediction of Retina Stress and Strain Patterns in Soccer-Related Ocular Injury: Integrating Finite Element Analysis with Machine Learning Approach: The study addresses the growing concern of soccer-related ocular injuries, particularly retinal injuries, by proposing a framework that combines finite element (FE) analysis and machine learning (ML) for rapid prediction of retinal mechanics. The mechanics of a flying soccer ball cause abnormally high retinal stresses and strains, which were previously characterized using the FE method, but this method demands significant mechanical expertise and computational time, making it impractical for clinical settings. This study utilized FE simulations of different impact scenarios to obtain stress and strain patterns in the posterior retina, which were then used to train and test a partial least squares regression (PLSR) model for predicting soccer-induced retinal stress and strain. The model yielded prediction errors for peak von Mises stress and maximum principal strain between 3.03%

and 16.40% for different impact scenarios. This surrogate model of FE analysis enables fast prediction of retina dynamics following a soccer ball impact and could be a stepping stone towards developing a diagnostic tool for soccer-related ocular trauma.

Chapter 4: Deep Learning-based Prediction of Stress and Strain Maps in Arterial Walls for Improved Cardiovascular Risk Assessment: This study explores the potential of end-to-end deep learning tools as a more efficient alternative to finite element methods (FEM) for predicting stress-strain fields within 2D cross-sections of arterial walls, a key aspect of understanding atherosclerosis, a leading cause of global mortality. The researchers proposed a U-Net-based fully CNN and a conditional generative adversarial network (cGAN) to predict von Mises stress and strain distribution based on the spatial arrangement of calcification within arterial wall cross-sections. Ensemble approaches of U-Net and cGAN were also proposed to improve prediction accuracy further. The dataset used for this study was generated by implementing boundary conditions and extracting stress-strain field maps. The results indicated that the trained U-Net models and the cGAN models, combined with ensemble and transfer learning techniques, accurately predicted von Mises stress and strain fields with high structural similarity index scores (SSIM) and low mean squared errors (MSE) on a reserved test set. Overall, the study successfully developed a surrogate model for FEM analysis, which can accurately and efficiently predict stress-strain fields of arterial walls regardless of complex geometry and boundary conditions.

Chapter 5: Finite Element-Based Machine Learning Model for Predicting the Mechanical Properties of Composite Hydrogels: This study developed a finite element (FE)-based machine learning model to predict the mechanical properties of bioglass (BG)-collagen (COL) composite hydrogels. Microstructural images of BG-COL composite hydrogels were created based on scanning electron microscope observations, resulting in 2000 images with randomly distributed BG particles of varying diameters and volume fractions. These images and the mechanical properties calculated via FE simulations were used to train and test a convolutional neural network regression model. The resulting network effectively predicted the mechanical properties of the composite hydrogels, with R-squared

values of 95% and 83% for Young's modulus and Poisson's ratio, respectively. This model serves as a surrogate for finite element analysis, using microstructure images to predict BG-COL hydrogel mechanical properties, which is valuable for characterizing heterogeneous materials in big data-driven material designs.

Chapter 6: Conclusion and Future Work: In conclusion, the thesis is an interdisciplinary study combining machine learning, finite element analysis, and healthcare applications. It explores the role of Generative AI models in enhancing medical imaging, protein structure prediction, clinical documentation, and drug design. It introduces innovative frameworks for predicting soccer-related ocular injuries, cardiovascular risk, and mechanical properties of composite hydrogels. The study contributes to a deeper understanding of ocular and cardiovascular diseases and facilitates the development of novel therapeutic strategies and material designs. Future work includes expanding clinical diagnosis prediction tasks, integrating bioinformatics, advancing computer vision, developing deep learning pipelines, bridging finite element modeling and machine learning, and advancing Generative AI in healthcare. The study ultimately contributes to advancing healthcare applications of Generative AI, improved diagnosis and risk assessment in ocular and cardiovascular diseases, and the development of novel material designs. The closure emphasizes the resurgence and impact of deep learning and Generative AI on various sectors, including healthcare. It highlights the study's contribution to understanding and addressing cardiovascular disease by integrating finite element analysis with machine learning and developing innovative solutions.

Chapter 2
A Comprehensive Review of Generative AI in Healthcare

Abstract

The advancement of Artificial Intelligence (AI) has catalyzed revolutionary changes across various sectors, notably in healthcare. Among the significant developments in this field are the applications of generative AI models, specifically diffusion and transformers models. These models have played a crucial role in analyzing diverse forms of data, including medical imaging (encompassing image reconstruction, image-to-image translation, image generation, and image classification), protein structure prediction, clinical documentation, diagnostic assistance, radiology interpretation, clinical decision support, medical coding, and billing, as well as drug design and molecular representation. Such applications have enhanced clinical diagnosis, data reconstruction, and drug synthesis. This review paper aims to offer a thorough overview of the generative AI applications in healthcare, focusing on diffusion and transformer models. Moreover, we propose potential directions for future research to tackle the existing limitations and meet the evolving demands of the healthcare sector. Intended to serve as a comprehensive guide for researchers and practitioners interested in the healthcare applications of generative AI, this review provides valuable insights into the current state of the art, challenges faced, and prospective future directions.

Keywords: Healthcare; Generative AI; Diffusion models; Transformers, Review

1. Introduction

Generative models began their journey in the 1950s and have seen major improvements in the last decade, changing how we understand deep learning. In the early days, concepts like Hidden Markov and Gaussian Mixture models introduced simple data generation methods. However, when newer methods like neural networks came into the picture, they added a lot of energy and advancements to these traditional methods. Today, generative AI has shown its importance across various types of data Figure 2-1. Overview of Generative AI tasks. In the text, models like OpenAI's GPT series, as mentioned in (Brown et al. 2020a), can produce detailed and logical stories.

Regarding images, generative AI can be used for tasks such as creating new images, referenced in (Ramesh et al. 2021), improving the quality of existing ones, and even enhancing their resolution. Videos, which require understanding the sequence of scenes, can now be generated more effectively with the help of this technology, as seen in (Ho et al. 2022). Beyond just 2D images, creating 3D objects and scenes, crucial for technologies like virtual reality and video games, has seen revolutionary changes, as indicated by (Shen et al. 2021). One of the standout areas where generative AI has made a mark is in healthcare. It is now being used to predict how proteins are structured, a very important topic in biology, highlighted by (Madani et al. 2023) and supported by (Meyers, Fabian, and Brown 2021). This entire journey and evolution of generative models have been documented and studied by many researchers, including (Bao et al. 2017), (Razavi, Van den Oord, and Vinyals 2019), (Kong et al. 2020), (Oord et al. 2016), (X. Li et al. 2022), and (G. Yang et al. 2019).

Even though deep learning architectures can be quite complex, there has been renewed excitement about them recently. This excitement is mainly because of the introduction of models like Generative Adversarial Networks (GANs) (Goodfellow et al. 2020), Variational Autoencoders (VAEs) (Rezende, Mohamed, and Wierstra 2014), normalizing flows (Dinh, Sohl-Dickstein, and Bengio 2016), and NeRF (Neural Radiance Fields) (Mildenhall et al. 2021). There are also new models based on diffusion processes (Ho, Jain, and Abbeel 2020) that provide different ways of thinking and overcome some challenges found in the older

models, such as VAEs, EBMs, GANs, and normalizing flows. In Natural Language Processing (NLP), transformer-based models (Vaswani et al. 2017) perform better than older models which include Recurrent Neural Networks (RNNs) (Rumelhart, Hinton, and Williams 1985), Long Short-Term Memory networks (LSTMs) (Hochreiter and Schmidhuber 1997), and Gated Recurrent Units (GRUs) (Chung et al. 2014) Figure 2-2.

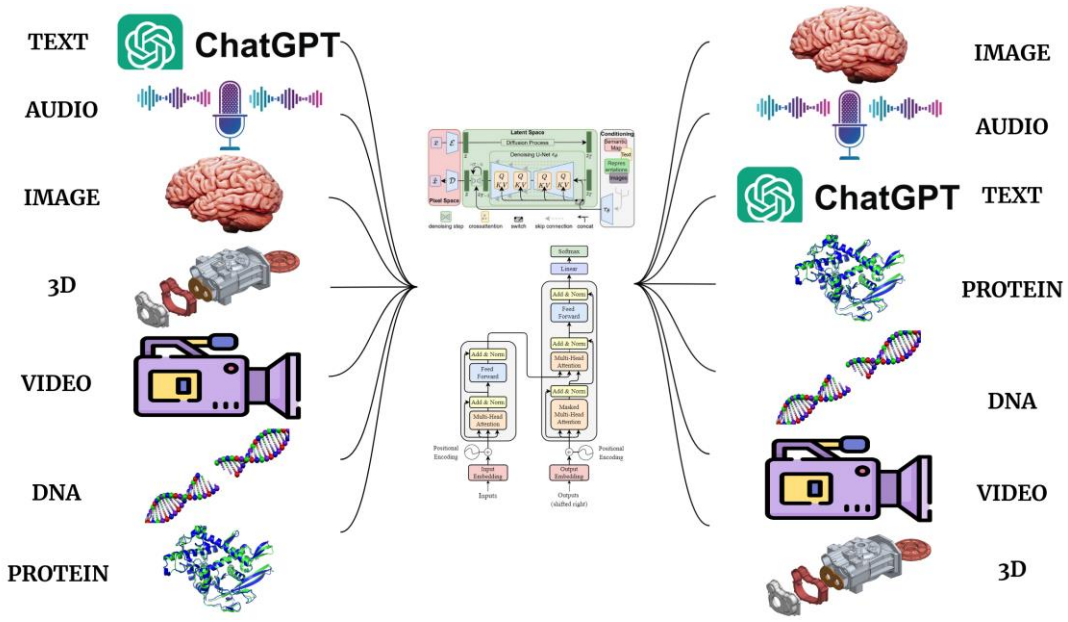


Figure 2-1. Overview of Generative AI tasks.

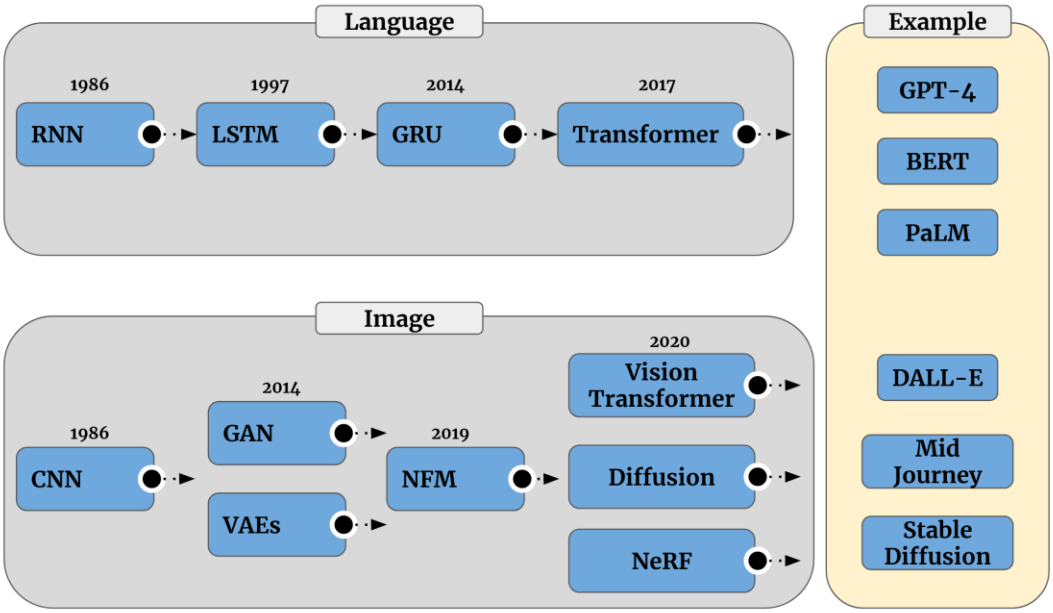


Figure 2-2. Timeline of Generative AI Family Types.

The healthcare industry is increasingly using generative AI, as shown in Figure 2-3. extensive research has been conducted on its diverse applications within the healthcare sector. Due to this growing academic attention, there have been many recent studies in this field. As the diffusion models in computer vision have showcased superior outcomes, and transformer-based models in the domain of large language models (LLMs) have consistently outperformed older models, we have predominantly selected these two models for discussion in the manuscript. It is now important to review this research. Our review comprehensively looks at the latest generative AI applications in healthcare. We believe ours is the first in-depth study of these AI techniques for healthcare, including diffusion models and transformer-based models and their combinations. We hope our work guides others and encourages more exploration of generative AI in this field. The main contributions of this paper are as follows:

- A pioneering review that extensively captures the application of generative AI in healthcare, encapsulating a comprehensive assessment of all pertinent studies up to September 2023.

- We present a clear classification of generative AI models in healthcare and divide them into two main types: diffusion models and transformer-based models. We further categorize their uses: diffusion models can complete tasks like image reconstruction, image to image translation, image generation, image classification, and other applications, while transformer-based models have been used for protein structure prediction, clinical documentation and information extraction, diagnostic assistance, medical imaging and radiology interpretation, clinical decision support, medical coding and billing, and drug design and molecular representation.
- Our focus is not solely on applications. We propose a novel classification wherein each study is broadly sorted based on its underlying algorithm and imaging methodology.
- AI types of models have been compared in various applications, with the pros and cons of each model being discussed. For each specific task in healthcare, the best model based on research and publications has been examined.

In conclusion, we addressed the looming challenges and potential issues and identified emerging trends. We posed thought-provoking questions regarding the future direction of generative AI in healthcare, covering both algorithms and practical considerations. **Moreover, we believe that generative AI has the potential to be recognized as a highly reliable assistant for medical professionals in the future, and its use could become widespread in healthcare.**

1.1. Rationale and Distinctiveness of Our Review:

In healthcare, generative AI has made significant progress over the years. Many experts have written detailed reviews about deep generative AI models designed specifically for healthcare purposes. These reviews include the studies by (Bohr and Memarzadeh 2020), (AlAmir and AlGhamdi 2022), (Ali et al. 2022), and (Kazerouni et al. 2023). Some of these reviews focus on specific uses or applications, while others explore different types of images produced by these models. Even though there have been previous reviews, numerous new advancements in healthcare have emerged since their publication.

Interestingly, there is a clear gap in the current reviews: they need to detail how generative AI is used in healthcare. This is a crucial topic because understanding this can push research in this area forward. Much potential in this field has not been fully explored yet. The healthcare community can get valuable insights by understanding the benefits and results of using generative AI, especially in producing images and LLMs.

This review examines past research comprehensively and what might be coming up regarding generative AI models in healthcare. We want to offer a complete viewpoint on this subject. Generative AI technology has shown great skill in creating artificial data. It is becoming a strong addition to our usual data and has a special role in some medical areas, which we will discuss in the next sections. With this paper, we want to guide medical experts in understanding and using these modern methods in their work. We look at these generative models and point out studies and writings that show how they are used in healthcare. However, we do not just stop explaining their use; we also explore how they work, the specific body parts they target, and the type of healthcare they relate to. By providing this deeper understanding, we hope to help researchers see how different studies and information connect, forming a clear story.

1.2. Search Methodology:

We searched several online platforms, namely DBLP, Google Scholar, PubMed, and Arxiv Sanity Preserver, using specific search queries designed for our study. These platforms allow for detailed search terms and provide lists of academic works. The types of content exhibited by these platforms include those from peer-reviewed journals, conference or workshop proceedings, articles that are not peer-reviewed, and early versions of papers. Our search used a set of terms related to advanced AI techniques and their healthcare applications. Specifically, we used search terms like (Generative AI* | transformer* | diffusion model*) (healthcare* | medicine* | medical* | medical imaging* | clinical* | diagnosis* | treatment*). After getting our search results, we filtered to ensure we only kept articles directly related to using generative AI models in healthcare. During our search, we came across many articles that, despite having relevant terms, were unrelated to our study. We did not include these in our final selection. When choosing which papers to focus on

more deeply, we looked for those that were innovative, made a major contribution to the field, were highly regarded, and set a precedent in the healthcare arena. After being thorough in our selection criteria, we chose a handful – specifically, two or three of the best papers – for a detailed review. However, it is important to mention that even though we aimed to give a comprehensive overview of the most important articles on this topic, there is a chance that we might have missed a few significant papers during our review.

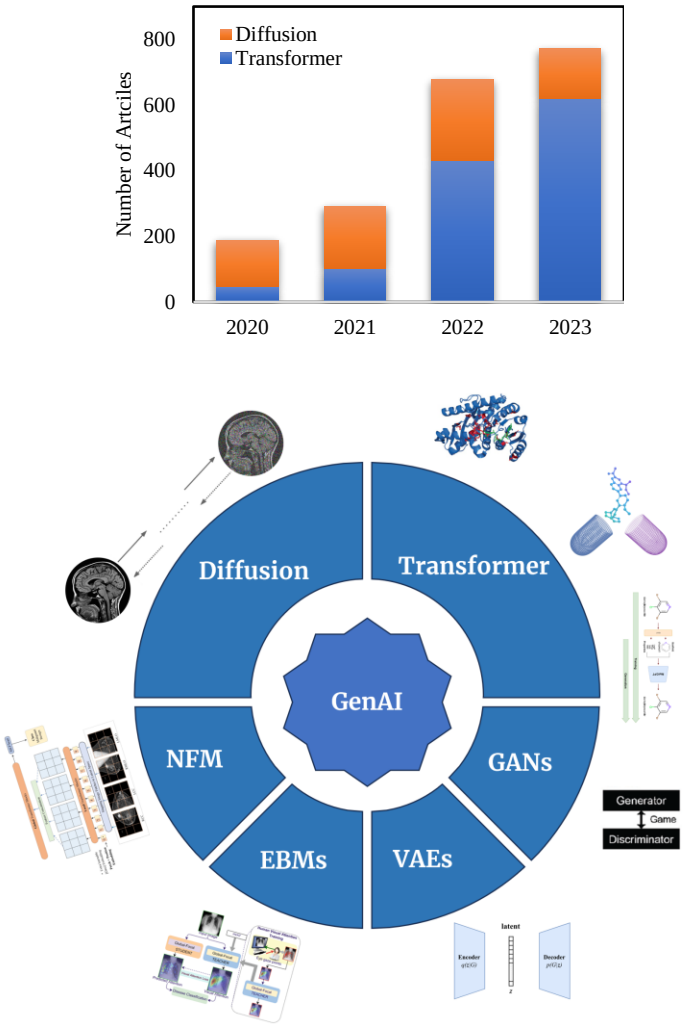


Figure 2-3. Generative AI in healthcare.

2. Generative AI models

As an advanced subclass of generative models, diffusion models have garnered significant attention for their adeptness at learning intricate data distributions. Their advent in the generative learning domain is recent, yet their utility is evident across various applications. The forthcoming section thoroughly explores the theoretical aspects underpinning diffusion models. We initiated this discussion by situating diffusion models within the wider context of generative learning, offering fresh comparative insight against other generative models. Subsequently, we categorize diffusion models from two principal vantage points: **the Variational and Score perspectives**. Taking a closer look at each perspective, we shed light on specific models classified under them. For instance, within the Variational Perspective, Denoising Diffusion Probabilistic Models (DDPMs) are notable, while Noise Conditional Score Networks (NCSNs) and Stochastic Differential Equations (SDEs) prominently feature within the Score Perspective. We aim to provide an in-depth, holistic comprehension of the theories that form the backbone of these methodologies. In the face of an exponential increase in accessible datasets and advancements in overarching deep learning architectures, a transformative shift in the landscape of generative modeling has emerged. Specifically, the three predominant generative frameworks, GANs (Goodfellow et al. 2014), VAEs (Rezende, Mohamed, and Wierstra 2014; Kingma and Welling 2013), and normalizing flows (Dinh, Sohl-Dickstein, and Bengio 2016), have come to the fore (refer to Fig. 2). Adoption of generative models in practical problem-solving contexts necessitates certain key characteristics. These include (i) the ability to produce high-quality samples, (ii) ensuring model coverage and diversity of samples, and (iii) guaranteeing quick execution times alongside computationally economical sampling (Xiao, Kreis, and Vahdat 2021). Generative models frequently strike a balance among different criteria. Specifically, GANs excel at rapidly generating high-quality samples, but they struggle with mode coverage and diversity of sampling (Wiatrak, Albrecht, and Nystrom 2019). In contrast, VAEs and normalizing flow models show proficiency in covering data modes but are often criticized for their low sample quality (Davidson et al. 2018; Asperti 2019).

GANs consist of two main components: a generator and a discriminator, also called a critic. They interact competently yet constructively, enhancing each other's performance. The generator's objective is to replicate the real data distribution, while the discriminator, a type of binary classifier, assesses the probability that a specific sample is from the real dataset. The role of the discriminator is to improve its capacity to distinguish between synthetic and real samples. Nonetheless, GANs frequently encounter difficulties related to unstable training dynamics, for example, mode collapse, disappearing gradients, and problems with convergence (Wiatrak, Albrecht, and Nystrom 2019). This has led to a change in research direction towards creating more effective GAN variants (Miyato et al. 2018; Motwani and Parmar 2020). VAEs, on the other hand, maximize the Evidence Lower Bound (ELBO) to optimize the data's log-likelihood. Despite significant progress, VAEs still face challenges related to theoretical and practical aspects, such as balancing problems and variable collapse issues (Davidson et al. 2018; Asperti 2019). Generative models based on normalizing flows utilize a series of reversible transformations to convert a basic distribution into a more intricate one, achieving the desired probability distribution for the final target variable using the change of variables theorem. Unlike GANs and VAEs, these models explicitly learn the data distribution, and their loss function is simply the negative log-likelihood. However, these models have their limitations due to the necessity of a particular architecture for the Likelihood-based method to create a normalized probability model.

Additionally, for VAEs, an alternative loss like ELBO is not directly computed for the generated probability distribution. The GAN learning process is inherently unstable due to the adversarial nature of the GAN loss. Recently, diffusion models have emerged as potent generative models, becoming a central area of focus in computer vision, challenging researchers and practitioners to keep up with the fast-paced advancements (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020).

2.1. Diffusion Models:

Diffusion models are a resilient class of probabilistic generative models designed to decipher intricate data distributions. This is possible through a bifurcated process involving forward and reverse diffusion. In the forward diffusion phase, noise is incorporated into the

input data, incrementally amplifying the noise level until it transforms into unadulterated Gaussian noise. This action alters the data distribution's architecture. Conversely, the reverse diffusion phase, also called denoising, is employed to restore the original architecture of the data from the altered data distribution. This action effectively reverses the deterioration instigated by the forward diffusion phase, culminating in a highly versatile and controllable generative model capable of precisely modeling intricate data distributions originating from random noise. The umbrella term 'Variational Perspective' includes models that leverage variational inference to approximate the desired distribution. This is commonly realized by minimizing the Kullback-Leibler divergence, a metric that quantifies the disparity between the approximate and desired distributions. A prime example of this class is the DDPMs. Initiated by (Sohl-Dickstein et al. 2015) and subsequently refined by (Ho, Jain, and Abbeel 2020), these models employ variational inference to approximate the parameters of a diffusion process (Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020).

2.1.1. Denoising Diffusion Probabilistic Models (DDPMs)

The Forward Diffusion Process. As outlined by DDPM, the forward diffusion process is depicted as a Markov Chain, characterized by the inclusion of Gaussian noise in a series of stages, culminating in the generation of noisy samples. It is important to note that the uncorrupted or original data distribution is represented as $q(x_0)$. With a data sample x_0 drawn from this distribution, $q(x_0)$, a forward noising operation, denoted as p, is employed. This operation introduces Gaussian noise iteratively at various time points, represented by t, resulting in a series of latent states x_1 through x_T. The process can be mathematically defined as follows:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t : \sqrt{1 - \beta_t}.\, x_{t-1}, \beta_t.\, \mathrm{I}\right), \forall_t \in \{1, \ldots, T\}, \quad (1)$$

In the context of this discussion, T denotes the number of diffusion steps, while $\beta 1, \ldots, \beta T$, each within the interval of [0, 1), signify the variance schedule spread throughout the diffusion steps. The identity matrix is symbolized by **I**, and $\mathcal{N}(x;\ \mu,\ \sigma)$, characterizes the normal distribution possessing a mean of μ and a covariance of σ. By introducing a_t as $1 -$

β_t and defining $\overline{a_t} = \prod_{s=0}^{t} a_s$, it becomes feasible to directly sample from any step of the noised latent, conditioned on the initial input x_0 as follows:

$$q(X_t \mid X_0) = N\left(X_t : \sqrt{\overline{a_t}}.X_0, (1 - \overline{a_t})\mathrm{I}\right), \forall_t \in \{1, \ldots, T\}, \tag{2}$$

$$X_t = \sqrt{\overline{a_t}}.X_0 + \sqrt{1 - \overline{a_t}\epsilon} \tag{3}$$

Reverse Process. Based on the definitions, we can construct an estimated inverse process for obtaining a sample from $q(x_0)$. This reverse process can be parameterized by initiating from $p(X_T) = \mathcal{N}(X_T; 0, \mathrm{I})$, as detailed below:

$$p_\theta(X_{0:T}) = p(X_T) \prod_{t=1}^{T} p_\theta\ (X_{t-1} \mid X_t) \tag{4}$$

$$p_\theta\ (X_{t-1} \mid X_t) = \mathcal{N}(X_{t-1}; \mu_\theta(X_t, t), \sum_\theta(X_t, t)) \tag{5}$$

The objective of training this model is to allow $p(x_0)$ to accurately learn and mimic the actual data distribution $q(x_0)$. This can be achieved by optimizing a specific variational bound associated with negative log-likelihood. This approach aligns with established methodologies and theories in the field (refer to sources).

$$\mathbb{E}\left[-\log p_\theta(X_0)\right] \leq\ \mathbb{B}_q\ \left[-\log \frac{p_\theta(X_{0:T})}{q(X_{1:T} \mid X_0)}\right] =$$

$$\mathbb{E}_q\ \left[-\log p(X_T) - \sum_{t \geq 1} \log \frac{p_\theta(X_{t-1} \mid X_t)}{q(X_t \mid X_{t-1})}\right] =\ -L_{VL.B} \tag{6}$$

(Ho, Jain, and Abbeel 2020) proposed an innovative approach to parameterization. Instead of directly parameterizing $\mu_\theta(x_t, t)$ via a neural network, they developed a model $\epsilon_\theta(x_t, t)$ to predict ϵ. This approach allowed them to reparametrize Equation (6), consequently simplifying the objective function as:

$$L_{simple} = E_{t, x_0, \epsilon}\left[\|\epsilon -\ \epsilon_\theta(x_t, t)\|^2\right] \tag{7}$$

where the authors draw a connection between the loss in Eq. (6) to generative score networks in (Song and Ermon 2019).

2.1.2. Score Perspective

The models based on scores employ a technique grounded in maximum likelihood estimation, using the data's log-likelihood score function to ascertain the diffusion process parameters. There are two main subsets in this category: Noise-conditioned Score Networks (NCSNs), as studied by (Song and Ermon 2019), and SDEs, researched by (Song et al. 2020). The primary objective of NCSNs is to calculate the derivative of the log density function for data distributions affected by noise at different levels. Conversely, SDEs serve as a broader approach, encompassing the characteristics of DDPMs and NCSNs. In the subsequent sections, a comprehensive analysis of the unique attributes of each subset is provided.

2.1.2.1 Noise Conditioned Score Networks (NCSNs)

The score function of a given data distribution denoted as $p(x)$, is characterized as the gradient of the log density concerning the input, expressed mathematically as $\nabla_x \log p(x)$. This score function can be approximated using a shared neural network trained via a process known as score matching. This involves the utilization of a score network, represented as s_θ, a neural network model parameterized by θ. The core aim of s_θ is to mimic the score of $p(x)$ (depicted as $p(x)(s_\theta(x) \approx \nabla_x \log p(x))$ by minimizing the corresponding objective:

$$\mathbb{E}_{x\sim p(x)} \left\| s_\theta(x) - \nabla_x \log p(x) \right\|_2^2 \tag{8}$$

Computational complexity associated with calculating $\nabla_x \log p(x)$ tends to limit the scalability of score matching, particularly in deep networks and high-dimensional data. To address this issue, (Song and Ermon 2019) suggest the utilization of denoising score matching (Vincent 2011) and sliced score matching (Song et al. 2020). Moreover, (Song and Ermon 2019) underline significant obstacles that prevent a straightforward implementation of score-based generative modeling on real-world data. The primary issue is the inaccuracy of the estimated score functions in low-density regions since real-world data often cluster on

low-dimensional manifolds in a high-dimensional space, aligning with the manifold hypothesis. To counter these challenges, the authors demonstrate the efficacy of data perturbation using Gaussian noise at varying scales. This process renders the data distribution more compatible with score-based generative modeling. A proposal is made to calculate the score corresponding to all noise levels via the training of a singular noise-conditioned score network (NCSN). The authors derive $\nabla_x \log p(x)$ as $\nabla_{xt} \log p_{\sigma_t}(x_t \mid x) = -\frac{x_t - x}{\sigma_t}$, where the noise distribution is chosen to be $p_{\sigma_t}(x_t \mid x) = \mathcal{N}(x_t : x, \sigma_t^2 . \mathrm{I})$, with x_t being a noisy version of x. Thus, given a sequence of Gaussian noise scales $\sigma_1 < \sigma_2 < \ldots < \sigma_T$, the equation can be formulated as follows:

$$\frac{1}{T}\sum_{t=1}^{T} \lambda(\sigma_T)\, \mathbb{E}_{p(x)}\, \mathbb{E}_{x_t \sim p_{\sigma_t}(x_t \mid x)} \left\| s_\theta(x_t, \sigma_t) + \frac{x_t - x}{\sigma_t} \right\|_2^2 \tag{9}$$

The weighting function is denoted as $\lambda(\sigma_t)$. To carry out the inference, an iterative method known as "Langevin dynamics" is utilized, which was originally discussed in the studies by (Parisi 1981) and later by (Grenander and Miller 1994). The Langevin dynamics constructs an MCMC procedure intended to procure samples from a specific distribution $p(x)$, utilizing its exclusive score function $\nabla_x \log p(x)$. In precise terms, the process starts by taking a random sample x_0 from the distribution $\pi(x)$. The subsequent step entails the iteration of a predefined sequence of operations to facilitate a gradual shift from x_0 towards other samples that follow the distribution $p(x)$.

$$x_i = x_{i-1} + \frac{\gamma}{2} \nabla_x \log p(x) + \sqrt{\gamma} . \omega_i \tag{10}$$

In the established process, ω_i is considered to follow a normal distribution, $\mathcal{N}(0, \mathrm{I})$, where i ranges from 1 to N. It is further observed that when γ approaches 0 and N tends towards infinity, the x_i samples derived from this procedure ultimately converge to a sample from $p(x)$. Song and Ermon (2019) introduced a refinement to this algorithm, designating it as the "annealed Langevin dynamics" algorithm. This amendment was proposed to alleviate specific difficulties and failure points associated with score matching. In this amended

algorithm, the noise scale σ_i is progressively reduced or "annealed" over time (Song and Ermon 2019).

2.1.2.2. Stochastic Differential Equations (SDEs):

Building upon the concepts introduced by the prior two methodologies, score-based generative models (SGMs) proposed by (Song et al. 2020), employ a unique transformation of the data distribution, $q(x_0)$, into noise. Notably, the distinguishing feature of SGMs is their ability to generalize the number of noise scales to infinity. This property allows us to conceptualize the previously discussed probabilistic models as discretized forms of an SGM. It is noteworthy that many stochastic processes, including the diffusion process, are found to be solutions to a specific type of stochastic differential equation (SDE), formulated as follows:

$$dX = f(X,t)dt + g(t)dW \tag{11}$$

The function $f(.,t)$ signifies the drift coefficient of the Stochastic Differential Equation (SDE), while $g(t)$ symbolizes the diffusion coefficient. The variable w is indicative of standard Brownian motion. Within this framework, x_0 represents the untainted data sample and x_0 stands for the disturbed data mirroring the standard Gaussian distribution. A unique characteristic of the SDE is a corresponding reverse time SDE, which functions in the opposite direction. Starting with a sample from p_T and reversing this diffusion SDE process allows us to procure samples from our original data distribution p_0. The formulation for the reverse-time SDE is as follows:

$$dX = [f(X,t)dt - g^2(t)\nabla_x \log p_t(x)]\, dt + g(t)d\bar{w} \tag{12}$$

The terms dt and $\bar{w}$ represent the infinitesimally small negative time step and retrograde Brownian motion. To numerically resolve the reversed-time Stochastic Differential Equation (SDE), it is feasible to employ a neural network to approximate the actual score function. This approach, known as score matching, has been previously suggested and utilized in literature by (Song and Ermon 2019) and (Song et al. 2020). It allows for the estimation of $s_\theta(x,t)$ which can be represented as $\approx \nabla_x \log p_t(x)$

(highlighted in red in Equation (12)). This score model's training relies on a specific objective, which is givin as follows:

$$\mathcal{L}(\theta) = \mathbb{E}_{X_t \sim p(X(t)|\, X(0)), X(0) \sim p_{data}} \times \left[\frac{\lambda(t)}{2} \left\| s_\theta(X(t), t) - \nabla_{x(t)} \log p_t(X(t) \mid X(0)) \right\|_2^2 \right] \quad (13)$$

In the provided context, λ represents a weighting function, while t is drawn from a uniform distribution over the interval [0, T]. Notably, the term $\nabla_x \log p_t(x)$ is substituted by $\nabla_x \log p_{0t}(X(t) \mid X(0))$ to overcome certain technical challenges that may arise. It is also worth noting that the sampling process from Stochastic Differential Equations (SDEs) can be efficiently executed by applying various numerical methods, as demonstrated in Equation (12). The subsequent sections comprehensively discuss three prevalent techniques typically employed in this context.

1. **Euler-Maruyama (EM) Approach:** This strategy employs a straightforward discretization method where the infinitesimal time increment, dt, is replaced with a finite time increment, Δt, and the infinitesimal Wiener process, $d\overline{w}$, is substituted with a Gaussian random variable, z, that follows a normal distribution with mean 0 and variance Δt. I. This allows for the resolution of Equation (12).

2. **Prediction-Correction (PC) Strategy:** This technique revolves around a nested loop of predictive and corrective procedures. The initial data is projected forward before being corrected in a series of steps. The Euler-Maruyama method can be used to solve the prediction aspect. As for the corrective phase, any score-based Markov Chain Monte Carlo (MCMC) procedure can be applied, including but not limited to annealed Langevin dynamics. Consequently, Langevin dynamics can be utilized to resolve Equation (10).

3. **Probability Flow Ordinary Differential Equation (ODE) Approach:** The stochastic differential equations represented in Equation (11) can be transcribed into ordinary differential equations according to the following method:

$$dX = \left[f(X,t) - \frac{1}{2} g^2(t) \nabla_x \log p_t(x) \right] dt \quad (14)$$

Therefore, through the resolution of the Ordinary Differential Equations (ODE) problem, the value of $x0$ can be obtained. However, despite its solution efficiency, ODE lacks a stochastic component necessary for error rectification, leading to a marginal compromise in its performance.

2.2. Transformers-based Models:

Language performs multiple roles as the foundation for human communication, cooperation, and education. It is a medium for articulating various emotions, encompassing affection, apprehension, and melancholy. Additionally, it aids in comprehending intricate notions such as justice, politics, and morality and proves essential in tackling complex problem-solving activities. Importantly, language is intrinsic to the creative arts, as illustrated by our extensive literary legacy and varied musical creations. Language significantly acts as a medium for transmitting and conservating knowledge across generations.

Nevertheless, with the emergence of breakthroughs in AI, the Cartesian assertions about the exclusive human capacity for language are being challenged. The conventional human exclusive control over language is progressively diminishing, and AI is evolving to occupy this void. This scenario highlights the critical importance of LLMs, a groundbreaking progression in AI. To thoroughly understand the transformative capabilities of these models, it is essential to examine some key elements of related technologies, especially NLP.

Convolutional neural networks (CNNs) have typically been found to be less effective in NLP due to their inability to handle sequential data like text adequately. A more fitting alternative is the RNN, which has garnered widespread use due to its capacity to process such data. RNNs (Rumelhart, Hinton, and Williams 1985), specifically LSTM networks (Hochreiter and Schmidhuber 1997) and GRU networks (Chung et al. 2014), have been robustly recognized as leading-edge methodologies in sequence modeling and transduction challenges. This includes, but is not limited to, language modeling and machine translation (Bahdanau, Cho, and Bengio 2014; Cho et al. 2014; Sutskever, Vinyals, and Le

2014). In the aftermath of their development and successful implementation, a series of rigorous efforts have been launched to enhance further the capabilities of these recurrent language models and encoder-decoder architectures (Jozefowicz et al. 2016; Luong, Pham, and Manning 2015; Wu et al. 2016). Such endeavors underscore the growing potential and persistent refinement in this area of computational linguistics and AI. Recurrent architectures generally allocate computational processes according to the positional alignment of the input and output symbols within sequences. This implies that such models generate a series of hidden states, denoted as h_t, contingent on the prior hidden state h_{t-1} and the input at position t. However, the sequential characteristic inherent in these models obstructs within-sample parallelization, becoming particularly impactful when managing lengthy sequences due to memory restrictions constraining batching across different examples. Nevertheless, recent research has demonstrated considerable enhancements in computational efficiency via factorization methodologies (Kuchaiev and Ginsburg 2017) and conditional computation techniques (Shazeer et al. 2017). These advancements have not only led to efficiency improvements but also enhanced model performance. Yet, the intrinsic limitation associated with sequential computation continues to persist.

Commencing in 2014, AI researchers embarked on explorations of new methodologies. Among these was the concept of 'attention', analyzing data sequences (such as text) and accentuating the most pertinent information while dismissing the remainder. This method proved instrumental in improving task accuracy. Despite its benefits, attention mechanisms were predominantly employed to boost the performance of RNNs, which implied that lingering performance issues remained unaddressed. However, a significant paradigm shift occurred in 2017. (Vaswani et al. 2017) proposed an innovative neural network architecture known as the 'transformer', transcending the capabilities of both CNNs and RNNs. After testing this architecture on eight GPUs and training the model for three and a half days, the researchers demonstrated promising results, particularly in English-to-French translations. During the period under discussion, there was an increasing recognition among scholars that conventional deep learning models were nearing their saturation point. This perspective hinged on the argument that large models yielded diminishing dividends.

However, the advent of the transformer model particularly galvanized the AI industry. At the heart of the transformer model lies the principle of self-attention, a concept that diverges considerably from a standard neural network. The training phase in self-attention involves simultaneous processing of all observations rather than employing a gradual progression. This model facilitates comprehension of the contextual essence of words. Notably, a transformer is typically a pre-trained model, indicating that the training of the dataset begins from a zero-knowledge state. The model does not benefit from any form of pre-existing knowledge. However, a pre-trained model is not immutable. It permits fine-tuning to accommodate modifications and enhancements, such as adding additional data. Consequently, we initiate our discussion by delineating the fundamental precept of the attention mechanism.

2.2.1 Self-attention:

Within the domain of a self-attention layer, as illustrated in **Figure 2-4**, the initial operation involves the transformation of the input vector into three distinct vectors: the query vector (q), the key vector (k), and the value vector (v), each maintaining a constant dimensionality. These individual vectors are subsequently integrated into three distinct weight matrices W^Q, W^K and W^V, respectively. A generalized representation of Q, K, and V can be derived from the input X:

$$K = W^K X, Q = W^Q X, V = W^V X, \tag{15}$$

W^Q, W^K and W^V signify mutable parameters within this context. The Scaled Dot-Product Attention mechanism is subsequently delineated as follows:

$$Attention\ (Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V, \tag{16}$$

The term $\sqrt{d_k}$ represents a scaling factor and a softmax function is utilized on the produced attention weights, effectively converting them into a mathematically normalized distribution.

2.2.2. Multi-Head Self-Attention:

The multi-head self-attention (MHSA) structure, as depicted in Fig. 4, has been introduced by (Vaswani et al. 2017) as a method for modeling intricate relationships among token entities from diverse perspectives. The MHSA module particularly aids the model in simultaneously focusing on information from numerous representation sub-spaces, given the comparatively rough modeling proficiency of a single-head attention block. The functional mechanism of the MHSA can be theoretically outlined as follows:

$$MultiHead(Q, K, V) = [Concat(head_1, \dots, head_h)]W^O, \quad (17)$$

The term $head_i$ is representative of the Attention mechanism, where the mathematical expressions $Attention\left(QW_i^Q, KW_i^K, VW_i^V\right)$ are operational representations. Furthermore, W^O denotes a linear mapping function applied for integrating the multi-headed representation. It is worth noting that h serves as a hyper-parameter, which, by the initial paper, was established at a value of h=8.

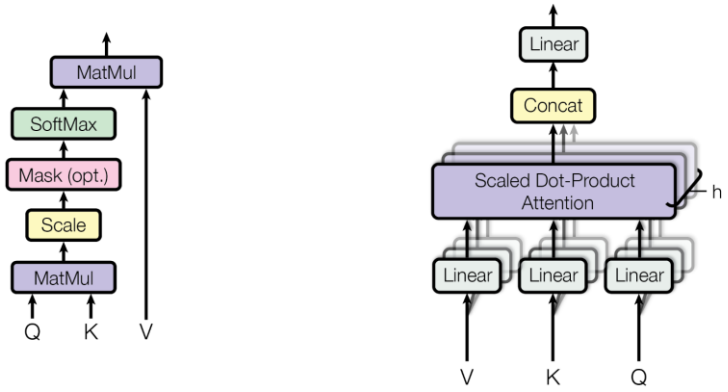


Figure 2-4. (left) we observe the representation of self-attention; (right) the illustration delineates the configuration of Multi-Head Attention, which comprises multiple concurrent attention layers (image by (Vaswani et al. 2017)).

The process under discussion involves an encoder utilizing a mechanism known as multi-head attention, which extends beyond the singular contextual comprehension found in self-attention. Specifically, multi-head attention can autonomously discern relationships among grammatical categories, such as nouns, adjectives, adverbs, and pronouns, without

explicit programming by identifying emergent patterns. Upon completion of this phase, the resultant vectors are transferred to the decoder, undergoing a process analogous to the encoder's operation. A series of decoders process the data. In addition to multi-head attention, another technique, masked multi-head attention, is utilized. It functions by predicting subsequent words or tokens by considering the text preceding and following the masked location. It is noteworthy that the functions of encoders and decoders can be disentangled in the following ways: **(1) Encoder-only models:** These models are advantageous for text classification tasks such as sentiment analysis. An exemplar of an LLM utilizing an encoder-only model is BERT. **(2) Decoder-only or autoregressive models**: These models are suitable for text generation tasks, akin to the predictive text functionality in a smartphone chat application. For instance, as you input text, the AI predicts the subsequent word or phrase. An example of this model is GPT-3. **(3) Encoder-decoder models**: These models facilitate generative AI tasks such as language translation and summarization. Notable LLMs employing this methodology include Facebook's BART and Google's T5.

2.2.3. Various Applications of Deep Generative AI Models

Inspired by transformer-based models' successes in NLP, many research efforts have been made to develop their counterparts for computer vision tasks. Among them, ViT (Dosovitskiy et al. 2020) is the first attempt demonstrating that a solely transformer-based architecture can achieve superior performance in image recognition with pre-training on a substantial dataset. Recently, Swin Transformer (Z. Liu et al. 2021) attracted great attention due to its exceptional performance on several benchmarks for tasks such as image classification, object detection, and semantic segmentation. Compared to many previous transformer-based models, Swin Transformer proposed a hierarchical architecture whose representation is computed with shifted windows. This strategy enhances efficiency by restricting self-attention computation to non-overlapping local windows while also allowing for cross-window connection. A comprehensive overview of transformer-based models in medical imaging domain can be found in (Shamshad et al. 2023).

3. Various Applications of Deep Generative AI Models

3.1. Diffusion Models in action:

Diffusion models, like other medical imaging methods, can be categorized in various ways. Here, we talk about ways to use diffusion in healthcare. These are broken down into five areas: (I) Image reconstruction, (II) Image to image translation, (III) Image generation, (IV) Image classification, (V) Other applications. **Figure 2-6** illustrates a compilation of studies within each category, detailing information like the specific algorithm used in the reverse process of the diffusion model for each study.

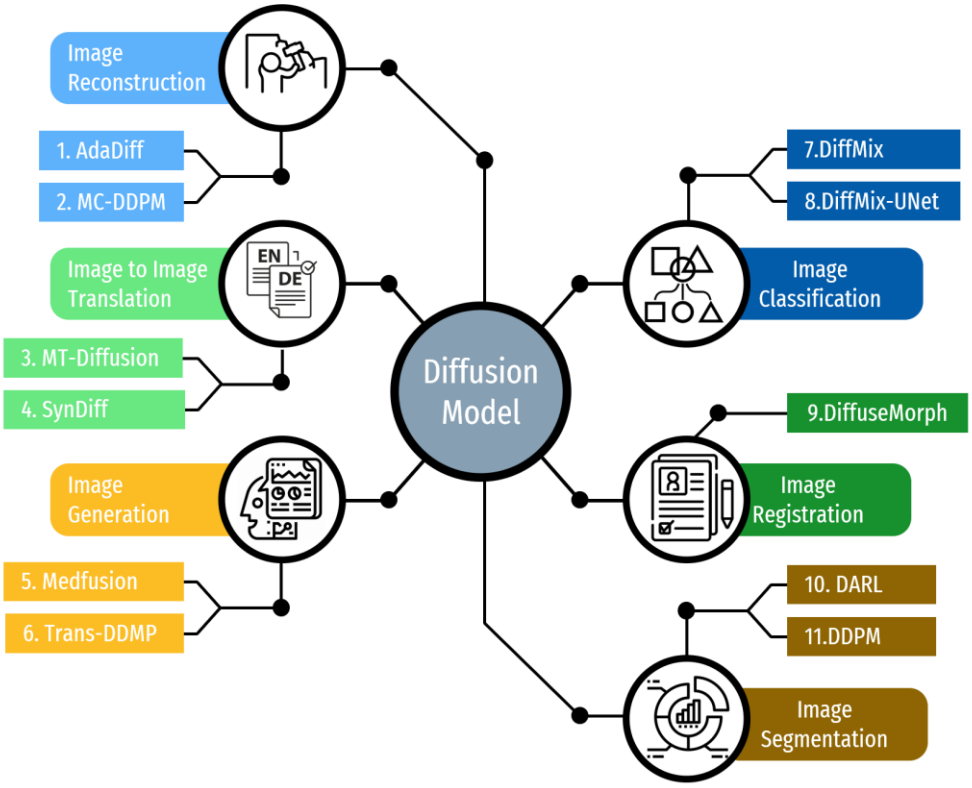


Figure 2-5. The proposed taxonomy for diffusion-based models in health care in six sub-fields, (I) Image Reconstruction 1. (Özbey et al. 2023), 2. (Xie and Li 2022), (II) Image to Image Translation 3. (Lyu and Wang 2022), 4. (Özbey et al. 2023), (III) Image Generation 5. (Müller-Franzes et al. 2023), 6. (Pan et al. 2023), (IV) Image Classification 7. (H.-J. Oh and Jeong 2023), 8. (Y. Yang et al. 2023), (V) Image Registration 9. (Kim, Han, and Ye 2022), (VI) Image Segmentation 10. (Kim, Oh, and Ye 2022), 11. (Azad et al. 2022).

3.1.1. Image Reconstruction:

The essential role of medical image reconstruction in the domain of medical imaging is indisputable, especially in producing high-quality images for clinical use while concurrently striving to minimize expenses and risks to patients, as emphasized by (Levac, Jalal, and Tamir 2023) and (C. Cao et al. 2022). Widely used medical imaging techniques, such as Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI), face challenges due to the fundamental physics involved, affecting their effectiveness, impeding their function, and potentially undermining the desired results. Increased radiation doses and extended tube occupancy times are necessary to obtain detailed, high-resolution patient data. However, these are only partially achievable due to health precautions and patient comfort considerations, underscoring the need for accelerated data collection in medical imaging methods like CT, PET, and MRI. Quicker imaging techniques can reduce examination times, enhance the availability of imaging services, and decrease waiting periods. They also improve the images' precision, particularly in dynamic studies that require rapid imaging sequences (Hyun et al. 2018; Korkmaz et al. 2021; C.-M. Feng et al. 2021). Consequently, strategies have been developed to decrease the radiation exposure from the standard dose or apply under-sampled or sparse-view imaging processes.

Nevertheless, these approaches have limitations, such as reduced Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR). Addressing these obstacles and solving this ill-posed inversion issue is the responsibility of medical image reconstruction, as pointed out by (Gothwal, Tiwari, and Shivani 2022). The subsequent discussion will offer a synopsis of diffusion-based medical image reconstruction and enhancement frameworks.

In their recent study, (Özbey et al. 2023) introduced an innovative technique named Adaptive Diffusion Priors (AdaDiff) for the reconstruction of MRI. This approach involves

a series of diffusion processes that enhance the authenticity of the generated images. Unlike traditional approaches utilizing static diffusion priors, AdaDiff dynamically adjusts its priors during the inference stage to align more closely with the distribution of the test data. The researchers demonstrated that this adjustment leads to superior results compared to existing methods regarding reconstruction quality and speed. Specifically, AdaDiff achieved a peak signal-to-noise ratio (PSNR) of 34.5 dB and a structural similarity index (SSIM) of 0.95 in a mere 1000 steps. Furthermore, the method demonstrated robustness against changes in the MRI acquisition protocol, maintaining consistent performance across various imaging operators **Figure 2-6**. Consequently, the authors deduced that AdaDiff holds great promise in enhancing the efficacy of MRI by facilitating swifter and more dependable image reconstruction.

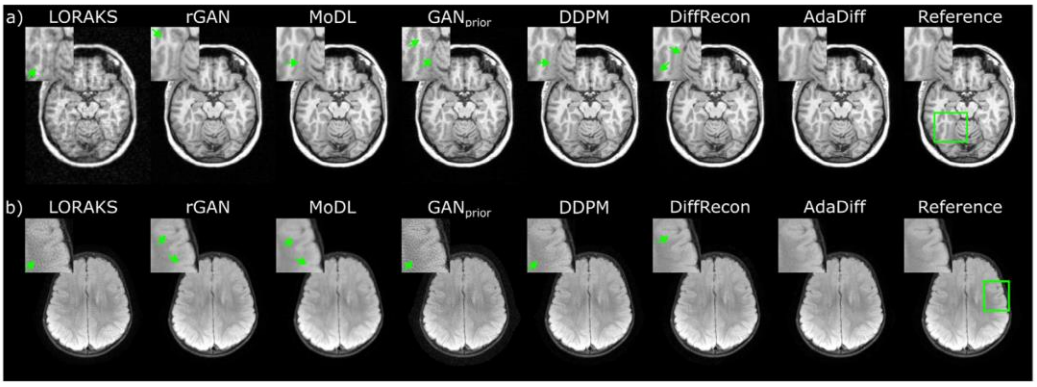


Figure 2-6. Results are shown for (a) T1-weighted acquisitions in the IXI dataset and (b) FLAIR-weighted acquisitions in the fast MRI dataset. Reconstructed images are given along with the reference image derived from fully sampled acquisitions, and zoom-in windows and arrows are included to highlight differences among methods. LORAKS and GAN prior show high noise amplification, rGAN shows residual aliasing, and MoDL shows visible spatial blurring despite high performance in quantitative metrics. DDPM has relatively higher noise among diffusion models, and DiffRecon shows local ringing artifacts near tissue boundaries. AdaDiff produces high-quality reconstructions with lower artifacts/noise and clearer tissue depiction than competing methods, images generated by (Özbey et al. 2023).

In a recent study by (Xie and Li 2022), a new integrated approach termed measurement-conditioned denoising diffusion probabilistic model (MC-DDPM) was

introduced to reconstruct under-sampled medical images using the DDPM framework. This innovative technique demonstrated superior performance compared to existing methods in metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). It provided the ability to assess the uncertainty of the reconstruction. The researchers conducted a series of experiments using PD and PDFS datasets, which revealed that the MC-DDPM method surpassed other techniques by a considerable margin. Furthermore, they investigated the impact of varying the number of sampling steps and the total number of samples on the quality of the reconstructed images.

Interestingly, they discovered that a minor reduction in sampling steps only led to a slight decrease in PSNR. Additionally, they found that increasing the number of samples improved the quality of the sample mean and appeared to reach a plateau. Therefore, the authors suggested that 20 samples and 250 sampling steps might be an ideal selection for optimal efficiency. Moreover, the study also delved into the potential applications of MC-DDPM in areas beyond MRI reconstruction, including CT and PET reconstruction. In summary, the article proposes a promising technique for reconstructing under-sampled medical images, potentially enhancing the accuracy and efficiency of medical imaging procedures.

3.1.2. Image To Image Translation:

Recent advancements in image-to-image translation, a critical component in many applications, hinge on two fundamental architectures: pix2pix (Isola et al. 2017) and CycleGAN (Zhu et al. 2017; Shokrollahi et al. 2023). These systems have played a significant role in the medical field, particularly in diagnosis and therapy, where acquiring multi-modality images is crucial. There are instances, however, when certain modalities may be missing or unobtainable due to various conditions. To address this, diffusion models have been developed, demonstrating impressive capabilities in generating missing modalities by utilizing cross-modalities. An exemplary application of this technology can be seen in the translation from MRI to CT, enhancing the flexibility and completeness of medical imaging (Lyu and Wang 2022). CT and MRI are commonly used imaging techniques. However, CT

struggles to capture detailed imagery of soft tissue injuries, often necessitating an MRI follow-up for a precise diagnosis. This two-step process can be lengthy and expensive and might lead to misalignment between the two sets of images.

To bridge this gap, harness the power of recently developed DDPMs and score-based diffusion models. Their approach includes a conditional DDPM and a conditional SDE, where the latter's reverse mechanism leverages T2w MRI images. The authors deploy DDPM and SDE in their research using three unique sampling methods: EM, PC, and ODE. When benchmarked against existing GAN (Gulrajani et al. 2017) and CNN (Ronneberger, Fischer, and Brox 2015) methodologies, their diffusion models show superior performance on the Gold Atlas male pelvis dataset (Nyholm et al. 2018). Key performance metrics include both SSIM and PSNR. Moreover, to gauge the reliability of diffusion models, (Lyu and Wang 2022) employ the Monte Carlo (MC) technique. Here, the model generates outputs ten times, with an average determining the conclusive result.

(Özbey et al. 2023) proposed a novel adversarial diffusion model called SynDiff for the unsupervised medical performance of medical image translation. SynDiff is designed to address the limitations of traditional GAN and diffusion models, which suffer from noise, spatial warping, and blurring. The authors conducted extensive experiments on multi-contrast MRI and MRI-CT translation and showed that SynDiff outperforms competing models regarding accuracy and artifact reduction **Figure 2-8**. The proposed method is based on a diffusion process that models the conditional distribution of the target image given the source image. The diffusion process is trained using a maximum likelihood objective, which is optimized using a stochastic gradient descent algorithm. The authors also introduce a novel adversarial loss function that encourages the generated images to be indistinguishable from the real images. The authors conducted several experiments to evaluate the performance of SynDiff on different medical imaging protocols. They compared SynDiff with several state-of-the-art GAN and diffusion models and showed that SynDiff outperforms these models regarding accuracy, artifact reduction, and computational efficiency. They also conducted a sensitivity analysis to evaluate the robustness of SynDiff to different hyperparameters and showed that SynDiff is relatively insensitive to these parameters. Overall, the proposed

SynDiff method is a promising approach for unsupervised medical image translation that can improve the accuracy and efficiency of medical imaging protocols.

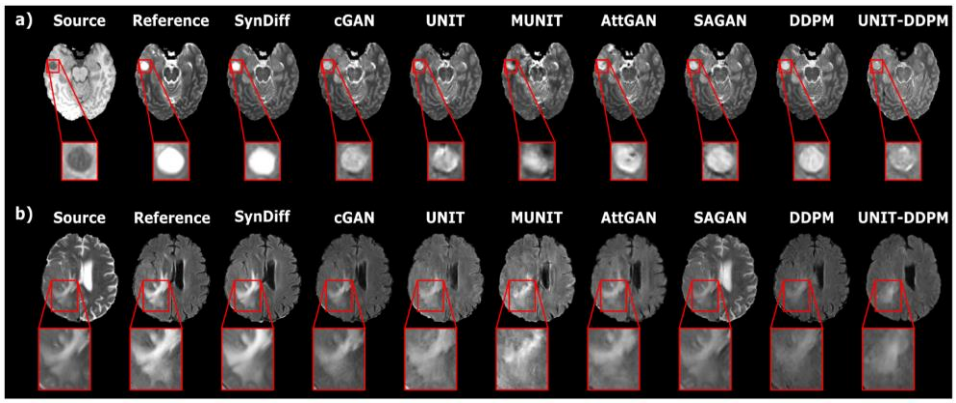


Figure 2-7. SynDiff showcased its capability for MRI contrast conversions On the BRATS dataset (Menze et al. 2014). For illustrative purposes, source images, synthesized outputs, and the actual reference images are presented for the following tasks: a) T1 to T2 and b) T2 to FLAIR (Fluid Attenuation Inversion Recovery). The visualization scales utilized are a) [0, 0.75] and b) [0, 0.80]. SynDiff effectively minimizes noise and artifacts, offering a clearer depiction of intricate structures than other baseline methods (Özbey et al. 2023).

3.1.3. Image Generation:

Diffusion models primarily focus on creating images and have been used extensively across different styles. Examples include the production of artificial 2D/3D medical visuals, as mentioned by (Pinaya et al. 2022), (Moghadam et al. 2023), (Dorjsembe, Odonchimed, and Xiao 2022), and (Kim and Ye 2022). Additionally, they have been utilized in converting 2D cell photographs into 3D representations, as cited by (Waibel et al. 2022). This segment delves into the techniques that use diffusion for generating medical imagery. (Müller-Franzes et al. 2023) Introduced Medfusion, a novel conditional latent DDPM, and assessed its performance against GANs using three publicly accessible datasets. These datasets comprised many RGB eye fundus images, chest CT scans, and 3D MRI brain scans from various subjects. The study involved training Medfusion on the 3D MRI brain scans and evaluating its effectiveness using several metrics, such as the PSNR, SSIM, and Fréchet inception distance (FID). The results indicated higher PSNR and SSIM values and lower FID scores for Medfusion, suggesting it outperformed GANs regarding image quality.

Additionally, a sensitivity analysis was conducted to examine the influence of different hyperparameters on Medfusion's performance, revealing that some hyperparameters were more crucial than others. Ultimately, the researchers concluded that Medfusion is a promising technique for medical image synthesis, with advantages such as improved stability and interpretability over GANs.

Similarly, (Pan et al. 2023) developed a technique for synthesizing 2D medical images using a transformer-based denoising diffusion probabilistic model **Figure 2-9**. This innovative method employed a diffusion process modeled by a neural network trained to denoise images at each stage to generate samples from a probability distribution matching the target image distribution. The model, trained on a vast collection of medical images, was tested on several benchmark datasets using a transformer-based encoder-decoder network. A unique loss function was also introduced to encourage the generation of realistic and diverse samples. The authors conducted experiments on benchmark medical image datasets, including chest X-rays, heart MRIs, pelvic CT images, and abdomen CT images. They compared their approach with existing state-of-the-art methods. Their method outperformed the others in terms of image quality and fidelity. Using the Inception score and Fréchet Inception Distance score, quantitative evaluations confirmed that the synthetic images generated by their model belonged to the same data distribution as real images. Overall, this approach presents a promising avenue for enhancing diagnostic and treatment planning processes by creating high-quality medical images. The authors plan to extend this approach to 3D medical volume synthesis and explore more advanced network architectures to improve image synthesis quality further.

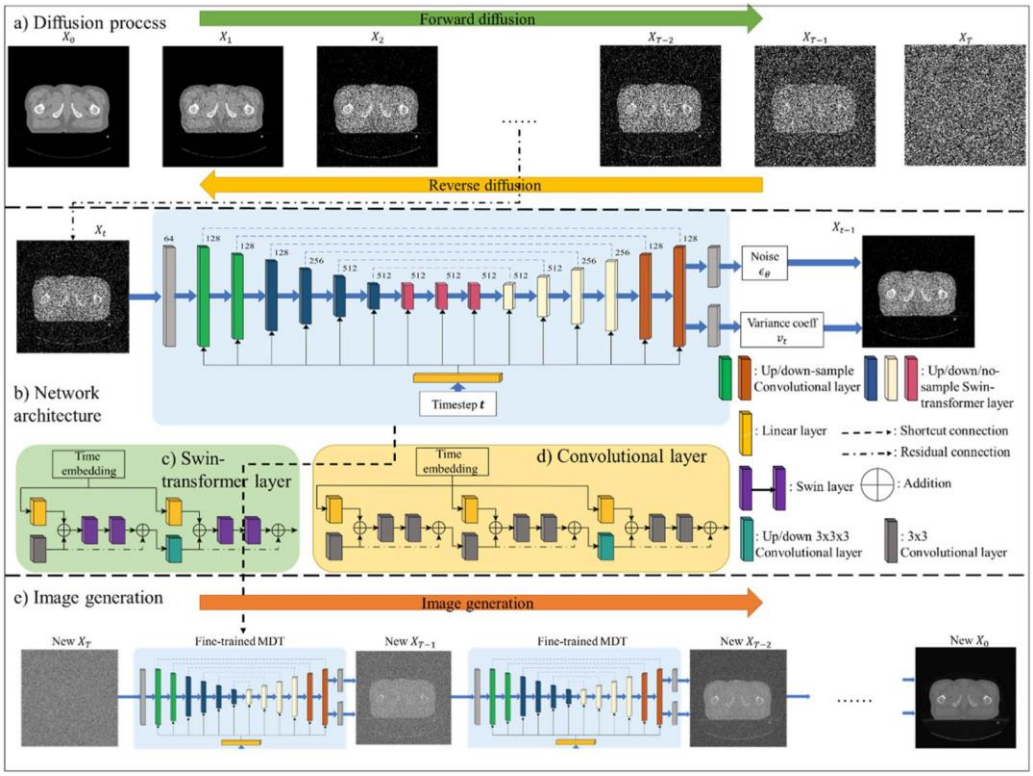


Figure 2-8. MT-DDPM Framework's Diffusion Methodology (a): Medical imagery is progressively transformed into pure Gaussian noise through incremental noise addition in the forward diffusion. For the backward process, a system is tasked to filter the Gaussian noise back into a pristine image continuously. (b): MT-DDPM Network Design: This network uses a balanced encoder-decoder design to master the backward process. The image devoid of noise is ascertained by forecasting the noise and its variance coefficient. (c): Component of Swin-transformer: The violet Swin segment comprises both window-based self-attention and a subsequent shifted window self-attention module. (d): Convolution Segment: This part houses three convolutional residual configurations that capture local image details. (e): Process of Image Creation: Upon advanced training, the MT-DDPM system can convert a fresh Gaussian noise image into a novel, synthesized image not previously found in the dataset (Pan et al. 2023).

3.1.4. Image Classification

The classification process is vital in the realm of medical image analysis. It paves the way for precise detection and description of various elements and irregularities found in medical imagery. This technique holds promise in transforming the medical sector by assisting healthcare experts in deciphering vast sets of intricate data (Shehab et al. 2022). However, there is still a pressing need to refine the integration of diffusion models for

improved classification outcomes. In a recent study, (Y. Yang et al. 2023) proposed a DiffMIC method for classifying diverse medical imaging types via diffusion algorithms **Figure 2-9**. The process starts by translating the given image into a distinct feature space. The Dual-granularity Conditional Guidance (DCG) strategy is employed to gather overarching and minute details. After this, the diffusion of the actual image and the two prior sets produces a triad of noisy variables. These noisy entities are merged with their respective priors, taking them to a hidden layer resulting in three separate feature sets. To extract the noise-related characteristics of these sets, a denoising U-Net model is integrated with the image's original features. These refined feature sets are then reverted to their primal state.

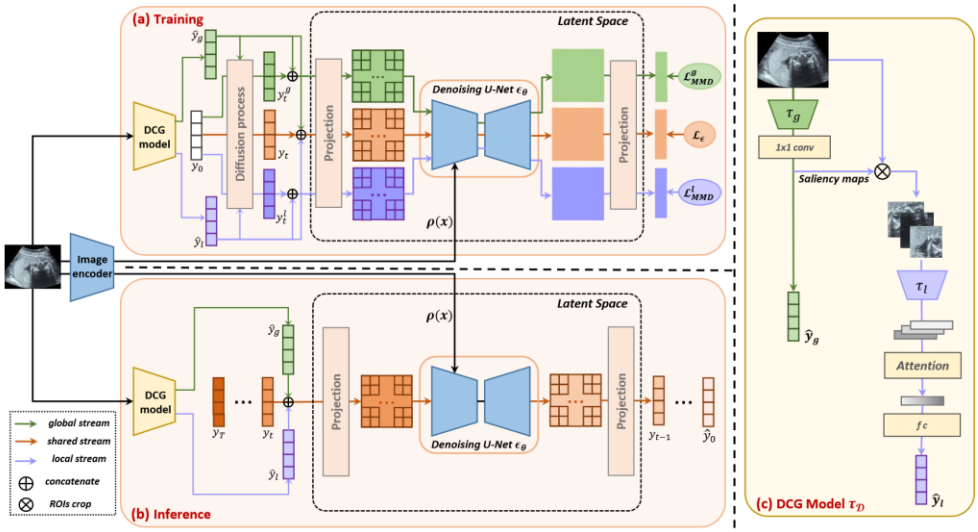


Figure 2-9. An overview of the DiffMIC framework includes (a) The forward process during the training phase (b) The reverse process for inference. (c) The DCG Model $\boldsymbol{\tau_D}$ directs the diffusion using dual priors from the raw image and ROIs (Y. Yang et al. 2023).

In the Image Generation section, we discussed Diffusion models' potency in image synthesis. Recent advancements in machine learning have witnessed the emergence of state-of-the-art models such as GPT-3 (Brown et al. 2020b), DALL-E (Ramesh et al. 2022), Imagen (Saharia et al. 2022), and Stable Diffusion (Rombach et al. 2022), all of which have been fine-tuned on vast and varied internet datasets. Termed as "foundation models" (Bommasani et al. 2021), they epitomize exceptional generative capabilities, aiding in code development, producing art, and formulating text. In the medical domain, diffusion models

have been harnessed to synthesize images, thereby augmenting datasets for tasks such as classification, among others. For example, In a recent study by (H.-J. Oh and Jeong 2023), the authors introduced a new data augmentation approach, termed 'DiffMix', specifically designed for segmenting and classifying nuclei in imbalanced pathology image datasets. This innovative framework leverages a semantic-label-conditioned diffusion model to create synthetic data samples, thereby enhancing the classification efficacy of rare nuclei types and exhibiting superior performance in both segmentation and classification tasks involving imbalanced datasets of pathology nuclei. A comprehensive evaluation of the proposed approach was conducted on two widely recognized imbalanced nuclei datasets, namely CoNSeP and MoNuSeg. The performance of DiffMix was benchmarked against well-established networks such as HoVer-Net and SONNET, as well as another data augmentation technique, GradMix. The findings from the experimental analysis revealed that DiffMix surpassed the other approaches with respect to classification accuracy, segmentation efficacy, and minimizing performance disparities across different class categories. Hence, the proposed approach holds significant potential to optimize the performance of various applications involving medical image processing tasks.

3.1.5. Other Applications of Diffusion Models:

Deformable **image registration** is a key method in medical imaging that finds flexible connections between two images that might move or change. It is important when the shape of images shifts because of reasons like the patient, when it is scanned, or the type of imaging used. (Kim, Han, and Ye 2022) presented a new method named DiffuseMorph. It consists of two primary systems: one for diffusion and one for deformation. Both systems are trained directly. The first system measures how the images change, and the second system estimates how they deform based on this measurement. Using spatial data, this method can smoothly show how one image transforms into another. **Image segmentation** is crucial in computer vision as it breaks down an image into important parts. This helps in medical studies by giving useful details about body structures. However, deep learning systems need more varied data with detailed labeling to work well. There are limitations in the number of medical images and their labels due to the time, expense, and expertise needed

(Azad et al. 2022). As a solution, diffusion models have become popular because they can create labeled data and reduce the need for detailed image annotations. (Kim, Oh, and Ye 2022) a new DARL method was introduced for self-guided blood vessel segmentation to identify vascular disorders. This DARL approach consists of two primary components: one that learns the background image pattern and another that creates vessel segmentation visuals or artificial angiograms through a versatile SPADE technique (Park et al. 2019).

3.2. Transformer-based Models in Actions:

Transformer-based models can be categorized in various ways. Here, we talk about ways to use transformers in healthcare. These are broken down into seven areas: (I) clinical documentation and information extraction, (II) medical coding and billing, (III) diagnostic assistance, (IV) medical imaging and radiology interpretation, (V) clinical decision support, (VI) protein structure prediction, (VII) drug design and molecular representation. **Figure 2-10** depicts a compilation of various studies for each category, detailing information like the particular algorithm employed in the transformer-based models.

Figure 2-10. The proposed taxonomy for Transformer-based models in health care in seven sub-fields, (I) protein structure prediction 1. (Vig et al. 2020), 2. (Behjati et al. 2022), 3. (Abdine et al. 2023), 4. (Boadu, Cao, and Cheng 2023), 5. (Y. Cao and Shen 2021), 6. (Geffen, Ofran, and Unger 2022), 7. (Castro et al. 2022), 8. Ferruz, Schmidt, and Höcker 2022, 9. (Ferruz, Schmidt, and Höcker 2022), 10. (Oliveira, Pedrini, and Dias 2023), (II) clinical documentation and information extraction 11. (Gérardin et al. 2023), 12. (Lentzen et al. 2022), 13. (Y. Li et al. 2022), 14. (Moon, He, and Liu 2022), 15. (S. H. Oh, Kang, and Lee 2022), 16. (Searle et al. 2023), 17. (Sivarajkumar and Wang 2022), 18. (Solarte-Pabón et al. 2023), 19. (Wei et al. 2022), 20. (Yogarajan et al. 2021) , (III) diagnostic assistance 21. (Azizi, Hier, and Wunsch Ii 2022), 22. (S. Chen et al. 2023), 23. (X. Chen et al. 2022), 24. (Dhinagar et al. 2023), 25. (Dong et al. 2023), 26. (Garaiman et al. 2023), 27. (Hosain et al. 2022), 28. (D. Hu et al. 2022), 29. (Mo et al. 2023), 30. (Zhou et al. 2023), (IV) medical imaging and radiology interpretation 31. (Balouch and Hussain 2023), 32. (Bhattacharya, Jain, and Prasanna 2022), 33. (Chaudhari et al. 2022), 34. (Jacenkow, O'Neil, and Tsaftaris 2022), 35. (J. Li et al. 2022), 36. (Moezzi et al. 2022), 37. (Mohsan et al. 2023), 38. (Nimalsiri et al. 2023), (V) clinical decision support 39. (J. Feng, Shaib, and Rudzicz 2020), 40. (W. Hu and Wang 2022), 41. (G. Huang 2022), 42. (Meng et al. 2021), 43. (Wang et al. 2023), (VI) medical coding and billing 44. (Liu et al. 2022), 45. (López-García et al. 2023), 46. (Ng,

Santos, and Rei 2023), 47. (Shang et al. 2019), 48. (Tchouka et al. 2023), (VII) drug design and molecular representation 49. (Bagal et al. 2022), 50. (Fabian et al. 2020), 51. (K. Huang et al. 2021), 52. (H. Li, Zhao, and Zeng 2022), 53. (Rong et al. 2020).

3.2.1. Clinical documentation and information extraction:

To enhance the effectiveness of subsequent NLP tasks, (Gérardin et al. 2023) devised and validated an algorithm for analyzing the layout of PDF clinical documents, extracting only clinically relevant text **Figure 2-11**. The algorithm involves several steps: text extraction using a PDF parser, text classification into categories like body text, left notes, and footers using a Transformer deep neural network architecture, and then aggregation of text lines associated with a given label. The algorithm was assessed technically and medically by applying it to a representative sample of annotated texts and examining the extraction of pertinent medical concepts. It was tested end-to-end in a medical use case involving automatic identification of acute infections detailed in a hospital report. (Gérardin et al. 2023) demonstrated that this preprocessing step improved the results in downstream tasks, such as medical concept extraction, proving its utility in a clinical context.

(Lentzen et al. 2022) thoroughly assessed the effectiveness of transformer-based AI models on German clinical notes. They employed a novel biomedical corpus for pre-training three new models. They systematically compared eight existing transformer-based models using a new dataset of clinical notes integrated with four other corpora. The newly trained BioGottBERT model outperformed the GottBERT model in clinical named entity recognition (NER) tasks. Additionally, (Y. Li et al. 2022) introduced two domain-specific language models, ClinicalLongformer and Clinical-BigBird, pre-trained on a large corpus of clinical text, improving several downstream clinical NLP tasks, such as question answering, NER, and document classification. (Moon, He, and Liu 2022) analyzed a corpus of 500,000 clinical texts for sublanguage characteristics across different practice environments and document types for the ten most common clinical sections. They used BERT, fine-tuned for bio-clinical characteristics, to extract named entities for problem, test, and therapy concepts, and Sentence-BERT for rapid clustering. The analysis revealed narrow scopes in evaluation and discharge summary documents and similar cluster distributions in Family Medicine and Primary Care practice environments, suggesting comparable sublanguages. Conversely,

Emergency Medicine exhibited a unique sublanguage, characterized by higher phrase frequencies in discontinuous clusters.

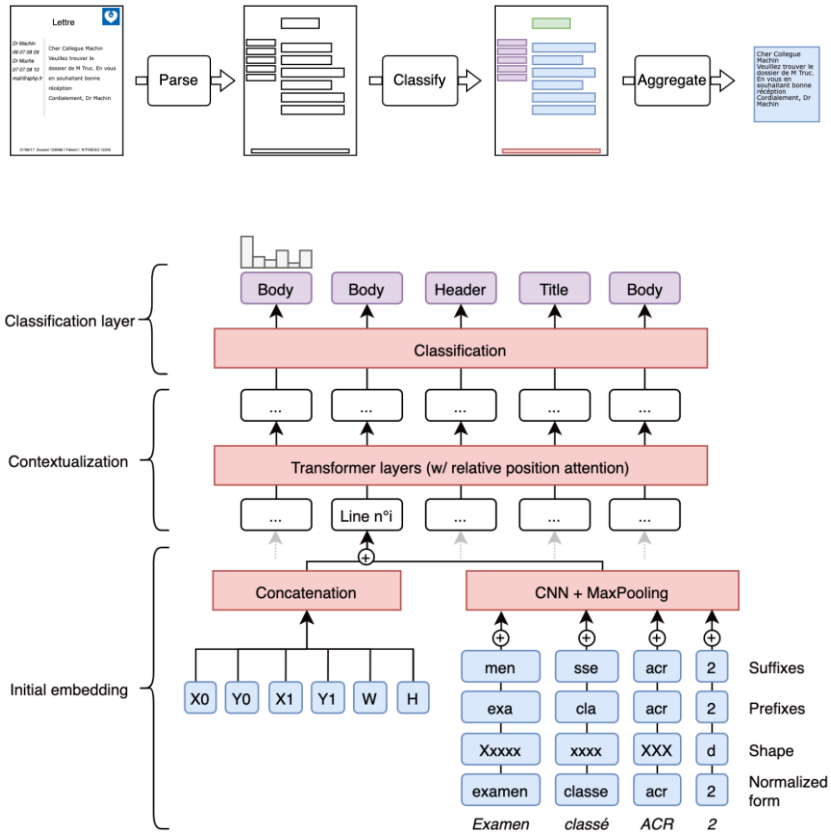


Figure 2-11. Structure of the Deep-Learning Line Classification Model. Textual and layout characteristics of each line are embedded to generate a unique representation for each line, which is subsequently contextualized using a four-layer Transformer that employs self-attention with relative position information. Finally, the representations are classified using a linear layer and a SoftMax function to determine the probability of each label (Gérardin et al. 2023).

(S. H. Oh, Kang, and Lee 2022) addressed the de-identification of protected health information (PHI) in medical documents and analyzed the efficacy of three prominent pre-training models (XLNet, RoBERTa, and BERT) for PHI recognition using the i2b2 2014 dataset. The dataset was tokenized and processed using an inside-outside-beginning (IOB) tagging approach and Word Piece tokenization. XLNet outperformed the other models in

PHI recognition. Additionally, (Searle et al. 2023) explored summarization techniques for Biomedical Health Care (BHC) data and evaluated a new ensemble model that integrates extractive and abstractive summarization methods, using the SNOMED medical concept ontology as a clinical guidance signal. They proposed an advanced abstractive summarization model based on BART that includes a clinical oncology-aware guidance signal for key terms, facilitating the creation of problem-list-oriented abstractive summaries. (Sivarajkumar and Wang 2022) proposed a novel approach to clinical texts by implementing prompt-based learning, resulting in developing a groundbreaking prompt-based clinical NLP framework called 'HealthPrompt'. This approach customizes task definitions via prompt template specification, eliminating the need for training data. (Solarte-Pabón et al. 2023) proposed a transformer-based approach for extracting named entities from Spanish clinical notes related to breast cancer, facilitated by a specialized breast cancer corpus and a schema for annotating clinical notes with breast cancer-related concepts. Their findings suggest that transformers are potentially effective in extracting information from Spanish clinical texts and can be enhanced by training in biomedical literature.

(Wei et al. 2022) introduced a multimodal technique called ClinicalLayoutLM for categorizing scanned clinical documents into specific categories, such as laboratory reports and CT scans. This technique combined text obtained from optical character recognition (OCR) with layout or image information. It outperformed the baseline model (which relied solely on OCR text) in classifying scanned clinical documents into 16 categories. (Yogarajan et al. 2021) considered transfer learning in the context of multi-label problem formulation for pre-trained models for predicting medical codes. They found that domain-specific transformers performed better for texts with fewer labels and/or documents shorter than 300 words. However, conventional neural networks were more effective for less frequently occurring labels and documents exceeding 300 words.

These studies highlight the utility of transformer-based models and other advanced techniques in various clinical NLP tasks, such as document classification, named entity recognition, and information extraction from clinical texts in multiple languages. However, the effectiveness of these models varies based on the task, language, and specific characteristics of the data, suggesting the need for a tailored approach to each application.

3.2.2. Medical coding and billing:

Medical coding and billing is a critical aspect of healthcare services, involving assigning standardized codes for illnesses and procedures in clinical documents. Recent advances in transformer-based models have shown promise in improving the efficiency and accuracy of this task. Several studies have proposed innovative approaches and models to facilitate the interpretable prediction of International Classification of Diseases (ICD) codes and to make the process more explainable. (Liu et al. 2022) the Hierarchical Label-Wise Attention Transformer (HiLAT) model utilizes a two-tier hierarchical label-wise attention mechanism to generate label-specific document representations. These representations are then used to predict the likelihood of assigning a particular ICD code to a clinical document. The model was evaluated using the Medical Information Mart for Intensive Care III (MIMIC-III) database and outperformed previous state-of-the-art models for the 50 most frequently observed ICD-9 codes.

Similarly, (López-García et al. 2023) assessed several multilingual transformer-based models, including XLM-RoBERTa, mBERT, and BETO, which were tailored and evaluated for explainable clinical coding. The authors compared a multitask approach and a hierarchical task approach for training the models. They found that the hierarchical-task approach yielded superior results, comprising medical named entity recognition (MER) and medical-named entity normalization (MEN). (Ng, Santos, and Rei 2023) introduced a groundbreaking model called Hierarchical Transformers for Document Sequences (HTDS), designed for temporal modeling of document sequences. The model processes textual and metadata content from a series of documents and outperforms the previously established state-of-the-art model when the entire collection of clinical notes is used as input. (Shang et al. 2019) proposed the G-BERT model, which combines the strengths of Graph Neural Networks (GNNs) and BERT for articulating medical codes and recommending medications. The model captured the underlying hierarchical structures in medical codes and outperformed all baseline models regarding prediction accuracy for medication recommendation tasks. (Tchouka et al. 2023) developed a model specifically designed for ICD10 code association, leveraging the latest advancements in Natural Language Processing (NLP) and multi-label classification. The model was rigorously evaluated on a French

clinical dataset and demonstrated its efficacy as the most accurate solution for ICD10 code association in the French language to date.

In conclusion, recent studies have demonstrated the effectiveness of transformer-based models in improving the efficiency and accuracy of medical coding and billing. These models leverage advanced NLP techniques, hierarchical attention mechanisms, and multi-label classification to process voluminous clinical documents and accurately assign ICD codes. In contrast, the models show promise in various languages and clinical contexts; ongoing research and development are necessary to optimize their performance further and make them more widely applicable in the healthcare industry.

3.2.3. Diagnostic assistance:

Recent studies have showcased the potential of transformer-based models in various diagnostic applications. (Azizi, Hier, and Wunsch Ii 2022) developed a model for named entity identification, converting text segments with neurological signs and symptoms into clinical concepts using deep learning techniques. A comparative analysis between a transformer-based bidirectional encoder representations model and a convolutional neural networks model indicated the former's superiority in identifying signs and symptoms. This study is corroborated by (S. Chen et al. 2023), who introduced a feature interaction Transformer network (FIT-Net) for diagnosing pathologic myopia through Optical Coherence Tomography images, demonstrating superior performance over traditional deep learning methods. (X. Chen et al. 2022) leveraged Multi-view Vision Transformers (MVTs) for long-range correlations between mammograms, showcasing superior reproducibility, robustness, and case-based malignancy classification accuracy over other block combination alternatives.

In another study, (Dhinagar et al. 2023) investigated Vision Transformer (ViT) architecture iterations for high-stakes neuroimaging downstream tasks, focusing on gender and Alzheimer's disease classification using 3D brain magnetic resonance imaging. The study evaluated the effects of different ViT training strategies and emphasized the importance of these strategies in neuroimaging applications. Similarly, (Dong et al. 2023) proposed an innovative end-to-end multi-label arrhythmia classification model, CNN-DVIT,

combining a vision transformer structure with deformable attention and CNNs, outperforming the most recent transformer-based ECG classification methods. (Garaiman et al. 2023) aimed to develop a deep-learning model based on a vision transformer for identifying specific microangiopathy indicators in nailfold capillaroscopy images of patients diagnosed with Systemic Sclerosis. Although rheumatologists exhibited higher average accuracy, the ViT model was proficient in detecting various microangiopathic changes, highlighting the variability in performance among annotators (**Figure 2-12**).

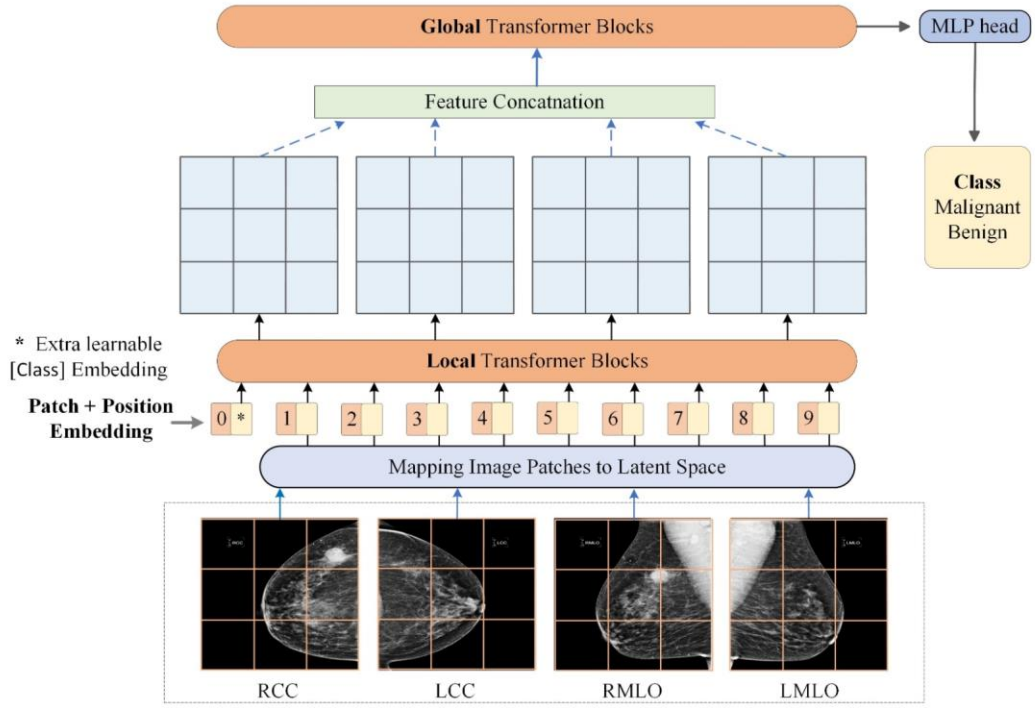


Figure 2-12. Structure of the Deep-Learning Line Classification Model. Textual and layout characteristics of each line are embedded to generate a unique representation for each line, which is subsequently contextualized using a four-layer Transformer that employs self-attention with relative position information. Finally, the representations are classified using a linear layer and a SoftMax function to determine the probability of each label (Gérardin et al. 2023).

(Hosain et al. 2022) proposed an innovative approach for categorizing endoscopic imagery attributes using a vision transformer and transfer learning model, outperforming the DenseNet201 in identifying gastrointestinal diseases from wireless capsule endoscopy

images. (D. Hu et al. 2022) introduced a cross-modal retrieval framework for executing retrieval tasks involving histopathology whole slide images and diagnostic reports, demonstrating the effectiveness of the proposed strategy in performing cross-modal retrieval tasks. (Mo et al. 2023) introduced an innovative region-of-interest-free model, HoVer-Transformer, for identifying breast cancer in ultrasound images, outperforming other convolutional neural network-based models and experienced sonographers in interim analysis. Lastly, (Zhou et al. 2023) introduced a transformer-based representation-learning model for consolidating multimodal information processing as a clinical diagnostic tool, demonstrating superior performance in identifying lung illnesses and predicting adverse clinical outcomes in COVID-19 patients compared to image-only and non-unified multimodal diagnostic models.

In summary, transformer-based models have shown significant potential in various diagnostic applications, including named entity identification, pathologic myopia diagnosis, mammogram correlations, neuroimaging downstream tasks, arrhythmia classification, microangiopathy indicator identification, endoscopic imagery attribute categorization, histopathology image retrieval, breast cancer identification, and multimodal clinical diagnostic tool development. These studies collectively indicate the effectiveness and versatility of transformer-based models in diagnostic assistance, providing a foundation for future research and implementation in clinical settings to optimize clinical decision-making processes and patient triage.

3.2.4. Medical Imaging and Radiology Interpretation:

Medical imaging and radiology interpretation have seen significant advancements in recent years, with transformer-based models playing a crucial role in various aspects, such as summarization, disease diagnosis, error identification, image classification, information extraction, and report generation. Several recent studies have explored different aspects and applications of transformer-based models in medical imaging and radiology interpretation. One of the key challenges in radiology is generating pertinent summaries from the intricate and voluminous information in radiology reports. (Balouch and Hussain 2023) devised a special mechanism for abstractive text summarization using the Biobart-V2 model, which

achieved state-of-the-art performance in summarizing medical texts, as evidenced by the ROUGE score. Additionally, (Bhattacharya, Jain, and Prasanna 2022) introduced the RadioTransformer. This groundbreaking framework assimilates radiologists' visual search patterns within a cascaded global-focal transformer framework, enhancing confidence in decision-making (**Figure 2-13**).

Moreover, several studies have highlighted the importance of fine-tuning transformer models for specific tasks. (Chaudhari et al. 2022) developed a specialized version of BERT, Radiology BERT, fine-tuned artificially generated errors, substantially enhancing the identification of actual clinical speech recognition errors in radiology reports. Similarly, (Jacenkow, O'Neil, and Tsaftaris 2022) fine-tuned a BERT model for multimodal classification using dual image-text input, which improved image classification in radiology reports. Another study by (J. Li et al. 2022) utilized bidirectional encoder representations from transformer (BERT)-based models and in-domain pre-training (IDPT), along with a sequence adaptation strategy, to develop an approach for the classification of actionable radiology reports in tinnitus patients, yielding promising results.

Furthermore, in clinical information extraction, (Moezzi et al. 2022) employed a transformer-based fine-grained Named Entity Recognition (NER) architecture that surpassed the performance of previously utilized techniques, delivering a coherent and structured report. Also, (Mohsan et al. 2023) proposed a comprehensive transformer-based model, TrMRG, for report generation, which achieved noteworthy results compared to prevailing methods. Another novel approach, MERGIS, proposed by (Nimalsiri et al. 2023), utilized image segmentation and a modern transformer-based encoder-decoder model to enhance the accuracy of automated report generation. In conclusion, applying transformer-based models, often with task-specific fine-tuning, has shown promising results in various medical imaging and radiology interpretation aspects. These advancements in text summarization, disease diagnosis, error identification, image classification, information extraction, and report generation indicate that transformer-based models are vital tools for enhancing the accuracy and efficiency of radiology reports, ultimately contributing to better patient care. As these models continue to evolve and become publicly accessible, they will play a significant role in the advancement of the field of medical imaging and report generation.

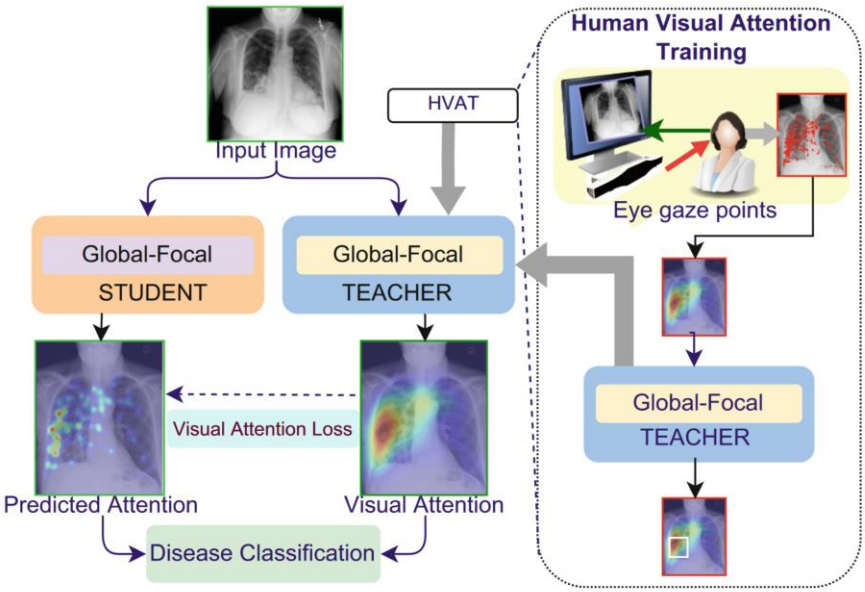


Figure 2-13. Schematic of the Proposed Methodology. The visual search patterns of radiologists on chest radiographs serve as the initial input for training a global-focal teacher network, denoted as Human Visual Attention Training. Subsequently, this pre-trained teacher network instructs the global-focal student network to acquire visual attention utilizing a newly devised Visual Attention Loss. Implementing the student-teacher network is meticulously designed to incorporate radiologist visual attention explicitly, thereby enhancing the accuracy of disease classification on chest radiographs (Bhattacharya, Jain, and Prasanna 2022).

3.2.5. Clinical Decision Support:

Transformer-based models have become increasingly prevalent in clinical decision support, as they hold significant potential for practical application in various aspects of clinical care. One novel approach involves the integration of a hierarchical Convolutional Neural Network (CNN) transformer with ClinicalBERT, a transformer-based model pre-trained on clinical corpora. This approach, proposed by (J. Feng, Shaib, and Rudzicz 2020), aims to capture the unique phrase-level patterns and global contextual linkages of medical language by leveraging the MIMIC-III database and investigating hidden attention layers. The study emphasizes the pragmatic utility of the model in sepsis prediction and ICU mortality by exclusively relying on textual data, exploring correlations between latent features extracted from structured and unstructured data, and recommending an evaluation

framework for model efficacy. The findings, substantiated by domain expert evaluations, indicate that the model generates meaningful justifications that align with expert assessments, thus demonstrating its potential as a transparent decision-support tool.

Additionally, BERT-based models have been evaluated for their efficacy in predicting glaucoma progression using clinical free-text progress notes. A study by (W. Hu and Wang 2022) marked the initial efforts to apply BERT-based models, including the original BERT, BioBERT, RoBERTa, and DistilBERT, to ophthalmology clinical text, indicating potential in forecasting glaucoma progression. (G. Huang 2022) proposed a novel online surgical action detection and prediction model, termed SA, which leverages the attention mechanism of the Transformer to facilitate contextual decision-making in operating theatres (**Figure 2-14**). The SA model, an encoder-decoder framework, concurrently classifies current and subsequent surgical actions by ascertaining relationships between image components, surgical activities, and preceding images. Empirical results indicate superior performance in recognizing current actions compared to predicting future actions. Another innovative approach was proposed by (Meng et al. 2021), who developed a Bidirectional Representation Learning model (BRLTM) employing a Transformer architecture for multimodal Electronic Health Records (EHR) data. This model facilitates a two-stage approach of pretraining and finetuning for EHR data modeling and enhances interpretability by quantitatively elucidating relationships between different EHR codes in sequences. The results indicate that bidirectional learning outperforms forward-only approaches in sequence modeling, thereby enabling the identification of individuals requiring depression screening and those at risk of developing depression within a specified time frame.

Lastly, (Wang et al. 2023) implemented a transformer-based approach in developing a family history (FH) lexical resource using a corpus of clinical notes from primary care settings. The study involved developing and evaluating rule-based and deep learning-based systems for FH information extraction. The integration of both systems enhanced the recall of FH information, although the F1 score remained variable yet comparable. In summary, the application of transformer-based models in clinical decision support has demonstrated significant potential in various aspects of clinical care, including sepsis and ICU mortality

prediction, glaucoma progression forecasting, surgical action detection and prediction, depression diagnosis prediction, and family history information extraction. These innovative approaches, which leverage the capabilities of transformer-based models, have enhanced model interpretability, provided meaningful justifications that align with expert assessments, and improved the prediction and classification of clinical outcomes. Although challenges remain, such as variable F1 scores and superior performance in recognizing current actions compared to predicting future actions, the overall results indicate that transformer-based models hold promise for practical application in clinical decision support.

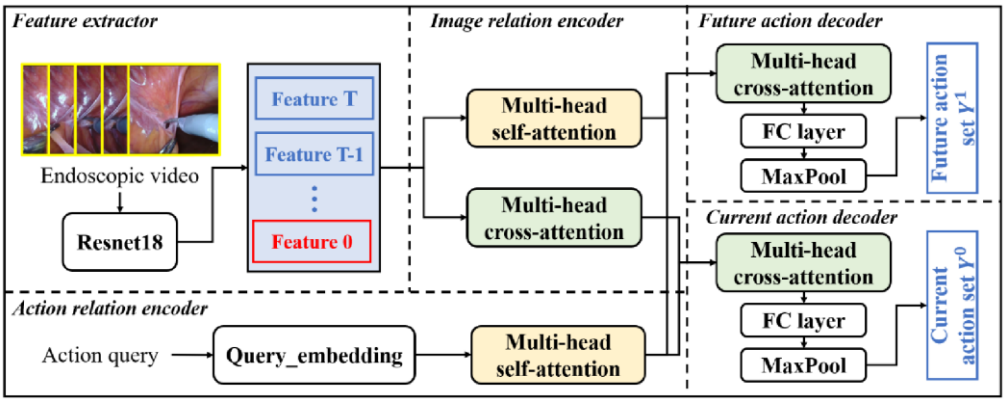


Figure 2-14. Depiction of the Proposed SA Model for Online Surgical Action Recognition and Prediction.(G. Huang 2022).

3.2.6. Protein structure prediction:

Protein structure prediction is a critical area of research with significant implications for understanding protein functionality and drug design. Recent studies have employed transformer-based models, specifically designed for protein sequences, to predict various aspects of protein structure and function. (Vig et al. 2020) presented a comprehensive analysis of the interpretability of transformer models in the context of protein structures. They assessed five transformer models, including TapeBert, ProtTrans, ProtBert, ProtBert-BFD, ProtAlbert, and ProtXLNet, using attention mechanisms to elucidate proteins' structural and functional attributes. Their findings highlighted that attention mechanisms could reveal proteins' three-dimensional folding patterns and binding sites, which are crucial for their functionality.

They further developed this work by employing the ProtAlbert transformer and introducing novel methods for interpreting attention weights, leading to more accurate predictions of protein sequence profiles. In a different approach, (Abdine et al. 2023) proposed Prot2Text **Figure 2-15**. This model generates textual descriptions of protein functions by combining Graph Neural Networks (GNNs) and LLMs, thus providing detailed and accurate descriptions of protein functions.

Other recent advancements include the work by (Boadu, Cao, and Cheng 2023), who developed TransFun, a technique that combines protein sequences and 3D structures for accurate protein function prediction. (Y. Cao and Shen 2021) proposed TALE, a deep learning approach that uses sequence information for proteins and embeds protein function labels and sequence features into a joint latent space, enhancing the model's ability to generalize to novel sequences and functions. (Geffen, Ofran, and Unger 2022) streamlined the ProtBert model to create DistilProtBert, which performed exceptionally well in distinguishing actual protein sequences from random ones while minimizing computational resources. (Castro et al. 2022) developed Regularized Latent Space Optimization (ReLSO), a transformer-based autoencoder that optimizes protein sequences for fitness landscape navigation. At the same time (Ferruz, Schmidt, and Höcker 2022) designed ProtGPT2, a language model trained on protein sequences to generate novel protein sequences that mimic natural ones. Lastly, (Brandes et al. 2022) developed ProteinBERT, a specialized deep language model for protein sequences that amalgamates local and global representations for comprehensive end-to-end processing. (Oliveira, Pedrini, and Dias 2023) proposed TEMPROT, a technique that fine-tunes and extracts embeddings from pre-trained architectures on protein sequences, and TEMPROT+, an extension of TEMPROT integrated with BLASTp for enhanced performance. Together, these studies represent significant advancements in protein structure prediction and function annotation. They highlight the potential of transformer-based models and other innovative techniques to provide detailed and accurate insights into protein structures and functions, ultimately contributing to advancing drug design and other biomedical applications.

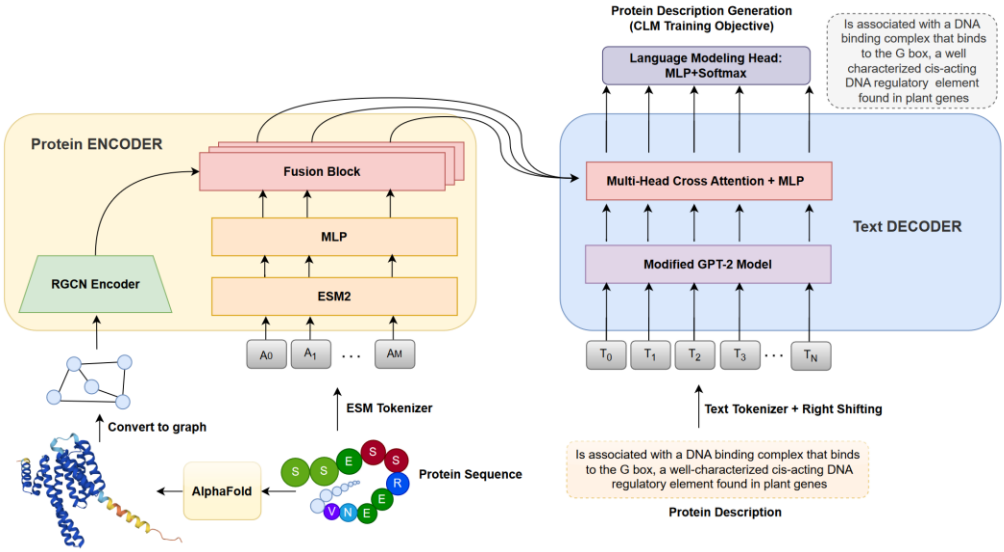


Figure 2-15. The proposed architecture of the Prot2Text framework is designed for predicting protein function descriptions in the free-form text (picture by (Abdine et al. 2023)).

3.2.7. Molecular Representation and Drug Design:

The recent advancements in transformer-based models have shown remarkable potential in molecular representation and drug design. These models leverage the capabilities of transformer decoders, masked self-attention, and self-supervised learning to generate valid, distinctive, and innovative compounds, predict drug-target interactions, and manage molecular properties. (Bagal et al. 2022) introduced MolGPT, a model designed to generate compounds with targeted scaffolds and chemical characteristics by utilizing scaffold SMILES strings. The model could construct molecules with property values that deviate from the provided values while maintaining the ability to produce molecules with user-specified scaffolds (**Figure 2-16**). (Fabian et al. 2020) implemented a BERT-based model, MOLBERT, which surpassed the prevailing state-of-the-art models on benchmark datasets by utilizing learned molecular representations. The study underscored the importance of selecting appropriate self-supervised tasks during pre-training to enhance the dependability

of the acquired representations and improve performance in downstream tasks such as Virtual Screening.

(K. Huang et al. 2021) proposed MolTrans, an approach that combines a knowledge-inspired sub-structural pattern mining algorithm, an interaction modeling module, and an enhanced transformer encoder for more precise and interpretable drug-target interaction (DTI) predictions. This approach outperformed leading-edge baseline models in a comparative analysis using real-world data. (H. Li, Zhao, and Zeng 2022) introduced the KPGT framework, which utilizes a knowledge-guided pre-training approach to learn molecular graph representations and overcome the current limitations of self-supervised learning methods in molecular property prediction. This approach surpassed existing state-of-the-art self-supervised learning approaches, demonstrating the effectiveness of the KPGT architecture. (Rong et al. 2020) developed GROVER, which interprets extensive quantities of unlabeled molecular data to glean intricate structural and semantic details about molecules. GROVER combines Message Passing Networks with Transformer-style architecture to create more expressive encoders for complex information. It identifies semantic motifs in molecular networks and predicts their existence in a molecule using graph embeddings and integrating domain knowledge.

In conclusion, transformer-based models such as MolGPT, MOLBERT, MolTrans, KPGT, and GROVER have shown promising results in generating innovative compounds, predicting drug-target interactions, and managing molecular properties. These models leverage scaffold SMILES strings, self-supervised tasks, knowledge-inspired sub-structural pattern mining algorithms, and graph embeddings to achieve superior performance in various molecular representation and drug design aspects. The integration of domain knowledge, appropriate self-supervised tasks during pre-training, and the incorporation of knowledge-guided pre-training approaches have been pivotal in overcoming the existing limitations and enhancing the performance of these models.

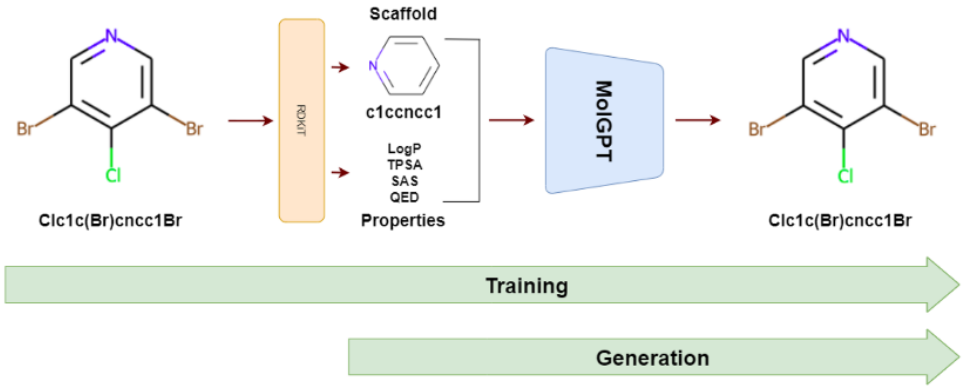


Figure 2-16. Pipeline for training and generation using the MolGPT model (Bagal et al. 2022).

4. Comparative Review

Table 1 comprehensively categorizes the assessed diffusion model papers by their application, key findings, and the employed or inspired algorithms, including DDPMs, NCSNs, and SDEs. It highlights each algorithm's fundamental concepts and objectives and the potential practical applications that can be explored and implemented in future studies based on the reviewed literature. The literature encompasses a spectrum of applications such as image reconstruction, image-to-image translation, image generation, image classification, image registration, and image segmentation. For example, (Özbey et al. 2023) utilized AdaDiff for MRI reconstruction and discovered it surpassed existing methods. Similarly, (Xie and Li 2022) used MC-DDPM for reconstructing under-sampled medical images and concluded it was a promising technique. These studies underscore the efficacy of diffusion models in diverse medical imaging applications, from image reconstruction and translation to generation, classification, registration, and segmentation. Despite these advancements, it is necessary to investigate the reasons for their popularity in medical imaging and the success of certain tasks in adopting diffusion models. Primarily, their popularity is attributed to their effectiveness, ease of implementation, and high output quality, which are critical for obtaining high-resolution images with precise local information for disease detection. This has contributed to the growing popularity of diffusion models in medical imaging.

The table indicates that specific applications, such as image reconstruction, denoising, and generation, have garnered more attention than segmentation, text-to-image

translation, and registration. This is primarily due to the alignment of diffusion model theory with the objectives of reconstruction and denoising tasks. Diffusion models involve adding noise to data and denoising it until the original data is reconstructed, which aligns well with the requirements of these tasks. Additionally, diffusion models can effectively model the complex interactions between signals and noise in data, resulting in more accurate image reconstruction. While diffusion models have potential across different tasks, some may necessitate further modifications for adaptation. For instance, text-to-image translation requires an auxiliary network with robust text encoding capabilities. Although initial works mainly focus on image generation, reconstruction, and denoising, more research is anticipated to address a broader range of tasks over time, as the table suggests a promising future for this field academically.

Furthermore, integrating transformers with specialized models and architectures has facilitated progress in various applications, from protein structure and function prediction to clinical documentation, diagnostic assistance, molecular representation, and drug design. For example, (Abdine et al. 2023) developed ProtAlbert, which leverages novel methods for interpreting attention weights, resulting in more accurate protein sequence profile predictions. Similarly, the TransFun model by (Y. Cao and Shen 2021) combines protein sequences and 3D structures for accurate protein function prediction. In clinical documentation, (Lentzen et al. 2022) developed a transformer deep neural network algorithm for extracting clinically relevant text from PDF clinical documents, enhancing downstream task performance, such as medical concept extraction. These studies exemplify the versatility and effectiveness of transformer-based models and architectures in addressing diverse challenges across various biomedicine domains. In conclusion, the reviewed studies demonstrate the potential of diffusion and transformer-based models and architectures in various biomedical applications. Despite the significant advancements, there is a need for continued research to explore further applications and refine the models for optimal performance in specific tasks.

Table 2-1: Summary of the evaluated diffusion models and transformer-based models in healthcare applications

Application	Paper	Networks	Application	Key Findings
Image Reconstruction	(Özbey et al. 2023)	AdaDiff	MRI Reconstruction	AdaDiff leads to superior results compared to existing methods
	(Xie and Li 2022)	MC-DDPM	Undersampled Medical Images Reconstruction	MC-DDPM is a promising technique for reconstructing under-sampled medical images
Image To Image Translation	(Lyu and Wang 2022)	DDPM, SDE	Translation from MRI to CT	The approach includes a conditional DDPM and a conditional SDE. The model generates outputs ten times, with an average determining the conclusive result. The approach shows superior performance when benchmarked against existing GAN and CNN methodologies. The authors employ the Monte Carlo technique to gauge the reliability of the diffusion models.
	(Özbey et al. 2023)	SynDiff	Unsupervised medical performance of medical image translation	The proposed SynDiff method is based on a diffusion process and a novel adversarial loss function. The authors conducted extensive experiments and showed that SynDiff outperforms state-of-the-art GAN and diffusion models regarding accuracy, artifact reduction, and computational efficiency. Sensitivity analysis showed that SynDiff is relatively insensitive to different hyperparameters.
Image Generation	(Müller-Franzes et al. 2023)	Medfusion (DDPM)	Medical imagery: RGB eye fundus images, chest CT scans, 3D MRI brain scans	Medfusion outperformed GANs regarding image quality (higher PSNR and SSIM values, lower FID scores). Some hyperparameters were more crucial than others for Medfusion's performance. Medfusion has improved stability and interpretability over GANs.
	(Pan et al. 2023)	Transformer-based denoising diffusion probabilistic model	Synthesizing 2D medical images: chest X-rays, heart MRIs, pelvic CT images, abdomen CT images	The method outperformed existing state-of-the-art methods in image quality and fidelity. The Inception score and Fréchet Inception Distance score confirmed that the synthetic images generated belonged to the same data distribution as real images.
Image Classification	(H.-J. Oh and Jeong 2023)	DiffMix	Segmenting and classifying nuclei in imbalanced	DiffMix is a semantic-label-conditioned diffusion model that creates synthetic data samples,

			pathology images	enhancing the classification efficacy of rare nuclei types.
	(Y. Yang et al. 2023)	DiffMIC	Classifying medical imaging via diffusion algorithms	The DiffMIC method classifies medical images by translating them into distinct feature spaces and employing a denoising U-Net model.
Image Registration	(Kim, Han, and Ye 2022)	DiffuseMorph	Deformable Image Registration	Introduced a new method consisting of two primary systems: one for diffusion and one for deformation. This method can smoothly show how one image transforms into another.
Image Segmentation	(Azad et al. 2022)	DDPM	Image Segmentation	Diffusion models can create labeled data and reduce the need for detailed image annotations.
	(Kim, Oh, and Ye 2022)	DARL	Blood Vessel Segmentation	Introduced a new DARL self-guided blood vessel segmentation method to identify vascular disorders. This approach creates vessel segmentation visuals or artificial angiograms.
Protein structure prediction	(Vig et al. 2020)	TapeBert, ProtTrans, ProtBert, ProtBert-BFD, ProtAlbert, ProtXLNet	Protein structure and function prediction	Attention mechanisms can reveal proteins' three-dimensional folding patterns and binding sites.
	(Behjati et al. 2022)	ProtAlbert	Protein sequence profile prediction	Novel methods for interpreting attention weights led to more accurate predictions of protein sequence profiles.
	(Abdine et al. 2023)	Prot2Text	Protein function description	Combining Graph Neural Networks (GNNs) and LLMs provides detailed and accurate descriptions of protein functions.
	(Boadu, Cao, and Cheng 2023)	TransFun	Protein function prediction	Combining protein sequences and 3D structures leads to accurate protein function prediction.
	(Y. Cao and Shen 2021)	TALE	Protein function prediction	Embedding protein function labels and sequence features into a joint latent space enhances the model's generalization ability to novel sequences and functions.
	(Geffen, Ofran, and Unger 2022)	DistilProtBert	Protein sequence distinction	The streamlined ProtBert model performed exceptionally well distinguishing actual protein sequences from random ones while

				minimizing computational resources.
	(Castro et al. 2022)	ReLSO	Protein sequence optimization	Transformer-based autoencoder optimizes protein sequences for fitness landscape navigation.
	Ferruz, Schmidt, and Höcker 2022	ProtGPT2	Novel protein sequence generation	Language models trained on protein sequences can generate novel proteins that mimic natural ones.
	(Ferruz, Schmidt, and Höcker 2022)	ProteinBERT	Protein sequence processing	A specialized deep language model amalgamates local and global representations for comprehensive end-to-end processing.
	(Oliveira, Pedrini, and Dias 2023)	TEMPROT, TEMPROT+	Protein sequence embedding extraction	Fine-tuning and extracting embeddings from pre-trained architectures and integrating with BLASTp enhances performance.
	(Brandes et al. 2022)	ProteinBERT	Protein sequence processing	A specialized deep language model amalgamates local and global representations for comprehensive end-to-end processing.
Clinical documentation and information extraction	(Gérardin et al. 2023)	Transformer deep neural network	Analyzing the layout of PDF clinical documents	Developed and validated an algorithm for extracting clinically relevant text from PDF clinical documents. The algorithm improved the results in downstream tasks such as medical concept extraction, proving its utility in a clinical context.
	(Lentzen et al. 2022)	BioGottBERT	German clinical notes	Newly trained BioGottBERT model outperformed the GottBERT model in clinical named entity recognition (NER) tasks.
	(Y. Li et al. 2022)	ClinicalLongformer, Clinical-BigBird	Clinical text	Introduced two domain-specific language models pre-trained on a large corpus of clinical text, improving several downstream clinical NLP tasks such as question answering, NER, and document classification.
	(Moon, He, and Liu 2022)	BERT, Sentence-BERT	Clinical texts across different practice environments	Analyzed a corpus of 500,000 clinical texts and found narrow scopes in evaluation and discharge summary documents and similar cluster distributions in Family Medicine and Primary Care practice environments. Emergency Medicine exhibited a unique sublanguage.

	(S. H. Oh, Kang, and Lee 2022)	XLNet, RoBERTa, BERT	De-identification of protected health information (PHI)	XLNet outperformed the other models in PHI recognition using the i2b2 2014 dataset.
	(Searle et al. 2023)	BART	Summarization of Biomedical Health Care (BHC) data	Proposed an advanced abstractive summarization model based on BART that includes a clinical oncology-aware guidance signal for key terms, facilitating the creation of problem-list-oriented abstractive summaries.
	(Sivarajkumar and Wang 2022)	HealthPrompt	Clinical texts	Proposed a groundbreaking prompt-based clinical NLP framework called 'HealthPrompt' that customizes task definitions via prompt template specification, eliminating the need for training data.
	(Solarte-Pabón et al. 2023)	Transformer-based approach	Spanish clinical notes related to breast cancer	Proposed a transformer-based approach facilitated by a specialized breast cancer corpus and a schema for annotating clinical notes, suggesting that transformers are potentially effective in extracting information from Spanish clinical texts.
	(Wei et al. 2022)	ClinicalLayoutLM	Categorizing scanned clinical documents	Introduced a multimodal technique that combined text obtained from optical character recognition (OCR) with layout or image information, outperforming the baseline model (which relied solely on OCR text) in classifying scanned clinical documents into 16 categories.
	(Yogarajan et al. 2021)	Domain-specific transformers	Predicting medical codes	Found that domain-specific transformers performed better for texts with fewer labels and/or documents shorter than 300 words, while conventional neural networks were more effective for less frequently occurring labels and documents exceeding 300 words.
Diagnostic assistance	(Azizi, Hier, and Wunsch Ii 2022)	Transformer-based bidirectional encoder representations model, CNNs	Named entity identification of neurological signs and symptoms.	Superiority of transformer-based model in identifying signs and symptoms.
	(S. Chen et al. 2023)	Feature Interaction Transformer network (FIT-Net)	Diagnosing pathologic myopia through Optical	Superior performance over traditional deep learning methods.

		Coherence Tomography images.	
(X. Chen et al. 2022)	Multi-view Vision Transformers (MVTs)	Long-range correlations between mammograms.	Superior reproducibility, robustness, and case-based malignancy classification accuracy over other alternatives.
(Dhinagar et al. 2023)	Vision Transformer (ViT) architecture	Gender and Alzheimer's disease classification using 3D brain magnetic resonance imaging.	Importance of ViT training strategies in neuroimaging applications.
(Dong et al. 2023)	CNN-DVIT (combination of vision transformer structure with deformable attention and CNNs)	Multi-label arrhythmia classification.	Outperformed the most recent transformer-based ECG classification methods.
(Garaiman et al. 2023)	Vision Transformer (ViT)	Identifying specific microangiopathy indicators in nailfold capillaroscopy images of patients with Systemic Sclerosis.	Proficient in detecting various microangiopathic changes despite higher average accuracy by rheumatologists.
(Hosain et al. 2022)	Vision transformer and transfer learning model	Categorizing endoscopic imagery attributes for identifying gastrointestinal diseases from wireless capsule endoscopy images.	Outperformed DenseNet201 in identifying gastrointestinal diseases.
(D. Hu et al. 2022)	Cross-modal retrieval framework	Executing retrieval tasks involving histopathology whole slide images and diagnostic reports.	Effective strategy in performing cross-modal retrieval tasks.

	(Mo et al. 2023)	HoVer-Transformer	Identifying breast cancer in ultrasound images.	Outperformed other CNNs-based models and experienced sonographers in interim analysis.
	(Zhou et al. 2023)	Transformer-based representation-learning model	Consolidating multimodal information processing as a clinical diagnostic tool for identifying lung illnesses and predicting adverse clinical outcomes in COVID-19 patients.	Superior performance compared to image-only and non-unified multimodal diagnostic models.
Medical imaging and radiology interpretation	(Balouch and Hussain 2023)	Biobart-V2	Summarizing medical texts	The ROUGE score showed a state-of-the-art performance in summarizing medical texts.
	(Bhattacharya, Jain, and Prasanna 2022)	RadioTransformer	Assimilating radiologists' visual search patterns	Enhanced confidence in decision-making within a cascaded global-focal transformer framework.
	(Chaudhari et al. 2022)	Radiology BERT	Identification of clinical speech recognition errors	Enhanced the identification of actual clinical speech recognition errors in radiology reports.
	(Jacenkow, O'Neil, and Tsaftaris 2022)	BERT model	Multimodal classification using dual image-text input	Improved image classification in radiology reports.
	(J. Li et al. 2022)	BERT-based models, IDPT, sequence adaptation strategy	Classification of actionable radiology reports in tinnitus patients	Yielded promising results.
	(Moezzi et al. 2022)	Transformer-based fine-grained NER architecture	Clinical information extraction	The performance of previously utilized techniques was surpassed, delivering a coherent and structured report.
	(Mohsan et al. 2023)	TrMRG	Report generation	Achieved noteworthy results compared to prevailing methods.
	(Nimalsiri et al. 2023)	MERGIS	Automated report generation	Utilized image segmentation and a modern transformer-based encoder-decoder model to enhance the accuracy of automated report generation.

	(J. Feng, Shaib, and Rudzicz 2020)	Hierarchical CNN transformer, ClinicalBERT	Sepsis prediction, ICU mortality	The model captures phrase-level patterns and global contextual linkages of medical language, generating meaningful justifications that align with expert assessments, demonstrating its potential as a transparent decision-support tool.
	(W. Hu and Wang 2022)	BERT, BioBERT, RoBERTa, DistilBERT	Predicting glaucoma progression	The models indicate potential in forecasting glaucoma progression.
	(G. Huang 2022)	SA model (encoder-decoder framework)	Surgical action detection and prediction	Superior performance in recognizing current actions compared to predicting future actions, facilitating contextual decision-making in operating theatres.
Clinical Decision Support	(Meng et al. 2021)	BRLTM (Bidirectional Representation Learning model)	EHR data modeling, depression screening, and risk prediction	Bidirectional learning outperforms forward-only approaches in sequence modeling, enhancing interpretability by elucidating relationships between different EHR codes in sequences, enabling the identification of individuals requiring depression screening and those at risk of developing depression within a specified time frame.
	(Wang et al. 2023)	Transformer-based approach	Family history information extraction	The integration of rule-based and deep learning-based systems enhanced the recall of FH information, although the F1 score remained variable yet comparable.
	(Liu et al. 2022)	Hierarchical Label-Wise Attention Transformer (HiLAT)	Medical coding and billing	HiLAT model uses a two-tier hierarchical label-wise attention mechanism to generate label-specific document representations and predict the likelihood of assigning a particular ICD code to a clinical document. It outperformed previous state-of-the-art models for the 50 most frequently observed ICD-9 codes.
Medical coding and billing	(López-García et al. 2023)	Multilingual transformer-based models (XLM-RoBERTa, mBERT, BETO)	Explainable clinical coding	A hierarchical task approach for training the models, comprising medical-named entity recognition (MER) and medical-named entity normalization (MEN), yielded superior results compared to a multitask approach.
	(Ng, Santos, and Rei 2023)	Hierarchical Transformers for	Temporal modeling of	HTDS processes textual and metadata content from a series of documents and outperforms the

		Document Sequences (HTDS)	document sequences	previously established state-of-the-art model when the entire collection of clinical notes is used as input.
	(Shang et al. 2019)	G-BERT model	Articulating medical codes and recommending medications	G-BERT combines the strengths of Graph Neural Networks (GNNs) and BERT for articulating medical codes and recommending medications. It captured the underlying hierarchical structures in medical codes and outperformed all baseline models regarding prediction accuracy for medication recommendation tasks.
	(Tchouka et al. 2023)	A model leveraging NLP and multi-label classification	ICD10 code association	The model was rigorously evaluated on a French clinical dataset and demonstrated its efficacy as the most accurate solution for ICD10 code association in the French language to date.
Molecular Representation and Drug Design	(Bagal et al. 2022)	MolGPT	Generate compounds with targeted scaffolds and chemical characteristics	MolGPT utilizes scaffold SMILES strings to construct molecules with property values that deviate from the provided values while maintaining the ability to produce molecules with user-specified scaffolds.
	(Fabian et al. 2020)	MOLBERT	Predict drug-target interactions, manage molecular properties and Virtual Screening.	MOLBERT utilized learned molecular representations and outperformed prevailing state-of-the-art models on benchmark datasets. The study highlighted the importance of selecting appropriate self-supervised tasks during pre-training.
	(K. Huang et al. 2021)	MolTrans	More precise and interpretable drug-target interaction (DTI) predictions	MolTrans combines a knowledge-inspired sub-structural pattern mining algorithm, an interaction modeling module, and an enhanced transformer encoder. It outperformed leading-edge baseline models in a comparative analysis using real-world data.
	(H. Li, Zhao, and Zeng 2022)	KPGT	Learn molecular graph representations, molecular property prediction	KPGT utilizes a knowledge-guided pre-training approach to overcome the current limitations of self-supervised learning methods and surpasses existing state-of-the-art self-supervised learning approaches.
	(Rong et al. 2020)	GROVER	Interpreting structural and semantic details about molecules, predict the	GROVER combines Message Passing Networks with Transformer-style architecture to create more expressive encoders for complex information. It identifies semantic

	existence of semantic motifs in molecules.	motifs in molecular networks and predicts their existence in a molecule using graph embeddings and integrating domain knowledge.

5. Future direction and open challenges

Generative AI, including diffusion models and transformer-based models, has showcased remarkable potential in the healthcare domain, particularly in medical imaging and disease diagnostics, by overcoming hurdles encountered by earlier models. It does not necessitate labeled data, making it a potent tool for numerous medical applications, and it excels in efficiently representing high-dimensional data like images. However, despite its merits, generative AI encounters challenges, such as slow generation processes, limited adaptability to certain data types, and the incapability of performing dimensionality reduction. This article intends to thoroughly assess recent medical research utilizing generative AI and classify studies to underscore its potential. We aim to emphasize the significance of generative AI in refining healthcare methodologies and pinpointing areas warranting further exploration.

Key challenges encompass investigating various medical imaging modalities, acquiring semantically meaningful data representations, enhancing architecture design, and addressing privacy concerns. Generative AI is apt for probing diverse modalities for specific downstream tasks. Although most extant studies concentrate on CT and MRI modalities, other modalities like ultrasound imaging could also be gained from generative AI. Despite VAEs and GANs being designed to retain and learn meaningful data representations, generative AI models struggle to form semantically meaningful data representations in their latent space (L. Yang et al. 2022). This issue is crucial as semantically meaningful representations facilitate superior image reconstructions and semantic interpolations. The network structure of generative AI models is a pivotal design choice influencing their capacity to comprehend intricate data relationships and yield high-quality outcomes. Recent research has delved into transformer models, but a comprehensive understanding of their capabilities and limitations is still lacking. Generative AI models can discern underlying

probability distributions of datasets and generate novel data points, which is beneficial for intricate causal inference, discovery, and counterfactual generation tasks. However, privacy remains a major concern in the medical community (Carlini et al. 2023). AI image synthesis models, including generative AI, face scrutiny for potentially infringing copyright laws and jeopardizing training data privacy.

Federated learning combined with generative AI can establish a robust learning platform in the medical domain, addressing privacy concerns and enhancing the quality of learned models. Additionally, reinforcement learning can be employed to estimate the optimal inversion path for the inverse problem-solving of generative AI models, where conventional optimization methods become computationally prohibitive or unfeasible. Future research should concentrate on these challenges to maximize the potential of generative AI in healthcare. Addressing these challenges will enhance the capabilities of medical imaging techniques, improve patient care, and contribute to the exciting and rapidly evolving field of generative AI in healthcare. Moreover, integrating diffusion and transformer models holds promise for addressing current challenges and limitations in healthcare applications. While diffusion models are potent tools for simulating disease spread, comprehending brain networks, and predicting drug efficacy, they struggle with managing high-dimensional data and capturing complex interactions. Conversely, transformer models, rapidly adopted in healthcare for tasks like clinical report generation, medical image segmentation, and drug-drug interactions, face challenges associated with interpretability, environmental impact, computational costs, fairness and bias, AI alignment, and data privacy and sharing.

Interpretability is a significant concern as deep learning systems are often perceived as "black box" models, impeding the systemic acceptance of AI-aided diagnostics in the medical domain. Although transformers provide some transparency through attention weights visualization, the interpretation is often fragmented and not robust (Chefer, Gur, and Wolf 2021). There have been efforts to enhance the interpretability of transformer models, such as proposing B-cos transformers and generating attention visualizations and cosine similarity between learned clinical diagnoses embeddings (Böhle, Fritz, and Schiele 2023). However, there is still a need for novel techniques to enhance interpretability tailored

towards healthcare AI. The environmental impact of AI advancements has been substantial, with large-scale deep learning models generating significant carbon dioxide emissions. Initiatives have been undertaken to promote energy-efficient hardware and algorithms, such as the United Kingdom National Health Service's net-zero emissions goal by 2040. However, global initiatives are needed to make AI sustainable and accessible worldwide. Computational costs associated with the high parametric complexity of transformers and the necessity for vast datasets present challenges for healthcare settings that typically require lightweight models for real-time predictions with minimal maintenance costs (Strubell, Ganesh, and McCallum 2019). Techniques such as pruning, knowledge distillation, and quantization can provide more efficient model implementations for deployment within practical hardware constraints (Lagunas et al. 2021; Sun et al. 2019; Yao et al. 2022). Fairness and bias in AI models can lead to unfavorable treatment of specific patient groups, often arising from the under-representation of certain populations in training datasets. Continuous monitoring and auditing for fairness and bias post-deployment are necessary to ensure equitable healthcare outcomes. AI alignment involves designing AI systems that align with human values and goals while minimizing unintended consequences and harmful outcomes (Moor et al. 2023). It is especially critical in healthcare to ensure that large-scale foundation models are ethical, responsible, respectful of patient privacy, and not causing harm. A clear set of standards and guidelines is needed to establish the ethical use of AI in healthcare.

Data privacy and sharing are also significant concerns, with federal regulations like HIPAA impacting the development of large models that require vast amounts of data. The federated learning paradigm presents a potential solution by developing a shared training model that leverages data from multiple fragmented sources without divulging sensitive patient information (Gostin 2001; Van Panhuis et al. 2014). However, technical challenges in building an operational federated learning workflow must be addressed, such as inhomogeneous data distributions and inconsistent privacy preservation settings. In conclusion, integrating diffusion and transformer models presents a promising future direction in healthcare applications. Addressing the challenges related to interpretability, environmental impact, computational costs, fairness and bias, AI alignment, and data privacy

and sharing is essential for successfully deploying these models in healthcare settings. Novel techniques and approaches, such as federated learning and model compression methods, are needed to overcome these challenges and make AI sustainable, accessible, and ethical in healthcare applications.

6. Conclusion

In this study, we explored the literature surrounding transformers and diffusion models, emphasizing their use in healthcare. We delved into diffusion models' roles in tasks like image reconstruction, image translation, generation, classification, etc. Additionally, we discussed transformer models' applications in protein structure prediction, clinical documentation, diagnostic assistance, radiology interpretation, clinical decision-making, medical billing, and drug design. We offered a structured categorization and a broad overview of key methods for each use case. We also differentiated the diffusion models into three main categories: DDPMs, NCSNs, and SDEs, and elaborated on the attention mechanisms within transformers. As a closing remark, we highlighted future research directions. While we shed light on the burgeoning application of transformers and diffusion techniques in healthcare, we recognize that the domain is still nascent and evolving. As these models rise in prominence and undergo further research, our review is a crucial guide for those aiming to leverage them. We aspire for our analysis to spark further interest and investigation into the capabilities of transformers and diffusion models in healthcare. Some references in our study are pre-prints.

Nevertheless, we have been meticulous in choosing top-tier research from trustworthy sources. Incorporating pre-prints offers a holistic view of this swiftly progressing arena. Our review furnishes insightful perspectives on the potential of transformers and diffusion models in healthcare, pinpointing exciting prospects for forthcoming research.

Chapter 3
Rapid Prediction of Retina Stress and Strain Patterns in Soccer-related Ocular Injury: Integrating Finite Element Analysis with Machine Learning Approach

Abstract

Soccer-related ocular injuries, especially retina injuries, attract increasing attention. Mechanics of a flying soccer ball induced abnormally higher retinal stresses and strains, and its correlation with retinal injuries has been characterized using the finite element (FE) method. However, FE simulations demand solid mechanical expertise and extensive computational time, which is difficult to adopt in clinical settings. This study proposes a framework that combines FE analysis with a machine learning (ML) approach for the fast prediction of retina mechanics. Different impact scenarios were simulated using the FE method for obtaining the von Mises stress map and maximum principal strain map in the posterior retina. These stress and strain patterns and their input parameters were used to train and test a partial least squares regression (PLSR) model for predicting the soccer-induced retina stress and strain in terms of distributions and peak magnitudes. The peak von Mises stress and maximum principal strain prediction errors were 3.03% and 9.94% for the frontal impact and 9.08% and 16.40% errors in the diagonal impact, respectively. The average prediction error of von Mises stress and the maximum principal strain was 15.62% and 21.15% for frontal impact and 10.77% and 21.78% for diagonal impacts, respectively. This work provides a surrogate model of FE analysis for fast predicting the dynamic mechanics of the retina in response to the soccer impact, which could be further utilized for developing a diagnostic tool for soccer-related ocular trauma.

Keywords: Soccer-related ocular injuries, finite element analysis, retinal stress, retinal strain, machine learning, partial least squares regression.

Introduction

Soccer is the fastest growing youth sport, and soccer-related ocular injuries in young players attract increasing attention (Patel et al. 2021). A rapid-moving soccer ball hitting the eye could cause hyphemia, corneal abrasions, traumatic retinal edema, retinal hemorrhage, retinal detachment, macular hole, choroidal hemorrhages, and even impaired vision (Ghosh and Bauer 1995; MacEwen 1989; Reed et al. 2002; J.A.C. Filipe, Fernandes, et al. 2003; J.C. Filipe, Rocha-Sousa, et al. 2003; Vinger and Filipe 2004; Yan 2020; Kent et al. 2007; Gökçe et al. 2013; Horn et al. 2000; Hassan et al. 2018). Horn et al. (Horn et al. 2000) reported thirteen soccer-related retinal injury cases, with half of these cases requiring surgical intervention. Filipe et al. (J.A.C. Filipe, Fernandes, et al. 2003) reported severe ocular injuries in soccer players at all skill levels, sometimes without earlier symptoms. Retinal hemorrhage on the posterior segment of the eye is one common ocular injury in soccer youth players (Leshno et al. 2021). It is critical to quantify the flying soccer ball-induced forces in the eyeball to better understand the mechanism of soccer-related ocular injuries.

The finite element (FE) method has been a popular tool to quantify the physics and mechanics of lesions and provide insights into injuries. Weaver et al. characterized the ocular lesion's stresses, energy, and pressure associated with the injury risk (Weaver et al. 2011). Stress and strain distributions were also obtained from FE models in abuse head trauma (AHT) (Stitzel et al. 2002; Tong et al. 2019; Saffioti 2014; Yamazaki, Yoshida, and Mizunuma 2014; Suh et al. 2021; Rangarajan et al. 2009). Clemente et al. validated the FE simulation by in vitro experiments of bullet impact to the porcine eyeballs. They demonstrated that retinal injuries were caused by the tension resulting from the reflection of pressure waves (Clemente et al. 2014). Liu et al. performed an FE analysis of the plastic pellets impact on the eyeball. They showed that the pressure wave propagation in the retina caused the retina tear, and negative pressure contributed to the retina detachment (Liu et al. 2013). Our group has also developed FE models to capture the pressure wave propagation in the vitreous and the impact on the retina for understanding the mechanism of soccer-related retina injuries (M.R. Lam et al. 2022b). These quantitative analyses have enhanced

the mechanistic understanding of soccer-related retina injuries. However, the computational cost hinders the direct application of FE analysis in clinical settings.

The machine learning (ML) method, integrated with the FE method, could solve the problem of computational cost, and instantly predict the FE results. The ML methods have been applied in biomedical research focusing on image processing and disease diagnosis, while the integration of FE and ML methods is still in its early stage. Gharaibeh et al. have developed a deep learning approach for segmenting calcified coronary plaque from optical coherence tomography images (Gharaibeh et al. 2019). ML has also been utilized to categorize glaucoma, diabetes, etc., based on fundus images (Schmidt-Erfurth et al. 2018; Shoba and Therese 2020). FE data has been used to provide enriched datasets for ML models for predicting the Young's Modulus and Poisson's ratio of composite materials (X. Li et al. 2019; Ye et al. 2019; Ford et al. 2021). Our group also developed simulation-driven ML models for predicting the lumen area following the implantation of a medical device, stent. Our earlier work showed that the mechanistic understanding of the FE method could enhance the feature selection and prediction accuracy of ML models (Dong et al. 2020). Liang et al. have developed a deep learning approach to estimate the stress distribution in the thoracic aorta based on FE datasets (Liang et al. 2018). This has demonstrated that the ML method could serve as a fast and reliable surrogate of the FE method.

This work will develop a surrogate model of FE analysis of soccer ball-induced retinal mechanics by combining the machine learning method with FE analysis. We will construct a streamlined three-dimensional eye model, including its essential components such as the sclera, vitreous, retina, and retinal vessels, anatomically located in a rigid human skull, which is then subjected to the impact from a deformable soccer ball with six different velocities (30mph, 32mph, 34mph, 38mph, 50mph, and 60mph) along two different impact orientations (frontal and diagonal). FE analysis of 12 different impact scenarios will be performed. The maximum principal strain and von Mises stress distributions in the posterior retina will be extracted from FE analysis as ML outputs. The FE input parameters (impact velocity and orientation) and outputs (distribution and values of von Mises stresses and maximum principal strains) are used to train the ML models and then predicting/visualizing the FE outputs in another three cases.

Materials and Methods

The overall workflow of our framework that integrates FE analysis with ML model for rapid prediction of soccer-related retinal injuries is illustrated in **Figure 3-1**. There are two major steps: 1) FE analysis of soccer-related ocular injuries with streamlined models: 2) Statistical model training with results from FE analysis for rapid prediction of ocular injury patterns.

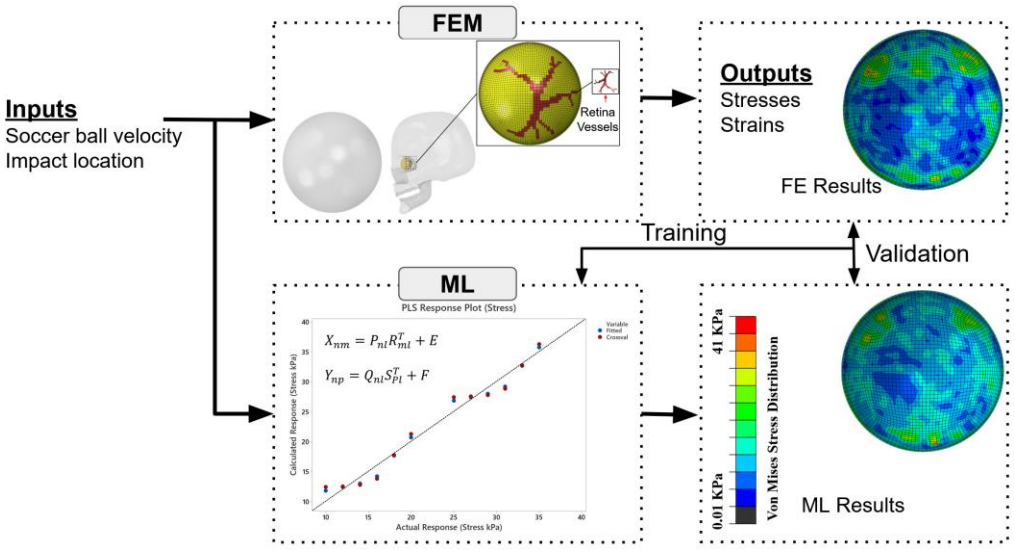


Figure 3-1. Overall framework to combine FE analysis with an ML-based statistical model training method for rapidly predicting soccer-induced posterior retina stress/strain patterns.

FE Analysis

A streamlined eye model of a young athlete was developed that retained the sclera, vitreous, retina, and retinal vessels (M.R. Lam et al. 2022b; M. Lam et al. 2021; Shokrollahi, Dong, et al. 2022b; Song et al. 2022). The eyeball, located in the skull, was subjected to a deformable soccer ball (**Figure 3-2**a). The sclera and retina were simulated as hollow spheres with outer diameters of 26mm and 24.5mm, and thicknesses of 0.8 mm and 0.25 mm, respectively. The retina was filled with the vitreous (Yamazaki, Yoshida, and Mizunuma 2014; Suh et al. 2021; Saffioti 2014). Retina vessels were incorporated along the posterior retina, and vitreous were highlighted in the posterior of the retina based on the

standard fundus photograph (Rangarajan et al. 2009; Saffioti 2014), as shown in **Figure 3-2**b. A rigid human skull model was used from the library of Grab cad (GrabCad 2018). The eyeball position in the skull was 16 mm (Pivonka 2018; Stephan 2013) between the corneal apex and the lateral orbital rim. A spherical shell with an outer diameter of 220 mm and a thickness of 0.8 mm (Hassan et al. 2018) was considered the standard soccer ball. We simulated both frontal and diagonal (45-degree angle) impacts. The soccer ball has an initial location of 15 mm in both frontal and diagonal to the apex of the eyeball and six different velocities (30mph, 32mph, 34mph, 38mph, 50mph, and 60mph). These results from 12 different impact scenarios were input to the statistical model training as described in the next section.

Material properties of the eye model are summarized in Table 1. Sclera, vitreous, and retina were considered incompressible materials with a Poisson's ratio of 0.49. The sclera was modeled as a hyperelastic material based on the published uniaxial test data (Tong et al. 2019). Vitreous, composed of mostly water (>90%), was considered as a viscoelastic material described by a Prony series expansion of the dimensionless relaxation modulus as:

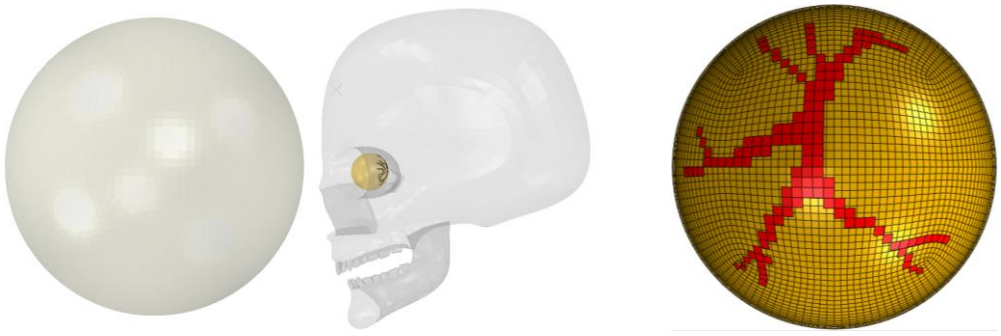
$$g_R(t) = 1 - \sum_{i=1}^{N} \bar{g}_i^P + (1 - e^{-t/\tau_i^G}) \quad (1)$$

Where materials constants $\bar{g}_i^P$ is 0.97, and τ_i^G is 0.07. Retina was modeled as elastic materials with Young's modulus of 20 KPa (Wollensak and Spoerl 2004) and Poisson's ratio of 0.49, and the skull was considered a rigid body. The surface-based fluid cavity technique was used to simulate the air pressurization of the soccer ball (Price, Jones, and Harland 2006). The pressure in the cavity of the soccer ball was filled with air A cavity reference node with a single degree of freedom was chosen at the soccer ball's center of gravity to represent the pressure inside the cavity (Systèmes 2013). Following the mesh convergence study, the number of elements for each component is shown in Table 1. The rigid skull was fixed (Razaghi, Biglari, and Karimi 2020). Four points on the frontal plane of the sclera, which is 7.8 mm in front of the equator of the eyeball, were fixed to mimic the constraints of extraocular rectus muscles, which stabilize and control the movement of the eye in humans (Rangarajan et al. 2009; Saffioti 2014). A penalty method with a friction coefficient of 0.3

was implemented among the soccer ball and eyeball. Frictionless contact was enforced between the retina and vitreous. Tie constraints were prescribed between the retinal vessels area and vitreous and between the retina and sclera. We used the final von Mises stress and maximum principal strain value at the end of the simulation in these 12 impact scenarios. The solutions were obtained using the dynamic explicit solver of a commercial Abaqus/Explicit software version 2019 (Dassault Systemes Simulia Corporation., Providence, RI, USA).

Table 3-1: Material parameters of the FE model and the number of elements.

Model component	Element type	Material model	Number of elements	Density (kg/ m^3)	Material parameters
Skull	R3D3	Rigid Body	18103	-	-
Sclera	C3D8R	Hyperelastic	28032	1243	(Tong et al. 2019)
Retina	C3D8R	Elastic	17856 (8322 located at the posterior retina)	1000	E= 20 KPa, $\nu = 0.49$ (Wollensak and Spoerl 2004)
Vitreous	C3D8R	Viscoelastic	103968	1009	E= 43 Pa, $\nu = 0.49$, $\bar{g}_i^P$ =0.97 , τ_i^G = 0.07 (Suh et al. 2021)
Rubber soccer ball	S4R	Elastic shell (isotropy and homogeneity)	5001	1160	E= 48 MPa, $\nu = 0.49$, A=28.97 g. mol^{-1}, B=29.19 g. mol^{-1}, C=0.9 bar, T=0.8mm (Price, Jones, and Harland 2006)
$\bar{g}_i^P$, τ_i^G= Prony series coefficients, A= Ideal Gas Molecular Weight, B= Molar Heat Capacity, C= Fluid Cavity Pressure, T= Thickness, 3-node, R3D3= 3-D rigid triangular facet, C3D8R= An 8-node linear brick, reduced integration, hourglass control, S4R= A 4-node doubly curved thin or thick shell, reduced integration, finite membrane strains.					



(a) (b)

Figure 3-2. Finite element model of the soccer ball trauma injuries. (a): Initial setup of a flying soccer ball hitting the face, including the eyeball; (b): Retinal vessels in the posterior retina (Frontal view).

Statistical Model Training

We trained a Partial Least Squares Regression (PLSR) model using the data acquired from FE simulations of 12 different eye impact scenarios. For this study, the goal is to illustrate the effectiveness of our framework for one eye's geometry. Therefore, the location and impact velocities are required as the geometry is the same for all simulations. Specifically, we used six different impact velocities from 30 mph to 60 mph and two different impact locations (frontal, diagonal).

PLSR is a multivariate technique that predicts response variables from predictor variables. The two fundamental equations in PLSR are the predictor matrix (X) and the response matrix (Y) given by (Shim et al. 2020)

$$X_{nm} = P_{nl}R_{ml}^{T} + E \tag{2}$$

$$Y_{np} = Q_{nl}S_{Pl}^{T} + F \tag{3}$$

where P and Q are the projection matrices, R and S are the transposed orthogonal loading matrices (where the rows are created from eigenvectors or principal components), and E and F are the error terms. X and Y are estimated using linear regression through

$$Y = X\widetilde{B} + \tilde{B}_0 \tag{4}$$

where $\widetilde{B}$ is the least squares regression estimate and $\widetilde{B}$ is the prediction error. In our study, the predictor variables are the initial velocity and the impact location. The response variables are the von Mises stress and maximum principal strain within the elements of the posterior retina. The model prediction accuracy was tested by doing a 'leave-one-out' analysis where one simulation from 12 was left out of the PLSR model and predicted independently and then repeated for each scenario. Specifically, the average errors between PLSR predictions and FE model results were determined by comparing the von Mises stress values from PLSR and FE models at every element (total of 8322 elements) within the posterior retina. Statistical model training was done with the PLSR plugins from the Python SciPy

(www.scipy.org) and scikit learn (machine learning) modules. Also, we used python script in the Abaqus to extract all elements' stress and strain values and automatically write predicted stress and strain results in the Abaqus ODB file to compare visualization of the FE simulation results and PLSR predicted results. We used MySQL (www.mysql.org) database connected to the PyCharm as an IDE for python programming to store all data,

Results

The distributions of von Mises stress and maximum principal strain in the posterior retina in different impact scenarios are shown in **Figure 3-3** and **Figure 3-4**. Different stress/strain distribution patterns caused by frontal and diagonal impact indicate that the impact direction plays an essential role in the stress/strain distribution following impact. In both frontal and diagonal impact scenarios, as the soccer ball velocity increased to 60 mph, the maximum value of von Mises stress and maximum principal strain in the posterior retina reached 50 KPa and 2.5, respectively. The frontal impact caused higher stress and strain in the posterior retina than diagonal impact, especially at the retina vessels bifurcation.

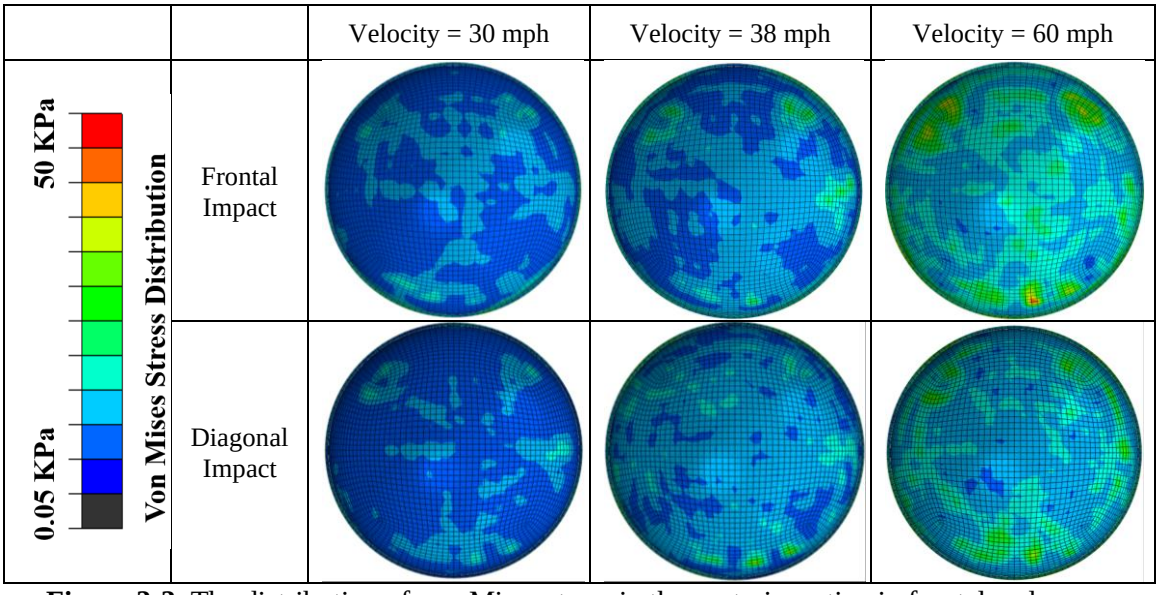


Figure 3-3. The distribution of von Mises stress in the posterior retina in frontal and diagonal impact scenarios at three different velocities.

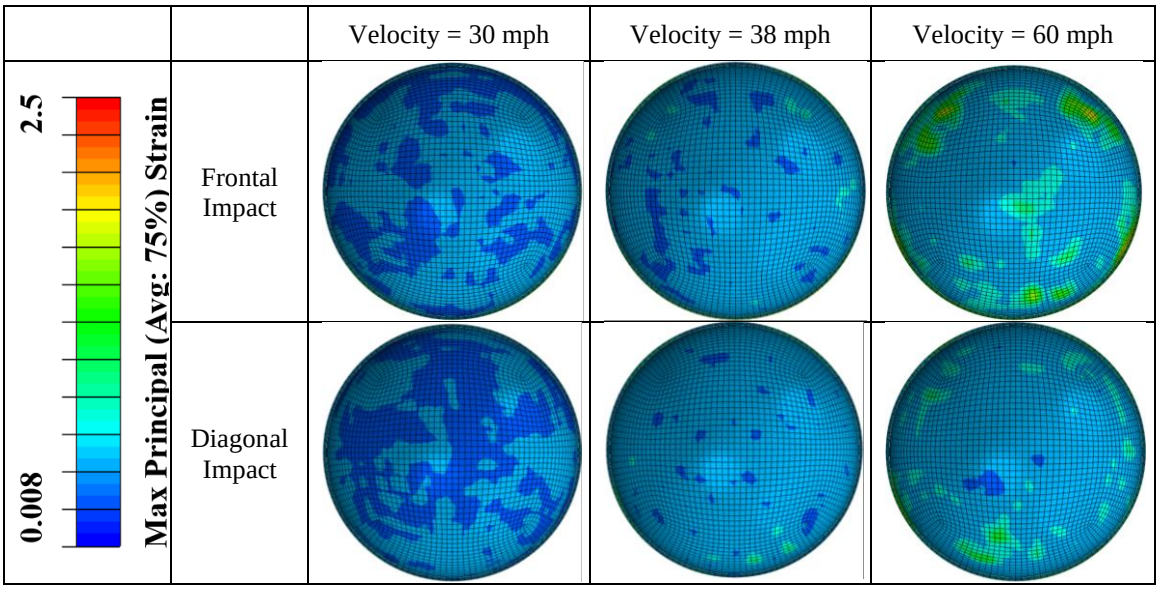


Figure 3-4. The distribution of maximum principal strain in the posterior retina in frontal and diagonal impact scenarios at three different velocities.

The distribution of von Mises stress and maximum principal strain predicted by our PLSR trained model was compared with the FE simulation (**Figure 3-5** and **Figure 3-6**) in the scenario of diagonal impact with soccer ball velocity of 35mph and frontal impact with soccer ball velocity of 65mph, respectively. Qualitative comparisons of the front, side and back views show that the PLSR trained model captured the overall von Mises stress distribution pattern, including the locations of the greatest values. The prediction error and relative prediction error of the maximum principal strain at each element in the retina are also shown in Fig. 6. Only a small portion (less than 5%) of the elements have a relative prediction error larger than 30% due to the low-stress value from FE analysis (nearly 0 value). In addition, our PLSR prediction also captured the von Mises stress distribution for a 30-degree impact (**Figure 3-7**), which angle is not included in the training dataset. This illustrates that our model can predict the stress with new velocity and impact location that had never been seen before by our trained model.

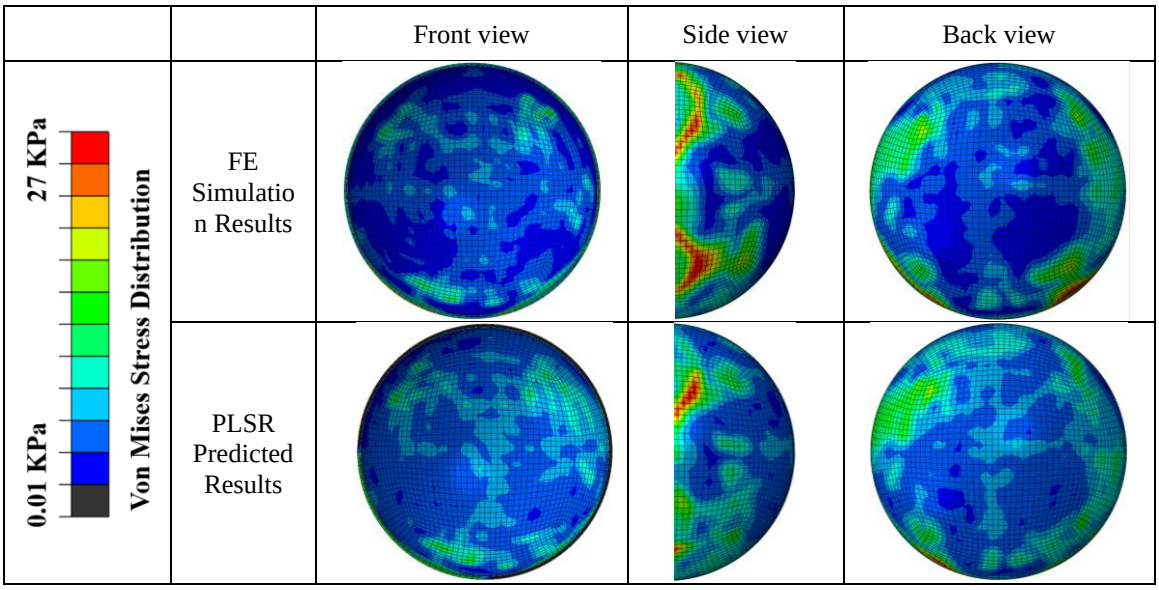
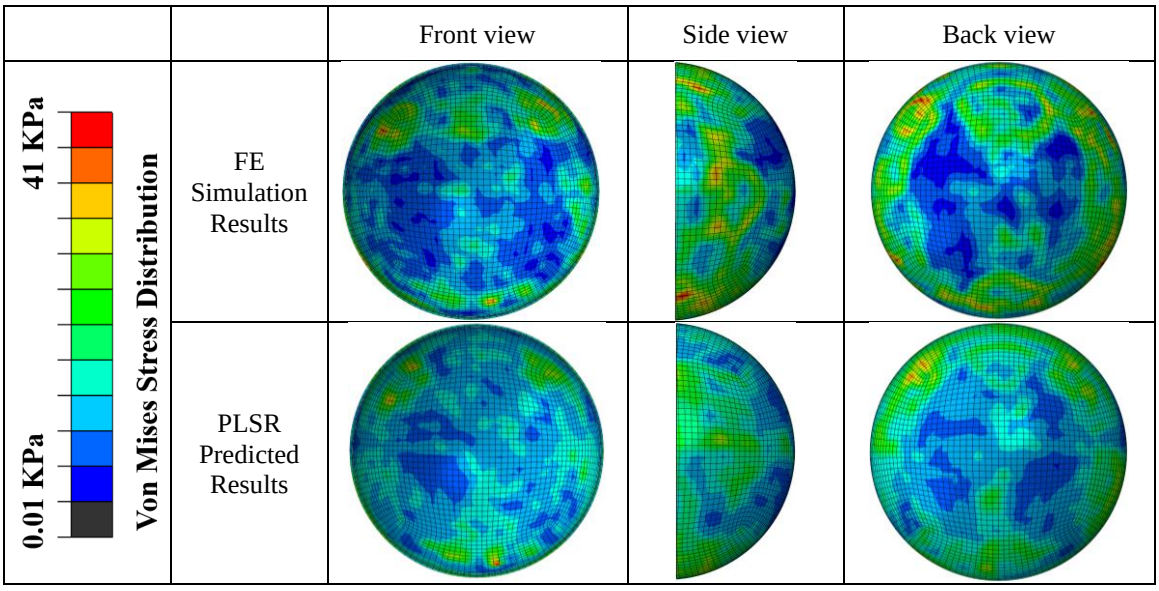


Figure 3-5. Comparison of FE simulated von Mises stress pattern and PLSR predicted von Mises stress (soccer ball velocity was 35mph and diagonal impact location). General agreements can be seen in the location and magnitude of maximum stress.



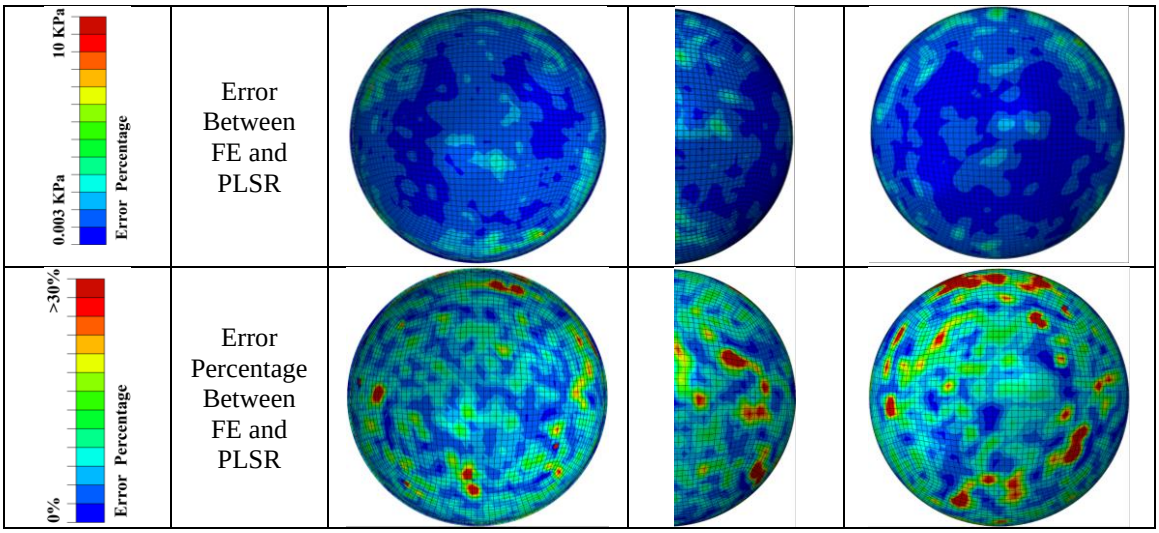


Figure 3-6. Comparison of FE simulated von Mises stress pattern and PLSR predicted von Mises stress (soccer ball velocity was 65 mph and frontal impact location). General agreements can be seen in the location and magnitude of maximum stress.

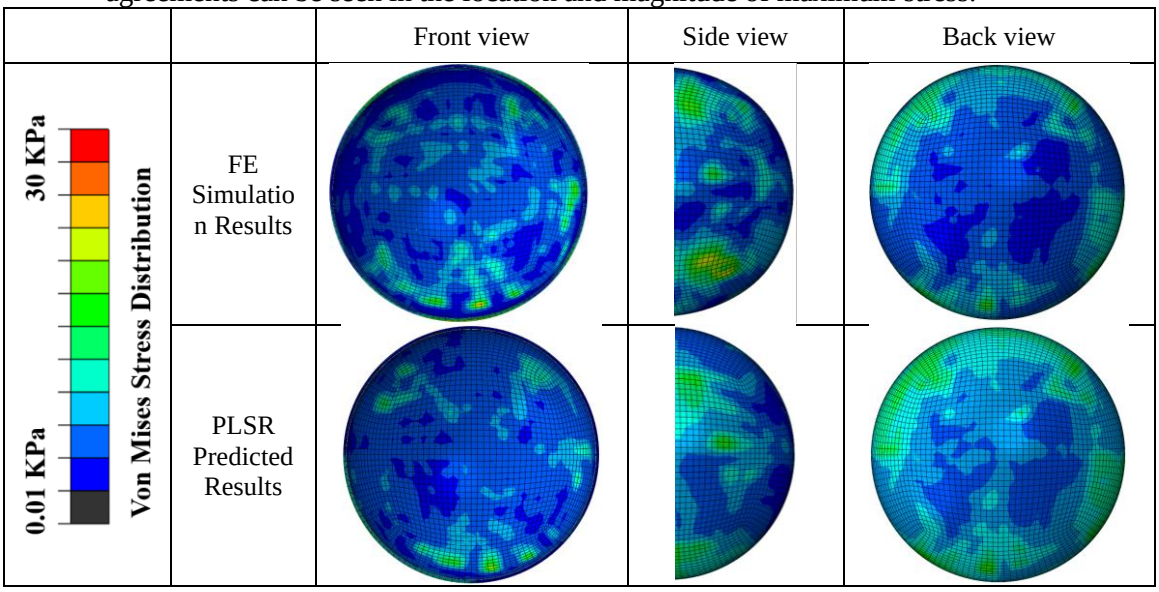
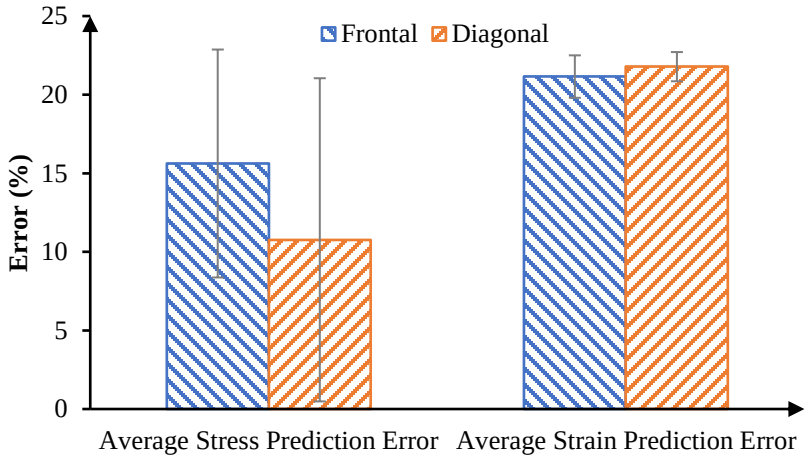


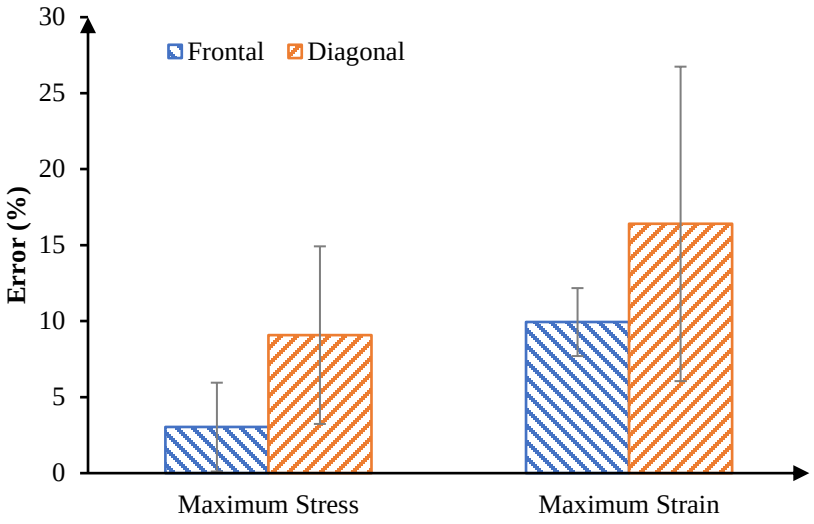
Figure 3-7. Comparison of FE simulated von Mises stress pattern, and PLSR predicted von Mises stress (soccer ball velocity was 40 mph and 30-degree angle impact location). General agreements can be seen in the location and magnitude of maximum stress.

The accuracy of our PLSR model was estimated by two different measures. First, we computed the absolute difference of stress and strain between FE and PLSR results in all

elements to calculate the average error. Also, an essential criterion for retina injuries is maximum stress and strain. To measure how well our PLSR model predicts the maximum von Mises stress and maximum principal strain, we compared the maximum von Mises stress and maximum principal strain from the FE and PLSR results for all simulations. The results are shown in **Figure 3-8**. As can be seen, the average error of von Mises stress and maximum principal strain between the PLSR and FE models were 15.62% and 21.15% for frontal impact and 10.77% and 21.78% for diagonal impacts. The error to estimate peak von Mises stress was 3.03% and 9.08% for the frontal and diagonal impact. The error of greatest maximum strain was 9.94% and 16.40% in frontal and diagonal impact. These results show that the location, magnitude, and distribution, of both principal strain and Von Mises stress, can accurately be predicted by our PLSR model.



(a)



(b)

Figure 3-8. (a): Comparison of average stress and strain error between FE results and PLSR prediction results and (b) comparing maximum stress and strain predicted error.

Discussions

This study illustrated the framework using FE analysis of soccer ball impact onto eyeball to train ML models, which can serve as a surrogate model for rapid prediction of posterior retina mechanics following the soccer ball impacts. Specifically, we used the PLSR method, which predicts the stress and strain distribution pattern by training a statistical model with FE simulation results. The impact velocity and orientation were used to predict the retinal stress and strain distributions, including the location and magnitude of peak von Mises stress and maximum principal strain. The prediction error of the peak retinal von Mises stress is approximately 3.03% for frontal impact and 9.08% for diagonal impact, respectively. The corresponding prediction error of the peak principal strain is 9.94% and 16.40%, respectively.

Our FE results demonstrated that the von Mises stresses in the posterior retina were noticeably increased, especially with a larger impact velocity and in regions interfaced with retina vessels and retina vessels bifurcations. In addition, the retinal stress and strain were aggregated by increasing the angle of impact from 45º (i.e., diagonal) to 0º (i.e., frontal). This agrees with the observations by Karimi et al. (Karimi et al. 2018). We have observed a peak von Mise stress of 50 kPa and average von Mises stress over the posterior retina of 7.5 kPa with a flying soccer ball of 50 mph at impact angle of 0º. Abnormally high von Mises stress and maximum principal strain have been linked with the risk of injuries (Wollensak and Spoerl 2004; Shim et al. 2020). Wollensak et al. (Wollensak and Spoerl 2004) reported the porcine retina fracture thresholds in engineering stress of 10 kPa and strain of 42% based on 30 enucleated porcine eyes. We have observed higher stress and strain value than the reported threshold values. The advantage of our approach lies in that we extracted stresses and strains in all discretized elements of the posterior retina. This allowed us to utilize the stresses and strains in every element as training of the ML model, then map all ML-predicted stresses and strains back to the posterior retina and examine the prediction errors in any element individually or averagely.

To overcome the major drawback of FE analysis, i.e., the computational cost, we utilized sophisticated FE models of various soccer-related eye impact scenarios to train ML-based models and then used the ML models for rapidly predicting the retinal von Mises stress and maximum principal strain distributions. Specifically, we adopted the PLSR method for training and prediction. The stress and strain locations and magnitude prediction for all elements had an average accuracy of 84.38% and 78.85% for stress and strain in the frontal impact and an average accuracy of 89.23% and 78.22% for stress and strain in the diagonal impact. Also, the accuracy of the maximum stress and strain predictions were 96.97% and 90.06% for frontal impact and 90.92% and 83.6% for diagonal impact, respectively. The reason that we got better accuracy in the prediction of stress than strain related to the deviation of the strain values, most of the strain values were less than 0.1, and comparing these strain values with stress values showed higher fluctuation than the stress values, so we got better accuracy in predicting stress than strain. This study provides a reasonable basis for

developing an objective diagnostic tool for eye injuries that incorporate stress and strain as retina damage markers.

While our PLSR model predicted the stress and strain with satisfactory accuracy, the major limitation of this work is the relatively small number of FE simulations used for the training. Additional FE simulations with different anatomical sizes, lesion properties, and impact orientations could enhance our training datasets. A complicated eye model considering the cornea, iris, lens, orbital fat (intraconal and extraconal), and the anterior segment will change the stress magnitude in the retina. It would be more practical to describe results in a comparative or qualitative trend instead of the absolute number. Still, we know that it could rely on patients' variations in terms of ages, disease conditions, anatomies, boundary conditions, and material properties. The skull was simulated as a rigid body. A deformable skull might lead to more minor retinal stresses. We are confident that integrating the FE and ML approach can rapidly predict posterior retina stress and strain patterns. Despite these simplifications and limitations, the present work demonstrated that integrating the FE and ML approach could fast predict posterior retina stress and strain patterns.

Chapter 4
Deep Learning-based Prediction of Stress and Strain Maps in Arterial Walls for Improved Cardiovascular Risk Assessment

Abstract

Conducting computational stress-strain analysis using finite element methods (FEM) is a common approach when dealing with the complex geometries of atherosclerosis, which is a leading cause of global mortality and complex cardiovascular disease. The considerable expense linked to FEM analysis encourages the substitution of FEM with a considerably faster data-driven machine learning (ML) approach. This study investigated the potential of end-to-end deep learning tools as a more effective substitute for FEM in predicting stress-strain fields within 2D cross sections of arterial wall. We first proposed a U-Net based fully convolutional neural network (CNN) to predict the von Mises stress and strain distribution based on the spatial arrangement of calcification within arterial wall cross-sections. Further, we developed a conditional generative adversarial network (cGAN) to enhance, particularly from the perceptual perspective, the prediction accuracy of stress and strain field maps for arterial walls with various calcification quantities and spatial configurations. On top of U-Net and cGAN, we also proposed their ensemble approaches, respectively, to further improve the prediction accuracy of field maps. Our dataset, consisting of input and output images, was generated by implementing boundary conditions and extracting stress-strain field maps. The trained U-Net models can accurately predict von Mises stress and strain fields, with structural similarity index scores (SSIM) of 0.854 and 0.830 and mean squared errors of 0.017 and 0.018 for stress and strain, respectively, on a reserved test set. Meanwhile, the cGAN models in a combination of ensemble and transfer learning techniques demonstrate high accuracy in predicting von Mises stress and strain fields, as evidenced by SSIM scores of 0.890 for stress and 0.803 for strain. Additionally, mean squared errors of 0.008 for stress and 0.017 for strain further support the model's performance on a designated

test set. Overall, this study developed a surrogate model for finite element analysis, which can accurately and efficiently predict stress-strain fields of arterial walls regardless of complex geometries and boundary conditions.

Keywords: Finite element methods (FEM); cardiovascular disease; convolutional neural network (CNN), U-Net; conditional generative adversarial neural network (cGAN); stress-strain field maps.

Introduction

Atherosclerosis is a complex cardiovascular disease characterized by plaque accumulation in the arterial walls, leading to the narrowing and hardening of the arteries. It is a major cause of heart disease, stroke, and other cardiovascular diseases, making it one of the leading causes of death worldwide. The disease is caused by genetic, environmental, and lifestyle factors, including high cholesterol, high blood pressure, smoking, obesity, and diabetes (Weber and Noels 2011; Libby, Ridker, and Maseri 2002). The mechanical properties of the plaque, such as its stiffness and strength, are important determinants of its stability and the risk of rupture. Therefore, predicting the stress-strain field maps of the plaque can provide valuable insights into its mechanical behavior and help identify regions that may be prone to rupture. Several studies have demonstrated the potential of computational models and imaging techniques for predicting the stress-strain field maps of atherosclerotic plaques and assessing their risk of rupture (Church and Miller 2016; Krams et al. 1997). Cheng et al. (J.M. Cheng et al. 2014) found that maximum von Mises stress was significantly higher in ruptured plaques compared to stable plaques. Therefore, predicting the von Mises stress could be an essential tool for preventing plaque rupture and reducing the incidence of cardiovascular events. The FEM, known as the finite element method, is the conventional numerical technique employed for stress-strain analysis of structures. It revolves around solving partial differential equations to evaluate the system's behavior (Bathe 2006; Reddy 2019; M.R. Lam et al. 2022a; Shokrollahi, Dong, et al. 2022a). However, FEM simulations can be expensive, especially for highly nonlinear analyses or for complex geometries. Consequently, substantial endeavors have been devoted to substituting FEM with machine learning (ML) techniques, commonly employed for surrogate modeling

(Cristianini and Shawe-Taylor 2000; Williams 1998; Bhaduri et al. 2018; Gholami, Ege, and Barzegar 2023; Shokrollahi, Dong, Gamage, et al. 2022; Shokrollahi, Dong, Kaya, et al. 2022; Mollaei Ardestani et al. 2023) of pertinent quantities of interest.

Madani et al.(Madani et al. 2019) presented a method for bridging FEM and ML for predicting the von Mises stress distribution in arterial walls affected by atherosclerosis. The authors utilized a FEM to simulate the mechanical behavior of arterial walls and generated a large dataset of stress values for varying degrees of plaque build-up. The dataset was then used to train an ML model that could predict stress values based on input parameters such as plaque thickness and the diameter of the arterial lumen. Liu et al.(Y.-C. Li et al. 2021) used a CNN model to automatically segment atherosclerotic plaques in intravascular ultrasound (IVUS) images. The authors used a dataset of IVUS images and achieved an accuracy of 90% in plaque segmentation. The segmentation results were then used to simulate the mechanical properties of the plaque using FEM. The study showed that the mechanical properties of the plaque were highly dependent on its morphology and composition. Chau et al.(Chau et al. 2004) performed Optical Coherence Tomography (OCT) imaging of atherosclerotic plaques in human cadaveric coronary arteries and used finite element analysis to assess the mechanical properties of the plaques. They found that the fibrous cap's thickness and the lipid core's size were important factors in determining the mechanical stability of the plaque. Plaques with thinner fibrous caps and larger lipid cores were found to have higher stress concentrations, which may contribute to plaque rupture. Cilla et al.(Cilla et al. 2012) presented a study on applying ML techniques to determine plaque vulnerability, an important factor in predicting the risk of heart attack and stroke. The authors discuss the limitations of traditional methods for assessing plaque vulnerability and the potential benefits of using ML techniques to analyze patient imaging data. The study involved analyzing patients with carotid artery disease, and the results show that ML techniques can effectively determine plaque vulnerability. The authors used conventional ML algorithms, such as support vector machines (SVM) and random forests (RF), to analyze imaging data and extract relevant features. They also developed a new "plaque volume score" method combining different features to predict plaque vulnerability better. The results show that ML algorithms can accurately predict plaque vulnerability, with an accuracy of up

to 90%. The authors suggest that these techniques can be used in clinical practice to improve risk assessment and treatment planning for patients with carotid artery disease. While all these papers discuss predicting stress distribution, none of them address the issue of how well their predictions would hold up in the face of variability in arterial wall geometries, such as changes in calcification angle, thickness, and number, or shifts in lumen location and fibrous thickness. Our innovative approach considers the variability in calcification numbers while predicting stress and strain in arterial walls.

This study aims to employ deep learning models for predicting stress and strain distribution in arterial walls under blood pressure loading, specifically considering various spatial calcification distributions. Encoder-decoder networks, particularly U-Net networks (Ronneberger, Fischer, and Brox 2015), have demonstrated effective image mapping due to their ability to capture high-resolution details and low-level features by propagating context information from lower to higher layers. Moreover, skip connections help prevent the vanishing gradient problem during model training. Conversely, cGAN, which learns a mapping from an observed image and random noise vector to an output image, is utilized for predicting stress-strain maps using a U-Net generator architecture and a PatchGAN discriminator, highlighting the superiority of cGAN in predicting strain maps of arterial walls compared to the U-Net model, particularly from the perceptual perspective (Ledig et al. 2017). This research implements deep learning models incorporating U-Net and cGAN, showcasing their generalization ability in predicting stress and strain maps of arterial walls. Moreover, we demonstrate the benefits of using ensemble and transfer learning strategies for improved accuracy. Impressively, our results display high accuracy even when trained on a mere 4,000 images, which we achieve through data augmentation grounded in the underlying physics of the issue. The structure of the manuscript is as follows: In Section 2, the proposed methodology is discussed. Section 3 showcases the prediction results and analysis, while Section 4 offers discussions and conclusions.

Materials and Methods

Generating Arterial Walls Model

Our training database comprises finite element simulations of different arterial geometries and boundary conditions. Our collaborators provided clinical data obtained through a two-step deep learning process for segmenting calcified coronary plaque in intravascular OCT images. This method involved combining a U-Net-based model with a GAN to identify calcified plaques in OCT images effectively. We utilized this segmentation data to construct a 3D model of a calcified artery based on OCT data. To streamline the process and generate various artery geometries rapidly, we first developed a 2D parametric model using Python scripting in Abaqus/Explicit software version 2019, drawing from the simplified clinical data and our observations. Python scripts in Abaqus can automate repetitive tasks, customize the software's behavior, and integrate it with other tools in a software pipeline. For example, a script could generate a series of similar simulations with varying parameters or post-process simulation results and generate customized visualizations. As shown in **Figure 4-1**, we simplified the assumption that the cross-section consists of a series of circles and splines. The specifications of the circles, including a lumen, fibrous, artery, calcium deposits, and the number of calcium, were randomly varied according to Table 1. To produce calcification in a free-form shape, we began by selecting two circles as the inner and outer boundaries of the calcification, each with a random radius. We then randomly chose a calcification angle between 0 and 180 degrees. Next, we generated 10 points with equal intervals within this angle on the calcification's inner and outer boundaries. Finally, we utilized spline tools in Python to connect all these points and form a closed surface as the calcification. The number of calcifications may also be either one or two.

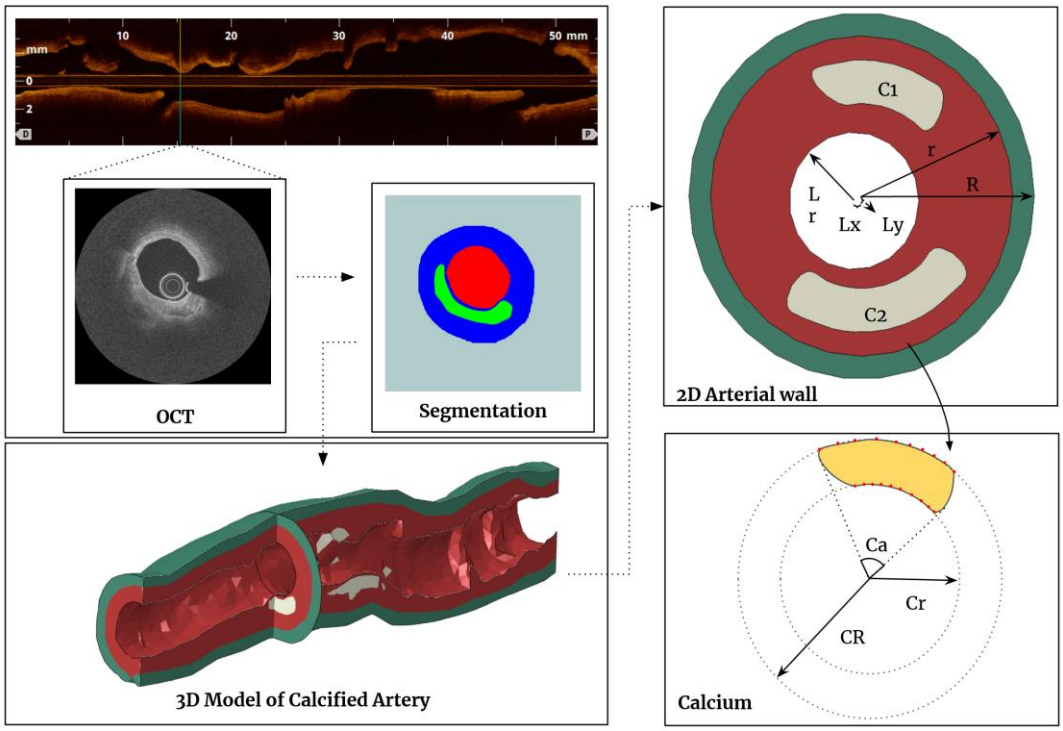


Figure 4-1. Using segmentation data, a 3D model of a calcified artery was constructed based on OCT data. To quickly generate various artery geometries, a 2D parametric model was developed using Python scripting in Abaqus/Explicit software. A simplified assumption of cross-sections consisting of circles and splines was made. Circle specifications, such as lumen, fibrous, artery, and calcium deposits, were randomly varied. Calcification was created by selecting two random circles as inner and outer boundaries, with a random angle between 0 and 180 degrees. Ten points were generated on inner and outer boundaries, and spline tools in Python were used to form a closed surface for calcification.

Table 4-1: Features for random generation of idealized 2D artery geometry.

Features	Symbol	Range
Artery Outer Radius	R	2 mm
Artery Inner Radius	r	1.75 mm
Lumen Radius	Lr	0.75 mm
Lumen Dislocation X	Lx	-0.25-0.25 mm
Lumen Dislocation Y	Ly	-0.25-0.25 mm
Number of Calcification	Cn	1-2
Calcification Inner radius	Cr	1-1.25 mm
Calcification Outer radius	CR	1.3-1.5 mm
Calcification Angle	Ca	0-180°

Finite Element Method Simulation Database

This study considers a two-dimensional plane strain cross-section of an arterial wall, depicted in **Figure 4-2**, where the material is assumed to be hyperelastic (Table 2). The FE mesh used in the simulation is shown in **Figure 4-2**, and a four-node constant strain element is discretized. To simulate the effect of blood pressure, a static pressure load of 140 mmHg (18.7 kPa)(Cilla et al. 2012) is applied to the boundary elements in the lumen, while nodes on the outside boundaries are restrained from displacement. The problem is solved using Abaqus software (Abaqus 2011), and a typical converged stress and strain distribution is shown in **Figure 4-2**.

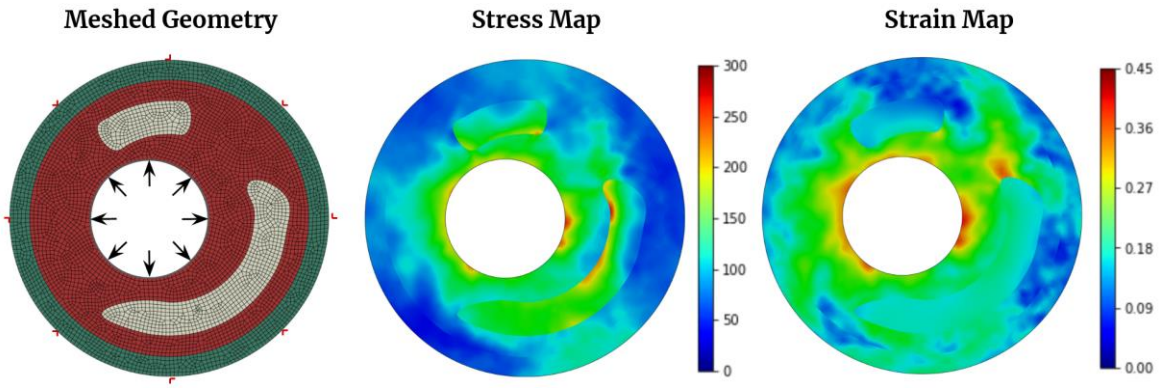


Figure 4-2. 2D arterial wall cross-section geometry meshed with four-node constant strain element and corresponding von Mises stress and strain distribution. The maximum stress and strain are considered 300 kPa and 0.45, respectively.

Table 4-2: Material properties of arterial wall features (Dong et al. 2021).

	C_{10}(MPa)	C_{01}(MPa)	C_{11}(MPa)	C_{20}(MPa)	C_{02}(MPa)	C_{30}(MPa)	C_{03}(MPa)
Artery	0.108	-0.101	-0.179	0.088	0.062		
Fibrous	0.040				0.003		0.0297
Calcium	-0.495	0.506	1.193	3.637		4.737	

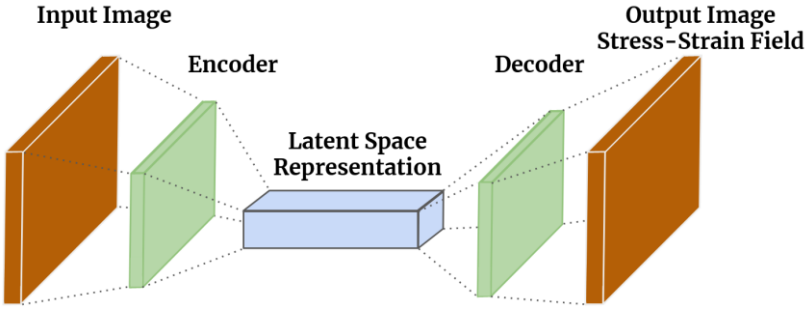
Models and Methods

U-Net Architecture

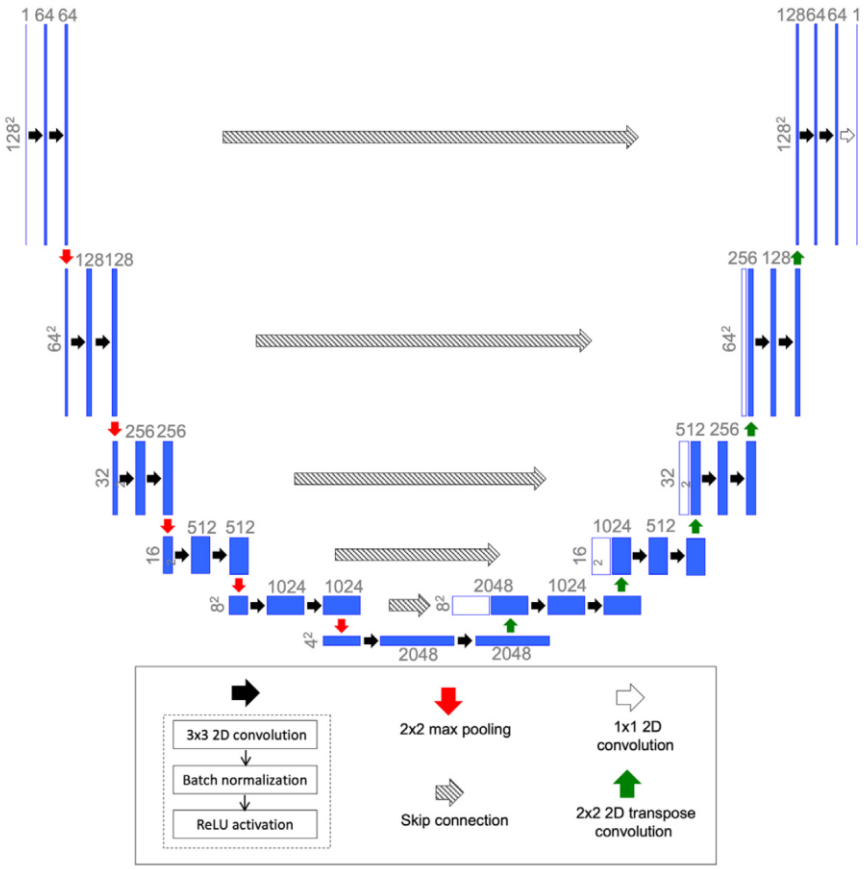
By executing the FEM model on a range of arterial wall geometries, one can acquire a collection of input and output image data. These data can then be utilized to train a deep learning model in a supervised learning manner. An encoder-decoder network maps the input image to the output stress and strain map (Nie, Jiang, and Kara 2020), depicted in **Figure 4-3**(a). The binary maps provided as input depict the positions of calcium, lumen, and fibrous tissue. The encoder-decoder network converts these input images into a latent space with lower dimensions, which is subsequently mapped back to the stress and strain field. This approach assumes that the input and target spaces share a common latent space. The U-Net architecture (Ronneberger, Fischer, and Brox 2015), based on the FEM mapping data of images to stress and strain field, has been utilized for this purpose. The weights of architecture are trained through learning. Initially introduced in 2015 for medical image segmentation, the U-Net architecture is named after its U-shape, consisting of an encoder and a decoder. It has proven effective in capturing latent representations for various types of images. The encoder extracts feature from the input image using convolutional and max pooling layers.

In contrast, the decoder generates the output segmentation map using up-convolutional and concatenation layers. One of the significant advantages of the U-Net architecture is its ability to produce high-quality segmentation results even when trained on a limited amount of data. U-Net has been widely used in various medical image segmentation tasks, such as segmenting brain tumors, retinal blood vessels, and cell membranes. The architecture has also been extended and modified for other applications, such as semantic segmentation, instance segmentation, and object detection. The standard architecture in this study incorporates contracting and expanding layers along with skip connections to propagate context information and improve output resolution. The U-Net architecture utilized here is a slightly modified version of the original design. The encoder section consists of six repeating blocks, each containing a 2×2 max pooling operation, two consecutive 3×3 2D convolutions (except for the first and last blocks), batch normalization,

and ReLU activation. The first decoder block employs a transpose convolution layer, while the last block consists of two consecutive convolution-batch norm-ReLU layers without a transpose convolution. The encoder and decoder blocks are followed by a final 1×1 convolutional layer, which reduces the 64-channel decoder output to a single channel. During weight training, the loss function is the weighted mean squared error (MSE) between the predicted and true von Mises stress and strain maps derived from the training data. The U-Net architecture was implemented using the TensorFlow and Keras libraries (Abadi et al. 2016a). **Figure 4-3**(b) illustrates a diagram of the modified U-Net architecture.



(a)



(b)

Figure 4-3. (a) An encoder–decoder based network that can serve as an efficient surrogate for the FEM mapping and (b) the U-Net architecture(Ronneberger, Fischer, and Brox 2015).

The training loss function is defined using the MSE, which is denoted as:

$$MSE = L[Y, f(X)] = \frac{1}{n} \sum_{i=1}^{n} [Y - f(X)]^2 \quad (1)$$

Where X denotes the input image describing the 2D arterial wall images, n denotes the batch size, $f(X)$ indicates the prediction of the U-Net model, and Y denotes the output

images. Performing verification to prevent overfitting during every training epoch also had been done in our network to select the best model without overfitting.

cGAN Architecture

The Generative Adversarial Network (GAN) is a deep neural network type that generates new data based on the data statistics of the training set (Creswell et al. 2018). GANs have two essential components, the generator, and the discriminator, which are trained against each other using game theory. The generator produces candidates that the discriminator evaluates. Though GANs were originally developed for unsupervised learning, incorporating labels as constraints can result in a subcategory known as conditional GANs (cGANs), as depicted in **Figure 4-4**. Our research focuses on developing a deep learning model that utilizes a conditional generative adversarial network (cGAN) and paired image data (Isola et al. 2017). The cGAN model comprises two crucial components: the generator, known as U-Net, and the discriminator, referred to as PatchGAN (Isola et al. 2017). The role of the generator is to take geometric images (labels or constraints) as input and produce field images of interest by incorporating random noise. Subsequently, the discriminator compares these generated field images to authentic images obtained from finite element modeling (FEM). The generator's objective is to increase the error rate of the discriminator.

In contrast, the discriminator aims to optimize its ability to distinguish between fake images generated by the generator and real ones. To accomplish this, we utilize TensorFlow (Abadi et al. 2016a), a versatile machine learning framework, to perform calculations and implement the generative adversarial network (GAN) architecture for translating arterial wall geometries into strain fields (Isola et al. 2017). Our specific model employs U-Net as the generator and PatchGAN as the discriminator. U-Net generates synthetic strain field images based on arterial wall geometries, and PatchGAN evaluates the authenticity of these generated field images by comparing them to real strain field images. Our generator U-Net shares similarities with the CNN models discussed earlier, and the generator loss function is defined as:

$$Gen_{loss} = \lambda * gan_{loss} + L_1\_loss \quad (2)$$

The variable gan_loss represents the sigmoid cross-entropy loss between the generated images and an array of ones. At the same time, L_1_loss calculates the mean absolute error between the generated and target images. It should be noted that (2) is equivalent to (1) when λ is set to 0 and L_1_loss is replaced by L_2_loss, which calculates the mean squared error. We tried several hyperparameters, such as λ with different values ranging from 0.005 to 0.05. We saved the best models' weights to be imported to the new training model. We experimented with different architectures and numbers of layers with different filter sizes for the discriminator architecture and presented our final architecture.

The discriminator PatchGAN comprises five layers, which have around 11,000,000 trainable weights. The model receives two inputs and concatenates them using the concatenate layer. Afterward, the concatenated input is processed through five Conv2D layers with different filter sizes (64, 128, 256, 512, and 1024), a stride value of 2, and the same padding scheme. An activation function called LeakyReLU (He et al. 2015) is applied to the output tensors with an alpha value of 0.25. Finally, the last convolutional layer's output is sent to another Conv2D layer with a filter size of 1, a stride value of 1, and the same padding scheme, which generates a single output value. The l_1_regularizer is utilized as a kernel regularizer for each Conv2D layer, with a value of 1e-4. PatchGAN examines the generated field images by categorizing individual patches in the image as either real or fake. The Discriminator loss function is defined as

$$Dis\_loss = real\_loss + generated\_loss \tag{3}$$

The variable real_loss represents the loss calculated using the sigmoid cross-entropy function (Z. Yang, Yu, and Buehler 2021) between real images and an array of ones. On the other hand, generated_loss represents the loss calculated using the sigmoid cross-entropy function between the generated images and an array of zeros. Our model is trained for 500 epochs, using a batch size of 4, to achieve convergence despite the common training instability observed in GANs. However, conditional GANs (cGANs) are more stable when the input labels are properly constrained. We carefully selected the number of training epochs to ensure that the generator learns sufficient information without the discriminator consistently classifying the generated images as fake (Goodfellow 2016). To determine the

appropriate number of training epochs, we assessed the predictions of our ML model on the test set.

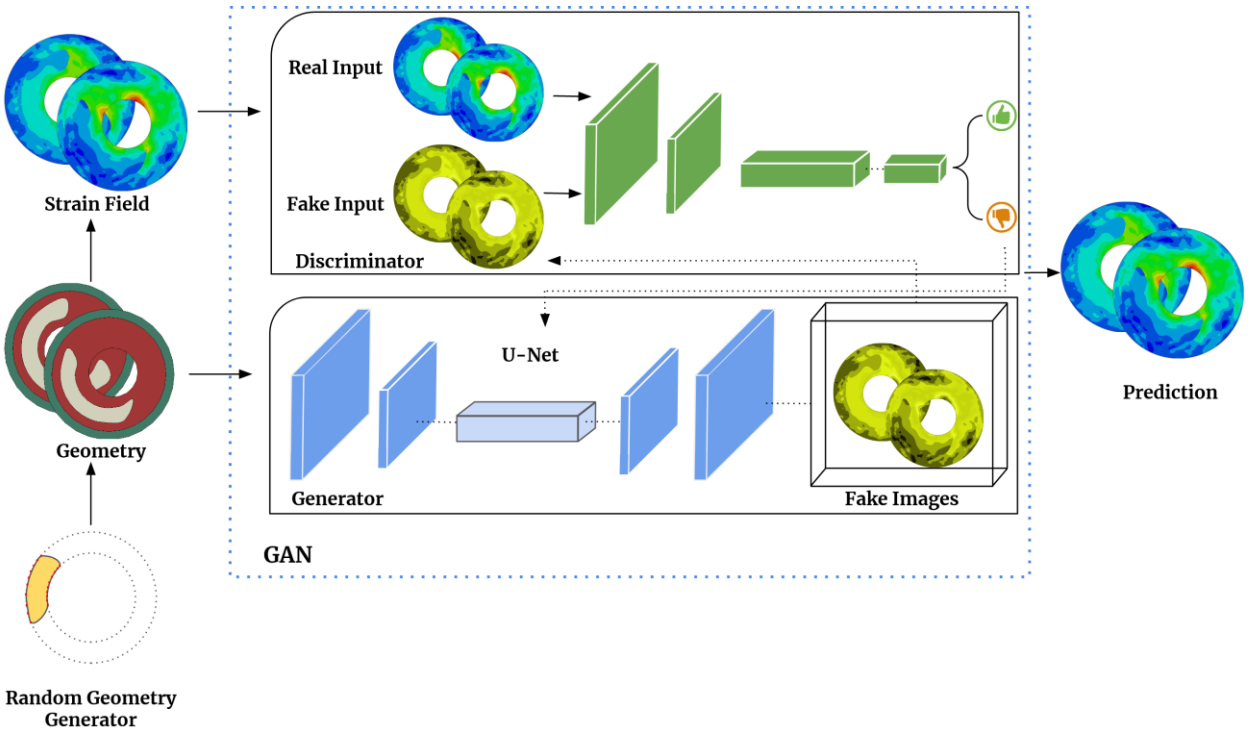


Figure 4-4. Workflow for the proposed method begins with a random generator using Python scripting to produce 2D arterial wall geometry images (256x256). Next, a FEM analysis is conducted to obtain arterial walls' true strain field information under blood pressure. ML model called Generative Adversarial Network (GAN), which includes a generator U-Net and a discriminator PatchGAN, is trained to predict strain fields from geometry images. The generator generates strain field maps using the geometry images as input. The discriminator then compares these generated images with real FEM-derived images. A well-trained model can accurately predict strain field maps, validated against high-fidelity FEM models and arterial walls with new geometries.

Ensemble learning

Combining multiple models in machine learning, known as ensemble learning (Dietterich 2000), can significantly enhance predictive accuracy and robustness. This technique is especially effective when applied to U-Net and cGAN architectures. One

method for implementing ensemble learning with U-Net/cGAN involves training various instances of the same architecture using different hyperparameters or random seeds. By averaging the predictions of these models, errors from any individual model can be minimized, leading to improved overall accuracy. Another approach explored in this paper involves combining different U-Net and cGAN generator architectures in an ensemble. An instance of an ensemble could comprise a conventional U-Net paired with a customized variation featuring supplementary skip connections. Alternatively, another ensemble may increase the number of filters in the convolutional layers of a standard U-Net and cGAN generator architectures by 32 channels. While this approach may improve feature extraction and result in more resilient predictions, it would also augment the model's parameter count and necessitate additional computational resources during training and inference. By capitalizing on the unique strengths of diverse U-Net and cGAN architectures through ensemble learning, it is frequently feasible to attain superior outcomes compared to solitary architecture.

This study employed two ensemble strategies for both U-Net and cGAN. The first strategy involved using the same U-Net architecture but with two different dropout rates in the layers - 0.2 and 0.35. After training the models with these rates, we averaged their results. This method was similarly applied to the cGAN generator architecture. The second strategy involved utilizing two different U-Net architectures, one with 10 layers and the other with 12 layers, as previously mentioned. We added a 32-channel layer at both the beginning and end of the network. Each model was trained with three dropout rates - 0.2, 0.3, and 0.35. In total, we trained 6 models using this approach. This same methodology was also applied to the cGAN architecture.

Deep transfer learning

Using transfer learning (Weiss, Khoshgoftaar, and Wang 2016), the previously learned weights of a deep learning model can be used as a starting point to retrain the same model on a new dataset, leading to faster convergence. To address the arterial walls problem, a combination of U-Net and cGAN architectures was utilized. The training process involved using 2D cross-section images along with their corresponding von Mises stress maps. This

approach aimed to minimize the training requirements for a separate network that predicts strain maps. Because generating different output maps is costly, transfer learning was employed as a more efficient approach to predict strain maps. The von Mises stress trained model is utilized to predict the strain map.

Image quality metrics and statistical analysis

We used SSIM (Hore and Ziou 2010) to measure the accuracy of our deep-learning models. In computer vision tasks, like image classification or object detection, the model's accuracy is often evaluated using metrics like precision, recall, and F1 score. However, these metrics do not provide information about the visual quality of the output images or the degree to which the model's predictions match the ground truth. SSIM can help fill this gap by measuring the structural similarity between the output and ground truth images. The output images from CNN can be compared to the ground truth images using SSIM, and the resulting value can be used to evaluate the model's accuracy. The range of SSIM is between 0 and 1, where a value of 1 indicates perfect similarity between two images, while a value of 0 indicates no similarity. The measure between two images x and y of common size N × N is:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (4)$$

Where x and y are the two images being compared, μ_x and μ_y are the mean pixel intensities of images x and y, σ_x^2 and σ_y^2 are the variances of pixel intensities in images x and y, σ_{xy} is the covariance of pixel intensities between x and y, and c_1 and c_2 are constants used to stabilize the division, typically $c_1 = (k_1 L)^2$ and $c_1 = (k_2 L)^2$, where L is the dynamic range of the pixel values (for example, 255 for 8-bit grayscale images) and k_1 and k_2 are constants that control the relative contribution of the luminance and contrast terms to the final SSIM score. The SSIM formula considers the compared images' luminance, contrast, and structure and produces a value between 0 and 1.

Results

Stress map prediction accuracy

This section assesses the precision of matching 2D cross-section images of arterial walls with their corresponding von Mises stress map. To enhance the dataset, we leverage the

physics of the problem by applying horizontal and/or vertical flipping to both the input images and output maps. This four-fold data augmentation is accomplished through three flipping operations: horizontal flip, vertical flip, and a combination of horizontal and vertical flip, as demonstrated in **Figure 4-5**.

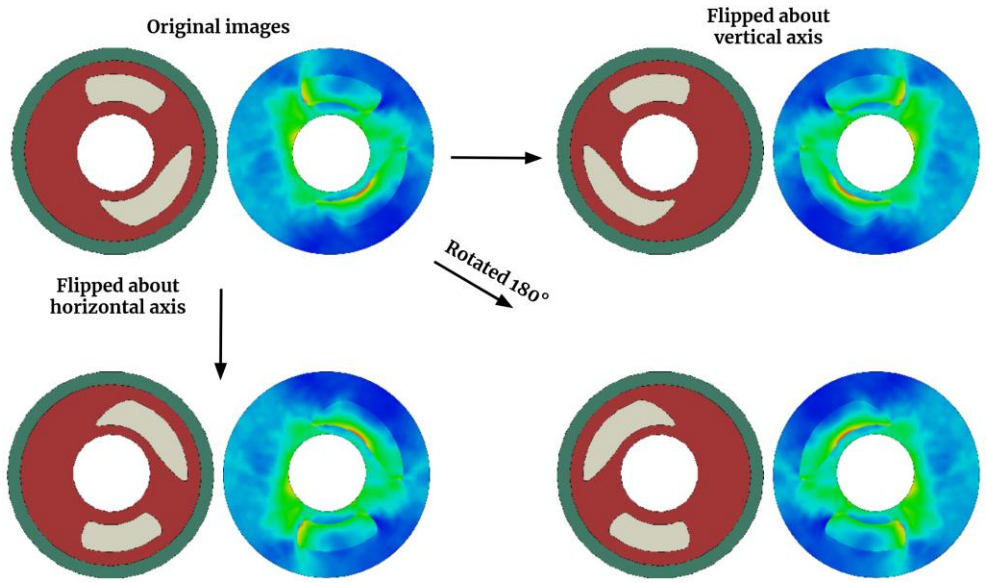
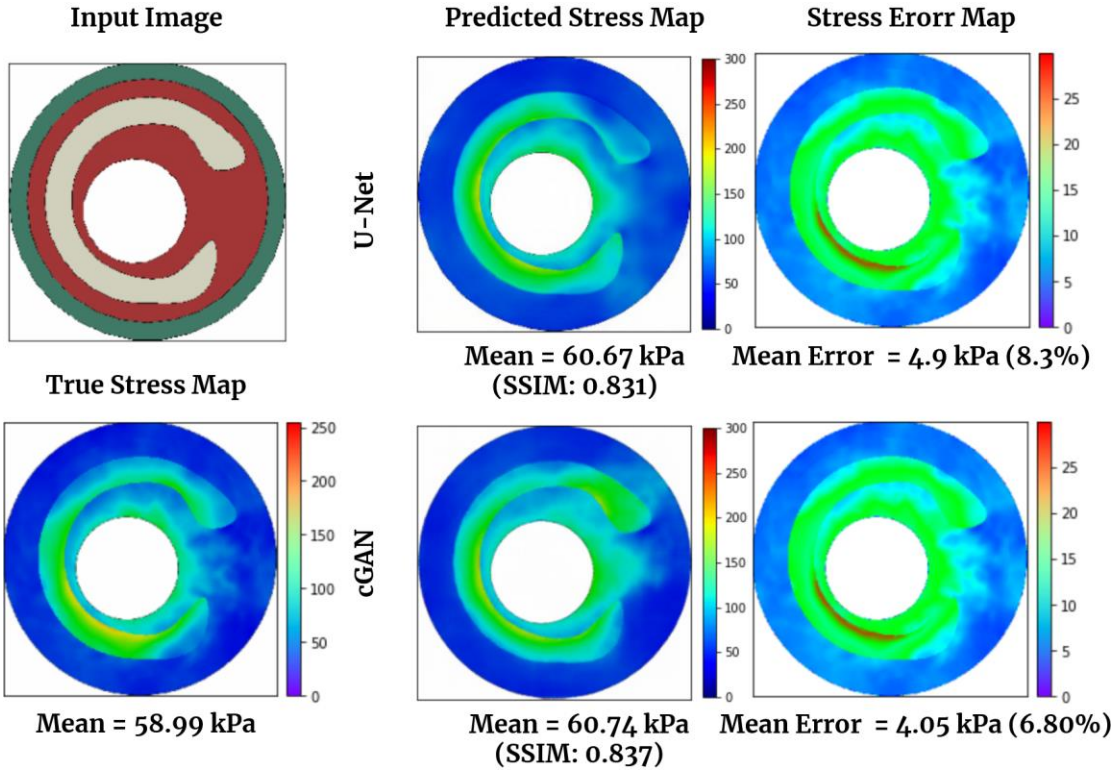


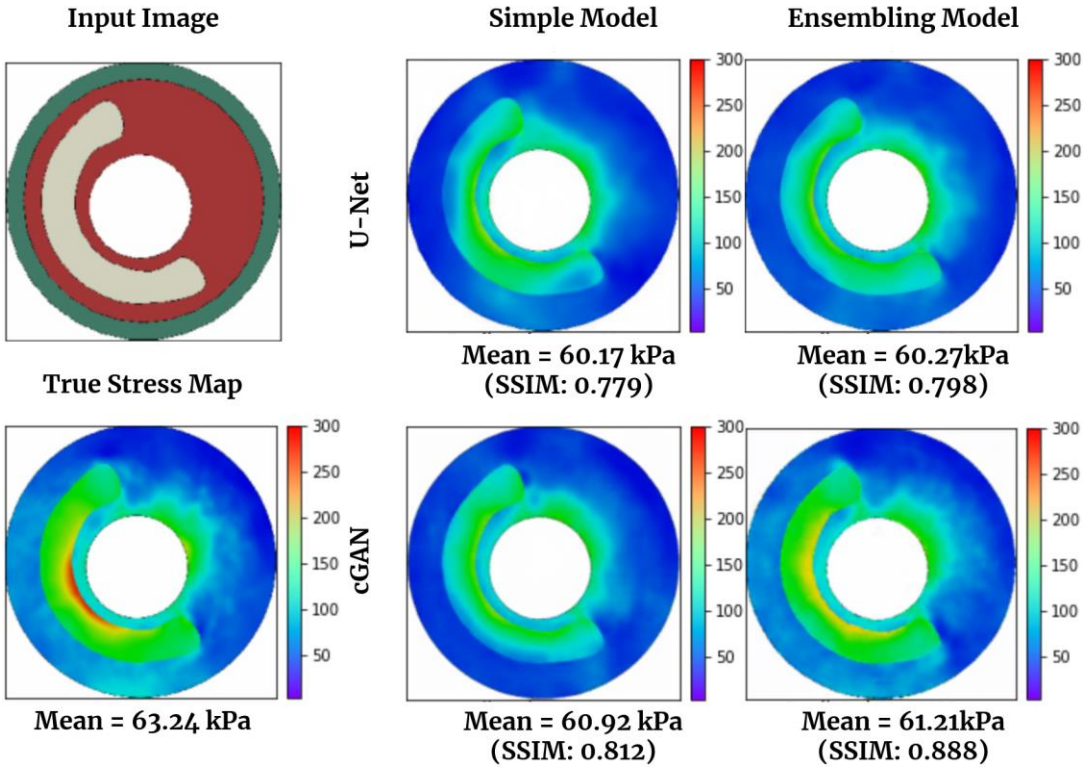
Figure 4-5. 4-fold data augmentation by image flipping.

After data augmentation, the training data consists of 4000 images of various spatial arrangements of calcium in the arterial wall and their corresponding von Mises stress maps for each model case (train: test 85%-15%). Fig. 6(a) and 6(b) displayed the actual (simulated via FEM) and predicted von Mises stress maps using U-Net and cGAN. In **Figure 4-6**(a), a test data image is presented, followed by a comparison between the real stress map, calculated using FEM, and those generated by U-Net and cGAN. CGAN provides a superior prediction of von Mises stress compared to U-Net, achieving an SSIM of 0.837 and a mean error percentage of 6.80, whereas U-Net attains an SSIM of 0.831 and a mean error percentage of 8.3. **Figure 4-6**(b) compares the stress map predictions made by the basic U-Net model and cGAN and the best U-Net and cGAN models obtained through ensemble techniques. The results for all models, including those using ensemble methods, are presented in Table 3. According to the table, the best ensemble model for U-Net was

achieved by combining different U-Net architectures, as previously described. In this study, we utilized one U-Net model with 8 layers and another with 10 layers (by adding a layer with 32 channels at the beginning and end of the networks). The ensemble of these networks yielded the best results for U-Net, with a mean SSIM of 0.854. For cGAN, we applied the same technique to the generator layers and then combined it with a basic cGAN. As shown, the ensemble of two architectures also resulted in the best performance for cGAN (mean SSIM of 0.890). We then selected the best models and compared their predictions to those of the basic U-Net model and cGAN in **Figure 4-6**(b). The ensemble technique for cGAN yielded superior predictions compared to the other models in this specific case, with an SSIM of 0.888.



(a)



(b)

Figure 4-6. Von Mises stress map predicted from U-Net and cGAN architecture is based on 1000 FEM analyses of 2D arterial wall images, augmented to 4000 training images, (a) Test data image with true stress map (FEM) compared to U-Net and cGAN-generated maps. cGAN outperforms U-Net with SSIM 0.837 and 6.80% mean error. In comparison, U-Net has SSIM 0.831 and 8.3% mean error, (b) Comparison of stress map predictions between the best ensemble models and the basic U-Net and cGAN models, highlighting the superior performance of the ensemble cGAN (SSIM = 0.888).

Table 4-3: Different approaches for predicting von Mises stress maps on the testing data.

	#of Models	Metrics	Min	Max	Mean
U-Net	1	MSE	0.008	0.039	0.019
		SSIM	0.773	0.902	0.842
cGAN	1	MSE	0.008	0.036	0.017
		SSIM	0.787	0.895	0.850
	2	MSE	0.008	0.038	0.018

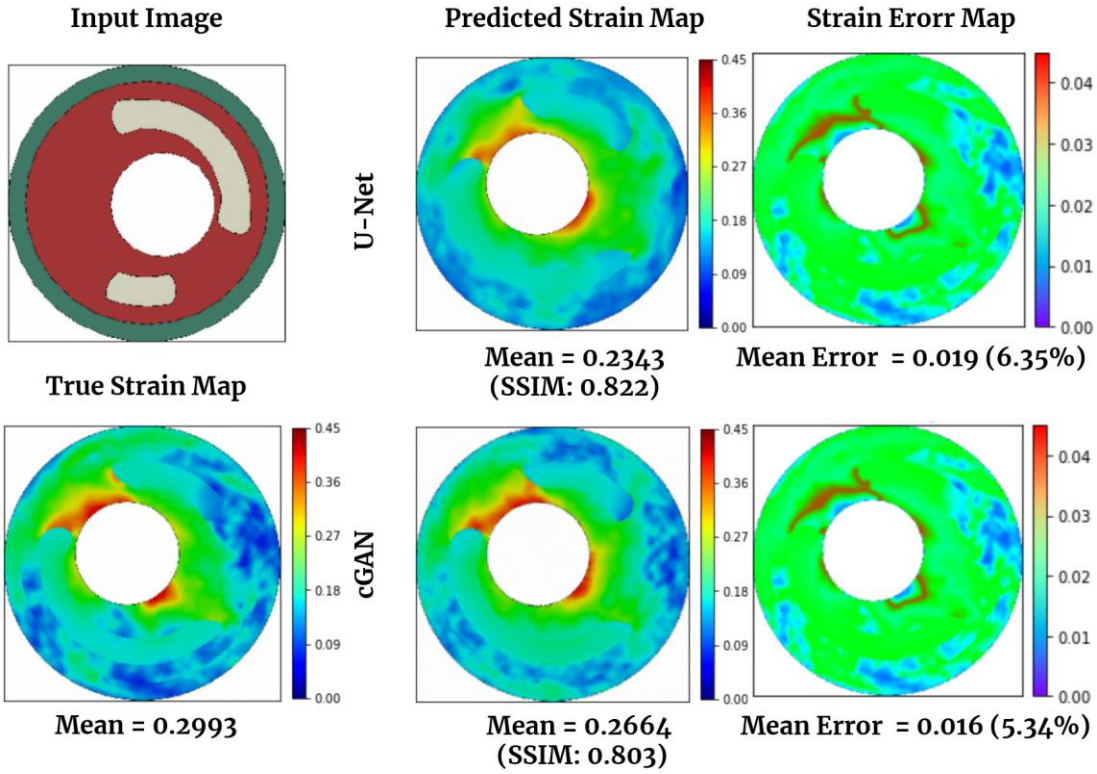
Ensembling (U-Net)	2×3	SSIM	0.771	0.903	0.845
		MSE	0.007	0.038	0.017
		SSIM	0.778	0.906	0.854
Ensembling (cGAN)	2	MSE	0.005	0.023	0.010
		SSIM	0.821	0.909	0.884
	2×3	MSE	0.003	0.024	**0.008**
		SSIM	0.823	0.924	**0.890**

Notably, the deep learning approach significantly reduces computational effort compared to FE results. While Abaqus requires approximately 2 minutes for a single analysis, our trained U-Net model predicts stress on a laptop CPU in less than a second. Our deep learning model has been trained on a GPU, specifically the NVIDIA GEFORCE RTX3090, within a Jupyter notebook environment. The utilization of this GPU significantly enhances computational speed compared to relying solely on a CPU. When running on GPU, each analysis takes approximately 0.06 seconds to complete.

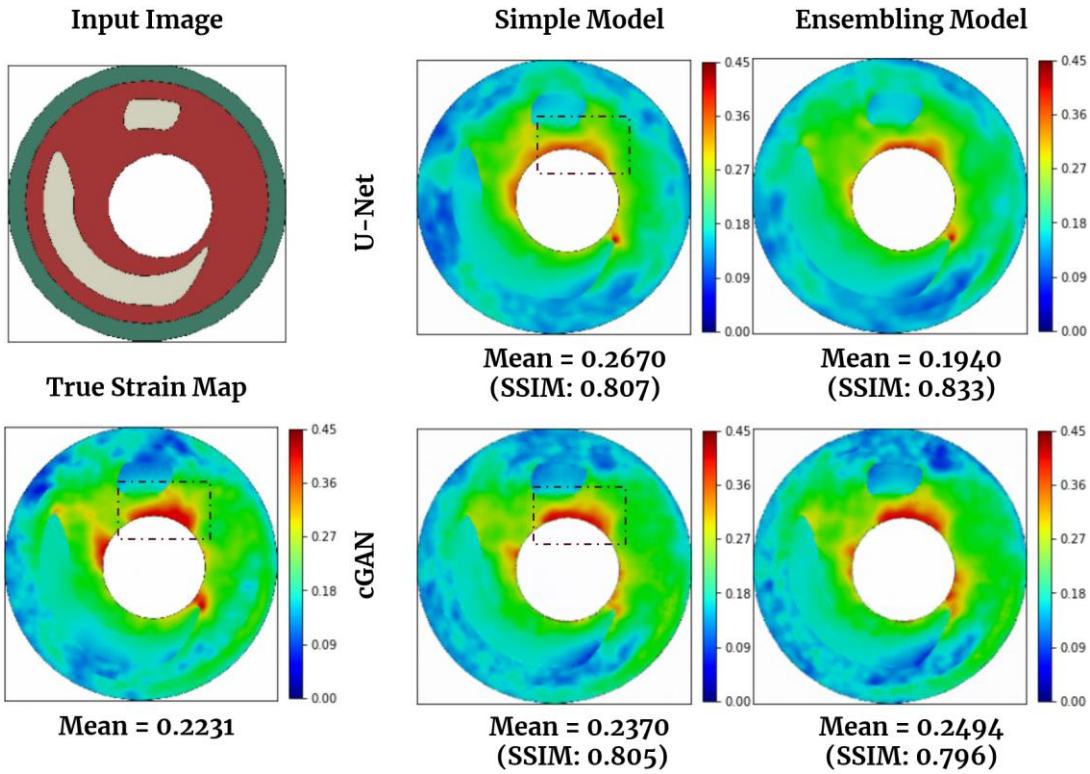
Strain map prediction accuracy

This section evaluates the accuracy of mapping 2D cross-sectional images of arterial walls to their corresponding strain maps. **Figure 4-7**(a) and **Figure 4-7**(b) showcased the actual (simulated via FEM) and predicted strain maps using U-Net and cGAN. **Figure 4-7**(a) displays a test data image, then compares the actual strain map (calculated using FEM) and the ones generated by U-Net and cGAN. U-Net outperforms cGAN in strain prediction, achieving an SSIM of 0.822 and a mean error percentage of 6.35, while cGAN achieves an SSIM of 0.803 and a mean error percentage of 5.34. **Figure 4-7**(b) compares strain map predictions made by the basic U-Net model and cGAN and the best U-Net and cGAN models obtained through ensemble techniques. The results for all models, including those utilizing ensemble methods, are presented in Table 4. The best ensemble model for U-Net is achieved by combining different U-Net architectures, as previously described. We employed one U-Net model with 8 layers and another with 10 layers (adding a layer with 32 channels at the beginning and end of the networks). This ensemble yields the best results for U-Net, with a mean SSIM of 0.830. For cGAN, we applied the same technique to the generator layers and then combined it with a basic cGAN. This ensemble of two architectures also results in the best cGAN performance (mean SSIM of 0.803). We then chose the best models and

compared their predictions to those of the basic U-Net model and cGAN in **Figure 4-7**(b). The ensemble technique for U-Net produces superior predictions in this case, with an SSIM of 0.833. Based on the numerical results showing here and more we observed, though cGAN and its ensemble version did not deliver the best numbers in terms of SSIM, they do have the best performance visually.



(a)



(b)

Figure 4-7. Strain map predicted from U-Net and cGAN architecture is based on 1000 FEM analyses of 2D arterial wall images, augmented to 4000 training images, (a) Test data image with true strain map (FEM) compared to U-Net and cGAN-generated maps. cGAN outperforms U-Net with SSIM 0.803 and 5.34% mean error. In comparison, U-Net has SSIM 0.822 and 6.35% mean error, (b) Comparison of strain map predictions between the best ensemble models and the basic U-Net and cGAN models, highlighting the superior performance of the ensemble U-Net (SSIM = 0.833), also, cGAN demonstrated superior performance in detecting higher strain values compared to U-Net, as highlighted by the black dashed box.

Table 4-4: Different approaches for predicting strain maps on the testing data.

	#of Models	Metrics	Min	Max	Mean
U-Net	1	MSE	0.0120	0.037	0.022
		SSIM	0.750	0.866	0.814
cGAN	1	MSE	0.010	0.029	0.019
		SSIM	0.751	0.842	0.793
	2	MSE	0.009	0.034	0.019

Ensembling (U-Net)	2×3	SSIM	0.800	0.877	0.825
		MSE	0.009	0.025	0.018
		SSIM	0.8153	0.877	**0.830**
Ensembling (cGAN)	2	MSE	0.011	0.026	0.017
		SSIM	0.785	0.839	0.800
	2×3	MSE	0.011	0.021	**0.017**
		SSIM	0.790	0.839	0.803
Transfer Learning (U-Net)		MSE	0.012	0.029	0.020
		SSIM	0.790	0.868	0.820
Transfer Learning (cGAN)		MSE	0.010	0.029	0.019
		SSIM	0.760	0.842	0.794

In Table 4, we display the SSIM and MSE values for the U-Net and cGAN models, which were trained using pre-trained weights from U-Net and cGAN models based on von Mises stress data. It was observed that both models led to an increase in SSIM and a decrease in MSE. Additionally, **Figure 4-8** compares the average SSIM and MSE values for stress and strain between the actual and predicted stress-strain maps across all 600 validation images, substantiating the high precision of our model's predictive capabilities. In **Figure 4-8**(a), we compared the MSE of predicted von Mises stress maps among six networks: a simple U-Net, an ensemble method with various hyperparameters, an ensemble using two distinct U-Net architectures, a simple cGAN, an ensemble method with various hyperparameters in the cGAN generator model, an ensemble using two distinct cGAN generator architectures. The box plots reveal that employing an ensemble with two cGAN architectures results in a lower MSE than other models. Moreover, **Figure 4-8**(b) indicates that the SSIM of an ensemble cGAN with two generator architectures is superior. **Figure 4-8**(c) examines the MSE of eight strategies for predicting strain maps. A comparison of the U-Net and cGAN with a basic network, ensembles, and transfer learning reveals that an ensemble method with various hyperparameters in the cGAN generator model exhibits a narrower MSE range than other approaches. Concerning the SSIM of predicted strain maps, **Figure 4-8**(d) illustrates that an ensemble using two distinct U-Net architectures outperforms other methods.

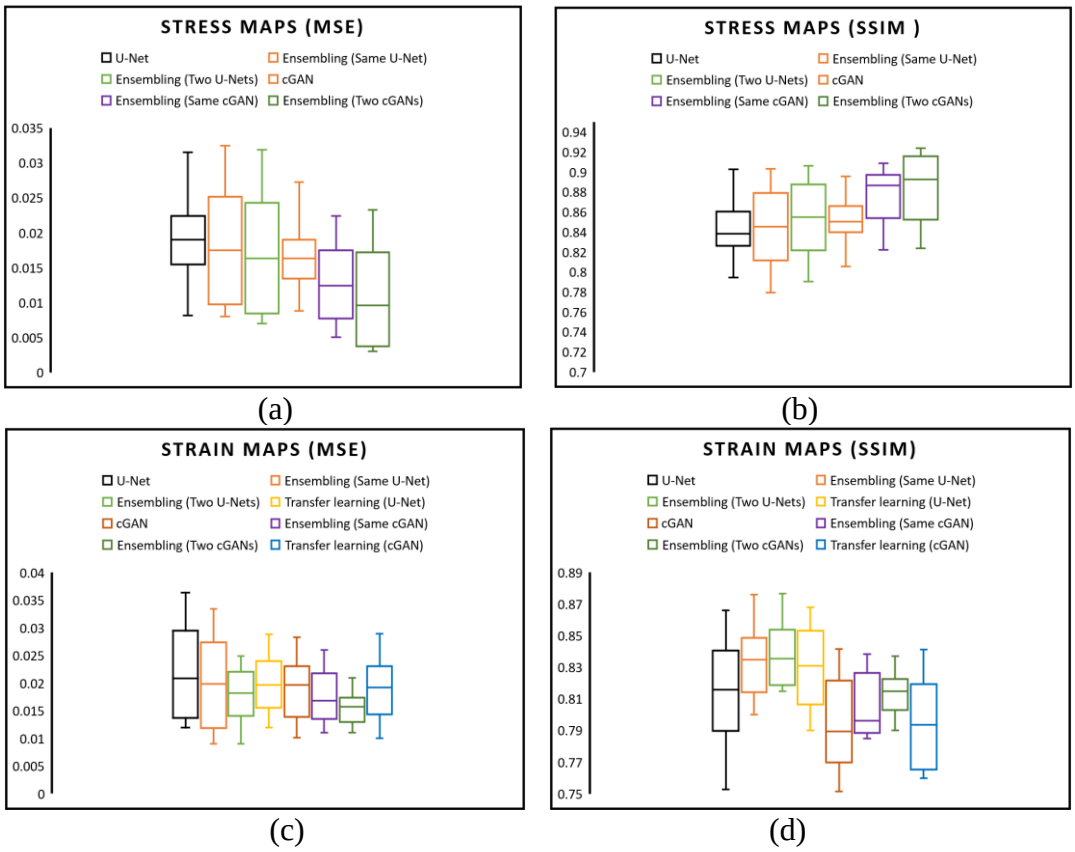


Figure 4-8. Comparison of MSE and SSIM for all 600 validation images, (a) MSE of all test data for prediction Stress maps using U-Net, Ensembling with same U-Net and different hyper parameters, and Ensembling with two U-Net architectures, cGAN, Ensembling with same cGAN generator model and different hyper parameters, and Ensembling with two cGAN generator architectures, (b) SSIM of all test data for prediction Stress maps using U-Net, Ensembling with same U-Net and different hyper parameters, and Ensembling with two U-Net architectures, cGAN, Ensembling with same cGAN generator model and different hyper parameters, and Ensembling with two cGAN generator architectures, (c) MSE of all test data for prediction Strain maps using U-Net, Ensembling with same U-Net and different hyper parameters, Ensembling with two U-Net architectures, U-Net using transfer learning, cGAN, Ensembling with same cGAN generator and different hyper parameters, Ensembling with two cGAN generator architectures, and cGAN using transfer learning (d) SSIM of all test data for prediction Strain maps using U-Net, Ensembling with same U-Net and different hyper parameters, Ensembling with two U-Net architectures, U-Net using transfer learning, cGAN, Ensembling with same cGAN generator and different hyper parameters,

Ensembling with two cGAN generator architectures, and cGAN using transfer learning.

Discussion and Conclusion

In summary, this research aims to explore the potential of deep learning tools in predicting stress-strain fields within 2D cross sections of arterial wall geometries, which can replace the traditional FEM method. To establish a relationship between the spatial distribution of calcification in arterial wall cross-sections and the von Mises stress and strain fields, two ML networks were investigated. The first network utilized a U-Net architecture within a CNN model, while the second employed a cGAN. The trained models can accurately predict stress-strain field maps for arterial walls with varying quantities and spatial configurations of calcification, which is a significant achievement. To generate the training dataset, a Python code was utilized to create various shapes of arterial walls with randomized features such as location, angle, thickness, number of calcifications, lumen area location, and diameters, drawing from the simplified clinical data and our observations. The dataset consists of input and output images obtained by applying boundary conditions and extracting stress-strain field maps. The trained U-Net models produce precise predictions of the von Mises stress and strain fields, with SSIM of 0.854 and 0.830 and MSE of 0.017 and 0.018 for stress and strain, respectively, on a held-out test set. Additionally, the stress and strain data-trained cGAN model outperforms the CNN model in prediction accuracy, demonstrating SSIM of 0.890 and 0.803 and MSE of 0.008 and 0.017 for stress and strain, respectively. However, the utilized image quality metrics have limitations. They cannot encompass all the image perception details significant for human visual systems. We expect the performance of cGAN can be evaluated more accurately and fairly if a comprehensive metric could be designed in future.

This research demonstrates the ability of a surrogate model for finite element analysis to accurately predict stress-strain fields of arterial walls, utilizing 2D cross-sectional images. Furthermore, we showcased the numerous benefits of utilizing ensemble and transfer learning methodologies to achieve superior levels of precision. By combining multiple models, an ensemble approach can significantly enhance predictive performance,

as the diverse models can compensate for each other's weaknesses and amplify their strengths. On the other hand, transfer learning can leverage the knowledge gained from a pre-trained model and apply it to a new task with a limited amount of data. This approach can drastically reduce the required training and improve the model's performance on the new task. Our findings indicate that by ensembling the same U-Net with various hyperparameters, the mean SSIM across all testing data can be boosted from 0.814 to 0.825. Moreover, by incorporating an additional layer into the standard U-Net and training two models thrice, we achieved the highest SSIM of 0.830 for predicting the strain map. Transfer learning was utilized in addition to the von Mises stress trained model to forecast strain maps. By comparing the loss of a U-Net model trained from scratch to that of a pretrained U-Net model trained with von Mises stress data over 100 epochs, it was noted that both models displayed convergence and exhibited similar behavior during training. Despite this, the transfer learning model demonstrated quicker convergence and produced an average SSIM of 0.820 for all test data, outperforming a U-Net trained from scratch. We employed a cGAN for stress and strain map predictions and compared the outcomes with those derived from a U-Net. The optimal stress map predictions were attained by assembling two generator architectures in the cGAN, achieving an SSIM of 0.890 (Table 3). Despite this, U-Net demonstrated superior performance in strain prediction, obtaining the highest SSIM of 0.830 when combined with two U-Net architectures. A comparison of the SSIM and MSE for cGANs trained with transfer learning and those trained from scratch revealed that transfer learning could marginally enhance SSIM from 0.793 to 0.794 while maintaining the same MSE. By incorporating ensemble and transfer learning techniques, our experiments demonstrated that we could achieve exceptional results even in scenarios with limited data and challenging problems.

In upcoming studies, transfer learning could predict additional properties, such as displacement or distribution maps of arterial walls since they share comparable distribution color patterns with stress and strain maps. Although this study focused on two-dimensional situations, the methodology presented can be extended to predict stress and strain distribution in arterial walls by utilizing a three-dimensional variant for CNN training and testing. Additionally, the proposed approach can estimate other arterial wall properties. It is

worth noting that this work did not incorporate lipids or other tissues in the arterial wall geometry. However, these can be easily included using Python scripting in Abaqus. Expanding the model to include multiple tissues and exploring additional properties, such as arterial wall displacement and deformation, would be valuable for future work. Another potential area of investigation could be incorporating stenting in the simulation to predict arterial wall deformation.

Chapter 5 Finite Element-Based Machine Learning Model for Predicting the Mechanical Properties of Composite Hydrogels

Abstract

In this study, a finite element (FE)-based machine learning model was developed to predict the mechanical properties of bioglass (BG)–collagen (COL) composite hydrogels. Based on the experimental observation of BG-COL composite hydrogels with scanning electron microscope, 2000 microstructural images with randomly distributed BG particles were created. The BG particles have diameters ranging from 0.5 μm to 1.5 μm and a volume fraction from 17% to 59%. FE simulations of tensile testing were performed for calculating the Young's modulus and Poisson's ratio of 2000 microstructures. The microstructural images and the calculated Young's modulus and Poisson's ratio by FE simulation were used for training and testing a convolutional neural net-work regression model. Results showed that the network developed in this work can effectively predict the mechanical properties of the composite hydrogels. The R-squared values were 95% and 83% for Young's modulus and Poisson's ratio, respectively. This work provides a surrogate model of finite element analysis to predict BG-COL hydrogel mechanical properties using microstructure images, which could be further utilized for characterizing heterogeneous materials in big data-driven material designs.

Keywords: Composite hydrogels; Mechanical properties; Machine learning.

Introduction

In tissue engineering, biocompatible materials (i.e., biomaterials) are employed to generate hydrogels or other scaffolds for use in the repair or replacement of damaged and diseased tissues. To reduce the risk of scar tissue formation at the interface with the host tissues, biomimetic materials that match the physicochemical properties of native tissue are commonly desired. Collagen (COL) is the most abundant protein in mammals (Eglin et al.

2006), which could be reinforced by bioglass (BG) to formulate composite scaffolds with improved mechanical properties for bone tissue engineering applications (Sarker et al. 2015). Specifically, Kajave et al. (Kajave et al. 2021) showed that incorporating BG into COL reduced the swelling and improved the stability and rheological properties (i.e., yield stress) of COL hydrogels. Gurumurthy et al. (Gurumurthy and Janorkar 2021) reported, in a review paper, that the stiffness of COL scaffolds significantly increased upon adding BGs. The mechanical properties of BG-COL composite scaffolds have been estimated from either a mathematical model or finite element (FE) method (Sousa et al. 2021). Homogenization methods, such as the double inclusion method, Mori–Tanaka mean field method, and self-consistent approaches, are generally practical for simple microstructures (Das et al. 2015; Hori and Nemat-Nasser 1993; C.-C. Yang and Huang 1996; Hua and Gu 2013). The representative volume element (RVE) technique in the FE method has been well utilized to estimate the effective properties of composite materials (Zhou et al. 2016; G.-D. Cheng, Cai, and Xu 2013; Omairey, Dunning, and Sriramula 2019). Prior work has shown that the mechanical properties of BG-COL scaffolds depend on the concentration, spatial distribution, and particle size of the BG, as well as the fabrication process (Swaminathan, Ghosh, and Pagano 2006).

Machine learning methods, especially deep learning methods, have been well utilized in solving engineering problems, e.g., predicting atomic and molecular properties, crystal structures and stability, stent expansion, and retinal mechanics (Guo et al. 2021; Shokrollahi, Dong, Kaya, et al. 2022; Dong et al. 2020; Garg et al. 2022). Specifically, deep neural network (DNN) approaches have proven efficient in uncovering unique structures. With sufficient training data, DNN uses the high-dimensional feature vector from the original data and learns the feature vector's nonlinear relationship with the expected output (Ziletti et al. 2018; Salmenjoki, Alava, and Laurson 2018). Ye et al. (Ye et al. 2019; Ye et al. 2021) demonstrated that DNN could efficiently provide an accurate mapping between the effective mechanical properties (Young's modulus and Poisson's ratio) and the microstructures of composites. A supervised ML was presented by Ford et al. (Ford et al. 2021) to predict the homogenized elastic properties of two-phase materials. Yang et al. (C. Yang et al. 2020) predicted the stress–strain curve of binary composites by using a

combination of principal component analysis and convolutional neural networks (CNN). Hamel et al. (Hamel et al. 2019) presented a FE-based evolutionary algorithm to design active composite structures for 4D printing that can achieve target shape shifting responses. Ponnusami et al. (Ponnusami 2021) introduced a deep learning approach for predicting unidirectional fiber-reinforced composites' transverse elastic and plastic properties. Wei et al. (Wei et al. 2018) utilized different machine learning methods, including CNN, support vector regression (SVR), and Gaussian process regression (GPR), to predict the effective thermal conductivity of composite materials. Rong et al. (Rong et al. 2019) employed 2D CNN to predict the effective thermal conductivity of 3D composites using 2D cross-section images. Kim et al. (Kim, Lim, and Lee 2021) predicted the transverse mechanical behavior of composites in terms of the stress–strain curves by implementing CNN. They considered the volume fraction (Vf) percentage of the particles as 40%, 50%, and 60%. However, the ML models have not been applied for predicting the mechanical properties of composite hydrogels. The current study is the first attempt to adopt an FE-based ML approach for predicting the mechanical properties of BG-COL. The integration of FE and ML methods, a.k.a. FE-based ML approach, is still in its early stage. The FE input (imaging) and outputs (Young's modulus and Poisson's ratio) are used to train the ML models and predict the FE outputs in test cases. This study can serve as a surrogate model for predicting Young's modulus and Poisson's ratio of composite hydrogels.

In this work, a CNN regression method was used to predict Young's modulus and Poisson's ratio of BG-COL composites. First, 2000 images of BG-COL microstructures were generated. Then, the mechanical properties of the BG-COL composite were calculated using the FE simulation software. These FE obtained data were used to train a CNN regression model for predicting the mechanical properties of BG-COL based on its microstructural image. We demonstrated that our CNN regression model could predict the mechanical properties of BG-COL and hence can aid in overcoming the challenges of predicting these properties with traditional homogenization methods. This work could guide the design of BG-COL and other composite hydrogels.

Materials and Methods

The overall workflow of our framework that integrates FE analysis with the CNN model to predict composite material properties is illustrated in **Figure 5-1**. Based on observing the microstructural images of BG-COL composite, we first generated 2000 virtual microstructural images of BG-COL with various structures consisting of circular shapes of BG. The microstructural images of BG-COL were then imported into the simulation software, Abaqus/Explicit software version 2019 (Dassault Systems Simulia Corporation, Providence, RI, USA). Then, their effective mechanical properties were extracted by using python script. The data acted as the foundation for developing our DNN prediction approach. Then, by training a CNN regression network, the mechanical properties of the composite were predicted.

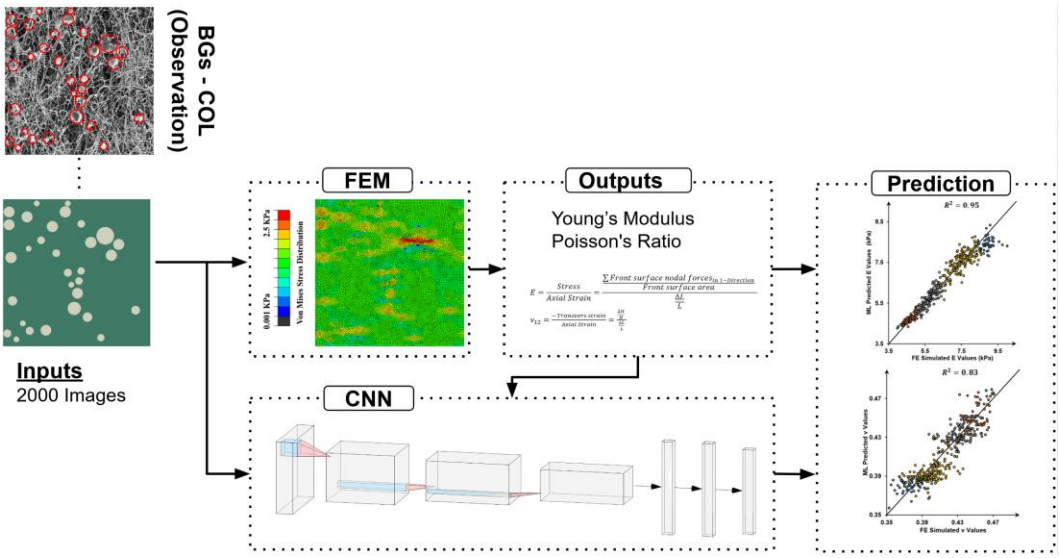


Figure 5-1. The overall framework combines FE analysis with an ML method to predict the mechanical properties of composite.

Simulation-Based Datasets

Bioglass particles were circular based on the experimental results and scanning electron microscope (SEM) images, as shown in **Figure 5-1** (depicted by red circles). Periodic two-phase microstructures were created using a synthetic two-dimensional (2D) microstructure generation algorithm we developed using Python (Python Software

Foundation, Delaware, United States). This algorithm can generate non-overlapping circles of varying sizes and Vf (Table 1). Additionally, a python script was developed to calculate the BG-COL material properties from preliminary RVE data. Python code developed in-house was used to organize, create, and link sets necessary for attaining deformable periodic boundary surfaces, which can distort and no longer remain plane, and compute homogenized properties, including Young's modulus and Poisson's ratio. Then, the periodic RVE homogenization method concepts were automatically assigned. First, the python code determines the boundary surfaces and RVE dimensions; then, by building nodal sets and applying the boundary conditions and the displacement needed for each property, FE analysis was performed. The interface between the BG and COL was perfectly bonded for simplicity.

The Young's modulus (E) and Poisson's ratio (ν) of the BG were adopted as E = 76.7 GPa and ν = 0.261, respectively (A.K. Srivastava, Pyare, and Singh 2012). The elastic properties of the COL were tested using our atomic force microscopy (AFM) as E = 3 kPa and ν = 0.49. Two-dimensional generalized plane stress elements (CPS4R) were used for the BG and COL meshing. A uniform displacement of 4 μm, i.e., an average strain of 20% (Wang et al. 2019), was applied at the boundary surface of the RVE.

To calculate the Young's modulus and Poisson's ratio of the BG-COL hydrogel, the reaction forces were computed as the sum of boundary nodal forces along the loading direction. The average stress was calculated as the reaction force divided by the boundary surface area. Then, Young's modulus of BG-COL was calculated by dividing the stress value by the applied axial strain of 20%. The transverse strain is simply a ratio of the change in height to the original height, and Poisson's ratio is estimated as the ratio of the transverse strain to the applied axial strain of 20%. The FE modeling framework was validated by the 3D RVEs in (Omairey, Dunning, and Sriramula 2019; Sun and Vaidya 1996).

Table 5-1: Microstructural images of BG-COL parameters.

Two-Phase Microstructure	Python Parameters	Material Parameters
COL	COL size = 20 × 20 μm	E = 3 kPa, ν = 0.49
BGs	BGs radius = 0.5 to 1.5 μm BGs V*f* % = 0.17 to 0.59 BGs number = 30 to 120	E = 76.7 GPa, ν = 0.261 (A.K. Srivastava, Pyare, and Singh 2012)

Machine Learning Approach

A CNN regression model was trained to estimate the effective mechanical properties of the BG-COL composite from microstructural images. The TensorFlow and Keras library were used to create the CNN model architecture (Abadi et al. 2016b). CNN is a class of DNN that applies a series of computationally nonlinear layers to analyze visual imagery. CNN can gradually extract representations of images with higher-level generalizations by operating on two functions that produce a third function, which expresses how the shape of one is modified by the other. The current CNN model was selected after trying many architectures and tuning hyperparameters. The tuned graph of the CNN regression used is depicted in **Figure 5-2**. It contains convolutional layers (first five layers) merged by fully connected layers (last four layers). The convolution features (or kernels) are memorized hierarchically and comprised of low-level features to build more complex patterns. The input to the CNN is the 2D BG-COL composite gray images of 200 × 200 pixels. A learnable kernel was applied to the input images to extract the convolved feature. This convolved feature is computed through a rectified linear unit (ReLU) (Agarap 2018) function and passed on to the following layers. The respective mechanical properties (Young's modulus and Poisson's ratio) are emanated at the outputs layer. The first five layers of architecture involve 3 × 3 convolutional layers using a stride of 2 in both the x and y directions; 2 × 2 maximum pooling layers; and 512, 256, 128, 64, and 32 feature maps, followed by four fully connected layers with 1000, 100, 50, and 10 units. To enhance the precision of the CNN's prediction, 20% dropout (N. Srivastava et al. 2014) was used in four fully connected layers to prevent overfitting. The Adam optimizer was used to speed up the convergence of the network (Kingma and Ba 2014). Linear activation function and sigmoid in the output layer of the CNN were chosen for Young's modulus and Poisson's ratio, respectively, by using functional API. A loss function was used to estimate the differences between CNN's prediction results and the real mechanical properties of the images. The loss function is minimized via backpropagation by optimizing the CNN parameters, such as biases and kernels, in the convolutional layers and the weights in the fully connected layers, as depicted

in Figure 2. The loss function is defined using the mean square error (MSE) since the prediction is a regression problem. MSE is denoted as:

$$MSE = L[Y, f(X)] = \frac{1}{n} \sum_{i=1}^{n} [Y - f(X)]^2 \quad (1)$$

where X denotes the input image describing the microstructure of the BG-COL, n represents the batch size, and $f(X)$ indicates the prediction of the CNN. Y denotes the mechanical properties of BG-COL, which includes both Young's modulus and Poisson's ratio. Performing verification to prevent overfitting during every training epoch was also done to select the best model. The CNN model was run on a workstation with the following specifications: Ryzen 9 5950X processor, 128 GB DDR4/2666 MHz memory, and Nvidia GeForce RTX 3090 GPU.

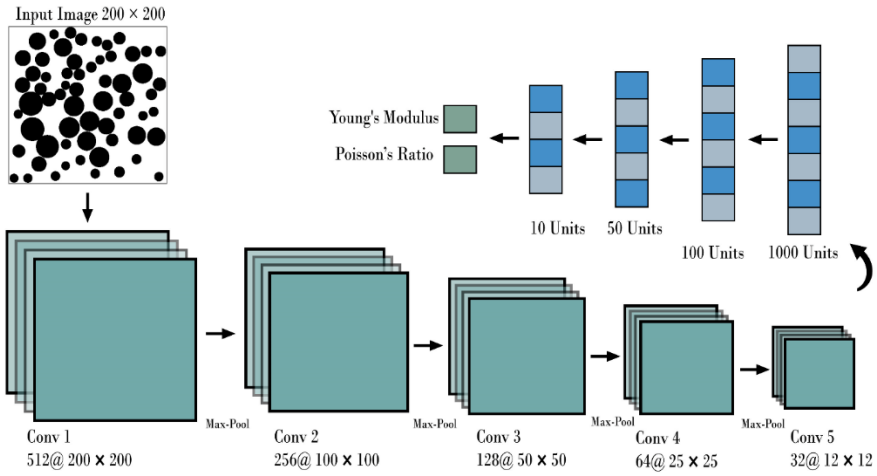


Figure 5-2. The CNN regression schematic for predicting Young's modulus and Poisson's ratio of BG-COL.

Results

The training and testing datasets contain 1600 and 400 images, respectively, and the corresponding mechanical properties (using an 80-20 split). We randomly chose 20% of the dataset as a validation set to assess performance when selecting model architecture and hyperparameters. By 20 repeated epochs, we trained our network, as demonstrated in **Figure**

5-3. The training and validation loss converge to a similar value, indicating insignificant overfitting.

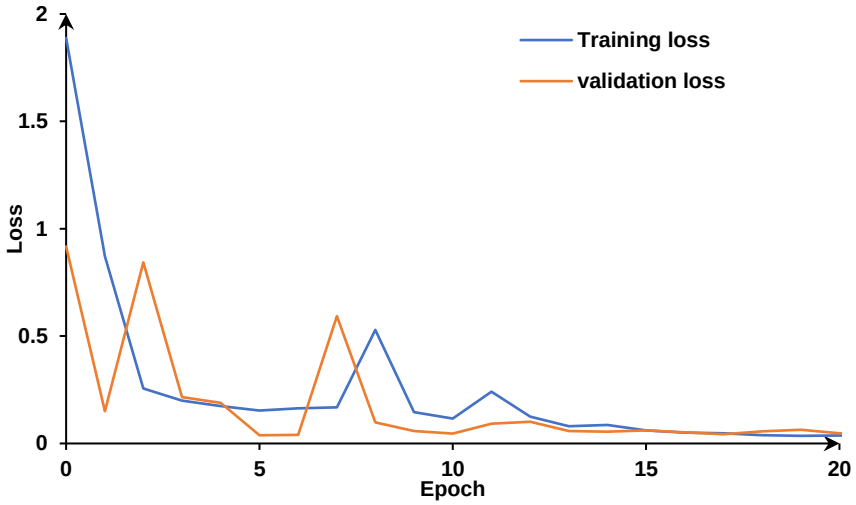
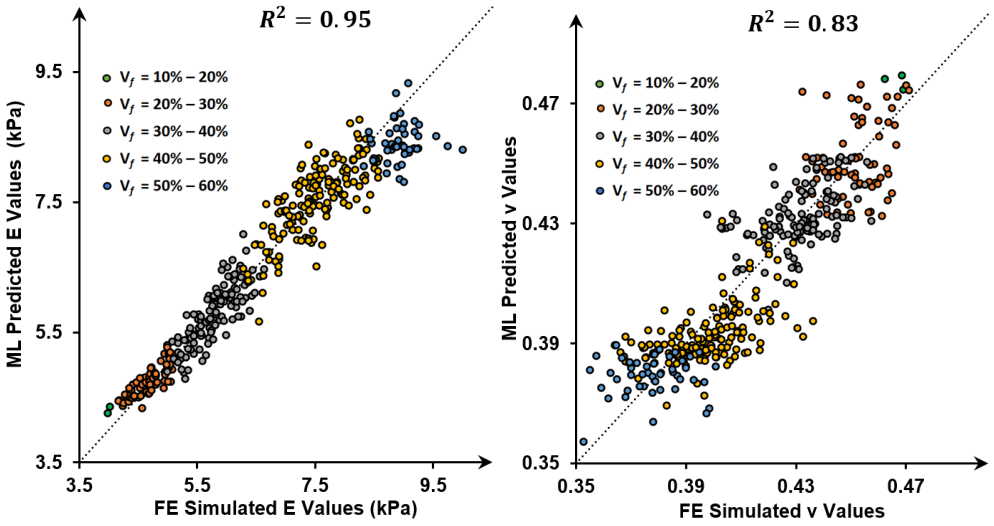


Figure 5-3. Training and validation loss over 20 epochs display convergence and behavior of model training.

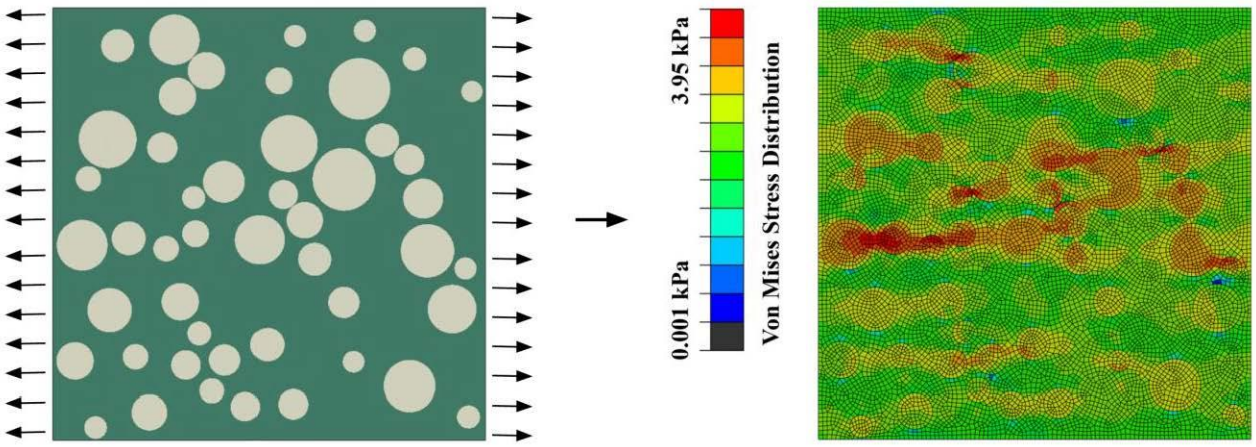
The Young's modulus and Poisson's ratio of the test set predicted by the CNN plotted against those obtained by the FE simulation are shown in **Figure 5-4**. Our network can efficiently learn and map the microstructure images to the mechanical properties. Additionally, these results indicate that our model works well for testing images. The ranges of Vf were illustrated in different colors to estimate the influence of BG on the mechanical properties. Results demonstrated that Young's modulus will increase with a larger Vf of BG. In contrast, Poisson's ratio decreased with a larger Vf. The MSE for Young's modulus is 0.135, and Poisson's ratio is 0.000162. The R-squared for Young's modulus is 0.95, and Poisson's ratio is 0.83, which shows how well our model fits. Our prediction is similar to the accuracy by Ye et al. (Ye et al. 2019).



(**a**) (**b**)

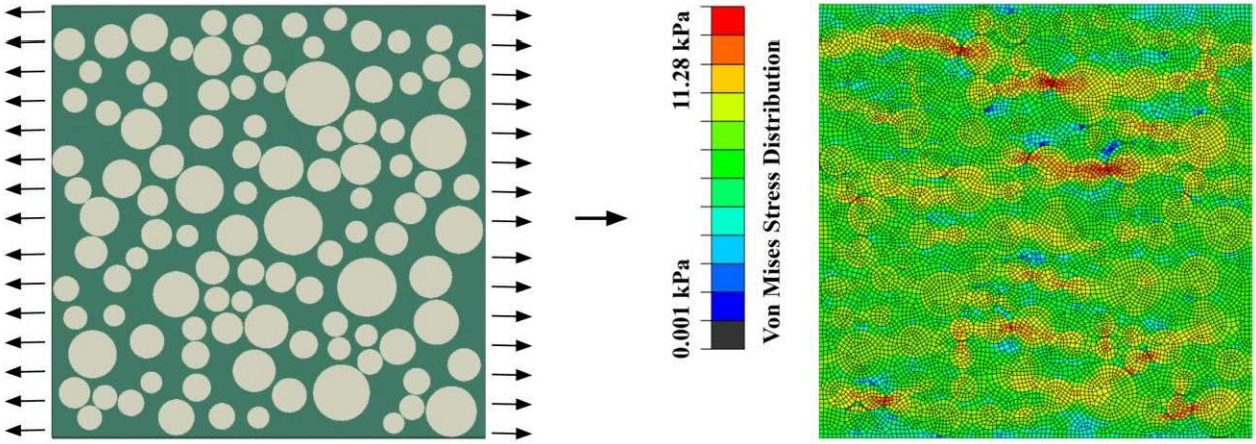
Figure 5-4: (a) The Young's modulus (b) and Poisson's ratio predicted by the CNN are plotted and compared to those generated by the RVE homogenization method and pulled by FE simulation.

Furthermore, out of all 400 testing cases, two representative cases were further delineated regarding prediction error in Young's modulus and Poisson's ratio of test data. As shown in **Figure 5-5**a, in this case, the prediction error of Young's modulus and Poisson's ratio is 0.32% and 3.61%, with 27.73% Vf, respectively. Moreover, in **Figure 5-5**b, the prediction error of Young's modulus and Poisson's ratio is 1% and 1.7%, with 47.85% Vf, respectively. Additionally, the legends of FE in **Figure 5-5** show the von Mises stress (kPa) distribution in the BG-COL composite, which is mostly a high magnitude of stress concentrated around BG, with maximum stress of 3.952 kPa and 11.28 kPa for Figure **Figure 5-5**a,b, respectively.



Mechanical properties	FEM	CNN	ERROR %	V_f
Young's modulus (kPa)	4.8425	4.8583	0.32%	27.73%
Poisson's Ratio	0.4494	0.4332	3.61%	

(a)



Mechanical properties	FEM	CNN	ERROR %	V_f
Young's modulus (kPa)	7.9161	7.8368	1.0%	47.85%
Poisson's Ratio	0.3740	0.3676	1.7%	

(b)

Figure 5-5. The prediction error of Young's modulus and Poisson's ratio for two cases with Vf at (a) 27.73% and (b) 47.85%.

Discussion

A FE-based CNN regression network method was used for predicting the effective mechanical properties of BG-COL. From the microstructure images, we can efficiently and accurately predict the mechanical properties of BG-COL, considering various BG numbers, sizes, and Vf. In total, 2000 2D RVE microstructural images were generated with various circular shapes of BG and Vf. This RVE of BG-COL was then imported into the FE simulation software to solve the BG-COL's effective mechanical properties. Subsequently, a python script in Abaqus was developed to extract the effective mechanical properties, including Young's modulus and Poisson's ratio. Finally, a CNN regression network was trained and tuned to predict the mechanical properties of the composite with 95% and 83% accuracy for Young's modulus and Poisson's ratio, respectively. The adopted framework can be used for nondestructive testing and big data-driven material design.

The FE input (imaging) and outputs (Young's modulus and Poisson's ratio) are used to train the ML models and then predict the FE outputs in test cases. A sufficiently large number of BG-COL images were generated to represent various composite configurations based on our experimental observations. Specifically, we effectively developed a framework to generate geometries, solve the FE computations, and then create a substantial database of input–output pairs. Similarly, Ford et al. (Ford et al. 2021) used three common two-phase materials (UD-CRTS (Khaled et al. 2018), UHP mortar (Arora et al. 2019; Mobasher et al. 2019), and MPR paste (Das, Kizilkanat, and Neithalath 2015)), including fibers of circular shape with the same diameter, but they used different fiber Vf, median sizes, and particles in a robust cementitious matrix and a metallic particulate-reinforced (MPR) cement mortar. They used artificial neural network (ANN) and forest ensemble MLs methods to predict Young's modulus and Poisson's ratio of materials by using geometrical features of particles and Vf inputs. CNN regression was chosen in this work considering that our dataset is 200 × 200 gray images with corresponding mechanical properties.

The FE-based CNN method to predict the mechanical properties of the composite hydrogel is still lacking in the literature (Pathan et al. 2017; Khaled et al. 2018; Ford et al.

2021). The prediction of BG-COL properties herein was motivated by the work of Ye et al. (Ye et al. 2019). They used a CNN regression method to predict the mechanical properties of an artificial composite consisting of two different components (matrix and inclusions). The uniqueness of our dataset is that it is based on the observations of the microstructural images of the BG-COL composites (**Figure 5-1**). In addition, we adopted the Vf ranging from 17% to 59% based on the published experiments (Kajave et al. 2021). The range of our Vf is wider than the one by Kim et al. (Kim, Lim, and Lee 2021). Our results have shown that a larger Vf significantly increased Young's modulus and reduced the Poisson's ratio. Subjected to the same deformation, a larger Vf is associated with a larger force load and a higher peak stress in the BG-COL (**Figure 5-5**). This trend was also observed by Yi et al. (Hua and Gu 2013).

The most critical parameters to achieve the prediction accuracy of 95% and 83% for Young's modulus and Poisson's ratio, respectively, were the number of feature maps in the convolutional layers and the number of units in the fully connected layer. A total of nine ML architectures with different layers (number of convolutional layers and feature maps in these layers, and fully connected layers and units in each layer), and different activation functions, including ReLU, sigmoid, and linear, were picked and trained. In the first architecture, our network was created by three convolutional layers with 256, 128, and 32 feature maps in each layer, respectively. The last convolutional layer was then connected to the first fully connected layer as the input in the vector form. Additionally, three fully connected layers with 500, 100, and 10 units were considered. The initial accuracy was 80% for Young's modulus and 72% for Poisson's ratio. By adding more layers and feature maps in the convolutional layers and units in the fully connected layers, our network had more trainable parameters and could learn more details and extract more features about images and their following material properties. It must be noted that overfitting can occur in a network with more layers, although more layers can increase the performance to a certain extent. The optimal number of layers and nodes in each configuration should be decided to avoid falling into the situation of overfitting.

The tuned model in this work consists of learnable kernels applied to the input images for extracting convolved features. These convolved features were computed through

a rectified linear unit (ReLU) function and passed on to the following layers. With 512 different kernels (the sizes of 3 × 3, using a stride of 2 in both x and y directions) at the first layer, similar procedures were repeated in layers 2, 3, 4, and 5 with 256, 128, 64, and 32 kernels, respectively. The last convolutional layer was then connected to the first fully connected layer as the input in the vector form. In the fully connected layers, learnable weights were calculated for each layer. We used the dropout technique after every fully connected layer. The last layer was a linear function for Young's modulus and a sigmoid function for Poisson's ratio. To minimize the prediction error, the kernels mentioned above were learned and will be activated when a similar feature appears in the input.

This study utilized two-dimensional RVE datasets. However, the framework could be expanded to predict the three-dimensional mechanical properties of BG-COL. Additional properties of the composite hydrogel, such as thermal conductivity, thermal expansion coefficients (Rong et al. 2019), fatigue life, toughness, and stress–strain curves (C. Yang et al. 2020), could be predicted using our framework. Moreover, it must be mentioned that this work did not consider the impact of COL crosslinking, which is typically done for these hydrogels (Kajave et al. 2021). When the COL and BG are crosslinked, the effect of BG on the mechanical properties of COL can change depending on what COL crosslinking method was used and how much crosslinking was done. For example, COL crosslinking can mask the effect of BG such that the BG incorporation does not impact the mechanical properties of the COL hydrogels.

Conclusions

In this work, the FE input (imaging) and outputs (Young's modulus and Poisson's ratio) were used to train a CNN regression network for predicting the mechanical properties of BG-COL composites, including Young's modulus and Poisson's ratio. Randomly distributed spherical BGs with diameters ranging from 0.5 μm to 1.5 μm and volume fractions from 17% to 59% were considered. The mechanical properties of 2000 images of the BG-COL microstructure were calculated in the FE simulation software utilizing an in-house python script. The prediction accuracy was 95% and 83% for Young's modulus and

Poisson’s ratio, respectively. The FE-based ML model is expected to facilitate the design of BG-COL and other composite hydrogels.

Chapter 6
Conclusion

Synopsis

The nature of the work outlined in this thesis is intrinsically interdisciplinary, bridging the gaps between machine learning, finite element analysis, and various crucial applications in healthcare. We initiated our exploration by evaluating the transformative role of Generative AI models, specifically transformers and denoising diffusion models, across a wide array of applications within the healthcare sector. From enhancing the accuracy of medical imaging and facilitating protein structure prediction to optimizing clinical documentation and revolutionizing drug design, our comprehensive review delineates how these cutting-edge models have significantly improved clinical diagnosis, data reconstruction, and drug synthesis processes. Additionally, our review provides a roadmap for future research by identifying ongoing challenges and suggesting potential avenues to tailor these models to the evolving needs of the healthcare industry. Following this, we address the critical issue of soccer-related ocular injuries by introducing an innovative framework that merges finite element (FE) analysis and machine learning (ML) to enable the rapid prediction of retinal mechanics. Given the rising incidence of soccer-related retinal injuries, timely and accurate diagnostics are paramount. Our proposed framework leverages FE simulations of various impact scenarios to derive stress and strain patterns in the posterior retina, which are then utilized to train and test a partial least squares regression (PLSR) model. This model facilitates swift prediction of soccer-induced retinal stress and strain, thus serving as a foundational step toward developing diagnostic tools for soccer-related ocular trauma. Subsequently, we delve into cardiovascular health by exploring the potential of end-to-end deep learning tools as efficient alternatives to finite element methods (FEM) for predicting stress-strain fields within arterial walls, a pivotal aspect of understanding atherosclerosis—a leading cause of global mortality. Our research proposes a U-Net-based fully CNN and a cGAN to predict von Mises stress and strain distribution based on the spatial arrangement of calcification within arterial wall cross-sections. Incorporating ensemble

approaches and transfer learning techniques further enhances the prediction accuracy of our models, thereby contributing to a more nuanced understanding of atherosclerosis and facilitating improved cardiovascular risk assessment. Lastly, we turn our attention to the development of a finite element (FE)-based machine learning model designed to predict the mechanical properties of composite hydrogels, specifically bioglass (BG)-collagen (COL) composite hydrogels. This innovative approach involves the creation of microstructural images based on scanning electron microscope observations, then utilizing these images and the mechanical properties calculated via FE simulations to train and test a convolutional neural network regression model. The resulting network effectively predicts the mechanical properties of the composite hydrogels, thereby serving as a valuable surrogate for finite element analysis in the characterization of heterogeneous materials in big data-driven material designs. The work presented in this thesis represents a significant contribution to the interdisciplinary fields of machine learning, finite element analysis, and healthcare. By developing innovative models and frameworks, we contribute to a deeper understanding of ocular and cardiovascular diseases, ultimately facilitating the development of novel therapeutic strategies and material designs.

Future Work

The future developments of this study can span a multitude of domains, from clinical and applied/engineering to scientific, and are outlined as follows:

• **Extension of Clinical Diagnosis Prediction Tasks:** This study has primarily concentrated on the diverse applications of machine learning and finite element analysis, including predicting stress-strain in cardiovascular vessels to minimize rupture risk, swiftly forecasting sports-related eye injuries, and biomaterials. However, there is significant potential for broadening the scope of clinical diagnosis prediction tasks, encompassing cardiovascular, ocular, tissue engineering, and other medical imaging domains. For example, deep learning algorithms for cardiovascular diseases could be further developed to forecast stent expansion, heart attack risk, and restenosis. Additionally, this research could be extended to other imaging modalities like MRI and CT scans and diseases such as coronary artery or

abdominal/thoracic aneurysms. Besides soccer, other sports such as basketball and handball could be considered regarding ocular injuries. Also, generative AI could create new healthcare data to address data insufficiency for training large models. For instance, generative AI could be used to generate new fundus eye images for diabetic retinopathy using diffusion models. In the context of material design for bone engineering, future studies could incorporate complex models and experimental verification into the current work. Furthermore, various material properties like thermal conductivity, thermal expansion coefficients, fatigue life, toughness, and stress-strain curves could be predicted using our framework.

• **Bioinformatics Integration:** Merging bioinformatics data, such as genomics and proteomics, with biophysical data using deep learning models can yield novel insights into cardiovascular and ocular diseases. Current healthcare guidelines are often based on limited study samples and may not capture the complexities of these diseases. However, leveraging deep learning's ability to recognize patterns in high-dimensional data can enable the integration of informatics and biophysical data, uncovering new insights crucial for the diagnosis and biomechanical analysis of cardiovascular diseases, ocular injuries, and the material properties of biological tissues. This approach can lead to more personalized and precise healthcare guidelines, addressing the limitations of existing models and contributing to the ongoing efforts in using AI for healthcare applications.

• **Advancements in Computer Vision for Anomaly Detection:** Medical imaging and automated diagnosis have needs that overlap with core machine learning and computer vision goals, such as anomaly detection and training with limited labeled data. Advancing computer vision techniques in these areas can help steer research in medical imaging and automated diagnosis.

• **Development of Full Deep Learning Pipelines:** Extending the atherosclerosis finite element work to develop a full deep learning pipeline from imaging to mechanical analysis can be a crucial next step. For example, intravascular imaging could be fed into a model for geometry and material property characterization, and the extracted parameters could then be

used for stress distribution prediction. This endeavor requires sufficient data and labeling for multiple prediction tasks and, ideally, could be trained end-to-end.

• **Bridging Finite Element Modeling and Machine Learning:** Bridging traditional FEM techniques for biomechanics with machine learning can enable high-throughput simulations for clinicians. This could involve further development in enabling multi-scale FEM, which suffers from computational inefficiency, and the advancement of machine learning for molecular dynamics simulations to alleviate computational inefficiency concerns and enable longer simulation timescales. Here are more suggestions based on current projects provided:

- FEM can be leveraged in the cardiovascular system to create more precise models derived from OCT images. Additionally, combining 3D vessel models and 3D machine learning methods with FEM can enhance the accuracy of the results.
- Regarding ocular injuries, employing intricate FEM techniques will facilitate a deeper understanding of the disease and its patterns in the eye. For future research, developing a comprehensive FEM eye model encompassing all details will contribute to refining the results.
- Concerning material design in tissue engineering, incorporating finer details into the FEM and extending it to 3D models can improve predictions of material properties based on actual materials.

• **Advancement of Generative AI:** The merging of generative AI and healthcare applications presents a myriad of opportunities for the future. As demonstrated in this study, the rapid progression of Generative AI models has already revolutionized several aspects of the healthcare sector, from medical imaging and protein structure prediction to clinical documentation and drug design. However, much work still needs to be done to harness these models' potential fully. Future efforts should focus on developing more sophisticated models capable of handling the vast and complex data inherent in healthcare, addressing current limitations such as data privacy and model interpretability, and expanding the application of Generative AI to other areas of healthcare that have yet to be explored. Additionally, there is a need for more comprehensive studies to assess the long-term impact and cost-

effectiveness of implementing Generative AI in healthcare settings. Ultimately, the advancement of Generative AI in healthcare could lead to more personalized and efficient patient care, streamlined operational processes, and innovative therapeutics and diagnostics.

In conclusion, the work presented in this thesis contributes to a deeper understanding of ocular and cardiovascular diseases and facilitates the development of novel therapeutic strategies and material designs. The outlined future directions provide a roadmap for extending this work across various domains, ultimately contributing to advancing healthcare applications of Generative AI, improved diagnosis and risk assessment in ocular and cardiovascular diseases and developing novel material designs.

Closure

Generative AI and machine learning research, though rooted in the mid to late twentieth century, have experienced a resurgence due to the availability of large, open datasets, advancements in computer hardware, and technical refinements in neural networks. This has propelled deep learning to the forefront, impacting multiple industries and transforming our interaction with the world. While it is important to remain cautious of excessive hype, inflated expectations, and unwarranted fears surrounding deep learning, its profound and enduring impact on various sectors, including healthcare, is undeniable.

In this study, I endeavored to harness the power of Generative AI and machine learning to contribute positively to human health and well-being. The healthcare sector, a colossal industry, is primed for technological innovation, and cardiovascular disease, a complex and pressing public health issue, presents an opportunity for impactful intervention. Approaching this challenge with biomechanics expertise, I aimed to establish a foundation for using deep learning and Generative AI in understanding and addressing cardiovascular disease, ocular injuries, and biomaterial design. This involves enhancing clinical diagnosis and risk assessment and predicting the mechanical properties of materials crucial for developing therapeutic interventions. By integrating finite element analysis with machine learning, predicting stress-strain fields within arterial walls, and developing models to predict the mechanical properties of composite hydrogels, this work contributes to a more

comprehensive understanding of cardiovascular disease. It paves the way for the development of innovative solutions.

References

Abdine, Hadi, Michail Chatzianastasis, Costas Bouyioukos, and Michalis Vazirgiannis. 2023. “Prot2Text: Multimodal Protein’s Function Generation with GNNs and Transformers.” arXiv. http://arxiv.org/abs/2307.14367.

AlAmir, Manal, and Manal AlGhamdi. 2022. “The Role of Generative Adversarial Network in Medical Image Analysis: An in-Depth Survey.” ACM Computing Surveys 55 (5): 1–36.

Ali, Hazrat, Md Rafiul Biswas, Farida Mohsen, Uzair Shah, Asma Alamgir, Osama Mousa, and Zubair Shah. 2022. “The Role of Generative Adversarial Networks in Brain MRI: A Scoping Review.” Insights into Imaging 13 (1): 98.

Asperti, Andrea. 2019. “Variational Autoencoders and the Variable Collapse Phenomenon.” Sensors & Transducers 234 (6): 1–8.

Azad, Reza, Moein Heidari, Moein Shariatnia, Ehsan Khodapanah Aghdam, Sanaz Karimijafarbigloo, Ehsan Adeli, and Dorit Merhof. 2022. “Transdeeplab: Convolution-Free Transformer-Based Deeplab V3+ for Medical Image Segmentation.” In International Workshop on PRedictive Intelligence In MEdicine, 91–102. Springer.

Azizi, Sima, Daniel B. Hier, and Donald C. Wunsch Ii. 2022. “Enhanced Neurologic Concept Recognition Using a Named Entity Recognition Model Based on Transformers.” Frontiers in Digital Health 4 (December): 1065581. https://doi.org/10.3389/fdgth.2022.1065581.

Bagal, Viraj, Rishal Aggarwal, P. K. Vinod, and U. Deva Priyakumar. 2022. “MolGPT: Molecular Generation Using a Transformer-Decoder Model.” Journal of Chemical Information and Modeling 62 (9): 2064–76. https://doi.org/10.1021/acs.jcim.1c00600.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2014. “Neural Machine Translation by Jointly Learning to Align and Translate.” arXiv Preprint arXiv:1409.0473.

Balouch, Bilal Ahmed Khan, and Fawad Hussain. 2023. “A Transformer Based Approach for Abstractive Text Summarization of Radiology Reports.” International Conference on Applied Engineering and Natural Sciences 1 (1): 476–86. https://doi.org/10.59287/icaens.1042.

Bao, Jianmin, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. 2017. “CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training.” In Proceedings of the IEEE International Conference on Computer Vision, 2745–54.

Behjati, Armin, Fatemeh Zare-Mirakabad, Seyed Shahriar Arab, and Abbas Nowzari-Dalini. 2022. “Protein Sequence Profile Prediction Using ProtAlbert Transformer.” Computational Biology and Chemistry 99 (August): 107717. https://doi.org/10.1016/j.compbiolchem.2022.107717.

Bhattacharya, Moinak, Shubham Jain, and Prateek Prasanna. 2022. “RadioTransformer: A Cascaded Global-Focal Transformer for Visual Attention–Guided Disease Classification.” In Computer Vision – ECCV 2022, edited by Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, 13681:679–98. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19803-8_40.

Boadu, Frimpong, Hongyuan Cao, and Jianlin Cheng. 2023. “Combining Protein Sequences and Structures with Transformers and Equivariant Graph Neural Networks to Predict Protein Function.” Preprint. Bioinformatics. https://doi.org/10.1101/2023.01.17.524477.

Böhle, Moritz, Mario Fritz, and Bernt Schiele. 2023. “Holistically Explainable Vision Transformers.” arXiv Preprint arXiv:2301.08669.

Bohr, Adam, and Kaveh Memarzadeh. 2020. “The Rise of Artificial Intelligence in Healthcare Applications.” In Artificial Intelligence in Healthcare, 25–60. Elsevier.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, and Emma Brunskill. 2021. “On the Opportunities and Risks of Foundation Models.” arXiv Preprint arXiv:2108.07258.

Brandes, Nadav, Dan Ofer, Yam Peleg, Nadav Rappoport, and Michal Linial. 2022. "ProteinBERT: A Universal Deep-Learning Model of Protein Sequence and Function." Edited by Pier Luigi Martelli. Bioinformatics 38 (8): 2102–10. https://doi.org/10.1093/bioinformatics/btac020.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, and Amanda Askell. 2020a. "Language Models Are Few-Shot Learners." Advances in Neural Information Processing Systems 33: 1877–1901.

———. 2020b. "Language Models Are Few-Shot Learners." Advances in Neural Information Processing Systems 33: 1877–1901.

Cao, Chentao, Zhuo-Xu Cui, Jing Cheng, Sen Jia, Hairong Zheng, Dong Liang, and Yanjie Zhu. 2022. "SPIRiT-Diffusion: SPIRiT-Driven Score-Based Generative Modeling for Vessel Wall Imaging." arXiv Preprint arXiv:2212.11274.

Cao, Yue, and Yang Shen. 2021. "TALE: Transformer-Based Protein Function Annotation with Joint Sequence–Label Embedding." Edited by Alfonso Valencia. Bioinformatics 37 (18): 2825–33. https://doi.org/10.1093/bioinformatics/btab198.

Carlini, Nicolas, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramer, Borja Balle, Daphne Ippolito, and Eric Wallace. 2023. "Extracting Training Data from Diffusion Models." In 32nd USENIX Security Symposium (USENIX Security 23), 5253–70.

Castro, Egbert, Abhinav Godavarthi, Julian Rubinfien, Kevin Givechian, Dhananjay Bhaskar, and Smita Krishnaswamy. 2022. "Transformer-Based Protein Generation with Regularized Latent Space Optimization." Nature Machine Intelligence 4 (10): 840–51. https://doi.org/10.1038/s42256-022-00532-1.

Chaudhari, Gunvant R., Tengxiao Liu, Timothy L. Chen, Gabby B. Joseph, Maya Vella, Yoo Jin Lee, Thienkhai H. Vu, et al. 2022. "Application of a Domain-Specific BERT for Detection of Speech Recognition Errors in Radiology Reports." Radiology: Artificial Intelligence 4 (4): e210185. https://doi.org/10.1148/ryai.210185.

Chefer, Hila, Shir Gur, and Lior Wolf. 2021. “Transformer Interpretability beyond Attention Visualization.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 782–91.

Chen, Shaobin, Zhenquan Wu, Mingzhu Li, Yun Zhu, Hai Xie, Peng Yang, Cheng Zhao, et al. 2023. “FIT-Net: Feature Interaction Transformer Network for Pathologic Myopia Diagnosis.” IEEE Transactions on Medical Imaging, 1–1. https://doi.org/10.1109/TMI.2023.3260990.

Chen, Xuxin, Ke Zhang, Neman Abdoli, Patrik W. Gilley, Ximin Wang, Hong Liu, Bin Zheng, and Yuchen Qiu. 2022. “Transformers Improve Breast Cancer Diagnosis from Unregistered Multi-View Mammograms.” Diagnostics 12 (7): 1549. https://doi.org/10.3390/diagnostics12071549.

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. “Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation.” arXiv Preprint arXiv:1406.1078.

Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.” arXiv Preprint arXiv:1412.3555.

Davidson, Tim R., Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M. Tomczak. 2018. “Hyperspherical Variational Auto-Encoders.” arXiv Preprint arXiv:1804.00891.

Dhinagar, Nikhil J., Sophia I. Thomopoulos, Emily Laltoo, and Paul M. Thompson. 2023. “Efficiently Training Vision Transformers on Structural MRI Scans for Alzheimer’s Disease Detection.” https://doi.org/10.48550/ARXIV.2303.08216.

Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. 2016. “Density Estimation Using Real Nvp.” arXiv Preprint arXiv:1605.08803.

Dong, Yanfang, Miao Zhang, Lishen Qiu, Lirong Wang, and Yong Yu. 2023. “An Arrhythmia Classification Model Based on Vision Transformer with Deformable Attention.” Micromachines 14 (6): 1155. https://doi.org/10.3390/mi14061155.

Dorjsembe, Zolnamar, Sodtavilan Odonchimed, and Furen Xiao. 2022. “Three-Dimensional Medical Image Synthesis with Denoising Diffusion Probabilistic Models.” In Medical Imaging with Deep Learning.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, and Sylvain Gelly. 2020. “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv Preprint arXiv:2010.11929.

Fabian, Benedek, Thomas Edlich, Héléna Gaspar, Marwin Segler, Joshua Meyers, Marco Fiscato, and Mohamed Ahmed. 2020. “Molecular Representation Learning with Language Models and Domain-Relevant Auxiliary Tasks.” https://doi.org/10.48550/ARXIV.2011.13230.

Feng, Chun-Mei, Yunlu Yan, Huazhu Fu, Li Chen, and Yong Xu. 2021. “Task Transformer Network for Joint MRI Reconstruction and Super-Resolution.” In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24, 307–17. Springer.

Feng, Jinyue, Chantal Shaib, and Frank Rudzicz. 2020. “Explainable Clinical Decision Support from Text.” In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1478–89. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.115.

Ferruz, Noelia, Steffen Schmidt, and Birte Höcker. 2022. “ProtGPT2 Is a Deep Unsupervised Language Model for Protein Design.” Nature Communications 13 (1): 4348. https://doi.org/10.1038/s41467-022-32007-7.

Garaiman, Alexandru, Farhad Nooralahzadeh, Carina Mihai, Nicolas Perez Gonzalez, Nikitas Gkikopoulos, Mike Oliver Becker, Oliver Distler, Michael Krauthammer, and Britta Maurer. 2023. “Vision Transformer Assisting Rheumatologists in Screening for Capillaroscopy Changes in Systemic Sclerosis: An Artificial Intelligence Model.” Rheumatology 62 (7): 2492–2500. https://doi.org/10.1093/rheumatology/keac541.

Geffen, Yaron, Yanay Ofran, and Ron Unger. 2022. “DistilProtBert: A Distilled Protein Language Model Used to Distinguish between Real Proteins and Their Randomly Shuffled Counterparts.” Preprint. Bioinformatics. https://doi.org/10.1101/2022.05.09.491157.

Gérardin, Christel, Perceval Wajsbürt, Basile Dura, Alice Calliger, Alexandre Moucher, Xavier Tannier, and Romain Bey. 2023. “Detecting Automatically the Layout of Clinical Documents to Enhance the Performances of Downstream Natural Language Processing.” https://doi.org/10.48550/ARXIV.2305.13817.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. “Generative Adversarial Nets.” Advances in Neural Information Processing Systems 27.

Gostin, Lawrence O. 2001. “National Health Information Privacy: Regulations under the Health Insurance Portability and Accountability Act.” Jama 285 (23): 3015–21.

Gothwal, Ritu, Shailendra Tiwari, and Shivendra Shivani. 2022. “Computational Medical Image Reconstruction Techniques: A Comprehensive Review.” Archives of Computational Methods in Engineering 29 (7): 5635–62. https://doi.org/10.1007/s11831-022-09785-w.

Grenander, Ulf, and Michael I. Miller. 1994. “Representations of Knowledge in Complex Systems.” Journal of the Royal Statistical Society: Series B (Methodological) 56 (4): 549–81.

Gulrajani, Ishaan, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C. Courville. 2017. “Improved Training of Wasserstein Gans.” Advances in Neural Information Processing Systems 30.

Ho, Jonathan, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, and David J. Fleet. 2022. “Imagen Video: High Definition Video Generation with Diffusion Models.” arXiv Preprint arXiv:2210.02303.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. “Denoising Diffusion Probabilistic Models.” Advances in Neural Information Processing Systems 33: 6840–51.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” Neural Computation 9 (8): 1735–80.

Hosain, A.K.M. Salman, Mynul Islam, Md Humaion Kabir Mehedi, Irteza Enan Kabir, and Zarin Tasnim Khan. 2022. “Gastrointestinal Disorder Detection with a Transformer Based Approach.” In 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 0280–85. Vancouver, BC, Canada: IEEE. https://doi.org/10.1109/IEMCON56893.2022.9946531.

Hu, Dingyi, Fengying Xie, Zhiguo Jiang, Yushan Zheng, and Jun Shi. 2022. “Histopathology Cross-Modal Retrieval Based on Dual-Transformer Network.” In 2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE), 97–102. Taichung, Taiwan: IEEE. https://doi.org/10.1109/BIBE55377.2022.00028.

Hu, Wendeng, and Sophia Y. Wang. 2022. “Predicting Glaucoma Progression Requiring Surgery Using Clinical Free-Text Notes and Transfer Learning With Transformers.” Translational Vision Science & Technology 11 (3): 37. https://doi.org/10.1167/tvst.11.3.37.

Huang, Guoling. 2022. “Surgical Action Recognition and Prediction with Transformers.” In 2022 IEEE 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI), 36–40. Xiamen, China: IEEE. https://doi.org/10.1109/SEAI55746.2022.9832094.

Huang, Kexin, Cao Xiao, Lucas M Glass, and Jimeng Sun. 2021. “MolTrans: Molecular Interaction Transformer for Drug–Target Interaction Prediction.” Edited by Zhiyong Lu. Bioinformatics 37 (6): 830–36. https://doi.org/10.1093/bioinformatics/btaa880.

Hyun, Chang Min, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, and Jin Keun Seo. 2018. “Deep Learning for Undersampled MRI Reconstruction.” Physics in Medicine & Biology 63 (13): 135007.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. “Image-to-Image Translation with Conditional Adversarial Networks.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1125–34.

Jacenkow, Grzegorz, Alison Q. O’Neil, and Sotirios A. Tsaftaris. 2022. “Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers.” In 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 1–5. Kolkata, India: IEEE. https://doi.org/10.1109/ISBI52829.2022.9761567.

Jozefowicz, Rafal, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. 2016. “Exploring the Limits of Language Modeling.” arXiv Preprint arXiv:1602.02410.

Kazerouni, Amirhossein, Ehsan Khodapanah Aghdam, Moein Heidari, Reza Azad, Mohsen Fayyaz, Ilker Hacihaliloglu, and Dorit Merhof. 2023. “Diffusion Models in Medical Imaging: A Comprehensive Survey.” Medical Image Analysis, 102846.

Kim, Boah, Inhwa Han, and Jong Chul Ye. 2022. “DiffuseMorph: Unsupervised Deformable Image Registration Using Diffusion Model.” In European Conference on Computer Vision, 347–64. Springer.

Kim, Boah, Yujin Oh, and Jong Chul Ye. 2022. “Diffusion Adversarial Representation Learning for Self-Supervised Vessel Segmentation.” arXiv Preprint arXiv:2209.14566.

Kim, Boah, and Jong Chul Ye. 2022. “Diffusion Deformable Model for 4D Temporal Medical Image Generation.” In International Conference on Medical Image Computing and Computer-Assisted Intervention, 539–48. Springer.

Kingma, Diederik P., and Max Welling. 2013. “Auto-Encoding Variational Bayes.” arXiv Preprint arXiv:1312.6114.

Kong, Zhifeng, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. 2020. “Diffwave: A Versatile Diffusion Model for Audio Synthesis.” arXiv Preprint arXiv:2009.09761.

Korkmaz, Yilmaz, Mahmut Yurt, Salman Ul Hassan Dar, Muzaffer Özbey, and Tolga Cukur. 2021. “Deep MRI Reconstruction with Generative Vision Transformers.” In Machine Learning for Medical Image Reconstruction: 4th International Workshop, MLMIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4, 54–64. Springer.

Kuchaiev, Oleksii, and Boris Ginsburg. 2017. “Factorization Tricks for LSTM Networks.” arXiv Preprint arXiv:1703.10722.

Lagunas, François, Ella Charlaix, Victor Sanh, and Alexander M. Rush. 2021. “Block Pruning for Faster Transformers.” arXiv Preprint arXiv:2109.04838.

Lentzen, Manuel, Sumit Madan, Vanessa Lage-Rupprecht, Lisa Kühnel, Juliane Fluck, Marc Jacobs, Mirja Mittermaier, et al. 2022. “Critical Assessment of Transformer-Based AI Models for German Clinical Notes.” JAMIA Open 5 (4): ooac087. https://doi.org/10.1093/jamiaopen/ooac087.

Levac, Brett, Ajil Jalal, and Jonathan I. Tamir. 2023. “Accelerated Motion Correction for MRI Using Score-Based Generative Models.” In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), 1–5. IEEE.

Li, Han, Dan Zhao, and Jianyang Zeng. 2022. “KPGT: Knowledge-Guided Pre-Training of Graph Transformer for Molecular Property Prediction.” In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 857–67. Washington DC USA: ACM. https://doi.org/10.1145/3534678.3539426.

Li, Jia, Yucong Lin, Pengfei Zhao, Wenjuan Liu, Linkun Cai, Jing Sun, Lei Zhao, et al. 2022. “Automatic Text Classification of Actionable Radiology Reports of Tinnitus Patients Using Bidirectional Encoder Representations from Transformer (BERT) and in-Domain Pre-Training (IDPT).” BMC Medical Informatics and Decision Making 22 (1): 200. https://doi.org/10.1186/s12911-022-01946-y.

Li, Xiang, John Thickstun, Ishaan Gulrajani, Percy S. Liang, and Tatsunori B. Hashimoto. 2022. “Diffusion-Lm Improves Controllable Text Generation.” Advances in Neural Information Processing Systems 35: 4328–43.

Li, Yikuan, Ramsey M. Wehbe, Faraz S. Ahmad, Hanyin Wang, and Yuan Luo. 2022. "Clinical-Longformer and Clinical-BigBird: Transformers for Long Clinical Sequences." https://doi.org/10.48550/ARXIV.2201.11838.

Liu, Leibo, Oscar Perez-Concha, Anthony Nguyen, Vicki Bennett, and Louisa Jorm. 2022. "Hierarchical Label-Wise Attention Transformer Model for Explainable ICD Coding." Journal of Biomedical Informatics 133 (September): 104161. https://doi.org/10.1016/j.jbi.2022.104161.

Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–22.

López-García, Guillermo, José M. Jerez, Nuria Ribelles, Emilio Alba, and Francisco J. Veredas. 2023. "Explainable Clinical Coding with In-Domain Adapted Transformers." Journal of Biomedical Informatics 139 (March): 104323. https://doi.org/10.1016/j.jbi.2023.104323.

Luong, Minh-Thang, Hieu Pham, and Christopher D. Manning. 2015. "Effective Approaches to Attention-Based Neural Machine Translation." arXiv Preprint arXiv:1508.04025.

Lyu, Qing, and Ge Wang. 2022. "Conversion between Ct and Mri Images Using Diffusion and Score-Matching Models." arXiv Preprint arXiv:2209.12104.

Madani, Ali, Ben Krause, Eric R. Greene, Subu Subramanian, Benjamin P. Mohr, James M. Holton, Jose Luis Olmos Jr, Caiming Xiong, Zachary Z. Sun, and Richard Socher. 2023. "Large Language Models Generate Functional Protein Sequences across Diverse Families." Nature Biotechnology, 1–8.

Meng, Yiwen, William Speier, Michael K. Ong, and Corey W. Arnold. 2021. "Bidirectional Representation Learning From Transformers Using Multimodal Electronic Health Record Data to Predict Depression." IEEE Journal of Biomedical and Health Informatics 25 (8): 3121–29. https://doi.org/10.1109/JBHI.2021.3063721.

Menze, B., A. Jakab, S. Bauer, J. Kalpathy-cramer, K. Farahani, and J. Kirby. 2014. “The Multimodal Brain Tumor Image Segmentation Benchmark (Brats). Medical Imaging.” IEEE Transactions On, 1–32.

Meyers, Joshua, Benedek Fabian, and Nathan Brown. 2021. “De Novo Molecular Design and Generative Models.” Drug Discovery Today 26 (11): 2707–15.

Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. “Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis.” Communications of the ACM 65 (1): 99–106.

Miyato, Takeru, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. 2018. “Spectral Normalization for Generative Adversarial Networks.” arXiv Preprint arXiv:1802.05957.

Mo, Yuhao, Chu Han, Yu Liu, Min Liu, Zhenwei Shi, Jiatai Lin, Bingchao Zhao, et al. 2023. “HoVer-Trans: Anatomy-Aware HoVer-Transformer for ROI-Free Breast Cancer Diagnosis in Ultrasound Images.” IEEE Transactions on Medical Imaging 42 (6): 1696–1706. https://doi.org/10.1109/TMI.2023.3236011.

Moezzi, Seyed Ali Reza, Abdolrahman Ghaedi, Mojdeh Rahmanian, Seyedeh Zahra Mousavi, and Ashkan Sami. 2022. “Application of Deep Learning in Generating Structured Radiology Reports: A Transformer-Based Technique.” Journal of Digital Imaging 36 (1): 80–90. https://doi.org/10.1007/s10278-022-00692-x.

Moghadam, Puria Azadi, Sanne Van Dalen, Karina C. Martin, Jochen Lennerz, Stephen Yip, Hossein Farahani, and Ali Bashashati. 2023. “A Morphology Focused Diffusion Probabilistic Model for Synthesis of Histopathology Images.” In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2000–2009.

Mohsan, Mashood Mohammad, Muhammad Usman Akram, Ghulam Rasool, Norah Saleh Alghamdi, Muhammad Abdullah Aamer Baqai, and Muhammad Abbas. 2023. “Vision Transformer and Language Model Based Radiology Report Generation.” IEEE Access 11: 1814–24. https://doi.org/10.1109/ACCESS.2022.3232719.

Moon, Sungrim, Huan He, and Hongfang Liu. 2022. “Sublanguage Characteristics of Clinical Documents.” In 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 3280–86. Las Vegas, NV, USA: IEEE. https://doi.org/10.1109/BIBM55620.2022.9995620.

Moor, Michael, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M. Krumholz, Jure Leskovec, Eric J. Topol, and Pranav Rajpurkar. 2023. “Foundation Models for Generalist Medical Artificial Intelligence.” Nature 616 (7956): 259–65.

Motwani, Tanya, and Manojkumar Parmar. 2020. “A Novel Framework for Selection of GANs for an Application.” arXiv Preprint arXiv:2002.08641.

Müller-Franzes, Gustav, Jan Moritz Niehues, Firas Khader, Soroosh Tayebi Arasteh, Christoph Haarburger, Christiane Kuhl, Tianci Wang, Tianyu Han, Teresa Nolte, and Sven Nebelung. 2023. “A Multimodal Comparison of Latent Denoising Diffusion Probabilistic Models and Generative Adversarial Networks for Medical Image Synthesis.” Scientific Reports 13 (1): 12098.

Ng, Clarence Boon Liang, Diogo Santos, and Marek Rei. 2023. “Modelling Temporal Document Sequences for Clinical ICD Coding.” https://doi.org/10.48550/ARXIV.2302.12666.

Nimalsiri, Wimukthi, Mahela Hennayake, Kasun Rathnayake, Thanuja D. Ambegoda, and Dulani Meedeniya. 2023. “Automated Radiology Report Generation Using Transformers.” In 2023 3rd International Conference on Advanced Research in Computing (ICARC), 90–95. Belihuloya, Sri Lanka: IEEE. https://doi.org/10.1109/ICARC57651.2023.10145699.

Nyholm, Tufve, Stina Svensson, Sebastian Andersson, Joakim Jonsson, Maja Sohlin, Christian Gustafsson, Elisabeth Kjellén, Karin Söderström, Per Albertsson, and Lennart Blomqvist. 2018. “MR and CT Data with Multiobserver Delineations of Organs in the Pelvic Area—Part of the Gold Atlas Project.” Medical Physics 45 (3): 1295–1300.

Oh, Hyun-Jic, and Won-Ki Jeong. 2023. “DiffMix: Diffusion Model-Based Data Synthesis for Nuclei Segmentation and Classification in Imbalanced Pathology Image Datasets.” arXiv Preprint arXiv:2306.14132.

Oh, Seo Hyun, Min Kang, and Youngho Lee. 2022. “Protected Health Information Recognition by Fine-Tuning a Pre-Training Transformer Model.” Healthcare Informatics Research 28 (1): 16–24. https://doi.org/10.4258/hir.2022.28.1.16.

Oliveira, Gabriel B., Helio Pedrini, and Zanoni Dias. 2023. “TEMPROT: Protein Function Annotation Using Transformers Embeddings and Homology Search.” BMC Bioinformatics 24 (1): 242. https://doi.org/10.1186/s12859-023-05375-0.

Oord, Aaron van den, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. “Wavenet: A Generative Model for Raw Audio.” arXiv Preprint arXiv:1609.03499.

Özbey, Muzaffer, Onat Dalmaz, Salman UH Dar, Hasan A. Bedel, Şaban Özturk, Alper Güngör, and Tolga Çukur. 2023. “Unsupervised Medical Image Translation with Adversarial Diffusion Models.” IEEE Transactions on Medical Imaging.

Pan, Shaoyan, Tonghe Wang, Richard LJ Qiu, Marian Axente, Chih-Wei Chang, Junbo Peng, Ashish B. Patel, Joseph Shelton, Sagar A. Patel, and Justin Roper. 2023. “2D Medical Image Synthesis Using Transformer-Based Denoising Diffusion Probabilistic Model.” Physics in Medicine & Biology 68 (10): 105004.

Parisi, Giorgio. 1981. “Correlation Functions and Computer Simulations.” Nuclear Physics B 180 (3): 378–84.

Park, Taesung, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. “Semantic Image Synthesis with Spatially-Adaptive Normalization.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2337–46.

Pinaya, Walter HL, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F. Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. 2022. “Brain Imaging Generation with Latent Diffusion Models.” In MICCAI Workshop on Deep Generative Models, 117–26. Springer.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. “Hierarchical Text-Conditional Image Generation with Clip Latents.” arXiv Preprint arXiv:2204.06125.

Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. “Zero-Shot Text-to-Image Generation.” In International Conference on Machine Learning, 8821–31. PMLR.

Razavi, Ali, Aaron Van den Oord, and Oriol Vinyals. 2019. “Generating Diverse High-Fidelity Images with vq-Vae-2.” Advances in Neural Information Processing Systems 32.

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. 2014. “Stochastic Backpropagation and Approximate Inference in Deep Generative Models.” In International Conference on Machine Learning, 1278–86. PMLR.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. “High-Resolution Image Synthesis with Latent Diffusion Models.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684–95.

Rong, Yu, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. 2020. “Self-Supervised Graph Transformer on Large-Scale Molecular Data.” https://doi.org/10.48550/ARXIV.2007.02835.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 234–41. Springer.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1985. Learning Internal Representations by Error Propagation. Institute for Cognitive Science, University of California, San Diego La ….

Saharia, Chitwan, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L. Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, and Tim Salimans. 2022. “Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding.” Advances in Neural Information Processing Systems 35: 36479–94.

Searle, Thomas, Zina Ibrahim, James Teo, and Richard J.B. Dobson. 2023. “Discharge Summary Hospital Course Summarisation of in Patient Electronic Health Record Text with Clinical Concept Guided Deep Pre-Trained Transformer Models.” Journal of Biomedical Informatics 141 (May): 104358. https://doi.org/10.1016/j.jbi.2023.104358.

Shamshad, Fahad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. 2023. “Transformers in Medical Imaging: A Survey.” Medical Image Analysis 88 (August): 102802. https://doi.org/10.1016/j.media.2023.102802.

Shang, Junyuan, Tengfei Ma, Cao Xiao, and Jimeng Sun. 2019. “Pre-Training of Graph Augmented Transformers for Medication Recommendation.” https://doi.org/10.48550/ARXIV.1906.00346.

Shazeer, Noam, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. 2017. “Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer.” arXiv Preprint arXiv:1701.06538.

Shehab, Mohammad, Laith Abualigah, Qusai Shambour, Muhannad A. Abu-Hashem, Mohd Khaled Yousef Shambour, Ahmed Izzat Alsalibi, and Amir H. Gandomi. 2022. “Machine Learning in Medical Applications: A Review of State-of-the-Art Methods.” Computers in Biology and Medicine 145: 105458.

Shen, Tianchang, Jun Gao, Kangxue Yin, Ming-Yu Liu, and Sanja Fidler. 2021. “Deep Marching Tetrahedra: A Hybrid Representation for High-Resolution 3d Shape Synthesis.” Advances in Neural Information Processing Systems 34: 6087–6101.

Shokrollahi, Yasin, Pengfei Dong, Xianqi Li, and Linxia Gu. 2023. “Deep Learning-Based Prediction of Stress and Strain Maps in Arterial Walls for Improved Cardiovascular Risk Assessment.” arXiv Preprint arXiv:2308.01771.

Sivarajkumar, Sonish, and Yanshan Wang. 2022. “HealthPrompt: A Zero-Shot Learning Paradigm for Clinical Natural Language Processing.” AMIA ... Annual Symposium Proceedings. AMIA Symposium 2022: 972–81.

Sohl-Dickstein, Jascha, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics." In International Conference on Machine Learning, 2256–65. PMLR.

Solarte-Pabón, Oswaldo, Orlando Montenegro, Alvaro García-Barragán, Maria Torrente, Mariano Provencio, Ernestina Menasalvas, and Víctor Robles. 2023. "Transformers for Extracting Breast Cancer Information from Spanish Clinical Narratives." Artificial Intelligence in Medicine 143 (September): 102625. https://doi.org/10.1016/j.artmed.2023.102625.

Song, Yang, and Stefano Ermon. 2019. "Generative Modeling by Estimating Gradients of the Data Distribution." Advances in Neural Information Processing Systems 32.

Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2020. "Score-Based Generative Modeling through Stochastic Differential Equations." arXiv Preprint arXiv:2011.13456.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP." arXiv Preprint arXiv:1906.02243.

Sun, Siqi, Yu Cheng, Zhe Gan, and Jingjing Liu. 2019. "Patient Knowledge Distillation for Bert Model Compression." arXiv Preprint arXiv:1908.09355.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. "Sequence to Sequence Learning with Neural Networks." Advances in Neural Information Processing Systems 27.

Tchouka, Yakini, Jean-François Couchot, David Laiymani, Philippe Selles, and Azzedine Rahmani. 2023. "Automatic ICD-10 Code Association: A Challenging Task on French Clinical Texts." https://doi.org/10.48550/ARXIV.2304.02886.

Van Panhuis, Willem G., Proma Paul, Claudia Emerson, John Grefenstette, Richard Wilder, Abraham J. Herbst, David Heymann, and Donald S. Burke. 2014. "A Systematic Review of Barriers to Data Sharing in Public Health." BMC Public Health 14 (1): 1–9.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." Advances in Neural Information Processing Systems 30.

Vig, Jesse, Ali Madani, Lav R. Varshney, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. 2020. “BERTology Meets Biology: Interpreting Attention in Protein Language Models.” https://doi.org/10.48550/ARXIV.2006.15222.

Vincent, Pascal. 2011. “A Connection between Score Matching and Denoising Autoencoders.” Neural Computation 23 (7): 1661–74.

Waibel, Dominik JE, Ernst Röoell, Bastian Rieck, Raja Giryes, and Carsten Marr. 2022. “A Diffusion Model Predicts 3d Shapes from 2d Microscopy Images.” arXiv Preprint arXiv:2208.14125.

Wang, Liwei, Huan He, Andrew Wen, Sungrim Moon, Sunyang Fu, Kevin J Peterson, Xuguang Ai, Sijia Liu, Ramakanth Kavuluru, and Hongfang Liu. 2023. “Acquisition of a Lexicon for Family History Information: Bidirectional Encoder Representations From Transformers–Assisted Sublanguage Analysis.” JMIR Medical Informatics 11 (June): e48072. https://doi.org/10.2196/48072.

Wei, Qiang, Xu Zuo, Omer Anjum, Yan Hu, Ryan Denlinger, Elmer V. Bernstam, Martin J Citardi, and Hua Xu. 2022. “ClinicalLayoutLM: A Pre-Trained Multi-Modal Model for Understanding Scanned Document in Electronic Health Records.” In 2022 IEEE International Conference on Big Data (Big Data), 2821–27. Osaka, Japan: IEEE. https://doi.org/10.1109/BigData55660.2022.10020569.

Wiatrak, Maciej, Stefano V. Albrecht, and Andrew Nystrom. 2019. “Stabilizing Generative Adversarial Networks: A Survey.” arXiv Preprint arXiv:1910.00927.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, and Klaus Macherey. 2016. “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.” arXiv Preprint arXiv:1609.08144.

Xiao, Zhisheng, Karsten Kreis, and Arash Vahdat. 2021. “Tackling the Generative Learning Trilemma with Denoising Diffusion Gans.” arXiv Preprint arXiv:2112.07804.

Xie, Yutong, and Quanzheng Li. 2022. “Measurement-Conditioned Denoising Diffusion Probabilistic Model for under-Sampled Medical Image Reconstruction.” In International Conference on Medical Image Computing and Computer-Assisted Intervention, 655–64. Springer.

Yang, Guandao, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. 2019. “Pointflow: 3d Point Cloud Generation with Continuous Normalizing Flows.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, 4541–50.

Yang, Ling, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Yingxia Shao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. 2022. “Diffusion Models: A Comprehensive Survey of Methods and Applications.” arXiv Preprint arXiv:2209.00796.

Yang, Yijun, Huazhu Fu, Angelica Aviles-Rivero, Carola-Bibiane Schönlieb, and Lei Zhu. 2023. “DiffMIC: Dual-Guidance Diffusion Network for Medical Image Classification.” arXiv Preprint arXiv:2303.10610.

Yao, Zhewei, Reza Yazdani Aminabadi, Minjia Zhang, Xiaoxia Wu, Conglong Li, and Yuxiong He. 2022. “Zeroquant: Efficient and Affordable Post-Training Quantization for Large-Scale Transformers.” Advances in Neural Information Processing Systems 35: 27168–83.

Yogarajan, Vithya, Jacob Montiel, Tony Smith, and Bernhard Pfahringer. 2021. “Transformers for Multi-Label Classification of Medical Text: An Empirical Comparison.” In Artificial Intelligence in Medicine, edited by Allan Tucker, Pedro Henriques Abreu, Jaime Cardoso, Pedro Pereira Rodrigues, and David Riaño, 12721:114–23. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-77211-6_12.

Zhou, Hong-Yu, Yizhou Yu, Chengdi Wang, Shu Zhang, Yuanxu Gao, Jia Pan, Jun Shao, Guangming Lu, Kang Zhang, and Weimin Li. 2023. “A Transformer-Based Representation-Learning Model with Unified Processing of Multimodal Input for Clinical Diagnostics.” Nature Biomedical Engineering 7 (6): 743–55. https://doi.org/10.1038/s41551-023-01045-x.

Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.” In Proceedings of the IEEE International Conference on Computer Vision, 2223–32.

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard. 2016a. "Tensorflow: a system for large-scale machine learning." Osdi.

---. 2016b. "{TensorFlow}: a system for {Large-Scale} machine learning." 12th USENIX symposium on operating systems design and implementation (OSDI 16).

Abaqus, G. 2011. "Abaqus 6.11." *Dassault Systemes Simulia Corporation, Providence, RI, USA*.

Agarap, Abien Fred. 2018. "Deep learning using rectified linear units (relu)." *arXiv preprint arXiv:1803.08375*.

Arora, Aashay, Asim Almujaddidi, Farrokh Kianmofrad, Barzin Mobasher, and Narayanan Neithalath. 2019. "Material design of economical ultra-high performance concrete (UHPC) and evaluation of their properties." *Cement and Concrete Composites* 104: 103346.

Bathe, Klaus-Jürgen. 2006. *Finite element procedures*. Klaus-Jurgen Bathe.

Bhaduri, Anindya, Yanyan He, Michael D Shields, Lori Graham-Brady, and Robert M Kirby. 2018. "Stochastic collocation approach with adaptive mesh refinement for parametric uncertainty analysis." *Journal of Computational Physics* 371: 732-750.

Chau, Alexandra H, Raymond C Chan, Milen Shishkov, Briain MacNeill, Nicusor Iftimia, Guillermo J Tearney, Roger D Kamm, Brett E Bouma, and Mohammad R Kaazempur-Mofrad. 2004. "Mechanical analysis of atherosclerotic plaques based on optical coherence tomography." *Annals of biomedical engineering* 32: 1494-1503.

Cheng, Geng-Dong, Yuan-Wu Cai, and Liang Xu. 2013. "Novel implementation of homogenization method to predict effective properties of periodic materials." *Acta Mechanica Sinica* 29 (4): 550-556.

Cheng, Jin M, Hector M Garcia-Garcia, Sanneke PM de Boer, Isabella Kardys, Jung Ho Heo, K Martijn Akkerhuis, Rohit M Oemrawsingh, Ron T van Domburg, Jurgen Ligthart, and Karen T Witberg. 2014. "In vivo detection of high-risk coronary plaques by radiofrequency intravascular ultrasound and cardiovascular outcome: results of the ATHEROREMO-IVUS study." *European heart journal* 35 (10): 639-647.

Church, Charles C, and Douglas L Miller. 2016. "A two-criterion model for microvascular bio-effects induced in vivo by contrast microbubbles exposed to medical ultrasound." *Ultrasound in medicine & biology* 42 (6): 1385-1398.

Cilla, Myriam, Javier Martinez, Estefania Pena, and Miguel Ángel Martínez. 2012. "Machine learning techniques as a helpful tool toward determination of plaque vulnerability." *IEEE Transactions on Biomedical Engineering* 59 (4): 1155-1161.

Clemente, Chiara, Luca Esposito, Nicola Bonora, Jerome Limido, Jean-Luc Lacome, and Tommaso Rossi. 2014. "Traumatic eye injuries as a result of blunt impact: computational issues." Journal of Physics: Conference Series.

Creswell, Antonia, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. 2018. "Generative adversarial networks: An overview." *IEEE signal processing magazine* 35 (1): 53-65.

Cristianini, Nello, and John Shawe-Taylor. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Das, Sumanta, Ahmet Kizilkanat, and Narayanan Neithalath. 2015. "Crack propagation and strain localization in metallic particulate-reinforced cementitious mortars." *Materials & Design* 79: 15-25.

Das, Sumanta, Pu Yang, Sudhanshu S Singh, James CE Mertens, Xianghui Xiao, Nikhilesh Chawla, and Narayanan Neithalath. 2015. "Effective properties of a fly ash geopolymer: Synergistic application of X-ray synchrotron tomography, nanoindentation, and homogenization models." *Cement and Concrete Research* 78: 252-262.

Dietterich, Thomas G. 2000. "Ensemble methods in machine learning." Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1.

Dong, Pengfei, Hozhabr Mozafari, Juhwan Lee, Yazan Gharaibeh, Vladislav N Zimin, Luis AP Dallan, Hiram G Bezerra, David L Wilson, and Linxia Gu. 2021. "Mechanical performances of balloon post-dilation for improving stent expansion in calcified coronary artery: computational and experimental investigations." *Journal of the Mechanical Behavior of Biomedical Materials* 121: 104609.

Dong, Pengfei, Guochang Ye, Mehmet Kaya, and Linxia Gu. 2020. "Simulation-Driven Machine Learning for Predicting Stent Expansion in Calcified Coronary Artery." *Applied Sciences* 10 (17): 5820.

Eglin, David, Sonia Maalheem, Jacques Livage, and Thibaud Coradin. 2006. "In vitro apatite forming ability of type I collagen hydrogels containing bioactive glass and silica sol-gel particles." *Journal of Materials Science: materials in medicine* 17 (2): 161-167.

Filipe, JA Capão, A Rocha-Sousa, F Falcão-Reis, and J Castro-Correia. 2003. "Modern sports eye injuries." *British journal of ophthalmology* 87 (11): 1336-1339. https://doi.org/http://dx.doi.org/10.1136/bjo.87.11.1336.

Filipe, Joao A Capao, Vitor L Fernandes, Henrique Barros, Fernando Falcao-Reis, and José Castro-Correia. 2003. "Soccer-related ocular injuries." *Archives of ophthalmology* 121 (5): 687-694.

Ford, Emily, Kailasnath Maneparambil, Subramaniam Rajan, and Narayanan Neithalath. 2021. "Machine learning-based accelerated property prediction of two-phase materials using microstructural descriptors and finite element analysis." *Computational Materials Science* 191: 110328.

Garg, Aman, Mohamed-Ouejdi Belarbi, Abdelouahed Tounsi, Li Li, Ankit Singh, and Tanmoy Mukhopadhyay. 2022. "Predicting elemental stiffness matrix of FG nanoplates using Gaussian Process Regression based surrogate model in framework of layerwise model." *Engineering Analysis with Boundary Elements* 143: 779-795.

Gharaibeh, Yazan, Pengfei Dong, David Prabhu, Chaitanya Kolluru, Juhwan Lee, Vlad Zimin, Hozhabr Mozafari, Hiram Bizzera, Linxia Gu, and David Wilson. 2019. "Deep learning segmentation of coronary calcified plaque from intravascular optical coherence tomography (IVOCT) images with application to finite element modeling of stent deployment." Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling.

Gholami, Kimia, Faraz Ege, and Ramin Barzegar. 2023. "Prediction of Composite Mechanical Properties: Integration of Deep Neural Network Methods and Finite Element Analysis." *Journal of Composites Science* 7 (2): 54.

Ghosh, Fredrik, and Birgitta Bauer. 1995. "Sports-related eye injuries." *Acta Ophthalmologica Scandinavica* 73 (4): 353-354. https://doi.org/https://doi.org/10.5694/j.1326-5377.1994.tb125941.x.

Gökçe, Gökçen, Osman Melih Ceylan, Fazil Cüneyt Erdurman, Ali Hakan Durukan, and Gungor Sobacı. 2013. "Soccer ball related posterior segment closed-globe injuries in outdoor amateur players." *Ulus Travma Acil Cerrahi Derg* 19 (3): 219-22. https://doi.org/https://doi.org/10.5505/tjtes.2013.74050.

Goodfellow, Ian. 2016. "Nips 2016 tutorial: Generative adversarial networks." *arXiv preprint arXiv:1701.00160*.

GrabCad. 2018. "Skull Model." https://grabcad.com/library/skull-43.

Guo, Kai, Zhenze Yang, Chi-Hua Yu, and Markus J Buehler. 2021. "Artificial intelligence and machine learning in design of mechanical materials." *Materials Horizons* 8 (4): 1153-1172.

Gurumurthy, Bhuvaneswari, and Amol V Janorkar. 2021. "Improvements in mechanical properties of collagen-based scaffolds for tissue engineering." *Current Opinion in Biomedical Engineering* 17: 100253.

Hamel, Craig M, Devin J Roach, Kevin N Long, Frédéric Demoly, Martin L Dunn, and H Jerry Qi. 2019. "Machine-learning based design of active composite structures for 4D printing." *Smart Materials and Structures* 28 (6): 065005.

Hassan, Mohd Hasnun Arif, Zahari Taha, Iskandar Hasanuddin, and Mohd Jamil Mohamed Mokhtarudin. 2018. *Mechanics of soccer heading and protective headgear*. Springer.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision.

Hore, Alain, and Djemel Ziou. 2010. "Image quality metrics: PSNR vs. SSIM." 2010 20th international conference on pattern recognition.

Hori, Muneo, and Sia Nemat-Nasser. 1993. "Double-inclusion model and overall moduli of multi-phase composites." *Mechanics of Materials* 14 (3): 189-206.

Horn, Erich P, H Richard McDonald, Robert N Johnson, Everett Ai, George A Williams, John M Lewis, Patrick E Rubsamen, Paul Sternberg Jr, Robert B Bhisitkul, and William F Mieler. 2000. "Soccer ball-related retinal injuries: a report of 13 cases." *Retina (Philadelphia, Pa.)* 20 (6): 604-609. https://doi.org/https://doi.org/10.1097/00006982-200011000-00003.

Hua, Yi, and Linxia Gu. 2013. "Prediction of the thermomechanical behavior of particle-reinforced metal matrix composites." *Composites Part B: Engineering* 45 (1): 1464-1470.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition.

Jalali, Subhadra. 2003. "Retinal detachment." *Community Eye Health* 16 (46): 25.

Kajave, Nilabh S, Trevor Schmitt, Thuy-Uyen Nguyen, Akhilesh K Gaharwar, and Vipuil Kishore. 2021. "Bioglass incorporated methacrylated collagen bioactive ink for 3D printing of bone tissue." *Biomedical Materials* 16 (3): 035003.

Karimi, Alireza, Reza Razaghi, Seyed Mohammadali Rahmati, Toshihiro Sera, and Susumu Kudo. 2018. "A nonlinear dynamic finite-element analyses of the basketball-related eye injuries." *Sports Engineering* 21 (4): 359-365. https://doi.org/https://doi.org/10.1007/s12283-018-0280-3.

Kent, Jerrod S, Ryan B Eidsness, Kevin M Colleaux, and Kenneth G Romanchuk. 2007. "Indoor soccer-related eye injuries: should eye protection be mandatory?" *Canadian journal of ophthalmology* 42 (4): 605-608. https://doi.org/https://doi.org/10.3129/can.j.ophthalmol.i07-093.

Khaled, Bilal, Loukham Shyamsunder, Canio Hoffarth, Subramaniam D Rajan, Robert K Goldberg, Kelly S Carney, Paul DuBois, and Gunther Blankenhorn. 2018. "Experimental characterization of composites to support an orthotropic plasticity material model." *Journal of Composite Materials* 52 (14): 1847-1872.

Kim, Do-Won, Jae Hyuk Lim, and Seungchul Lee. 2021. "Prediction and validation of the transverse mechanical behavior of unidirectional composites considering interfacial debonding through convolutional neural networks." *Composites Part B: Engineering* 225: 109314.

Kingma, Diederik P, and Jimmy Ba. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*.

Krams, R, JJ Wentzel, JAF Oomen, R Vinke, JCH Schuurbiers, PJ De Feyter, PW Serruys, and CJ Slager. 1997. "Evaluation of endothelial shear stress and 3D geometry as factors determining the development of atherosclerosis and remodeling in human coronary arteries in vivo: combining 3D reconstruction from angiography and IVUS (ANGUS) with computational fluid dynamics." *Arteriosclerosis, thrombosis, and vascular biology* 17 (10): 2061-2065.

Lam, Matthew, Pengfei Dong, Yasin Shokrollahi, Donny Suh, and Linxia Gu. 2021. "Computer-Assisted Analysis of Soccer Ball Trauma of the Eye."

Lam, Matthew R, Pengfei Dong, Yasin Shokrollahi, Linxia Gu, and Donny W Suh. 2022a. "Finite element analysis of soccer ball-related ocular and retinal trauma and comparison with abusive head trauma." *Ophthalmology science* 2 (2): 100129.

---. 2022b. "Finite Element Analysis of Soccer Ball-Related Ocular and Retinal Trauma and Comparison With Abusive Head Trauma." *Ophthalmology Science*: 100129.

Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, and Zehan Wang. 2017. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition.

Leshno, Ari, Amir Alhalel, Miri Fogel-Levin, Ofira Zloto, Joseph Moisseiev, and Orit Vidne-Hay. 2021. "Pediatric retinal damage due to soccer-ball-related injury: Results from the last decade." *European journal of ophthalmology* 31 (1): 240-244.

Li, Xiang, Zhanli Liu, Shaoqing Cui, Chengcheng Luo, Chenfeng Li, and Zhuo Zhuang. 2019. "Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning." *Computer Methods in Applied Mechanics and Engineering* 347: 735-753.

Li, Yi-Chen, Thau-Yun Shen, Chien-Cheng Chen, Wei-Ting Chang, Po-Yang Lee, and Chih-Chung Johnson Huang. 2021. "Automatic detection of atherosclerotic plaque and calcification from intravascular ultrasound images by using deep convolutional neural networks." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 68 (5): 1762-1772.

Liang, Liang, Minliang Liu, Caitlin Martin, and Wei Sun. 2018. "A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis." *Journal of The Royal Society Interface* 15 (138): 20170844.

Libby, Peter, Paul M Ridker, and Attilio Maseri. 2002. "Inflammation and atherosclerosis." *Circulation* 105 (9): 1135-1143.

Liu, Xiaoyu, Lizhen Wang, Chao Wang, Ganyun Sun, Songyang Liu, and Yubo Fan. 2013. "Mechanism of traumatic retinal detachment in blunt impact: a finite element study." *Journal of biomechanics* 46 (7): 1321-1327.

MacEwen, Caroline J. 1989. "Eye injuries: a prospective survey of 5671 cases." *British Journal of Ophthalmology* 73 (11): 888-894. https://doi.org/http://dx.doi.org/10.1136/bjo.73.11.888.

Madani, Ali, Ahmed Bakhaty, Jiwon Kim, Yara Mubarak, and Mohammad RK Mofrad. 2019. "Bridging finite element and machine learning modeling: stress prediction of arterial walls in atherosclerosis." *Journal of biomechanical engineering* 141 (8).

Mobasher, Barzin, Aashay Arora, Matthew Aguayo, Farrokh Kianmofrad, Yiming Yao, and Narayanan Neithalath. 2019. *Developing Ultra-High–Performance Concrete Mix Designs for Arizona Bridge Element Connections.* Arizona. Department of Transportation. Research Center.

Mollaei Ardestani, Alireza, Ghasem Azamirad, Yasin Shokrollahi, Matteo Calaon, Jesper Henri Hattel, Murat Kulahci, Roya Soltani, and Guido Tosello. 2023. "Application of Machine Learning for Prediction and Process Optimization—Case Study of Blush Defect in Plastic Injection Molding." *Applied Sciences* 13 (4): 2617.

Nie, Zhenguo, Haoliang Jiang, and Levent Burak Kara. 2020. "Stress field prediction in cantilevered structures using convolutional neural networks." *Journal of Computing and Information Science in Engineering* 20 (1): 011002.

Omairey, Sadik L, Peter D Dunning, and Srinivas Sriramula. 2019. "Development of an ABAQUS plugin tool for periodic RVE homogenisation." *Engineering with Computers* 35 (2): 567-577.

Patel, Parth A, Rhea Gopali, Anvith Reddy, and Kajol K Patel. 2021. "Trends in soccer-related ocular injuries within the United States from 2010 through 2019." Seminars in ophthalmology.

Pathan, MV, S Patsias, JA Rongong, and VL Tagarielli. 2017. "Measurements and predictions of the viscoelastic properties of a composite lamina and their sensitivity to temperature and frequency." *Composites Science and Technology* 149: 207-219.

Pivonka, Peter. 2018. *Multiscale mechanobiology of bone remodeling and adaptation*. Springer.

Ponnusami, Sathiskumar Anusuya. 2021. "From microstructural images to properties-an interpretable deep learning approach to predict elastic-plastic properties of fiber composites."

Price, DS, R Jones, and AR Harland. 2006. "Computational modelling of manually stitched soccer balls." *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications* 220 (4): 259-268. https://doi.org/https://doi.org/10.1243%2F14644207JMDA83.

Rangarajan, Nagarajan, Sarath B Kamalakkannan, Vikas Hasija, Tariq Shams, Carole Jenny, Irina Serbanescu, Jamie Ho, Matthew Rusinek, and Alex V Levin. 2009. "Finite element model of ocular injury in abusive head trauma." *Journal of American Association for Pediatric Ophthalmology and Strabismus* 13 (4): 364-369. https://doi.org/https://doi.org/10.1016/j.jaapos.2008.11.006.

Razaghi, Reza, Hasan Biglari, and Alireza Karimi. 2020. "Finite element modeling of the eyeglass-related traumatic ocular injuries due to high explosive detonation." *Engineering Failure Analysis* 117: 104835. https://doi.org/https://doi.org/10.1016/j.engfailanal.2020.104835.

Reddy, Junuthula Narasimha. 2019. *Introduction to the finite element method*. McGraw-Hill Education.

Reed, William F, Kenneth W Feldman, Avery H Weiss, and Alan F Tencer. 2002. "Does soccer ball heading cause retinal bleeding?" *Archives of pediatrics & adolescent medicine* 156 (4): 337-340. https://doi.org/https://doi.org/10.1001/archpedi.156.4.337.

Rong, Qingyuan, Han Wei, Xingyi Huang, and Hua Bao. 2019. "Predicting the effective thermal conductivity of composites from cross sections images using deep learning methods." *Composites Science and Technology* 184: 107861.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18.

Saffioti, Jami Marie. 2014. "CHARACTERIZATION OF PEDIATRIC OCULAR." The University of Utah.

Salmenjoki, Henri, Mikko J Alava, and Lasse Laurson. 2018. "Machine learning plastic deformation of crystals." *Nature communications* 9 (1): 1-7.

Sarker, Bapi, Jasmin Hum, Showan N Nazhat, and Aldo R Boccaccini. 2015. "Combining collagen and bioactive glasses for bone tissue engineering: a review." *Advanced healthcare materials* 4 (2): 176-194.

Saxena, Sandeep, Subhadra Jalali, Lalit Verma, and Avinash Pathengay. 2003. "Management of vitreous haemorrhage." *Indian Journal of Ophthalmology* 51 (2): 189-196.

Schmidt-Erfurth, Ursula, Amir Sadeghipour, Bianca S Gerendas, Sebastian M Waldstein, and Hrvoje Bogunović. 2018. "Artificial intelligence in retina." *Progress in retinal and eye research* 67: 1-29.

Shim, Vickie B, Samantha Holdsworth, Allen A Champagne, Nicole S Coverdale, Douglas J Cook, Tae-Rin Lee, Alan D Wang, Shaofan Li, and Justin W Fernandez. 2020. "Rapid prediction of brain injury pattern in mTBI by combining FE analysis with a machine-learning based approach." *IEEE Access* 8: 179457-179465.

Shoba, SJ Grace, and A Brintha Therese. 2020. "Detection of glaucoma disease in fundus images based on morphological operation and finite element method." *Biomedical Signal Processing and Control* 62: 101986.

Shokrollahi, Yasin, Pengfei Dong, Peshala T Gamage, Nashaita Patrawalla, Vipuil Kishore, Hozhabr Mozafari, and Linxia Gu. 2022. "Finite Element-Based Machine Learning Model for Predicting the Mechanical Properties of Composite Hydrogels." *Applied Sciences* 12 (21): 10835.

Shokrollahi, Yasin, Pengfei Dong, Mehmet Kaya, Donny W Suh, and Linxia Gu. 2022. "Rapid Prediction of Retina Stress and Strain Patterns in Soccer-Related Ocular Injury: Integrating Finite Element Analysis with Machine Learning Approach." *Diagnostics* 12 (7): 1530.

Shokrollahi, Yasin, Pengfei Dong, Matthew Lam, Donny W Suh, and Linxia Gu. 2022a. "Eye Protection for Mitigating Soccer-Related Ocular Injuries: A Finite Element Approach." *Journal of Engineering and Science in Medical Diagnostics and Therapy* 5 (4): 041003.

Shokrollahi, Yasin, Pengfei Dong, Matthew Lam, Donny Won Suh, and Linxia Gu. 2022b. "Eye Protection for Mitigating Soccer Related Ocular Injuries: A Finite Element Approach." *Journal of Engineering and Science in Medical Diagnostics and Therapy*.

Song, Helen H, Wallace B Thoreson, Pengfei Dong, Yasin Shokrollahi, Linxia Gu, and Donny W Suh. 2022. "Exploring the Vitreoretinal Interface: a Key Instigator of Unique Retinal Hemorrhage Patterns in Pediatric Head Trauma." *Korean Journal of Ophthalmology*.

Sousa, Thais, Nilabh Kajave, Pengfei Dong, Linxia Gu, Stephanie Florczyk, and Vipuil Kishore. 2021. "Optimization of Freeze-FRESH Methodology for 3D Printing of Microporous Collagen Constructs." *3D Printing and Additive Manufacturing*.

Srivastava, Ankesh Kumar, Ram Pyare, and SP Singh. 2012. "In vitro bioactivity and physical–mechanical properties of MnO2 substituted 45S5 bioactive glasses and glass-ceramics." *Journal of Biomaterials and Tissue Engineering* 2 (3): 249-258.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15 (1): 1929-1958.

Stephan, CN. 2013. "Facial approximation."

Stitzel, Joel D, Stefan M Duma, Joseph M Cormier, and Ian P Herring. 2002. *A nonlinear finite element model of the eye with experimental validation for the prediction of globe rupture*. SAE Technical Paper.

Suh, Donny W, Helen H Song, Hozhabr Mozafari, and Wallace B Thoreson. 2021. "Determining the Tractional Forces on Vitreoretinal Interface Using a Computer Simulation Model in Abusive Head Trauma." *American journal of ophthalmology* 223: 396-404. https://doi.org/https://doi.org/10.1016/j.ajo.2020.06.020.

Sun, Chin-Teh, and Rajesh S Vaidya. 1996. "Prediction of composite properties from a representative volume element." *Composites science and Technology* 56 (2): 171-179.

Swaminathan, Shriram, Somnath Ghosh, and NJ Pagano. 2006. "Statistically equivalent representative volume elements for unidirectional composite microstructures: Part I-Without damage." *Journal of Composite materials* 40 (7): 583-604.

Systèmes, Dassault. 2013. Abaqus 6.13 Documentation Collection; Dassault Systèmes Simulia Corp., Providence, RI, USA.

Tong, Junfei, Sachin Kedar, Deepta Ghate, and Linxia Gu. 2019. "Indirect traumatic optic neuropathy induced by primary blast: a fluid–structure interaction study." *Journal of Biomechanical Engineering* 141 (10). https://doi.org/https://doi.org/10.1115/1.4043668.

Vinger, Paul F, and JA Capao Filipe. 2004. "The mechanism and prevention of soccer eye injuries." *British journal of ophthalmology* 88 (2): 167-168. https://doi.org/https://doi.org/10.1136/bjo.2003.026229.

Wang, Yanqin, Yanan Xue, Jinghui Wang, Yaping Zhu, Yu Zhu, Xuehui Zhang, Jingwen Liao, Xiaona Li, Xiaogang Wu, and Yi-Xian Qin. 2019. "A composite hydrogel with high mechanical strength, fluorescence, and degradable behavior for bone tissue engineering." *Polymers* 11 (7): 1112.

Weaver, Ashley A, Eric A Kennedy, Stefan M Duma, and Joel D Stitzel. 2011. "Evaluation of different projectiles in matched experimental eye impact simulations." *Journal of biomechanical engineering* 133 (3).

Weber, Christian, and Heidi Noels. 2011. "Atherosclerosis: current pathogenesis and therapeutic options." *Nature medicine* 17 (11): 1410-1422.

Wei, Han, Shuaishuai Zhao, Qingyuan Rong, and Hua Bao. 2018. "Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods." *International Journal of Heat and Mass Transfer* 127: 908-916.

Weiss, Karl, Taghi M Khoshgoftaar, and DingDing Wang. 2016. "A survey of transfer learning." *Journal of Big data* 3 (1): 1-40.

Williams, Christopher KI. 1998. "Prediction with Gaussian processes: From linear regression to linear prediction and beyond." *Learning in graphical models*: 599-621.

Wollensak, Gregor, and Eberhard Spoerl. 2004. "Biomechanical characteristics of retina." *Retina* 24 (6): 967-970. https://doi.org/http://dx.doi.org/10.1097/00006982-200412000-00021.

Yamazaki, Junpei, Makoto Yoshida, and Hiroshi Mizunuma. 2014. "Experimental analyses of the retinal and subretinal haemorrhages accompanied by shaken baby syndrome/abusive head trauma using a dummy doll." *Injury* 45 (8): 1196-1206. https://doi.org/https://doi.org/10.1016/j.injury.2014.04.014.

Yan, Hua. 2020. *Sports-related Eye Injuries*. Springer Singapore.

Yang, Charles, Youngsoo Kim, Seunghwa Ryu, and Grace X Gu. 2020. "Prediction of composite microstructure stress-strain curves using convolutional neural networks." *Materials & Design* 189: 108509.

Yang, Chung-Chia, and R Huang. 1996. "Double inclusion model for approximate elastic moduli of concrete material." *Cement and Concrete Research* 26 (1): 83-91.

Yang, Zhenze, Chi-Hua Yu, and Markus J Buehler. 2021. "Deep learning model to predict complex stress and strain fields in hierarchical composites." *Science Advances* 7 (15): eabd7416.

Ye, Sang, Wei-Zhi Huang, Min Li, and Xi-Qiao Feng. 2021. "Deep learning method for determining the surface elastic moduli of microstructured solids." *Extreme Mechanics Letters* 44: 101226.

Ye, Sang, Bo Li, Qunyang Li, Hong-Ping Zhao, and Xi-Qiao Feng. 2019. "Deep neural network method for predicting the mechanical properties of composites." *Applied Physics Letters* 115 (16): 161901.

Zhou, X-Y, PD Gosling, CJ Pearce, Z Ullah, and L Kaczmarczyk. 2016. "Perturbation-based stochastic multi-scale computational homogenization method for woven textile composites." *International Journal of Solids and Structures* 80: 368-380.

Ziletti, Angelo, Devinder Kumar, Matthias Scheffler, and Luca M Ghiringhelli. 2018. "Insightful classification of crystal structures using deep learning." *Nature communications* 9 (1): 1-10.