HETEROGENEOUS FEATURE INTEGRATION FOR REGRESSION IN MULTIMODAL HEALTHCARE APPLICATIONS

by

Maryam Tavakoli Hosseinabadi

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2020

Approved by:

____

Dr. Yaorong Ge

____

Dr. Mirsad Hadzikadic

____

Dr. Reza Mousavi

____

Dr. Michael Dulin

____

Dr. Wlodek Zadrozny

DEDICATION

To my dear parents, brother and sister, Mohammad Bagher, Zahra, Ahmadreza, and Fatemeh.

ACKNOWLEDGEMENTS

I would like to express the most profound appreciation and gratitude to my supervisors Dr. Mirsad Hadžikadić and Dr. Yaorong Ge, for their invaluable support and patient leadership during my Ph.D. study. Special thanks to my committee members Dr. Wlodek Zadrozny, Dr. Reza Mousavi, Dr. Micheal Dulin, for their thoughtful and constructive suggestions and feedback. I also want to thank Dr. Samira Shaikh and Dr. Lisa Russell-Pinson for their insightful guidance whenever I needed it. I would also like to thank Sandra Krause, Julie Fulton, and Dillon Clausner, who always went beyond their responsibilities to assist everyone in every possible way; they are indeed role models.

This work would not have been possible without the generous support of the National Institute of Health (NIH), College of Computing and Informatics (CCI), Data Science Initiatives (DSI), and the Graduate Assistant Support Plan (GASP) award from the graduate school at UNC Charlotte. I want to appreciate them for supporting my study.

The five years of Ph.D. study introduced me to the brightest friends at UNC Charlotte, whom I am grateful for having them. You made this journey sweet and lovable. I would like to sincerely thank my friends at the Institute for Social Capital (ISC), Complex Systems Institute (CSI), Health Informatics (HI) lab. Having you by my side made these years sweet and memorable. David, Amy, Ashley, and Justin, thank you for your kindness, spirit, and care when I was very new to everything in this country. Tianyi, Jingyi, Yi, Jody, Chuqin, Yasaman, Narges, and Jessica, thank you for being by my side, for our insightful discussions, and for lots of fun in our time at HI and CSI labs.

Zara, Arezou, Maryam A., Elahe, Kobra, Sara L., Farzaneh S., Mahana, Fatemeh H., Haleh, Gizem, Nasrin, Elizabeth, Mariah, Lida, Malak, Rob, Riyi, Louai, Amir, Maryam K., Saeed, Yasna, Farzane, Mehrdad, Soroush, Hossein, Shervin, Samaneh,

Milad, Farzad, Yousra, and Sarah A., we shared our time through happiness, sadness, joy, and stress together. You know how much your presence in my everyday life means to me; far or close, you hold a special place in my heart.

I am forever grateful for the endless love and support of my parents and my brother and sister. Your thoughts, courage, strength, kindness, and love inspire me in every step of my life and motivate me to overcome any obstacle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The healthcare decisions and predictions are based on various data modalities with heterogeneous levels of description [1]. This semantic heterogeneity demands a task-specific representation, with different granularity and abstraction levels, for each modality of data [2]. In this context, integration of information with a semantically [3], quantitatively [4, 5], and computationally [6] reasonable approach is a challenging task. In this dissertation, we are proposing a domain-knowledge-driven pipeline for effective feature integration in the healthcare domain.

1.1 Research Motivation

The modality representation is determined by the features extracted from that modality of data for a particular task. In fact, it is common in the literature to use terms "representation" and "feature" interchangeably [7]. Designing an effective representation —i.e., extracting appropriate features, involves both data-driven feature extraction and domain-knowledge-driven feature engineering tasks [8].

The data-driven approaches extract features in a bottom-up fashion from the lowest abstraction-level of data —i.e., raw data. On the other, the knowledge-driven methods are based on hypotheses and experiments and eventually synthesizing the data. In other words, these features are built in a top-down fashion, starting from a feature with the highest level of abstraction [9]. While the former features have higher freedom of information extraction, it is easier for the human mind to memorize, interpret, validate, and feedback on the latter group of features [6].

In the healthcare domain, the decisions are made by a complex combination of high-level information and low-level data-driven findings from multiple modalities of data.

For example, in radiation therapy planning, the physician considers the patient a high-level physiological condition for prescribing the dose-at-target. However, the planner team tries the feasible plans based on the patient's low-level anatomical features to find the best feasible plan [10–12]. This example clearly shows how multiple modalities of data provide information on different levels of abstraction.

The multimodal machine learning has substantial literature for the integration of modalities with similar abstraction-level. The conventional methods include high-level feature integration in classic multivariate machine learning [13] and low-level feature fusion methods in deep learning models [14, 15]. However, the integration of features with heterogeneous representation and dissimilar abstraction-level is a long-time open question [7, 9].

The importance of learning from modalities with both low-level —i.e., data-driven— and high-level —i.e., hand-crafted— features in the healthcare domain and the complex inter-relation of these features among themselves and with respect to the target task motivated this research to study the integration of heterogeneous features systematically. We are specifically interested in studying this problem in healthcare and addressing the role of *domain knowledge* for designing an effective approach.

1.2 Research Scope

Many tasks in the healthcare domain have a regression nature [16–18]. However, the majority of research studies concentrated on tasks with a classification nature. This tendency is observable in prior research for scalar — e.g., disease diagnosis [19–24] — and high dimensional tasks — e.g., segmentation [25–27]. This disparity creates a research gap for tasks with a regression nature in this domain. This dissertation provides a framework for regression applications to address this gap in the healthcare domain.

For a regression task, we define a hypothetical framework with two types of modalities: *primary modality* with fine-grain information, which we call *modality A*, and an *auxiliary modality* which provides supporting information and is called *modality B* in this dissertation.

1.3 Research Questions

In the research scope mentioned in Section 1.2, we want to study the role of domain-knowledge in providing the best integration approach in multimodal healthcare applications. In particular, we want to answer the following research questions in a systematic way:

RQ1: , An increasingly common representation for *modality A*, is using the raw data — e.g., image, and letting a deep learning model learn the best representation. Is this **blind representation learning** for primary modality always results in the most accurate and generalizable performance?

RQ2: The most common approach for the integration of *modality B* is using multimodal machine learning with various levels of fusion. Are these **blind fusion/integration** always effective in improving model accuracy over single modality learning in all the health care applications with heterogeneous data?

RQ3: How can domain knowledge help us determine the best model architecture? Particular questions include (1) proper identification of multiple modalities; (2) best architecture for the primary modality(s); (3) the different ways the information in supporting modalities (e.g., modality B) may contribute to the accuracy of integrated models; and (4) the best architecture for integrating supporting modalities.

In addition to answering the above questions based on literature and in the context of real healthcare applications, we also study the impact of using domain-knowledge in the two applications of radiation treatment planning and disease progression prediction. With these experiments, we aim to answer the following research questions:

RQ4: Can domain-driven representation improve model accuracy and generalizability over blind representation methods — i.e., high-level and low-level, in predicting dose distribution for radiation treatment planning?

RQ5: Can domain-driven multimodal machine learning methods improve model accuracy over single modality one in predicting disease progression using ADNI data?

1.4 Research Method

We systematically review the researches that fit the scope we mentioned in Section 1.2 and analyze their learning methods and data modalities. We formulate the problem in a framework with two hypothetical modalities and discuss the possible solutions based on the first two questions based on the evidence from the literature. We further propose a pipeline in which we use the domain-knowledge to represent the modalities and integrate the information.

We use a domain-driven convolution neural network to extract fine-grained features from primary modality with an efficient abstraction-level. We also propose a tree-structure convolution neural network along this pipeline, a novel perspective for integrating the multimodal information and getting a more accurate result than the blind fusion architectures in some applications.

The pipeline is then applied to two healthcare applications for radiation treatment planning and disease progression prediction to evaluate the methods. We clarify the pipeline and validate the proposed methods by studying the implication in a scalar regression task in Alzheimer's Disease (AD) progression prediction domain [28], and a high dimension regression task for knowledge-based treatment planning (KBP) [10,29]. The results are validated by comparing the proposed approaches with the previously reported results.

1.5 Structure of the dissertation

In the following chapters of the dissertation, we first provide some background on multimodality and deep learning architectures in Chapter 3 and systematically review the recent publications in multimodal deep learning. In Chapter 4, we describe the proposed framework. We study the domain-driven representation in a framework with a 2D regression prediction for radiation treatment planning in Chapter 5, and we analyze the multimodal learning and proposed methods for the case of Alzheimer's Disease progression prediction using ADNI data in Chapter 6. Finally, in Chapter 7, we summarize the contributions and limitations of the current dissertation. We further discuss the potential direction of this work in the future.

CHAPTER 2: Multimodality and Feature Integration

Human understanding of the environment has always been multimodal. This means that a person relies on various sources of information for experience and interpreting an event. This information comes from visual, auditory, and olfactory senses and the prior stored knowledge in one's mind. This is an important note in real machine learning applications, where often similar features should be injected into the algorithm for a human-comparable performance. This chapter briefly reviews some of the previous machine learning strategies for handling multimodal features instead of single-modal ones.

2.1 Multimodality

Modality as a term is defined as how something is experienced [30] or the type of representation format in which information is stored [31]. Image, text, and audio are among the common modalities in machine learning tasks. Although each of these modalities has distinct statistical properties, the higher-level distributions of various modalities have some form of correlation when they represent a shared task. This situation is what we call multimodality.

A multimodal distribution in statistics is a distribution with multiple modes, which appears as multiple distinct peaks or local maxima in the probability density function. This distribution often is considered to be a combination of multiple uni-modal distributions. Therefore, typical summary statistics such as the mean, median, and standard deviation can be misleading and inappropriate descriptive measurements. Instead, the characteristics of each uni-modal distribution and the between-modals relationship offer better measures to describe the distribution. While a real-life multimodal problem has a far more complex distribution than this definition, for having an abstract imagination, a multimodal event can be perceived as a product of an unknown interaction of multiple unimodal events.

Datasets in real-life applications have various heterogeneity types, which is referred to as any in-homogeneity in the data [1]. Some examples in the literature are label heterogeneity, which can result from multiple inhomogeneous instances or label providers, data distribution or task heterogeneity, and feature source or view heterogeneity [1]. Multimodality falls under the category of feature source heterogeneity.

2.2 Multimodal Machine Learning

The multimodal machine learning area extends the potential of machine learning to take advantage of heterogeneous sources of information [1]. However, it introduces numerous novel areas and challenges.

A recent survey paper [30] addresses five core challenges in multi-modal machine learning applications, including *representation*, *translation*, *alignment*, *fusion*, and *co-learning*. Each application of machine learning may deal with one or a couple of these categories, as they demonstrated some examples in Table 1 of the paper [30]. A subset of these challenges was also previously mentioned in the multiview machine learning area [32, 33], which makes the two areas inter-related.

A major concentration of prior studies is on the modalities with high-dimensional or sparse encoding/representation such as image, audio, video, or text [34, 35]. However, the importance of low-dimensional modalities and the modal heterogeneity from the semantic axis, granularity, and dimension perspective is rarely addressed in the literature. This is despite a wide range of prior applied studies and application domains [19, 36]. One of the main categories of these applications is incorporating prior domain knowledge into learning from raw signals [1, 37, 38]. We discuss this aspect of modality representation in the following section.

2.3 Modality Representation

In this section, we review the modality dimensional and sparsity characteristics. Modality representation in a machine learning task is directly relevant to the features extracted from that modality, and these terms are sometimes used [7] interchangeably. We discuss multiple dimensions and abstracted related characteristics of modalities in the following subsections.

2.3.1 Knowledge-Driven vs. Data-Driven

Classic machine learning provides a set of tools for feature extraction, dimension reduction, and feature selection. These toolsets are a complimentary guide for knowledge-based feature engineering to extract a few features that can efficiently capture a target's information. According to domain knowledge, these features used to be hand-crafted and validated in a statistically approved approach.

A common characteristic of classic features is their coarse-grain description domain and high-level of information representation. The shortcomings appear when a task needs more local or more fine-grained information.

Deep learning methods and particularly convolution neural networks, unlike the hand-crafted features, learn features in a bottom-up fashion. This means that hierarchical levels of dimension reduction learn the local information much more efficiently than the typical classic approaches. This structure made a breakthrough in tasks such as image classification [39].

This learning model uses the available data resource efficiently for the structural-based situation through hierarchical weight sharing and the innate convolution regularization. Nevertheless, this efficiency causes unintended bias when a higher-level characteristic such as shape, depth, or distance is what the target function needs [40].

Furthermore, even with the efficient use of data similar to other non-parametric approaches, the predictions' quality still depends on the availability of reasonable variation and amount of data [6, 25].

2.3.2 Representation Abstraction

The term abstraction is used for an operation which hides or removes the less critical detail and preserve the desirable properties [41]. The desirability in machine learning is typically calculated with respect to the target function [42]. In other words, the abstraction shows the density of the expected *knowledge* in a particular representation [42, 43]. This means that a higher abstraction compressed the same knowledge in a lower dimension of representation.

Abstraction can be thought as a *feature* or *category* (such as car vs bike) or a *discrete* or *continuous* function of sensory data, such as past tense of a sentence or speed of an object in a video [44]. A feature with high-level of abstraction has a coarse-grain description domain, while a low-level feature potentially provides finer-grain information.

2.3.3 Representation Granularity and Sparsity

Representation is the format in which information is extracted from the source and encoded for the machine. Thus, it characterizes the type of features and the potential computational algorithms this information can be fed to. A *multi-modal* representation should also have the potential to integrate the information from multiple heterogeneous sources in a meaningful way.

We discussed in Section 2.3.2 that the information can be represented in various levels of abstraction. The immediate impact of leveraging abstraction of representation - i.e., lower-level representation to a higher-level one, is usually a dimension reduction. As a result, each element of representation holds more information. This causes a denser - i.e., less sparse representation. As a result, there is a common correlation between abstraction increase, dimension reduction, and a feature's coarser granularity. However, it is crucial to note that these are not equal concepts.

The features' sparsity is measured through the ratio of zero or near-zero values compared to the numeric volume of a representation. Literature suggested various metrics for the sparsity measurement. A comparison of these measures suggested that *Gini-sparsity-index* is the best among all with respect to six criteria [45]. Equation 2.1, demonstrates the formula for this metric. In this equation, k is the index of each feature-item when their values are sorted in an ascending form, $||c||_1$ is the sum of absolute values, and N is the number of feature-items.

$$S = 1 - 2\sum_{k=1}^{N} \frac{c_{(k)}}{||c||_1}(\frac{N - k + \frac{1}{2}}{N}), c_{(1)} \leq c_{(2)} \leq ... \leq c_{(N)} \tag{2.1}$$

The dimensional aspect of the representation and features are commonly discussed either in a high-dimensional sparse domain or in a low-dimensional dense projection. The feature dimensionality is measurable by the average size or the number of items in one feature/modality. For example MNIST dataset [46] contains image-pixels of size 28x28, which means each instance of the dataset contains 784 numbers. This is an example of a small image, yet it is a high-dimensional feature. It can be compared to Iris dataset [1] with only four pieces of information, i.e., sepal and petal width and length, that is saved in four integer numbers.

Apart from the superficial appearance of dimensionality, there is a slightly deeper difference between MNIST and Iris datasets. This difference is about the density and sparsity of the feature vectors. For example, each instance of MNIST train dataset on average contains only 19.12 % nonzero value, while Iris data has four nonzero values, with semi-uniform distributions between 0.1 to 7.9. This attribute of a feature or modality is called density.

[1] https://archive.ics.uci.edu/ml/datasets/iris

2.4 Deep Learning Representation

The hierarchical architecture of deep learning models and a general compression tendency in these networks [47] makes the concept of abstraction critical in the deep learning representations and feature extraction. The ideal expectation from deep learning is an auto-learning of the multi-level abstraction similar to the human [42].

The real hierarchical structure in deep learning usually decreases the granularity and dimensionality. However, after measuring correlation [42] or mutual information [48] between the target class and network layers [42], it has been shown that the representation density and abstraction, unlike the expectation, does not monotonically increase in this hierarchy [42]. Furthermore, even if the abstraction increases with respect to the train data, it is not always in a generalizable and meaningful hierarchy like human [41, 49].

In classic machine learning, it was common to extract dense low-dimensional features from all modalities. These features typically had a coarse-grained description of the status of that modality. The compressed nature of these representations, along with their applicability for the classic machine learning to this type of representation, is part of the reasons that a large portion of our data-driven prior domain knowledge is stored and used in this format [50].

Deep learning models' emergence with the current high-power computation and storage systems transformed the dialogues back into sparse representations. The idea is that in an ideal unlimited situation, a deep learning type of architecture can capture the information of the sparse modalities that are relevant to the target. However, in the absence of sufficient resources, we may need to constrain these assumptions.

The fundamental quantitative and semantic differences between these two representations are important in selecting the most efficient machine learning and deep learning approach. Deep learning, as one of the most popular machine-learning approaches in many areas, provides the capability of hierarchical levels of abstraction for the modality representation [39].

2.5 Modality Representation Quality

The existence of different representation approaches and various levels of abstraction, discussed in previous subsections, makes it critical to have an assessment metric. Some studies use the performance of a learning task to evaluate the representation quality for a particular learning model [51]; other studies directly analyze the representation-target relation to conduct this assessment [48, 52].

One direct method for quality assessment of the representations is mutual information (MI). Specifically, a good representation is the one that maximizes the mutual information between inputs and desirable targets. Formally, it is defined based on Shannon entropy as stated in Equation 2.2, where H is the Shannon entropy, and $H(X|Y)$ is the conditional entropy of Z given X. The calculation for high-dimensional and continuous feature space is complicated.

$$I(X;Y) = H(X) - H(X|Z) \tag{2.2}$$

In order to extend the concepts of entropy to a continuous feature space, it is common to use the target *standard deviation*. The problem with these measurements is that they only assess the quality with respect to global statistics.

There are more complex neural-network-based definitions such as Mutual Information Neural Estimation (MINE) [48] that measures more locally the amount of knowledge being retained in a neural-network representation. Deep InfoMax (DIM) method [53] demonstrates that in their assessment, local information produces better representation for classification, whereas generation tasks need more global information in representation to perform well.

The indirect evaluation of a particular representation is through comparing the model performance using that representation compared to a common representation method. In these evaluation methods, it is essential to 1) keep the other parameters and aspects of the model consistent [51]. 2) study the behavior of the representation upon changing other variables. 3) study the representation's generalizability by comparing its performance on seen data vs. unseen data [41].

2.6 Summary

In this chapter, we provided background about the concept of multimodal machine learning. We discussed important aspects of modality representation consisting of the generation method, abstraction level, representation dimensionality, and sparsity. We further described the deep learning-based representation and the importance of abstraction levels in the corresponding representations. Finally, we discussed two common approaches for comparing the quality of two different modality representation. In Chapter 3, we analyze the modality representations in the literature from a dimensionality perspective. Further, we use the indirect method, explained in Section 2.5, for assessing the suggested representation in Chapters 4 to 6.

CHAPTER 3: Multimodal Deep Learning: A Systematic Review

In this chapter, we conduct a systematic review of the recent publications to understand the common abstraction-level and integration approach for multimodal tasks in the healthcare domain. We limit our review to the papers with a form of deep learning method for two primary reasons: 1) The deep learning architectures are diverse and capable of representing and learning both high-level and low-level features. Also, the hierarchical structure of most of the deep learning models has an explicit representation of abstraction-levels. 2) The popularity of the deep learning models and toolboxes created a trend in healthcare publications to use deep learning models in their applications.

3.1 Review Questions

Through this systematic review, we want to answer the following research questions according to the study cohort:

- What are the common modalities and representation-levels in multimodal healthcare applications?
- What are the common modality integration methods in multimodal applications in the healthcare domain?

These primary questions help us understand the context of the problem and the common methods currently used in this field. Following these questions, we want to examine whether these are the best approaches, and the common methods are always effective. We will discuss these questions in the last sections of this chapter and Chapter 4. We will further address the role of domain knowledge in choosing the best approach.

3.2 Review Method

We conduct a systematic review by looking for the papers containing "multimodal deep learning" in their *topics* - i.e., title, abstract, and keywords, which are indexed in the web of science [1] and published between 2017 to 2019. These documents are then sorted by Publication Year and Number of Citation (descending), respectively. We also include some useful contents from previous related review papers, which are not captured in our pipeline but contained related information [54–61].

The selected cohort are categorized into Review Papers and Application and Method manuscripts. The method typically is those with innovation from the multimodality perspective, while the applications are the ones that used the previous methods to solve a practical problem. Nevertheless, the method papers typically contain one or multiple application use-cases. The method papers are not representative of a common approach, and we kept them regardless of their application domain. However, we did not include them in the analysis for Subsection 3.3.2. The review papers are the ones that review the previous works. While a few of them are specifically on multimodal deep learning, the majority have other concentration subjects. Figure 3.1 demonstrates the review pipeline.

3.3 Review Results

The final cohort contains 105 papers consisting of 57 application papers in healthcare, 15 review papers with either general multimodal deep learning concentration or healthcare-related, and 33 method papers for multimodal deep learning with any domain of interests. We reviewed the cohort for each of these categories, the following subsections, separately. This review intends to understand the primary data modalities and some representative tasks in this domain.

[1] http://www.webofknowledge.com/

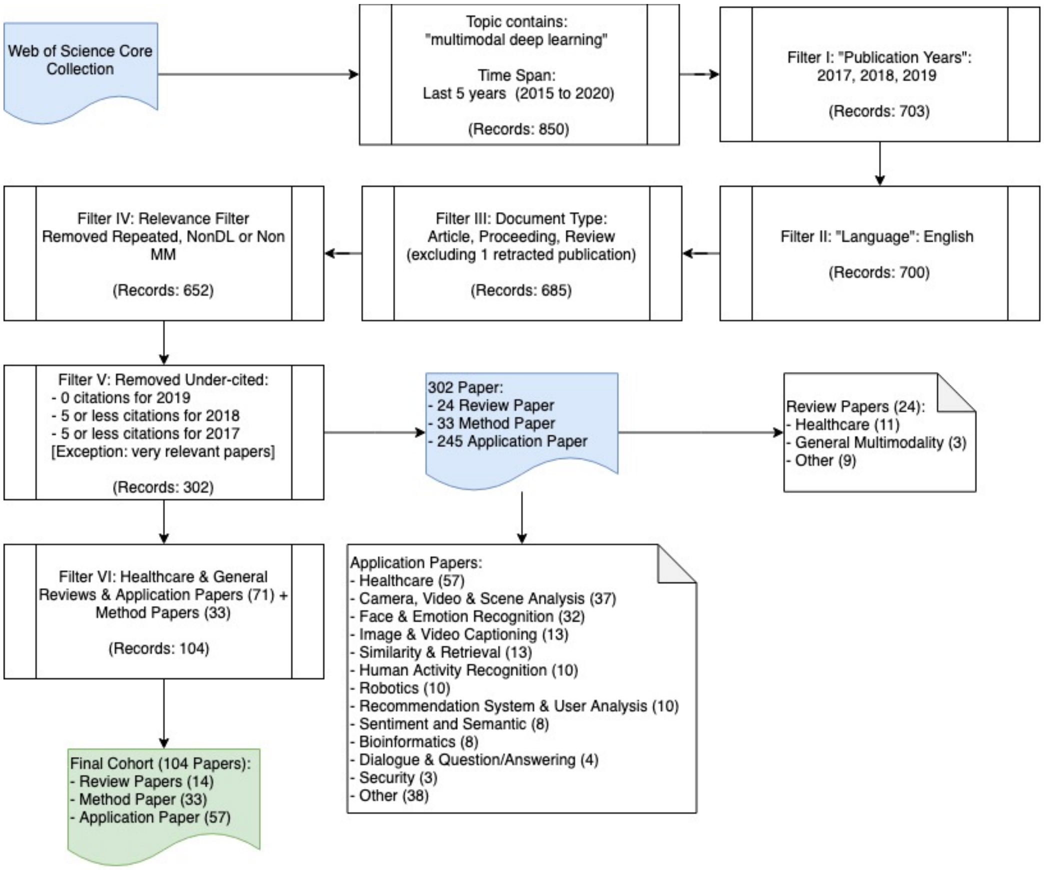

Web of Science Core Collection
Topic contains: "multimodal deep learning"
Time Span: Last 5 years (2015 to 2020)
(Records: 850)
Filter I: "Publication Years": 2017, 2018, 2019
(Records: 703)
Filter II: "Language": English
(Records: 700)
Filter III: Document Type: Article, Proceeding, Review (excluding 1 retracted publication)
(Records: 685)
Filter IV: Relevance Filter Removed Repeated, NonDL or Non MM
(Records: 652)
Filter V: Removed Under-cited:
- 0 citations for 2019
- 5 or less citations for 2018
- 5 or less citations for 2017
[Exception: very relevant papers]
(Records: 302)
302 Paper:
- 24 Review Paper
- 33 Method Paper
- 245 Application Paper
Review Papers (24):
- Healthcare (11)
- General Multimodality (3)
- Other (9)
Filter VI: Healthcare & General Reviews & Application Papers (71) + Method Papers (33)
(Records: 104)
Application Papers:
- Healthcare (57)
- Camera, Video & Scene Analysis (37)
- Face & Emotion Recognition (32)
- Image & Video Captioning (13)
- Similarity & Retrieval (13)
- Human Activity Recognition (10)
- Robotics (10)
- Recommendation System & User Analysis (10)
- Sentiment and Semantic (8)
- Bioinformatics (8)
- Dialogue & Question/Answering (4)
- Security (3)
- Other (38)
Final Cohort (104 Papers):
- Review Papers (14)
- Method Paper (33)
- Application Paper (57)

Figure 3.1: Overview of the method for systematic review of recent representative multimodal deep learning papers in healthcare. The search is done on Web Of Science platform, and the search entriy retrieval is done on April 22, 2020.

3.3.1 Survey Papers

Among the 302 manuscripts before being filtered-down to the healthcare domain, about 8%, i.e., 24, are review papers. This is a relatively high ratio, which is not unexpected, and it shows that multimodality and deep learning are a common concern in the recent review perspective in multiple domains.

As Figure 3.1 shows, these review studies come from various healthcare and biology domains, RGB-D and land-use image analysis, human activity recognition, emotion and facial recognition, and finally, mobile applications. Nevertheless, more than half of these studies belong to the healthcare and biological data. This can be interpreted as the significance of the multimodality and deep learning topic and concern in the recent healthcare publications, compared to the other areas.

The healthcare survey papers mainly review recent methods or deep learning approaches in specific health or medical domains. This can be disease-oriented [62–67], task-oriented [56, 68], or technology-oriented [55, 69]. While these papers are not specifically concern about multimodality, this is an inherent aspect of the data in this domain. Therefore most of them have either a short or a more extensive discussion on this perspective. Another group are modality oriented and discussing related approaches for processing one modality or coordination of multiple data modalities [55–58]. Finally, some surveys are devoted to deep learning [60] and multimodal [61] aspects of biological and medical data in general.

In addition to the domain-specific publications, two survey papers are dedicated to a more general perspective of multimodality in deep learning. The first one, published in 2017, overviews the popular datasets and applications in this domain [15]. These datasets and tasks are still prevalent in the application papers we review in Section 3.3.2. They further discussed the deep learning-based method and compared them with the conventional machine learning approaches [15].

The second one, [14], which is published more recently, discusses the representation aspect more in-depth. It classifies the current deep multimodal representation into three central frameworks of *joint representation*, *coordinated representation*, and *encoder-decoder*. They put recent multimodal approaches in this framework and analyze the advantage and disadvantages of each of them [14].

A relevant but broader study in this area [30], suggests a taxonomy for the topic of multimodal machine learning, which is consistent with the framework in prior surveys [14, 15]. The study puts the core challenges of this area into five folds of *representation*, *translation*, *alignment*, *fusion*, and *co-learning*. They discuss the conceptual frameworks and the conventional and deep learning methods for each of these challenges.

These reviews refer to the most significant challenges and approaches in the multimodal deep learning literature. However, the subject of feature semantic, dimension, and abstraction heterogeneity is not sufficiently discussed in any of them.

3.3.2 Application Papers: Modalities, Representation Levels, and Tasks

The level of abstraction has an inverse relationship with dimensionality. The dimensionality is a more straight-forward measure to quickly categorize the representation. While they don't provide a precise equivalent description of the abstraction-level, based on one we already discussed in the previous chapter - i.e., Section 2.3, dimensionality can be a quick and reasonable approximation for comparing representation abstraction across various researches.

As Figure 3.1 demonstrates, a large number of papers in the multimodal deep learning cohort are healthcare-related. We review these papers based on their data modalities and representative features, the type of machine learning task, and their general approaches.

The data modalities used in these manuscripts are classified in Table 3.1 based on the type of information they provide and the representation dimensionality. Each modality of data has various semantic aspects. However, the data representation is meant to reflect the most relevant semantic elements of the data to meet the computational and data availability constraints.

With the same approach, we also categorized the application papers according to their machine learning tasks. Table 3.2, demonstrates the task-oriented categorization of the application papers. The majority of publications analyzed their problem in the form of a classification task.

3.3.3 Application and Method Papers: Learning and Integration Methods

To analyze the methodology, we define a framework and will extend it in the next chapters with two hypothetical modalities; one has a primary role in the prediction

Table 3.1: The feature modalities used in the cohort are categorized based on the representation dimension and their temporal character for the defined task in the corresponding manuscript. The dimensionality of the representation compare to the raw-data demonstrates the abstraction-level of the corresponding representation.

	Contextual	Cross-Sectional	Temporal
0D	Demographics [70] Patient Meta-Data [21, 72]	Clinical [17, 19, 71]	Biomarkers [20] Hand-Crafted Handwriting Features [73]
1D	Social Features [74] Gene Expression [19, 71, 75] Copy Number Alteration [19] MicroRNA Expression [71]	Surface Linguistic Features [74] Hand-Crafted Image-Features [23, 76–79] EEG, fNIRS [82]	Text [70, 74] Hand-Crafted Image-Features [80, 81] EEG [83], sEMG [84] Finger Joints Angles [84] Motion Sensors [85] Audio [70, 73] Gait Signal [73]
2D		Image Slice [86–91] [18, 92–99] 2D Image [21, 72, 100] 2D Image Patches [22] Depth Image [92]	Video [70, 85] 2D Moving Images [101]
3D		3D Image [71, 102–105] [17, 24, 108–111] 3D image patches [112–114]	3D Image Transitions [106, 107]

task, which we call it *modality A*. This modality is sufficient for a reasonable prediction performance without extra information.

In this framework, we assume that we have an additional modality of information or another representation of the same modality that provides supplementary information to improve the original performance of modality A. This auxiliary modality is called, *modality B*, in this framework. The use of low and high dimension in this framework is relative. The homogeneously low-dimension features are the scalar or 1D vectors without neighborhood relation between the features closer together. The low/high heterogeneous features are categorized based on the relative differences of the dimension scale - i.e., scalar, 1D, 2D, etc., of the two modalities.

As we discussed in Section 3.2, to get a wider understanding of the common methods, we include the manuscripts from other areas if their methods are relevant to the framework we defined here. We also consider a few other manuscripts that were

Table 3.2: Tasks-based categorization of the application papers.

Category	Task	Papers
Classification	Scalar Classification	[19–21, 82, 90, 108, 115, 116] [22–24, 72, 77, 78, 83, 95, 106, 113] [74, 79, 80, 84, 85, 99, 109, 112]
	Retrieval & Similarity	[75, 92, 100]
	Segmentation	[86, 88, 89, 91, 102, 104] [87, 93, 105, 110, 111, 117] [81, 96–98, 107, 114]
Regression	Scalar Regression	[20, 70, 76]
	Generation	[18, 73, 108, 118]
	Registration	[89, 94, 101]
	Prognosis & Survival Analysis	[17, 71, 103]
	Retrieval & Similarity	[104]

relevant and did not appear in our systematic search. These typically were appeared in different forms of search-terms - e.g., "multi modal" instead of "multimodal" or before filtration based on year/citation.

Table 3.3 demonstrates examples of healthcare tasks, which fit our framework of a primary and auxiliary modality for a regression task. We further studied the approaches in each of these integration groups. Table 3.4, demonstrates some architectures that were used in these applications, and Table 3.5 shows the integration method that are popular in each integration group.

Table 3.3: Examples of applications, which fits the defined framework with modal A as the primary source of information and modal B as the auxiliary modality to provide supplementary information.

		Modal B (auxiliary)	
		Low Dimension	High Dimension
Modal A (primary)	Low Dimension	AD [20, 23, 77], Brain Age [76]	Survival Prediction [17]
	High Dimension	Skin [21, 72]	AD [106], Seg [17, 86]

3.4 Summary and Discussions

This review initially wanted to answer the two questions mentioned in Section 3.1. We demonstrated the results in Section 3.3. To summarize that, we review the two primary questions:

Table 3.4: Literature-based learning models for the framework with modal A as the primary source of information and modal B as the auxiliary modality to provide supplementary information.

Modal B (auxiliary)			
		Low Dimension	High Dimension
Modal A (primary)	Low Dimension	Recurrent Neural Network (RNN) [20] Deep Polynomial Network (DPN) [23] Restricted Boltzmann Machine (RBM) [77] Deep Neural Net (DNN) [76]	Random Forrest Regression [17]
	High Dimension	ConvNet [21,72]	ConvNet [86,107], W-Net [110], DenseNet [119]

Table 3.5: Literature-based integration methods for the framework with modal A as the primary source of information and modal B as the auxiliary modality to provide supplementary information.

Modal B (auxiliary)			
		Low Dimension	High Dimension
Modal A (primary)	Low Dimension	Bottom-Up: Early/Late [20,23,76], & Weighted Fusion [77]	Late Fusion [17]
	High Dimension	Top-Down: Loss-based [21,120] Bottom-Up: Early/Late [72,116]	Bottom-Up: Fusion Methods including: Early/Mid [86,107,119], Late [17,86,107,119] Decision [86,107,110,121] Top-Down: Attention-based [122], Adversary Method [18,94]

1. *What are the common modalities and representation-levels in multimodal healthcare applications?* Table 3.1 demonstrates the modality and dimensionality of the representations. We discussed that a dimensionality is a quick form of showing the abstraction level in our study.

2. *What are the common modality integration methods in multimodal applications in the healthcare domain?* Tables 3.4 and 3.5, demonstrate the architecture and integrated approaches in our framework, respectively. It is most common for the primary modality to be represented in high-dimension than the auxiliary.

Also, the diversity of studies and methods for the homogeneous features are more than heterogeneous ones.

Besides answering the primary questions, the review gave a better understanding of the healthcare domain, and its limitations and the importance of image modality, and particular characteristics of the medical image compare to regular images. We discuss about these topics in the Sections 3.4.1 and 3.4.2. We finally discuss a paragraph about the specific characteristic of regression tasks, in which we are more concentrated in this dissertation.

3.4.1 Healthcare domain

These application papers in the healthcare domain share some common themes and challenges that make them more relevant to this research area.

First and foremost, the features directly or indirectly hold information about the human body. The targets are better understanding human health aspects and assisting in a better prevention/treatment. Therefore, these application papers are related to a real-life problem, as opposed to machine learning hypothetical problems.

Secondly, the questions are complex and have multiple aspects, considerations, and constraints. Therefore, one modality or a few features are not sufficient to represent the case and learn the answers.

Thirdly, typical data sets are either small or very noisy as the data collection is a recent tendency in this domain. Furthermore, due to the privacy considerations and data-acquisition challenges, they are usually collected locally, which creates a natural bias in the data and label distribution.

Finally, unlike the scarcity of the data resource, there is a rich source of domain knowledge in this area, which is not fully reflected in the data. This makes the data-driven approaches hard to compete with the more traditional ones and hardly acceptable for the domain experts.

3.4.2 Medical image

Typically, features of interests are different in medical images than the RGB images of an object detection task. Firstly, concepts and objects of interest are more limited compare to regular object recognition.

Secondly, the geometrical features are much more important than the colors and intensities. Sometimes color and intensities are secondary tools for making the geometry more understandable for clinicians.

Thirdly, the required granularity of the image is completely task-dependent. This means that for one task such as Alzheimer's prediction, a coarse grain feature like region volume size is much more important than the details of regional changes, while in another task like brain vein segmentation, a much finer granularity is desirable.

Fourthly, granularity and importance vary across different regions. To some level, this happens in regular images, but it is more consistent for a particular context. In a medical image, the physician's intention, patient status, and other information outside of the image context have a central role in guiding the granularity level.

Finally and more importantly, unlike the typical image expected to be invariant with respect to orientation or scale change, the medical image is absolutely sensitive to these settings. By changing the scale or orientation, we may lose all the information we need from that image.

This is partly because the semantic origin of the geometrical measures depends on the purpose of imaging, and it is not the image origin. For example, in radiation treatment planning, the distance and measures matter with respect to the cancer area. Also, it is also constraint by the radiation beam source and bed placement.

This last characteristic mainly makes it hard to transfer the learned model between-patient and between-organizations. This is less problematic for the tasks involving entirely local features such as segmentation, where deep learning could achieve good performance (look at publications mentioned in Table 3.2). However, these models do not perform well for tasks that require higher-level geometrical features such as distances or volume.

3.4.3 Regression Task

The primary difference between regression and a classification task is in the target or dependent variable. There are a finite number of values for the prediction and an infinite number of possible separators to be captured by the classifier in a classification task. On the other hand, a regression task has an infinite number of possibilities for the target value.

An ideal regression model predicts a value that is as close to the actual target as possible. Therefore, a regression model needs to capture the relationship between the independent and the target variable more precisely than a classification model. In our review, we observed that the major portion of the multimodal deep learning models in healthcare and medical imaging [27, 123] are devoted to the classification tasks. This left a research gap for finer-grained target functions that have a regression nature.

3.5 Review Limitation

The systematic review in this section is definitely not exhaustive and could have missed a lot of important works in this domain. Some of these limitations are:

1. The web of science platform gives a great set of tools for a systematic review. However, the publications and indexes are not as up-to-date as other search engines such as google scholar. We reviewed many other manuscripts based on the google scholar results for researching this dissertation. However, the organized and easy to follow toolset in web-of-science makes it more reliable for this section's purpose. Still, not including the actual extensive set of publications in this area is part of the mentioned results' limitations.

2. The keyword search could be expanded. We searched multiple versions of the

keyword search, such as *multi-modal*, *deep-learning*, or other combinations of these phrases. However, we got the most relevant list of results using the current phrase "multimodal deep learning." While we tried to include the significantly relevant papers that did not appear in our search, we missed so many important works in this domain.

3. The concept of multimodality has various terminology in different domains. Therefore, the manuscripts that used this concept without mentioning the term is not included in this review. For example, multiview is a subclass of multimodal representation. While we reviewed some of the survey paper in this regard, the review process was not on those works.

4. The classic machine learning did also have a vast resource of multimodal-related works. Since we limited the search to the deep learning and recent approaches, we have not those works in our results.

5. We limited the scope to the healthcare domain for the application paper. We tried to briefly overview the other ones and keep them as a method paper if their approach is relevant. However, we still could miss some good works in this process.

CHAPTER 4: Feature Integration Framework

4.1 Learning models

In this section, we explain the two algorithms, which we used in designing the domain-driven learning pipeline. We describe model-tree, which is a form of decision-tree for a regression task and convolution neural network, which is a form of deep learning, which learns local filter-based patterns.

4.1.1 Model Tree

A decision tree is a simple data structure for categorizing the data into multiple branches. The discrete and nonlinear nature of this data structure makes it an important building block for several classic and modern learning algorithms.

In machine learning, a decision-tree is made in a top-down process of dividing the data into more homogeneous subsets with respect to the target. The homogeneity and heterogeneity are measured by multiple factors in various types of decision trees. Among those measuring heterogeneity by *entropy* for classification and *standard deviation* or variance for regression targets are the most popular ones.

In classification, the mutual information of feature A and target T is measured by the entropy decrements after branching on values of A. Equation 4.1 shows this formulation, where E_A is the expected information gain for feature A with branches a_i, $I(T, A)$ is the mutual information between target and feature A, $H(T)$ is the entropy of target before the branching, and $H(T|A)$ is the weighted sum of the branches' entropy.

$$E_A(IG(T, a_i)) = I(T, A) = H(T) - H(T|A) \tag{4.1}$$

The concept of entropy is replaced by a standard deviation or variance when the target function is continuous. Equation 4.2 demonstrates a typical formulation for a regression task using *Standard Deviation Reduction* (SDR) as a measure of homogeneity increase. The tree is created recursively, and a mean of the values in leaves are considered as the prediction of that branch. This is a simple form of regression tree, which is coded in the CART program [124].

$$\begin{aligned} SD(T|a_i) &= \sqrt{\frac{\sum(t-\bar{t})^2}{n}} \\ SD(T|A) &= \sum_{a_i \in A} P(a_i) SD(T|a_i) = \sum_{a_i \in A} \frac{|a_i|}{|T|} SD(T|a_i) \\ SDR(T,A) &= SD(T) - SD(T|A) \end{aligned} \tag{4.2}$$

A model-tree or m5 is a form of regression tree with an important improvement. In each leaf node, instead of value -i.e., the mean of data in that node, there is a linear model [124]. Despite the simplicity and limitation of the decision trees, they are very interpretable and easy to validate. The model trees are much more powerful for having the linear models in the tree leaves.

4.1.2 Convolution Neural Network

Convolution neural networks are one of the most successful machine learning methods, especially in the computer vision area. The architecture is inspired by the human visual system, particularly from the following two perspectives:

- The hierarchical architecture of perception and information processing.
- The ability to capture local patterns, which is similar to the receptive fields in the visual system.

If we compare a convolution neural network to a complete search with respect to the input features, there are multiple levels of regularization, which limits the search area. Each of these regularization methods is added to the architecture at a different time and made it more efficient for some applications:

1. The hierarchical structure, which limits the features' connection to a tree-shape form (similar to the decision tree).

2. The convolution weights, which changes the dependency of learned features from input size (ex. in an image each pixel one feature) to the smaller kernel size (ex. using k number of 3x3 kernels to convert any size of the input image into k features of the same or smaller size).

3. Pooling block (ex.max-pooling), which gradually reduces the dimension and resolution of the input, and keeps the most significant information.

4. Drop-out, which randomly turns on and off the connection to make the architecture robust to the small changes.

These regularization methods, along with other gradual refinements such as back-propagation, ReLU activation function, stochastic gradient-based optimizations, and normalization methods, made this architecture an important building block of many computer vision applications. One of the important features for us is its ability and performance in the extraction of local and spatial patterns. The Equation 4.3 demonstrates the forward path of a convolution neural network for on layer l with kernels k of initial weight W_k^l and bias b_k^l.

$$h_k^l = f(W_k^l * h^{l-1} + b_k^l) \tag{4.3}$$

4.2 Problem Framework

We define a hypothetical framework with two heterogeneous modalities. The assumption is that one modality provides the main information, and the other one holds high-level auxiliary information. This is similar to some of the applications we mentioned in the previous chapter [20, 21, 23, 72].

We call the primary modality A and the auxiliary one B. We also define a target regression curve R, which is a complex continuous curve for a healthcare regression task with a small data size.

4.2.1 Primary Modality

The primary modality, called A, has more than zero-dimension, i.e., it is not a scalar. This can be a one-dimensional vector or a two-dimensional matrix or higher. As discussed in Section 3.4, image is one of the important high-dimensional modalities in multimodal healthcare applications, so for simplicity in discussions, we consider this modality as an image type, but we provide examples from other modalities, as well.

$$A = (a_0, a_1, a_2, ...) \tag{4.4}$$

4.2.2 Auxiliary Modality

The second modality, called B, consists of one or multiple zero-dimension or scalar features. These features can be extracted from the same origin as A, or extracted from another high-dimensional modality or a piece of low-dimensional contextual information.

$$B = (b_0, b_1, b_2, ...) \tag{4.5}$$

4.2.3 Target Function

We formulate the regression tasks as in Equation 4.7, where $\hat{R}$ is the dependent variable of known variables in modalities A and B. The actual curve is a hypothetical function f in Equation 4.6, which depends on features of modalities A, B, and unknown Γ with small residual ϵ. An ideal regression model could minimize the distance of modeled $\hat{R}$ and actual values of R, which is called L for loss in Equation 4.8.

$$R = f(A, B, \Gamma) + \epsilon = (r_0, r_1, r_2, ...) \tag{4.6}$$

$$\hat{R} = \hat{f}(A, B) \tag{4.7}$$

$$Loss(R, \hat{R}) = dist(R, \hat{R}) \tag{4.8}$$

4.2.4 Framework Application

We reviewed the data types, applications, and approaches in multimodal healthcare applications in Chapter 3. Table 4.1 briefly mentions some of the applications, which fits the framework we defined in this section, and the corresponding representation extraction method. Particularly, we mention the spatial and temporal feature extraction method for primary modality.

Table 4.1: This table demonstrates some examples of the tasks in the literature, which fits the defined framework with corresponding modalities and the representation learning methods in those papers. We observe that the applications with smaller datasets are more careful about the features they use. Even using CNN is a tool for a more careful secondary analysis [17].

Task	Modal A	Modal B	A-Spat-Rep	A-Temp-Rep	Data-scale
Depression Scale Pred. [125]	Video	Demog.	CNN	RNN	M (671)
AD Prog. Pred. [20]	Med Image	Demog., CogTest	Handcrafted	RNN	M (1677)
Survival Pred. [17]	Med Image	Demog.	Handcrafted + CNN	-	S (285)
Prognosis Pred. [71]	Med Image Gene Exp.	Clinic.	CNN DNN	-	L (11000)

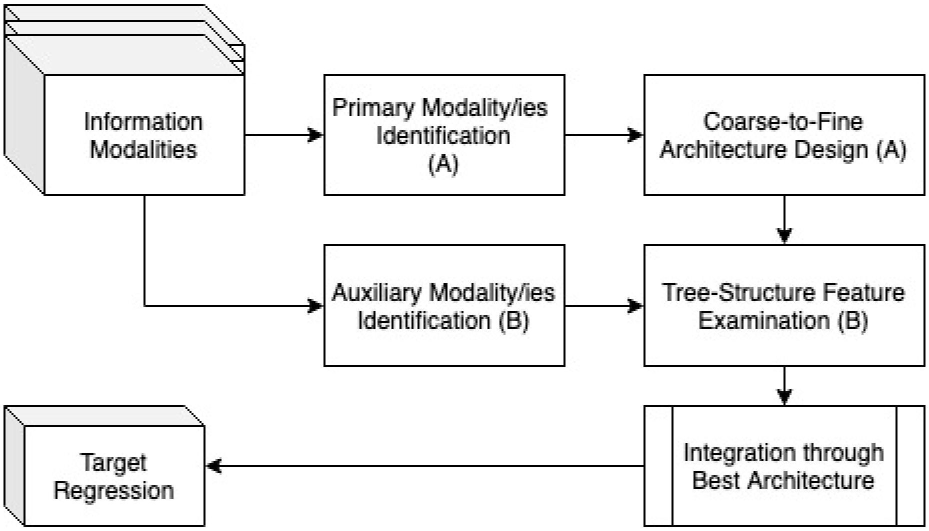

Information Modalities
Primary Modality/ies Identification (A)
Coarse-to-Fine Architecture Design (A)
Auxiliary Modality/ies Identification (B)
Tree-Structure Feature Examination (B)
Integration through Best Architecture
Target Regression

Figure 4.1: Proposed pipeline to used domain-knowledge for designing the multimodal information integration for a regression task.

4.3 Proposed Pipeline: Domain-driven Heterogeneous Feature Integration

There are different ways that the domain knowledge provides information to a machine learning task besides providing a resource of labeled data. A machine learning method, which is enriched by this prior knowledge, is referred to by an *informed* machine learning term [8]. To incorporate domain knowledge into the machine learning model, we need to answer three main questions [8,126]: (1) what is a good source for extracting the knowledge? (2) how to represent the knowledge? (3) how to integrate the knowledge with the learned model? We use the same set of questions to design a domain-driven multimodal machine learning model. Figure 4.1, demonstrate the domain-driven pipeline we designed.

The main building blocks of the proposed pipeline are listed below:

1. Primary Modality Identification: Selecting the primary modality/ies in healthcare applications needs a combination of domain-driven and data-driven parameters. The primary modality is the modality, which has both high and meaningful correlation with the target. Other factors selecting the modality is the quality of available data and the missing ratio. In our framework, the

primary modality has also another characteristics, which is the importance of local features.

   (a) Data-driven: relevance (correlation, co-variation), quality(missing-ratio).

   (b) Knowledge-driven: relevance (meaningfulness,causation analysis), and quality (precision of data acquisition).

2. Primary Modality Representation: We propose a domain-drive representation leverage along with convolution neural network for primary modality representation.

   (a) Data-driven: Convolutional Neural Network (Section 4.1.2.)

   (b) Knowledge-driven: domain-driven representation leverage (Section 4.3.2.)

3. Auxiliary Modality Identification:

   (a) Data-driven: relevance (correlation, co-variation), extra information (standard deviation reduction), quality(missing-ratio).

   (b) Knowledge-driven: relevance (meaningfulness,causation analysis), quality (precision of data acquisition).

4. Auxiliary Modality Representation: Depends on the availability of data, this modality can be also represented in different levels. In our framework, we assume that the highest level of abstraction (engineered features) provides sufficient information.

   (a) Data-driven: performance increase with different integration methods.

   (b) Knowledge-driven: the nature of information it provides about target function.

A common method in the literature is using various levels of fusion for combining the information from multiple modalities. We explain these methods in the form of a fusion framework in Section 4.3.1. We add two further approaches using a bottom-up and top-down knowledge incorporation methods in the Sections 4.3.2 and 4.3.3.

4.3.1 Literature-based Method: A Fusion Framework

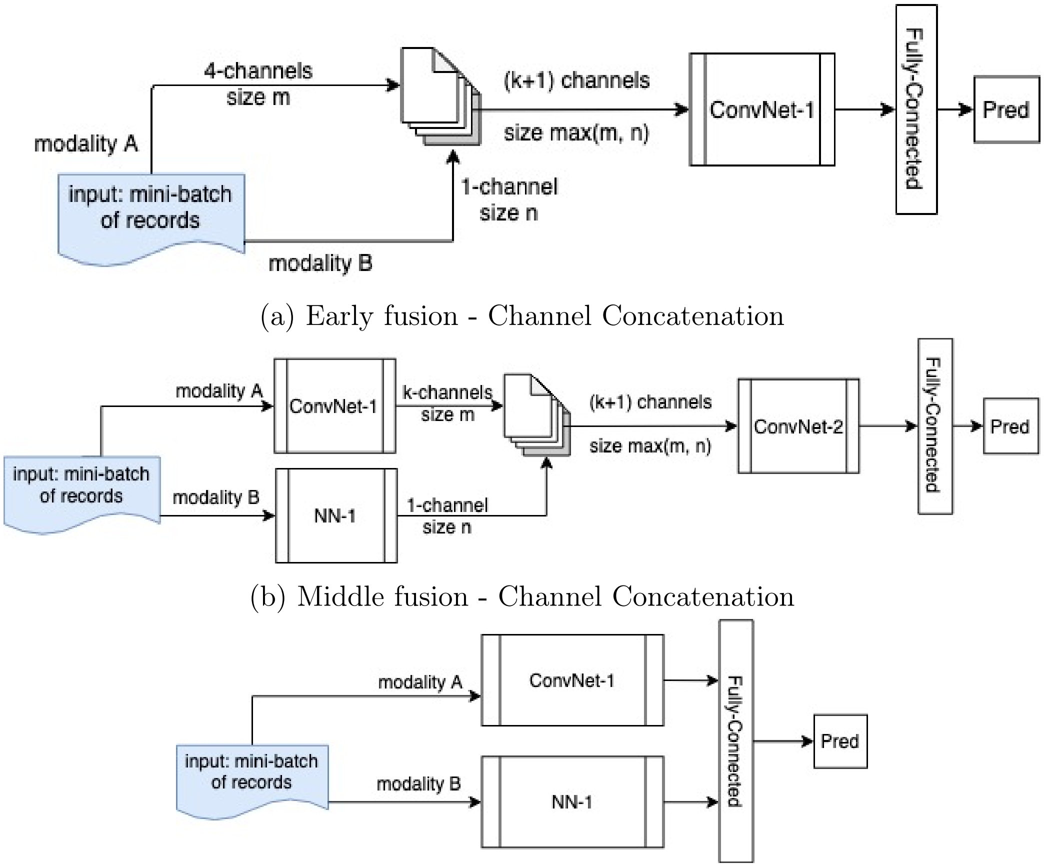
As we reviewed in Section 3, one of the most common approaches for information integration in multimodal applications is using fusion approaches. These approaches are usually based on blindly integrating all the features, including low and high levels, and let the network learn the relation between the features together and with the target. These methods are typically used with little or no discussion on the reasoning for the chosen level of representation and integration [17, 106].

A few works use multiple levels of fusion to find the best performance [86, 116]. We define a fusion framework as a systematic study of these explorations. Figure 4.2, demonstrates this literature-based framework, and Table 4.2 refers to some of the methods and corresponding applications in the literature.

Table 4.2: Fusion method and the corresponding applications according to the literature.

Method	Fusion level	Literature Applications
Feature level	Early	Representations have similar size, granularity & spatial semantic [86, 109, 127] [104, 119, 128]
Deep learning feature	Mid	Integration of heterogeneous high-dim modalities with high-level semantic correlation.
Classifier	Late	The most popular type of fusion for heterogeneous data, specially when the dimensionalities are also different [31, 129]. [99, 105, 121, 130]
Decision	Score	Modalities are completely uncorrelated or the same -i.e. multi-view/slicing [131] [105, 121].
Regularization Adversarial Models	Loss Fusion	Combining global & local information [37, 120] Modalities are complementary and correlated [18, 94].

A more automated approach is studied in some AutoML frameworks such as [132,



(a) Early fusion - Channel Concatenation

(b) Middle fusion - Channel Concatenation

(c) Late fusion architecture

Figure 4.2: Convolutional neural network based heterogeneous feature integration framework. In this framework a convolutional neural network extracts feature from primary (high-dim) modality and a one layer fully-connected network maps the hand-crafted features of the auxiliary modalities to appropriate dimension for that fusion level.

133]. These methods are comprehensive but more applicable for a generic application with a large set of data. In the majority of healthcare applications, understanding the causation is equally important as prediction accuracy. Additionally, the available size of data in these applications is typically on a small or medium scale.

4.3.2 Bottom-Up: Informed Representation

In Section 3.4, we discussed the complexities attached to healthcare imaging. We particularly mentioned that in a medical image, the intensity variation is not always the actual features that physicians need - e.g., in radiation therapy [10, 134, 135].

In fact, the medical image is a medium that is humanly understandable; therefore, information is transformed into a regular image to make it more convenient for the physician to understand that.

A physician interprets the image in the context of the other information s/he has, such as the scale, imaging setting, or the background information of the patient's disease. For a convolution neural network with a small set of data, it is highly possible to learn some irrelevant information and still have a good performance for a train or similar test cases [123]. However, this cannot be generalized to future cases without knowing the decision basis. Few approaches that are mentioned in the literature are:

1. Interpretation Models: These methods generally debug the learned networks to understand the important area/features for a particular decision [136–138]. This includes the use of packages such as LIME [139], Shapely [140]. The interpretation models would help physicians and experts validate whether the learned feature makes sense according to the domain knowledge or not [141].

2. Transfer Learning: Another approach for incorporating domain knowledge is transferring that knowledge from another domain. This is useful when we know the important features are common between a task τ_s with sufficient source of data and a target task τ_t with a small dataset - e.g., most of the healthcare applications. This is formulated as a function f_τ to learn task τ_t in domain D_t, to transfer latent knowledge from task τ_s and domain D_s, where $D_t \neq D_s$ and/or $\tau_t \neq \tau_s$. The assumption is that the size of D_s is much larger than D_t - i.e. $N_s >> N_t$ [142]. There is an extensive literature of transfer learning methods in classic machine learning [143] and deep learning [142] area. Using pre-trained models such as VGG16 [144] to transfer image filters to tasks like segmentation [145] is an example of this knowledge incorporation method. We use this method in vulnerable patient classification tasks for the radiation treatment planning task, which is explained in Section 5.4.2.

3. Data Augmentation: If instead of the important feature, the domain knowledge provides information about unimportant features, some forms of data augmentation could be helpful to improve the generalizability of the data-driven learning method [146]. There is a large literature on augmentation approaches, particularly in the image analysis domain [146–148] including various geometric methods and adversarial ones. This is reported to be an effective method, particularly in some deep learning image classification [148].

4. View Selection: In healthcare applications, it is common to select one slice or a selected number of slices as a representative for the information provided by the whole modalities [86, 116, 149, 150]. Despite the loss of information, the advantage of this approach is that it reduces the modality dimension in a straightforward, clear and understandable way. Depending on the task, the remaining information might also be sufficient.

These methods are effective for many tasks in regular and medical image processing. However, for more complex applications like disease progression, survival analysis [17], radiation treatment planning [134], or brain network activity analysis [118] these methods are not sufficient due to granularity of target function and the geometrical domain [134]. For example, in the case of Alzheimer's disease, the volume change of various regions across time is the bio-marker of interest or for the brain-network [151].

In the absence of sufficient data sources for these complex problems, more domain-dependent representation is needed. We argue that a form of feature engineering or feature expansion is helpful for these contexts. This can be a granularity increase or a form of feature disentanglement. An example iterative approach is shown in Figure 4.3. Another approach is to use the appropriate granularity level based on domain knowledge. We use this type of representation leverage approach for the applications of radiation treatment planning and disease progression prediction in Chapters 5 and 6.

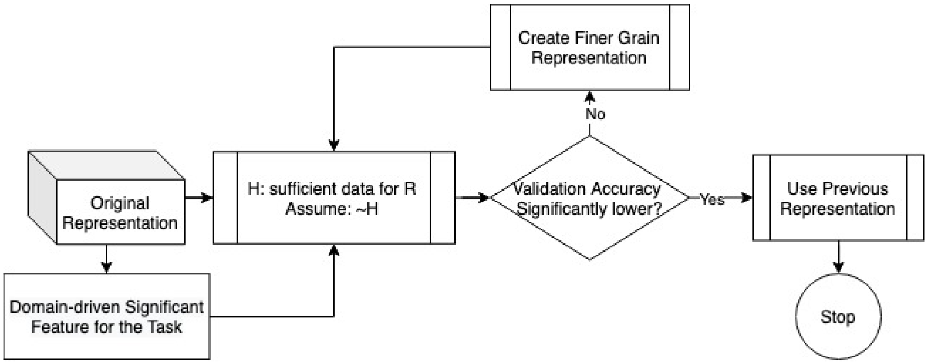

Create Finer Grain Representation
No
Original Representation
H: sufficient data for R Assume: ~H
Validation Accuracy Significantly lower?
Yes
Use Previous Representation
Domain-driven Significant Feature for the Task
Stop

Figure 4.3: An iterative method for representation granularity leverage; for this process, $H1$ is a hypothesis that with this representation and the available size of data, the convolution neural network can extract relevant features. The null hypothesis assumes that this representation will increase the validation accuracy compare to the coarser-grain representation. While this is not the case, we keep expanding the representation toward finer granularity. While this is a more systematic approach, in the following chapter, we did not actually go through this process.

4.3.3 Top-Down: Tree Structure Convolution Neural Network

To take advantage of global and local information, we define a tree structure CNN by combining the regression tree model and CNN. The idea is to use the macro-level features for dividing the regression function into its simpler components or sub-modalities. Figure 4.4 demonstrates two suggested approaches for tree-structure CNN. This architecture can be further refined using other prior probabilities, which is mentioned in the literature [37,152,153].

The simplest form of the architecture is dividing the complete network based on the value/s of modality B in a branching-block. This can be dividing only for feature extraction or the whole pipeline. Figures 4.4b and 4.4a, these models can be observed.

The branching block in the Figure 4.4, is where we divide the sample cases into multiple branches according to their values in one or multiple features in modality B. Similar to the model tree, we mentioned in Section 4.1.1. This block increases the sub-branches' homogeneity. For feature f_i, we select a break point b_i in a way that it minimizes the sum of branches' standard deviation. The simplest form of this structure is simply selecting one feature from modality B and dividing the data into two branches. This can be extended to more than two branches or a more than one feature in modality B.

4.4 Discussion

The main intention of the top-down architecture is to use the knowledge of the domain to distribute the learning process into more homogeneous sets. The main advantages of the model are:

1. Using global level features and local ones in different levels of information processing.

2. Consideration of feature correlation and overlap.

3. The learning model for each branch would be based on more balanced data.

4. A more interpretable performance and outcome based on the branching features and approach.

This architecture has, of course, some drawbacks, which makes its applications limited to some trade-offs:

1. The most important drawback is decreasing the amount of data in a situation that we already have a small set. The situation is even worse because of the parameter-sharing character of the convolution nets. In the regular form, the data of each category would have contributed to the learning parameters of all levels. If the data is divided into two equal categories, the number of data for the same size of the network is now half of the previous part.

2. Since this problem happens in the networks of all branches, it is multiplied by the number of branches.

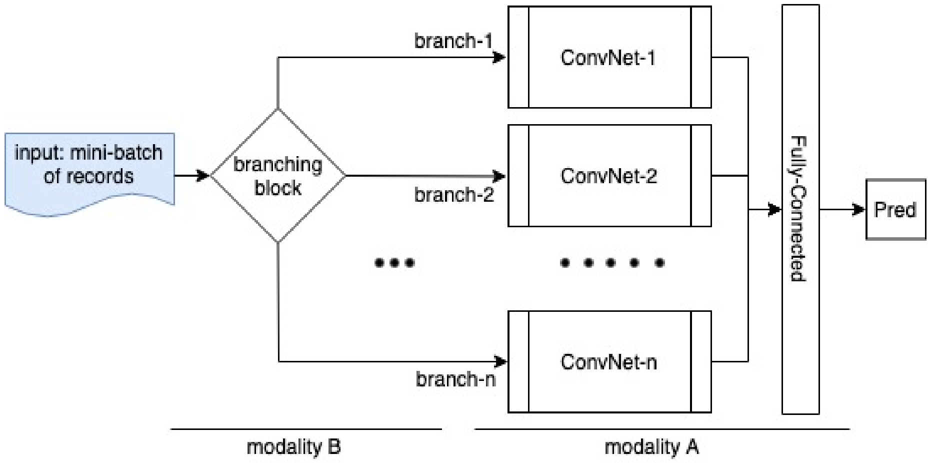
The above drawbacks show that for this architecture to be beneficial, the added homogeneity as a result of branching should be very high to worth this. This can be the case for smaller dimensions of modality A, with a strong feature from the domain. However, the extend-ability of the approach for a higher dimension should be further researched. The architecture, of course, is a very base form and can be refined with so many tools in the literature. We explore some of these refinements in Chapter 6, and discussed more suggestions that can be explored.

As we demonstrate in Chapter 3, demographic and genetic information and some clinical factors are among the frequent high-level, zero-dimensional features in healthcare. In a disease progression, for example, clinical factors intuitively tell the physician the stage of diseases, and the genetic or demographic would tell the pattern or intensity of the progression. Depend on the domain knowledge, if the effect is adding some intensity to the model, the factors can be used as a loss/regularizer [120] or prior probability [37,152,153]. However, when the effect causes different curve shapes or when the data of one category has a fundamentally different quality with the data of other categories, the prior information in the previous forms of the literature can be misleading for a regression task.

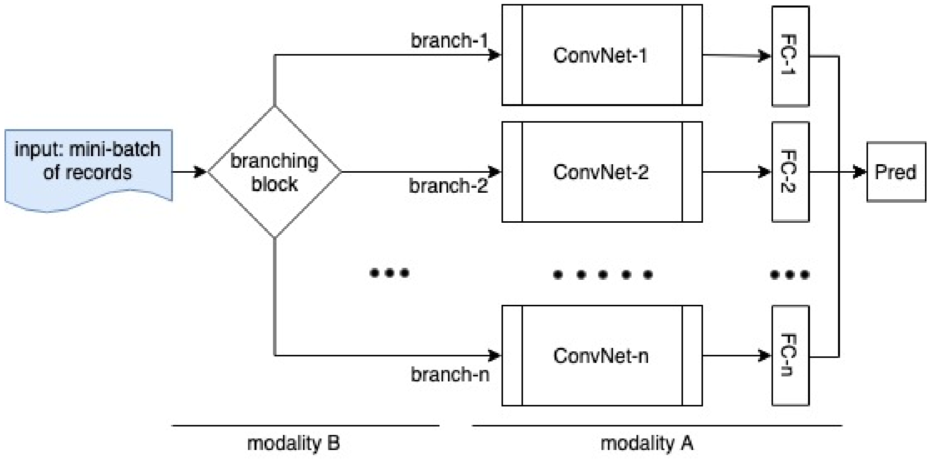
4.5 Summary

In this chapter, we proposed a framework for integrating information from heterogeneous modalities. The main concentration was extracting the information from the appropriate semantic level of abstraction for each modality. This semantic abstraction is defined by domain knowledge. We reviewed the current approaches in the literature and provided two other novel perspectives to this framework. In the following chapters, we will discuss the application of the framework in radiation treatment planning and Alzheimer's Disease progression in more detail in the following chapters. We introduce a distance-based representation on Chapter 5 and a region-based representation on Chapter 6. We evaluate the whole pipeline for disease progression prediction on Chapter 6.

.



(a) Conv-Net architecture



(b) Full-Net-architecture

Figure 4.4: Tree-Structure Convolutional Neural Network architectures. In this architecture, convolutional neural network is mainly predicting the regression function through information in modality A, and modality B is decomposing the curve into sub-groups of samples. In Conv-Net (top), the decomposition is on dimension reduction, while in Full-Net (bottom) the whole pipeline is decomposed into subsets of data.

CHAPTER 5: Case Study I: Informed Representation in Radiation Treatment Planning

The application of concentration for this chapter is knowledge-based treatment planning (KBP) for radiation therapy [1]. An ideal machine learning framework would be trained with the features of the previous (specific type of) cancer patients as well as their high-quality treatment plans, and can accurately predict a high-quality plan (or some information about that) for the future patients (according to their features) [29]. Despite the complexity of the task and various influential variables [12], the available high-quality data source is typically sparse. Therefore, an important part of the previous researches was the extraction of few (high-level) features - e.g., [10]) that can predict some information about the outcome (high-quality plan) [29]. Nonetheless, these methods cannot capture low-level and local features (for example, in a CT-image) properly. Therefore, this study adds to the domain knowledge by addressing the abstraction-level (or locality) of various features and providing a systematic means for selecting an appropriate level.

5.1 Background and Context of The Application

Radiation therapy is one of the treatment or control approaches for most types of cancer. The goal of this treatment approach is to destroy the cancerous tissues while saving the healthy ones. Intensity-modulated radiation therapy (IMRT) is one of the recent and flexible radiation therapy technology. However, this flexibility made the optimization problem complex from all mathematics, physics, and biology aspects. In other words, solving the current mathematical equations and optimization does not

[1] The content of this chapter is published in the proceeding of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) [134].

have a perfect answer.

This means that the ideal solution considering the physical and biological constraints cannot be prescribed. Therefore, there are various trade-offs involved in the process, making the process time-consuming with no *best for all* solution. Besides normal anatomical features that are traditionally considered in the mathematical formulation of this problem, there are many more parameters involved in the manual trade-off process. Besides, the complication of low-level cellular and molecular reactions left parts of the problem unknown, even from physical and biological aspects.

This study is not focusing on unknown features. Instead, it focuses on transferring the manual trade-off process into a more automated and systematic approach. It is meant to take one step toward facilitating the knowledge-based strategy for treatment planning. The intention of knowledge-based planning is efficiency and raise the overall quality across planners and cancer treatment centers with different levels of experience by learning from previous patients and their high-quality plans. Many studies were done in this category using a case-based and statistical, and machine learning approach. Following the recent achievements of deep learning, we are mostly focusing on exploring deep learning perspectives for addressing the learning tasks.

5.2 Background

Radiation therapy is one of the most effective techniques in prostate cancer treatment. It works by destroying or controlling the cancerous area. One of the main complications associated with the treatment process evolves around the intensity and the direction of the dose is applied. An insufficient dose in the cancer area fails to demolish the cancerous cells properly and increases the risk of tumor regrowth. Conversely, high dose radiation can damage the surrounding healthy organs or increase the risk of secondary cancers. Despite the planner team's effort in restraining the radiation to the target area, some patients are more vulnerable to receive excessive doses than others due to the patient anatomy and cancer geometry. Early assessment of the best achievable dose distribution's vulnerability and prediction can guide the physician and planner team to a better treatment strategy.

Quantitative analysis of normal tissue vulnerability and toxicity can be measured by the traditional Normal Tissue Complication Probability (NTCP) model [154,155]. This model uses dosimetric features, such as dose-volume histogram (DVH), of the proposed plan to evaluate the probability of dose-induced complications in noncancerous tissue. Later updates from Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) group suggest machine learning as the future pathway to develop more accurate predictive models [12, 156].

The conventional predictive models [10] utilize some hand-crafted anatomical features that can explain a significant amount of variations in inter-patient organ-at-risks (OAR)'s DVH after planning.

Recent studies [135, 157–159] claim better performance employing deep learning potentials to capture the features directly from 3D CT scan images. Due to the limited resources for training these models, most of the studies extract the contour information -i.e., segments, from the image before the learning process, and represent the patient anatomy using 2D slices of the contoured image. While this representation decreases the complexity of the learning task, it loses some significant information, such as the organ's three-dimensional position and orientation with respect to the cancer area. While augmentation methods in deep learning training can accommodate some variations, they may not be sufficient to handle patient positioning variations in clinical practices.

The distance-based representation we suggest in this study emphasizes the 3D distance of the OAR from planning target volume (PTV). This representation has a major advantage over existing approaches in that it maintains the volumetric information despite slicing the 3D images into 2Ds.

This chapter's experiments are designed to assess the potential of this distance-based deep learning framework for knowledge-based dose distribution prediction. Comparing the results from the distance-based deep learning representation, introduced in this study, with the contour-based [157] and hand-crafted features [10] in prior studies demonstrate comparable or higher performance.

The motive to substitute the image and contour-based representations with the distance one is its robustness to patient positioning. Hence, it increases the generalization potentials of knowledge-based models in transferring knowledge from one patient or institution to another.

Technical Significance: This work assesses distance-based representation for deep learning frameworks to predict dose distribution in radiation treatment planning. This novel use of distance representation combines prior knowledge of the radiation treatment domain with deep learning capabilities. Having limited data in deep learning is a major threat to the generalization of the learned models. Our work is in line with the argument that the prior and domain knowledge is a proper means to mitigate the generalizability challenge [160] specifically in the absence of sufficient resources of labeled data.

Clinical Relevance: Radiation therapy is a widely used and effective technology for cancer treatment. However, radiation treatment planning is currently a complex iterative process manually carried out by a team of physicians and planners. One of the reasons is that the clinically desirable dose distribution may not be achievable by the treatment system due to variations in patients' anatomical, tumor geometries, and other clinical factors. Therefore, it is important to inform physicians what is achievable for an individual or, at least, which patient is likely to fail the clinical criteria. This study addresses these questions by providing physicians with an early estimation of their vulnerability due to the anatomical limitation (Task 1). The method then provides an early estimation of the achievable dose distribution based on the organ features and knowledge from previous similar patients (Task 2).

5.3 Dataset: Distance-based Representation

5.3.1 Representation Motivation

An organ's distance to the target volume is one of the key anatomical constraints during treatment planning. The short distance of a healthy organ from a cancerous area makes the organ vulnerable to receive an excessive dose. Prior studies [161, 162] introduced metrics, such as Distance-to-Target Histogram (DTH) and Overlap Volume Histogram (OVH), to describe and analyze this factor quantitatively. Later, Yuan et al. demonstrated the significant contribution of the following measures in predicting DVH between-patients variability [10]:

- The median distance between Organ-at-Risk (OAR) and planning target volume (PTV)
- The portion of OAR volume within an OAR specific distance range
- The fraction of OAR volume that overlaps with PTV
- The portion of OAR volume outside the primary treatment field

Deep learning architectures are strong in capturing local shape features. This made it possible for the researchers to study the capability of local features in explaining further inter-patient dose distribution variations [135, 157–159]. Inspired by previous works [10, 161, 162], we propose the use of 3D distance matrices, extracted from contoured 3D CT scan images, as input representation for deep learning.

5.3.2 Data Extraction

The dataset was extracted from 216 prostate cancer patients who received radiation treatment. We have chosen prostate cancer for this study because of the closeness of the bladder and rectum as OARs to the prostate, which is the cancerous area. Patients may suffer if they receive an uncontrolled amount of radiation dose in these two organs. If we can determine the dose that these two structures will receive based on the geometry, proper measures can be taken during the treatment planning.

The raw data contains 3D structure contour values, 3D dose matrices, and the prescription doses. The raw data were further processed to meet the need for this study. First, the 3D dose matrices for bladder and rectum are created using the patients' contour information and planned dose matrices. The dose outside these two structures is removed. Then cumulative DVHs are calculated on dose bins increases by 50cGy interval. It demonstrates the percentage of the volume inside the structure that receives greater than or equal to each dose bins for each organ. The dose values in DVH and dose matrices are normalized by patient prescription dose.

The distance matrices are in three dimensions and extracted from 3D contours. Each of the 3D matrices is called *distance3d* throughout this chapter and represents one of the patient's OARs -i.e., bladder or rectum. Each element of the matrix is an assigned distance measure of the corresponding voxel. The assigned measure is defined as the shortest distance of that voxel to the 3D surface of the PTV -i.e., prostate contour. All the voxels outside of the OAR are assigned a value of zero, and the overlapping voxels that are inside both OAR and PTV contours get negative values. The voxels outside of the radiation field are artificially increased to reflect the out-of-field parameter [10]. No further information is stored about the prostate contour.

The dataset used in this study contains only distance and dose information and, thus, is completely anonymized.

5.4 Methodology: Deep Learning

Conventional knowledge-based planning methods extract engineered features from 3D CT-scan images to predict the best achievable dose in the cancerous area and other organs at risk [29]. Recent studies in medical imaging and radiation therapy demonstrated higher prediction power using deep learning features than the engineered ones [123]. However, the limited available resources of the labeled data compared to the number of learning variables cast doubt on the trained models' generalizability in the radiation treatment planning domain.

In medical image processing, dimension reduction - e.g., in [145], transfer learning - e.g., [145], and residual architectures - e.g., in U-Net [163], are among the popular techniques to improve deep learning performance despite the data scarcity. We evaluated the proposed representation by using the same strategies to design our learning model. In the following subsections, we explain these approaches and how we use them for our experiments.

5.4.1 Dimension Reduction: Slicing

The dimension reduction from 3D to 2D decreases the number of learning variables and reduces the scale of desired data for the learning task, causing some information loss. Nevertheless, the 3D information retained by the proposed representation -i.e., distance3d matrices and the structural/local information maintained by the chosen reduction approach -i.e., slicing, helped us minimize this information loss.

The formal statistical approaches for dimension reduction, such as principal component analysis (PCA), retain the greatest variations, perceived as information, in the data. However, the methods are optimized for retaining global information and sacrifice the local structural patterns. Since deep learning feature extraction is based on these local patterns, we avoided using those methods.

Instead, we consider the three anatomical axes as the main dimensions -i.e., features, and select two of these three features using *slicing*. Slicing along the anatomical planes maintains the imaging analogy and interpretability, as well. The anatomical planes are sagittal, which divides the body into left and right, coronal that divides the body into back and front, and axial that divides it into the head and bottom portions. One can imagine in this coordinate system, the x-axis goes from left to right, the y-axis goes from front to back, and the z-axis moves from top to bottom. The default slicing for prostate cancer CT images is along the axial plane and is indexed from top to bottom. We explore the effect of these three directions of slicing in the first experiment (Task 1).

Furthermore, the proposed representation in this study is based on 3D distance information. Thus, we expected the representation to prevent part of the volume-related information loss. We confirmed this assumption by observing better performance using the 2D distance matrices and the 3D distance ones for the same experiments (Task 1).

After slicing, we treat each of these slices as an independent 2D data point. This way, we increase the number of input samples and adds robustness to the network against the slice position. As mentioned, all the voxels outside of OARs are zero. Therefore, those full-zero slices are uninformative as independent samples and are trimmed for the classification experiment (Task 1).

5.4.2 Transfer Learning: VGG-16

In a typical machine learning task, the assumption is that the training set has the same distribution as the test-set. On the other hand, transfer learning is meant to reuse the learned knowledge on one training domain for a different task and distribution test. The deep learning need for a large amount of training data as well as salient results of the trained models made transfer learning popular in this domain.

We use transfer learning to demonstrate the proposed representation's potential to benefit from models pre-trained on (non-medical) image data to compensate for the data shortage in the planning domain (Task 1). The justification is that distance matrices can be perceived as gray-scale images. The pixel intensities in a gray-scale image carry information such as light, depth, and color. The distance matrices can be perceived as the depth of each point from the PTV perspective.

Inspired by Tran's work [145] we use VGG-16 [144] and transfer weights from that model. VGG-16, also called OxfordNet, is a deep convolutional neural network (CNN) with five convolution blocks, 16 layers (five convolution blocks followed by three fully connected layers), which was among the top winners of the ImageNet challenge in 2014 and the authors published the trained model and weights publicly [144].

The sequential format of the network gives the model some level of interpretability. The image signals travel in a sequential structure of the convolution blocks from low-level features to more abstract ones and eventually is used by the classification component. Thus, the highest blocks are the most abstract and task-specific representation of the image. Similar to prior work [145], we borrowed the first four convolution blocks of VGG-16 with their frozen weights. The extracted features, or the encoding, out of the last block, then goes to three consecutive fully connected layers to learn a classifier that can discriminate more vulnerable patients from less vulnerable ones in their OAR. The schematic diagram of the CNN architecture we utilized is represented in Figure 5.1. The distance values are repeated in three channels to make the input suitable for VGG-16 (image format).

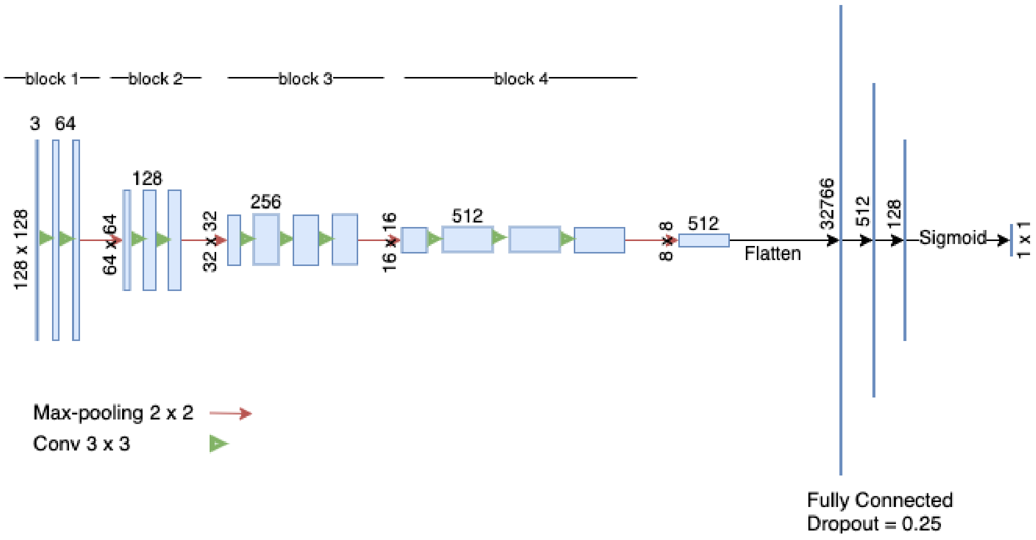


Figure 5.1: The architecture consists of four blocks of VGG-16 [144] with frozen weights, followed by three layers of fully connected with ReLU activation function and a final Sigmoid block for the binary classification.

5.4.3 Learning Architecture: U-net

U-net is an elegant architecture belonging to the fully convolutional network family [164]. The network was proposed for medical image segmentation [163]. Elimination of the fully-connected layer in these networks dropped the learning model variables and consequently demonstrated acceptable performance with the low number of training data points in encoder-decoder architectures.

In addition, U-net improved this performance by including the symmetric upsampling process after the last layer and adding the residual signals from the corresponding in the downsampling branch. This structure immunized the segmentation from losing critical information during the downsampling without adding higher order of complexity.

Recent related works [135, 157–159], used U-net based architectures for prediction of voxel-level dose distribution based on organs' contours. We implemented U-net architecture similar to the original work [163] and compared the prediction performance of the architecture from the distance representation to the contour ones. Figure 5.2 represents a schematic diagram of the model architecture we implemented for our prediction task.

5.5 Evaluation and Result: Dose Prediction

The distance-based representation was evaluated in two predictive tasks. The first task (Section 5.5.1) is the classification of patients based on their OARs' vulnerability to receive excessive dose level having a pre-trained model. The aim is to predict a high-level feature directly from the anatomical structure of each organ-at-risk, independently. The argument is that the representation keeps enough information from CT images that enables the learning models to benefit from the computer vision domain's features. Furthermore, there are informative features of the organ that can be processed independently of the other ones.

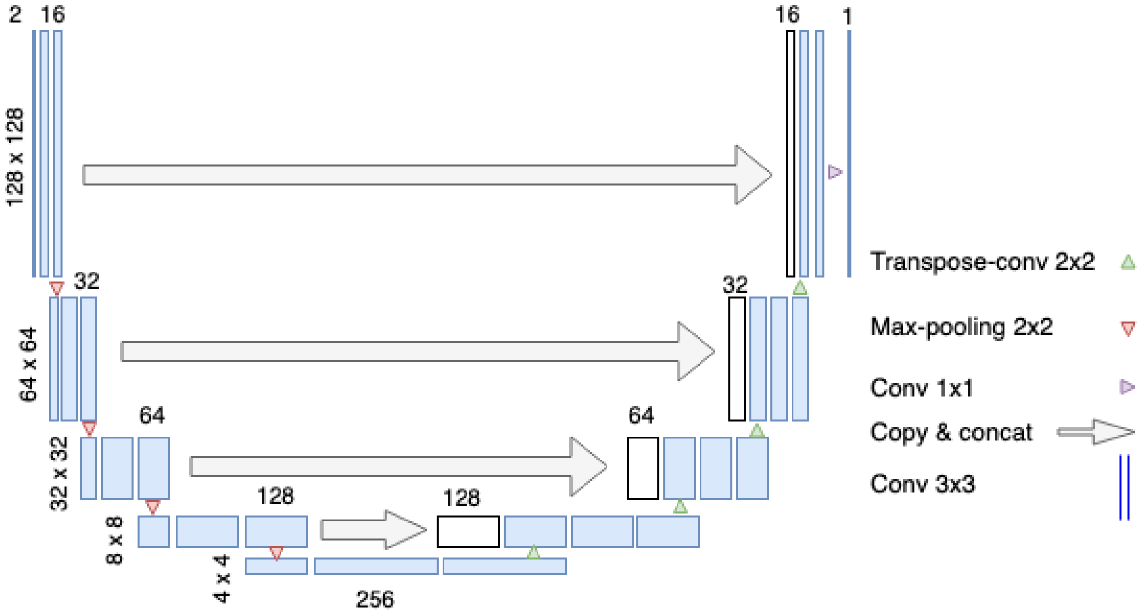


Figure 5.2: The U-net architecture takes a 2-channels 128 x 128 distance3d slice as an input and predicts the dose distribution on the output side. We follow the original U-net architecture paper for the implementation with a depth of five and a single last channel for the dose prediction [163]. The size of the feature map in each convolution block is written on the left side of the down-sampling direction. The size of feature maps in the right side is equal to their left counterparts.

The second task (Section 5.5.2) predicts voxel-level dose distribution by learning from high-quality plans of past patients. This task is implemented on a U-net architecture, which has been used successfully by several recent voxel-level dose prediction efforts using deep learning [157, 159].

The extracted data are preprocessed to meet the needs of the experiment. The 3D matrices are sliced into 2D ones, as mentioned in Section 5.4.1, and the slices are center cropped to the size of 128 x 128. Table 5.1 demonstrates the original size of the distance matrices for each of the organs at risk, as well as the size of the dataset after preprocessing.

The contour matrices have a similar format to the distance matrices except that the values are binary, zero for the voxels outside of the OAR, and one inside. Consequently, the cropping and slicing process on distance, contour, and dose matrices are parallel.

For the dose distribution prediction task (Task 2), bladder and rectum matrices of the same slice are treated as two channels of one data sample. It is important to note that the train, validation, and test splitting is done on the patient-level, making sure similar information won't appear in multiple cohorts. The slicing and other processing are done after that.

Figure 5.3 demonstrates an overview of the experiments' steps. Also, the hardware and software frameworks for the experiments are shown in Tables 5.2 and 5.3.

5.5.1 Task 1: Organ-at-Risk Vulnerability Prediction

As was discussed in the introduction, a high-quality plan minimizes the received dose in the OARs, but anatomically some patients are more vulnerable than the others. This experiment is meant to predict this vulnerability factor.

The task is a binary classification of distace3d slices using pre-trained VGG-16 architecture, explained in Section 5.4.2. Thus, the 2D distance3d slices are input, and the patient vulnerability labels are the expected output of the learning model.

Table 5.1: The original data set contains distance matrices of bladder and rectum (OARs) for 216 prostate cancer patients. Each matrix consists of 300 slices, each of which has a size of 192 x 252. For the patient vulnerability (classification) task five nonzero consecutive 2D samples extracted from the middle of each 3D matrix using sagittal, coronal and axial slicing direction. This increased the number of samples five times and decreased the complexity of the problem with the price of partial information loss. For the voxel-level dose prediction task, to compare the results with both representations in Nguyen et al. [157] and Yuan et al. [10], we designed multiple experiments and the number of selected slices are discussed in the corresponding section. Here, for simplicity, we put st and sv variables for the number of the selected slices in training and validation processes, respectively.

		Dimension of the Input Features
Original Dataset		2 * (216, 300, 192, 252, 1)
Classification Task	Train	2 * (720, 128, 128, 3)
Classification Task	Validation	2 * (180, 128, 128, 3)
Classification Task	Test	2 * (180, 128, 128, 3)
Dose Prediction Task	Train	st * (2, 172, 128, 128, 1)
Dose Prediction Task	Validation	sv * (2, 44, 128, 128, 1)

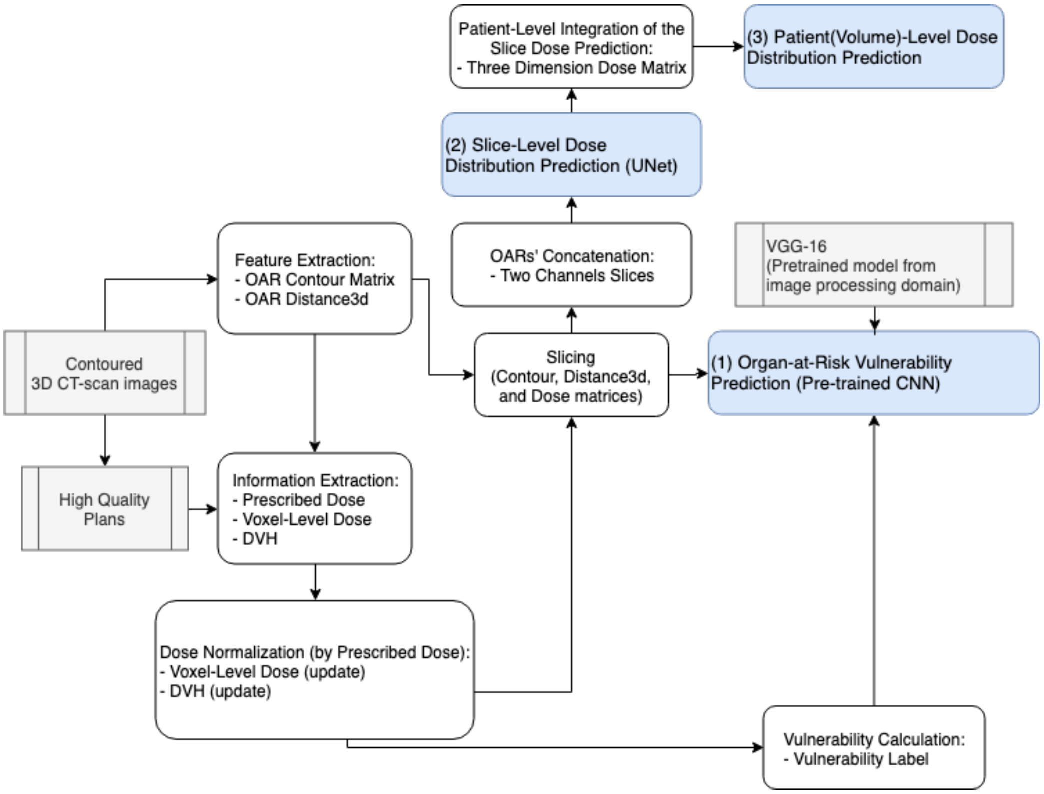

Patient-Level Integration of the Slice Dose Prediction:
- Three Dimension Dose Matrix
(3) Patient(Volume)-Level Dose Distribution Prediction
(2) Slice-Level Dose Distribution Prediction (UNet)
Feature Extraction:
- OAR Contour Matrix
- OAR Distance3d
OARs' Concatenation:
- Two Channels Slices
VGG-16
(Pretrained model from image processing domain)
Contoured 3D CT-scan images
Slicing (Contour, Distance3d, and Dose matrices)
(1) Organ-at-Risk Vulnerability Prediction (Pre-trained CNN)
High Quality Plans
Information Extraction:
- Prescribed Dose
- Voxel-Level Dose
- DVH
Dose Normalization (by Prescribed Dose):
- Voxel-Level Dose (update)
- DVH (update)
Vulnerability Calculation:
- Vulnerability Label

Figure 5.3: An overview of the data source, steps and predictions in this study. The evaluation blocks are colored in blue, and the external sources (data and model) are colored in gray.

Table 5.2: The experiments' platform specification.

Hardware	
Type of Machine	Google cloud n1-standard-8
Processors	8 vCPUs and 1 NVIDIA Tesla K80 GPU
Memory	30 GB
Local Disk	100 GB standard persistent disk
Software - Classification Task	
Operating System	Debian 9
Programming Language	Python 2.7
Deep learning back-end	tensorflow 1.10
Deep learning interface	Keras 2.2
Software - Dose Prediction Task	
Operating System	Debian 9
Programming Language	Python 3.6.4
Deep learning back-end	Pytorch 1.1.0

Table 5.3: Experiments' parameter setting.

Parameter	CNN setting	U-net setting
Optimizer	Adam	Adam
Loss function	Binary crossentropy	Mean squared error
Abstraction (downsampling)	4	4
Number of epochs	100	200
Batch size	32	16
Learning rate	0.000001	0.0001

In order to have a balanced set of labels, we compare the patients with the average case. We pick the dose-at-50 as the prediction index, but a similar model can be trained for the dose at other volumes. The patient's vulnerability in each OAR is then defined by a binary value that demonstrates either the patient is going to receive more than the average of the population or less.

Dose-at-50 is the maximum dose that at least 50 percent of the organ has received (or is going to receive) according to the high-quality plans. By dividing the dose value by the prescribed dose, the values are normalized and comparable across patients. Finally, the ratio is compared to the mean. Those with higher or equal than the mean would be considered more vulnerable (labeled as one), and those with lower ratios were considered less vulnerable (labeled as zero). An important note is that the vulnerability of a patient in the bladder does not mean s/he is also vulnerable in the rectum. Despite the possibility of correlation, the two organs-at-risk, bladder, and rectum, are independently labeled and evaluated in our experiment.

The experiments are run on cloud virtual machine using matrices extracted from 216 patients. The specifications of the experiment framework, settings and data are in Tables 5.2, 5.1, and 5.3, respectively.

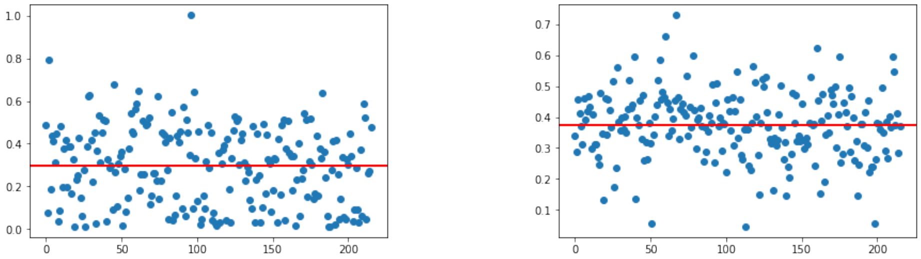
The results are evaluated using five-fold cross-validation, each containing random 173 patients as a training sample and 43 patients in the test set. While each slice is considered an independent data point, the slices extracted from one patient belong to the train set or the test set, not both. The binary-class evaluation metrics -i.e., accuracy, precision, recall, and f-score, are used as an evaluation measure of this experiment. The results are demonstrated in Table 5.4 for both validation and test rounds.

The average accuracy of vulnerability prediction on the best experiment is 84.89% for the bladder and 60.34% for the rectum. Considering the distribution of the vulnerability in Figure 5.4 and the populations on the borderline, the results specifically

Table 5.4: Experiment with distance3d as the input variable and vulnerability label as the prediction target. Average evaluation metrics for five-fold cross-validation are displayed along with the average of the five models on test-fold. Each fold is trained on 720 data point from 144 patients (67%) and validated on 180 data points from 36 patients (17%). The models are then tested on 180 data points from 36 patients (17%), which were not involved in the training process.

O-A-R	Slicing	Precision	Recall	F-score	Accuracy
		Validation Results			
Bladder	Axial	80.02 (+/- 8.11)	82.89 (+/- 1.89)	78.90 (+/- 4.21)	82.67 (+/- 2.47)
	Coronal	82.61 (+/- 10.63)	77.67 (+/- 1.99)	77.24 (+/- 6.12)	80.11 (+/- 4.80)
	Sagittal	80.51 (+/- 7.66)	82.42 (+/- 5.74)	79.44 (+/- 6.06)	81.44 (+/- 5.89)
Rectum	Axial	55.27 (+/- 15.81)	49.94 (+/- 11.23)	48.35 (+/- 9.38)	51.33 (+/- 5.78)
	Coronal	63.27 (+/- 12.42)	63.27 (+/- 10.99)	59.64 (+/- 6.64)	63.44 (+/- 3.05)
	Sagittal	65.82 (+/- 14.83)	66.13 (+/- 9.03)	63.07 (+/- 8.66)	66.67 (+/- 1.96)
		Test Results			
Bladder	Axial	79.14 (+/- 2.77)	81.16 (+/- 5.32)	78.55 (+/- 3.28)	77.89 (+/- 3.36)
	Coronal	**83.10 (+/- 3.74)**	**92.45 (+/- 3.00)**	**86.08 (+/- 2.66)**	**84.89 (+/- 3.13)**
	Sagittal	75.22 (+/- 3.05)	85.19 (+/- 2.29)	77.89 (+/- 3.25)	75.33 (+/- 3.38)
Rectum	Axial	57.34 (+/- 1.69)	44.81 (+/- 12.98)	48.39 (+/- 7.77)	51.33 (+/- 1.71)
	Coronal	50.46 (+/- 0.93)	46.79 (+/- 6.17)	47.64 (+/- 3.61)	48.56 (+/- 2.45)
	Sagittal	**60.48 (+/- 6.58)**	**60.61 (+/- 3.90)**	**58.43 (+/- 2.51)**	**60.34 (+/- 4.48)**
		U-net Validation			
Bladder	Axial	95.45	91.30	93.33	95.45
Rectum	Axial	73.91	77.27	75.56	75.00

for bladder are promising. The results lead the experiments to the next task for predicting the complete dose distribution matrices.



(a) Bladder

(b) Rectum

Figure 5.4: The patients' normalized dose-at-50 distribution in their bladder (left) and rectum (right). The red line is the mean, which divides patients into more (above the line) and less vulnerable.

5.5.2 Task 2: Dose Distribution Prediction

The complexity of radiation therapy planning and its dependency on various parameters made this process time and expertise intensive. On the other hand, the quality of this step and accurate assessment of the results are significant determinants of the planning process's success. Knowledge-based treatment planning aims to support this process by transferring knowledge from previous patients' high-quality plans to future ones. The intention is firstly to transfer knowledge from institutions with more expertise in planning to the less experienced ones. Secondly, provide an early assessment of the best achievable outcome to expedite the decision-making and planning process.

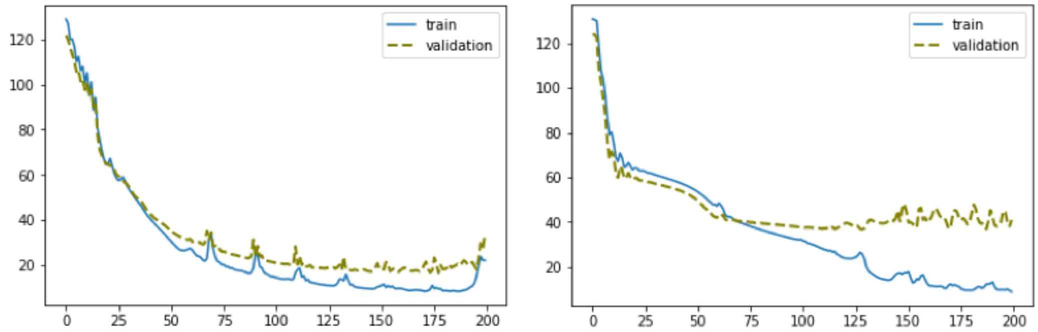
This task intends to assess the introduced representation in predicting dose distribution using a deep learning architecture. The U-net architecture, explained in Section 5.4.3, is selected for this task. U-net is one of the popular architectures, or components of the pipeline, in similar works. The evaluation for dose prediction in slice-level and volume(patient)-level are discussed in this section.

5.5.2.1 Evaluation of Dose Distribution using 2D Slices

This task's first experiment trains the U-Net model on one slice per patient (distance3d/contour matrix) from the training cohort and evaluates the dose prediction on one slice per patient in the validation set. This setting is similar to what the early deep learning dose prediction work [135] suggests.

Figure 5.5 demonstrates the plot of train vs. validation loss during the learning epochs. We intentionally did not use any data-dependent regularization to observe the generalization potential of the representation itself. The loss function is measured by the mean squared error of the voxels' dose. The validation loss did not have any impact on the learning parameters. As Figure 5.5 shows, while the training process is initially faster for the contour representation, from epoch-60, the over-fitting is much more significant for the contour representation than the distance3d ones.

The input/output for a random sample from validation set is demonstrated in Figure 5.6. The inverse relation of distance and received dose is apparent in Figure 5.6a, which gives a sense of the reasoning behind the proposed representation. This agrees with the prior knowledge of the domain [10].



(a) Distance representation

(b) Contour representation

Figure 5.5: Training/Validation Epochs for distance (a) and contour (b) matrices as the inputs of U-net for dose distribution prediction. From epoch-60, the over-fitting in learning via contour matrices is more significant than the distance ones.

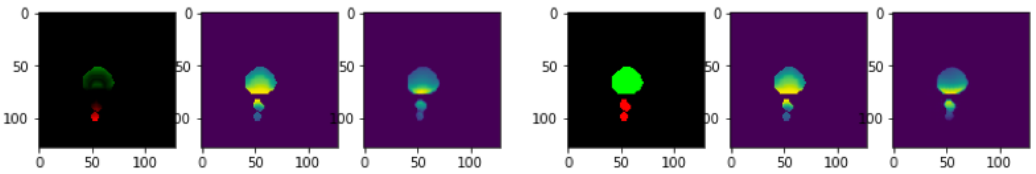
The slice-based mean loss information in Table 5.5 confirms the overfit on the second representation (contour matrices) compared to the first one (using distance matrices). Note that the values are *percentage* of the prescribed dose. While in a similar situation, the validation data is used for setting up parameters and regularization mechanisms to overcome the overfitting issue, due to the intention of this study in understanding the independent generalization potential of each representation, for neither of the representations, we did not modify the model according to the validation results. This experiment's observations demonstrated a better validation performance and generalizability on the learning task with distance matrices compared to the contour ones.

Table 5.5: Mean loss for Train and Validation Cohort containing 172 (80%) and 44 (20%) patients, respectively.

Feature	Train Loss	Validation Loss
Distance Matrix	28.24	33.18
Contour Matrix	8.73	40.82

5.5.2.2 Evaluation of Dose Distribution using 3D Volumes

To extend the evaluation on patient-level, we predicted all slices of patients' volume. We observed that the model trained on a single middle slice (for both distance and



(a) Distance matrix(input), actual dose distribution (expected output), and predicted dose distribution (predicted output).

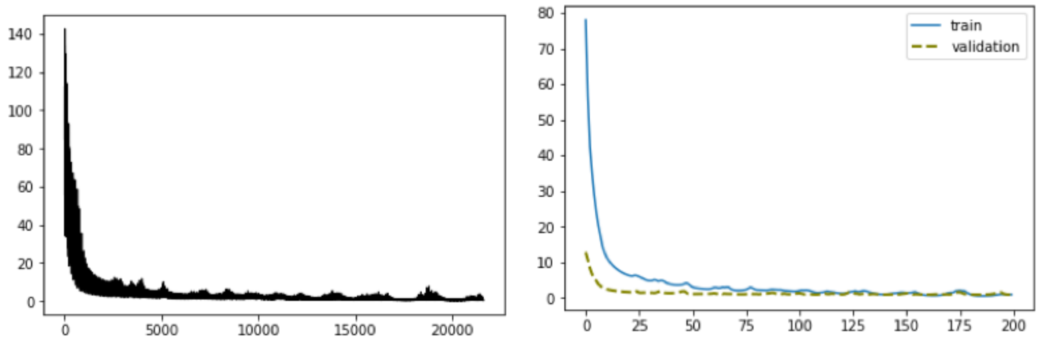
(b) Contoured image(input), actual dose distribution (expected output), and predicted dose distribution (predicted output).

Figure 5.6: input, expected and predicted output for one of the middle slices of a random patient's CT scan OARs' contour. The slice is selected from the validation set, and for the visualization, the bladder is set to the green channel, and the rectum pixels are assigned to the red channel. The intensity of pixels on the middle and right columns demonstrates the percentage of received by that voxel divided by the prescribed dose. Yellow demonstrates a higher receipted dose, and blue shows a lower level of dose.

contour representation) tends to overestimate the sparse slices' dose distribution. Therefore, we trained the model using ten slices (distance3d matrices) per patient to ensure the distribution is not biased toward the dense ones. We adopt a probabilistic selection method, which gives priority to the slices with more nonzero voxels for sample selection due to the higher number of zero slices. We then evaluated the model on all slices of the patients in the validation set to calculate their dose distribution in three dimensions. The train and validation loss of these slices across the epochs can be seen in Figure 5.7. The results are comparable with Figure 5 of the prior contour-based deep learning model with a similar setting [157].

Having the prediction for all slices of the validation cohort, we calculate the predicted DVH and compare it with the DVH extracted from high-quality plans. The results on Table 5.6 and Figure 5.9 are comparable with the results reported by knowledge-based planning with engineered features [10]. We can see that for the percent volume that received at least 85% of the prescription-dose (V85%) and 99% of the prescription-dose (v99%), our model outperforms the prior reported results in Yuan et al. [10].

Also, Figure 5.8 shows the predicted DVH compared to the planned DVH for two randomly selected patients from the validation cohort, and Figure 5.9 demonstrates the error for V50%, V85%, and V99%. These figures demonstrate a high correlation between the predicted and planned DVH. The figures are comparable with Figures 5 and 6 of the referenced study [10].



(a) Epoch-iteration loss

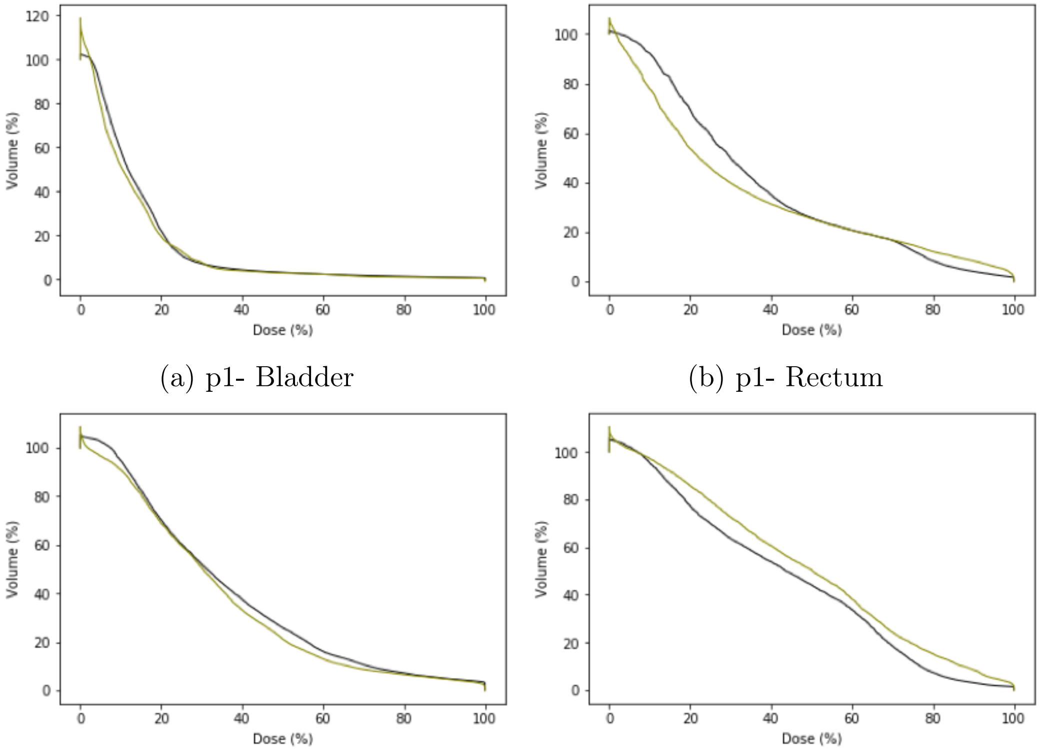
Figure 5.7: Patient-level evaluation: In order to evaluate the model on a patient level and compare with prior work [10], we trained the model on 10 slices of the patients rather than 1 similar to Nguyen et al. [157]).

Table 5.6: Difference of the volumes corresponding to 99%, 85%, and 50% of prescribed dose predicted vs. actual plans. Prior study of [10] reported 71%(17/24) within 6% error and 85% (21/24) within 10% error bound.

Bladder: Volume-at-dose	Error-bound	bladder	rectum
V50%	6%	**72.73** (32/44)	54.55 (24/44)
	10%	**84.09** (37/44)	70.45 (31/44)
V85%	6%	**84.09** (37/44)	**75.0** (33/44)
	10%	**93.18** (41/44)	**90.91** (40/44)
V99%	6%	**90.91** (40/44)	**81.89** (36/44)
	10%	**95.45** (42/44)	**95.45**(42/44)

5.6 Discussion and Future Work

There are two complementary approaches in feature extraction; using human knowledge to design hand-crafted features or letting the machine learn the patterns from various situations in data points. For example, in object classification applications, the second approach has shown strong results with the convolutional neural network.



(a) p1- Bladder

(b) p1- Rectum

(c) p2- Bladder

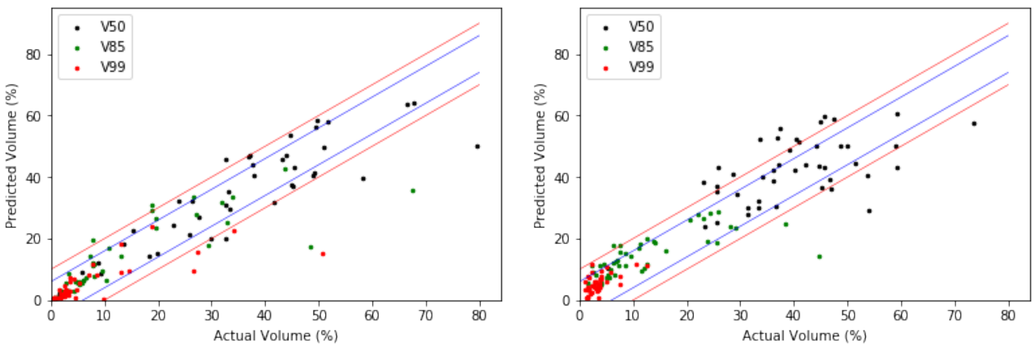
(d) p2- Rectum

Figure 5.8: A comparison of actual (black) vs. predicted (green) cumulative dose-volume histogram in the volume of Bladder (left) and Rectum (right) for two random patients (top and bottom) from validation set.

The state is different in the healthcare applications, where we do not have sufficient high quality labeled data points. Yet, there are large scale variables to learn due to the complexity of the images.

This study's contribution is the suggestion and evaluation of a strategy to combine the two mentioned approaches for extracting features in the domain of radiation treatment planning. We utilized distance matrices instead of the actual images, inspired by previous studies and expert knowledge. The method retains the structural information while removing the intensity representation (raw image data).

The reported results demonstrated comparable performance with state-of-the-art methods. It also opens multiple paths to move toward more generalizable and interpretable knowledge transfer in radiation treatment planning. Low-level feature



(a) Bladder

(b) Rectum

Figure 5.9: The predicted volume at V99%, V85%, and V50% on the DVH curves by predicted and actual plan DVH for (a) bladder and (b) rectum. The error bound corresponding to 6% and 10% OAR volume is shown to be comparable with [10].

engineering, proper use of pre-trained models, and informed dimension reductions are among the paths this work suggests to move on.

Despite the reasonable performance demonstrated in the experiments, there are several areas for further improvement. There are more relevant trained models in the transfer learning task to transfer features from - e.g., DeepLesion model [165]. We also intend to evaluate the representation of the more recently-used architectures, such as generative models [158].

Apart from the evaluation results, the substitution has other benefits. It encodes the CT scan images in a more anonymous format, which makes data sharing more convenient. Task 1 provides flexibility to analyze each of the OARs independent from other organs and PTV for the prediction and learning task. Furthermore, the invariant nature of the representation toward the displacement and orientation opens some potential for data augmentation in this domain. Finally, the data's interpretable format opens a space for reverse engineering of the deep learning extracted features. This can be significantly important not only from the validation perspective but also to unleash deep learning's potential in giving a simpler formulation for treatment planning optimization.

CHAPTER 6: Case Study II: Domain-Driven Feature Integration Pipeline in Alzheimer's Disease Progression Prediction

Alzheimer's disease causes neural damage, including brain atrophy in the patient [1]. Consequently, ventricles that contain cerebral fluid are expanded to filling those regions, which increases the proportional volume of ventricles in the brain. Therefore, abnormal growth of ventricle volume is an important indicator for estimating neural damage and, in turn, for the progression of Alzheimer's diseases. The rate of this volume-growth, i.e., neural damage, can be predicted by predictive and machine learning models using the patient's current status. These predictions help to assess the effectiveness of a particular treatment for a patient, in addition to providing some expectation of the disease timeline.

In this work, we propose and investigate the performance of multiple convolutional neural network (CNN) architectures for predicting ventricle volume biomarkers using TADPOLE competition data. We use engineered representation of a structural MRI (sMRI) scan as the prediction's primary modality. We further provide some auxiliary information from other modalities and investigate how supplementary information can benefit or harm the prediction task's performance. Finally, we demonstrate that single-modal CNN using engineered feature outperforms the winner of the TADPOLE competition, and the tree-structure multimodal CNN can improve the performance even further.

[1] Part of the content of this chapter is submitted to the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) Conference.

6.1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease, which causes brain atrophy, loss of cognitive functions, and death in severe cases. The cognitive impairment stage and AD progression are clinically measured by patient changes in minimental state examination (MMSE) or AD assessment scale-cognitive subscale (ADAS-Cog). Multiple other patterns in biomarker[2] abnormality are observed throughout research studies during the past two decades. These patterns enrich the assessment's confidence and accuracy for clinicians as well as computational and predictive models [166]. The progression models and predictive models predict the future severity of biomarkers based on patients' current feature values. These models assist physicians in quantifying their expectation of abnormality development and highlighting the effectiveness of a particular treatment for the patient.

One of the significant biomarkers in AD patients is the rate of ventricle volume change. As the gray matter in the brain deteriorates, the ventricles that contain cerebral fluid start filling those regions. This increases the proportional volume of ventricles in the brain. Therefore, the ventricle volume growth rate is an important representation of the brain atrophy rate.

This chapter suggests an engineered representation for structural MRI images to be fed to a convolutional neural network (CNN) model for ventricle volume biomarker prediction in ADNI TADPOLE competition data. The representation follows the idea in our prior work [134] in decreasing the complexity of the raw image input while maintaining the sub-regions' spatial structure. This is expected to prevent the network from capturing irrelevant and noisy patterns and increase the CNN model's generalizability potential in the absence of a large dataset. Our model has been applied to the ADNI-TADPOLE data and has shown that the overall accuracy metrics outperforming the current leaderboard's best results.

[2] Medical measurements that can indicate a disease.

In the following sections of the paper, we briefly review some of the previous works in this domain. Then, we describe our dataset and the reason for using MRI-based features. The method for representation, learning, along with the evaluation method, are discussed after that. Finally, we analyze the results and evaluate the robustness of the model.

6.2 Related Works: Alzheimer's Disease Progression Model

Ito et al. [167] suggested a regression-based disease progression model for Alzheimer's disease (AD). They demonstrate that disease severity in the baseline, along with age, APOE-ϵ4 genotype, and gender are among the strong co-variates affecting the rate of disease progression (called α). They emphasize that as patients move to the later stages, the brain deterioration gets faster.

The disease progression rate (α) is a hidden variable and is usually measured by the abnormal changes in one or multiple biomarkers. Ito et al. [167], for example, measured the progression by ADAS cognitive test. Other studies [168,169] estimated the progression using the volumetric measures of the brain subregions. Each of these biomarkers holds some information about the disease progression and stage of the patient. Nevertheless, they also contain noise and have some limitations. For example, the change in a biomarker can be caused by other disorders, not just Alzheimer's, or a frequent measurement for another biomarker might not be feasible because of its price or side effects.

The advantage of using a volumetric measurement over the cognitive test scores is that it is less subjective and more comparable across patients [168]. For this reason, we found it a more reliable index for the purpose of this paper, which is learning progression patterns from a population of patients. Therefore, throughout this study, it is important to note that we are estimating the *ventricle volume growth rate* part of α. However, to keep the notation simple, we use α for the ventricle volume growth rate.

6.2.1 AD Predictive Models: Deep Learning Approaches

An accurate prediction for the patient's biomarker progression is important to assess the patient's condition as well as evaluating the effectiveness of treatment. In this regard, multiple computational approaches are mentioned in the literature, including statistical, supervised, and unsupervised machine learning models [166,170].

Deep learning is among the recent methods suggested for developing machine learning and predictive models. In the context of disease trajectory modeling and stage prediction, typically, recurrent neural networks (RNN) are used. Mainly, in the absence of imaging data, RNNs are used for both regression, i.e., trajectory modeling [171] and classification tasks (ex. in [172]).

To have a quick evaluation of the performance of the RNN methods for ventricle volume prediction, we looked at the results in the TADPOLE competition. Despite having multiple teams with the RNN-based approach, none of them were among the winners in either longitudinal or cross-sectional biomarker progression prediction [28].

Convolutional neural network (CNN) is another category of architectures used in some previous works in this domain, mainly to deal with the image modalities [28]. Givon et al. used CNN to predict the future cognitive-scores having the current structural MRI (sMRI) image of the patient [173]. Bhagwat et al. also used a siamese neural-network (LSN) to predict future cognitive scores, i.e., MMSE and ADAS-13, having the multimodal MRI, genetic and clinical factors [174]. However, these studies use raw image features for their predictive task, which causes some concern regarding the learned model's generalizability considering the small/medium size of the dataset that is usually available in this domain.

We previously demonstrated that CNN has a good performance not just on raw image data but also on the engineered features extracted from them [134]. Intuitively, we keep some local patterns but using domain knowledge, and we remove some insignificant signals/features to reduce the possibility of the network learning unimportant patterns or noises.

In this chapter, we demonstrate the performance of CNN in predicting the ventricle volume change using the engineered features suggested by TADPOLE. According to the ventricle volume change, we apply a CNN architecture to the features extracted from MRI ROIs to predict $^{a}lpha$. The evaluation is done by comparing the predicted future ventricle volumes and the actual ones using a linear projection.

6.2.2 AD Predictive Models: Multi-modal Machine Learning

6.3 Dataset

Data used in the preparation of this paper are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[3] [175], and are pre-processed for The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) competition [28]. We compared our model's performance with the winner of the competition and other submissions with a form of deep learning approach [176]. These results are accessible through the live leaderboard up to the submission date[4].

The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD)[5] [175].

6.3.1 Gold Standard: TADPOLE Competition

TADPOLE competition was held in 2017 to compare the available predictive models' performance in predicting the future evolution of individuals at risk of Alzheimer's disease using the ADNI data. The competition evaluates the models and algorithms

[3] http://adni.loni.usc.edu/

[4] https://tadpole.grand-challenge.org/D4_Leaderboard/

[5] More information at http://www.adni-info.org/

on three tasks; a classification task for forecasting the patient's clinical diagnosis and two regression tasks for predicting the Alzheimer's Disease Assessment Scale Cognitive Subdomain (ADAS-Cog13) score and ventricle volume size (divided by intracranial volume) extracted from MRI [28]. This study concentrates on the third task. The performance of the proposed architectures in this study is compared with the outcomes of the competition winner and five other teams, who used neural-network-based approaches.

TADPOLE organizers offered three sets of data for training and validation, called D1 to D3, and a final dataset, called D4, for testing:

- D1 - a comprehensive longitudinal data set for training
- D2 - a comprehensive longitudinal data set on rollover subjects for forecasting
- D3 - a limited forecasting data set on the same rollover subjects as D2
- D4 - The test set contains data from rollover individuals, acquired after the challenge submission deadline, and used for evaluating the forecasts according to the challenge metrics.

We used D1 for training and validation, and the last record of D2 patients (D3) is used for prediction. A subset of these predictions are then evaluated with respect to the actual measurements in D4 using three error metrics [6]. The evaluation outcome was compared with the performance of other submissions. A detailed descriptive analysis of the data is demonstrated in Table 4 of the competition report paper [176]. A statistic overview of the TADPOLE data is demonstrated in Table 6.1. These numbers are reported in the live announcement presentation of the results [7].

[6] The source code for evaluation functions is available on GitHub: `https://github.com/noxtoby/TADPOLE/tree/master/evaluation`

[7] `https://www.youtube.com/watch?v=BFS9Sr0lhuM`

Table 6.1: Initial overview of TADPOLE data statistics. This numbers are reported in the presentation for the result announcement @ `https://www.youtube.com/watch?v=BFS9Sr0lhuM`. The first row, represents the records for Cognitively Normal (CN) patients, the second row is the records of patients with Mild Cognitive Impairment (MC) diagnosis, and the last row belongs to the patient with dementia or Alzheimer's Disease (AD) diagnosis.

	TADPOLE data set	**D1**	**D2**	**D3**	**D4**
	Number of Subjects	**1667**	**896**	**896**	**219**
CN	Number(%)	508 (30.5%)	369 (41.3%)	299 (33.4%)	94 (42.9%)
	Visits per subject	8.3 (4.5)	8.5 (4.9)	1.0 (0.0)	1.0 (0.2)
	Age (baseline)	74.3 (5.8)	73.6 (5.7)	72.3 (6.2)	78.4 (7.0)
	Gender(% male)	48.62%	47.15%	43.48%	47.90%
	MMSE (baseline)	29.1 (1.2)	29.0 (1.2)	28.9 (1.4)	29.1 (1.1)
	Converters **	17 (3.35%)	8 (2.17%)	-	-
MCI	Number(%)	841 (50.4%)	458 (51.1%)	269 (30.0%)	90 (41.1%)
	Visits per subject	8.2 (3.7)	9.1 (3.6)	1.0 (0.0)	1.0 (0.3)
	Age (baseline)	73.0 (7.5)	71.6 (7.2)	71.9 (7.1)	79.4 (7.0)
	Gender(% male)	59.33%	56.33%	57.99%	64.40%
	MMSE (baseline)	27.6 (1.8)	28.0 (1.7)	27.6 (2.2)	28.1 (2.1)
	Converters **	111 (13.20%)	34 (7.42%)	-	9 (10.0%)
AD	Number(%)	318 (19.1%)	69 (7.7%)	136 (15.2%)	29 (13.2%)
	Visits per subject	4.9 (1.6)	5.2 (2.6)	1.0 (0.0)	1.1 (0.3)
	Age (baseline)	74.8 (7.7)	75.1 (8.4)	72.8 (7.1)	82.2 (7.6)
	Gender(% male)	55.35%	68.12%	55.88%	51.70%
	MMSE (baseline)	23.3 (2.0)	23.1 (2.0)	20.5 (5.9)	19.4 (7.2)
	Converters **	-	-	-	9(31.0%)

6.3.2 Modality Selection

Ventricle volume biomarker progression estimates the damage to the nerve cells. The level of neural activity can measure this damage in brain regions, e.g., using PET or DTI images, or through the volume and shape shrinkage of the gray matter. Additionally, other biomarkers of the disease progression provide hints to a potential higher ventricle volume growth. Out of six modalities of features provided in TADPOLE data, we used the ones extracted from MRI modality with a cross-sectional processing pipeline as the primary predictors. The other modalities contribute to the prediction by providing auxiliary information to the MRI information.

Traditionally cognitive tests are the initial measurements for assessing the AD status and possible disease progression. However, more recent advances in technology and brain knowledge suggested other biomarkers to the assessment process of dementia progression. These biomarkers commonly belong to two categories; either they measure the *amyloid beta protein* or they estimate the *damage to nerve cells* [177]. MRI, PET, and DTI images and Cerebral Fluid test (CSF) are the main information modalities that can assist physicians in these measurements. Additionally, some individual risk-factors can impact the potential for developing dementia or the speed of the progression. These factors include, and are not limited to, demographic information, such as age, and genetic factors, such as Apolipoprotein E $\epsilon 4$ allele (APOE4) [178]. Table 6.2 describes the available information modalities in TADPOLE data, the number of features in each one, and the availability of values for the visits and patients.

Table 6.2: The feature sources and modalities provided by TADPOLE competition organizers, and the missing portions in the union of D1 and D2 data.

	Feature Source	**Modality**	**No. of Features**	**Missing Records (%)**	**Missing Patients (%)**
1	Key ADNI Features	Mixed (demo, cog-tests, etc)	83	33.31	0
2	Cross-Sectional (FreeSurfer V-4.3)	MRI-image	363	39.02	36.79
3	Banner Alzheimer's Institute PET NMRC Summaries	PET-image	339	83.12	34.12
4	UPENN CSF Biomarkers Elecsys	CSF-sample	10	82.97	49.80
5	Longitudinal FreeSurfer (FreeSurfer Version 4.4)	MRI-image	371	67.18	50.89
6	UC Berkeley - AV45 analysis	PET-image	238	83.47	57.63
7	DTI ROI summary measures	DTI-image	238	93.82	86.93
8	UC Berkeley - AV1451 analysis	PET-image	243	99.31	96.26

The MRI modality can quantify the nerve cells' atrophy by measuring the volume of gray (GM) and white matter (WM) of the brain. The GM is the brain tissue consists of nerve cells, and WM is the fibers connecting those nerve cells. When the neurons in a region die, the volume of that region shrinks. Thus, atrophy can be measured by volume loss in a region from an MRI scan in one session to the follow-up one. This quantification is important because MRI is a non-invasive and widely available modality.

6.4 Representation

Each record in the dataset belongs to one visit of a patient, which contains one or multiple biomarkers that are measured throughout the visit. From a representation perspective, some of these measures are low-dimension values and stand at a higher semantic level, while others are high-dimensional and semantically at a lower level. To make the notation easier, the rest of this manuscript referrers the high and low dimension features as *modality A* and *modality B*, respectively.

The primary measure we use for the prediction task is a structural MRI image. This feature modality is referred to as the *modality A* in the experiments, and we discuss the reasoning for this choice in the following subsections. Low-dimension values provide other auxiliary information from multiple sources, such as diagnosis, ADAS13 test score, TAU protein levels, and a higher level of knowledge than the modality A. As we mentioned in the previous paragraph, these features are referred to as *modality B*.

In the following subsections, we explain the features organization in each of these modalities and the pre-processing pipeline for each feature category.

6.4.1 Modality A: Low-level Feature Engineering

To extract information from the MRI images, instead of using the raw image data, we use the engineered features provided by the competition organizers for the regions of interest. This approach gives us more control over the network patterns based on our knowledge from the domain. Moreover, it decreases the size of the learning variables and complexity of the network. Finally, the spatial and semantic information remains in the data, and we can interpret the learned model easier.

According to the competition website, the MRI-based markers were measured after registering (i.e., aligning) the MRI images with each other and performing a segmentation of the relevant brain structures using an atlas-based technique. These markers were selected based on the domain literature and previous works [166]. The values are computed with an image analysis software called FreeSurfer using two pipelines: cross-sectional (each subject visit is independent) or longitudinal (uses information from all subject visits). While the longitudinal measures are more robust, as Table 6.2 shows, there are more missing values for those measures in TADPOLE data. Therefore, we used the cross-sectional features in our analysis.

We organize the MRI-based features in a 1D spatial tensor of 72 ROIs with 4-channels of features for each patient's visit. In Table 6.3, we describe a summary of the information in these four channels. We extract the label for that visit by calculating the ventricle volume change rate from the current visit to the patient's next visit.

The independent predictors we use are three categories of atrophy markers extracted from structural MRI and provided in the TADPOLE data. These markers include *ROI volumes*, *ROI cortical thicknesses*, and *ROI surface areas*, where ROI is a 3D sub-region of the brain such as the inferior temporal lobe. TADPOLE data provided volume information for both GM and WM, but we only used the GM volume in our analysis. We also did not use other MRI-features contained quality measurement and metadata. For the cortical thickness, we used both average and standard deviation of GM thickness across the region.

6.4.2 Modality B: High-level Domain Knowledge

Each record of the data is a snapshot of a patient's feature at a specific time. Therefore, as expected, a memory-free model such as CNN that we are using would miss the background information about disease evolution and stage. Besides, there are some particular risk-factors for each individual that can increase or decrease the patient progression pattern.

In an ideal situation with a sufficient amount of data, the MRI-based values in modality A could be sufficient to capture these patterns. However, in reality, not having a large set of data for a CNN model could cause overfitting to a non-generalizable pattern. Therefore, we provide the *stage-related* information as well as *risk-factors* in a high-level numerical representation as a supplement to the data from *modality A*. We selected these features according to the previous studies from the knowledge of the domain.

The stage-related features are selected based on a publication regarding the disease progression model [151], and the individual-based factors are selected based on the early study by Ito et al. [167]. These features and the descriptive analysis of their corresponding data are demonstrated in Table 6.5.

6.5 Preprocessing

6.5.1 Data Normalization

TADPOLE competition uses the Intracranial Volume (ICV) at their baseline - i.e., first visit (ICV_{bl}), to normalize the volume of the ventricles. This normalization makes the ventricle volume measurements comparable across patients and prevents the unintended bias in target values due to the patients' size of the skull.

We use a similar approach to normalize all the volume features by ICV_{bl}. We then remove the ICV from the cortical volume measures to prevent it from suppressing the other features. We normalize the other three categories of measures (surface area, cortical thickness average, and cortical thickness standard deviation) by the mean and standard deviation of the training population's values for all the corresponding category features. Equation 6.1 shows the normalization formula in which x is the original number belongs to feature f on record r, and f belongs to category Cat with

c features. The train-set contains n records.

$$\begin{aligned}
norm(x_{fr}) &= \frac{x_{fr} - Cat_{mean}}{Cat_{std}} \\
Cat_{mean} &= \frac{\sum_{f=1}^{c} \sum_{r=1}^{n} x_{fr}}{c * n} \\
Cat_{std} &= \sqrt{\frac{\sum_{f=1}^{c} \sum_{r=1}^{n} x_{fr} - Cat_{mean}}{(c * n) - 1}}
\end{aligned} \tag{6.1}$$

Eventually, we organize the features in a one-dimension tensor of 72 regions, in the four channels, explained in Table 6.3.

Table 6.3: Definition of the representation of modality A, i.e. MRI, using low-level engineered features. Features are define in four channels, and each channel is normalized separately. For applying the equation 6.1, the mean and standard deviation of the whole training set (D1) is being used.

ch.	Feature category	Normalization method
0	ROI cortical volume	Devision by ICV_{bl}
1	ROI surface area	Equation 6.1
2	ROI cortical thickness	Equation 6.1
3	ROI cortical thickness standard deviation	Equation 6.1
	Modality B Features	Equation 6.1

If for an ROI, we don't have a value of one or multiple features (ex. cortical volume for ICV, which was removed from the features), we simply fill it with zero. To prevent the dying signal issue during the learning process, we multiply all the values with a strength-factor of 1000. However, all the evaluations are reported after taking this factor off of the results.

The modality B features are similarly normalized by the Equation 6.1 regardless of being discrete or continuous. This way of normalization will retain the main characteristics of the data distribution.

6.5.2 Missing Data Imputation

A patient may have up to 19 records, where each record is demonstrating the information from one visit. This is shown in Figure 6.1. Each visit measured some of the features and may lack the values from others.

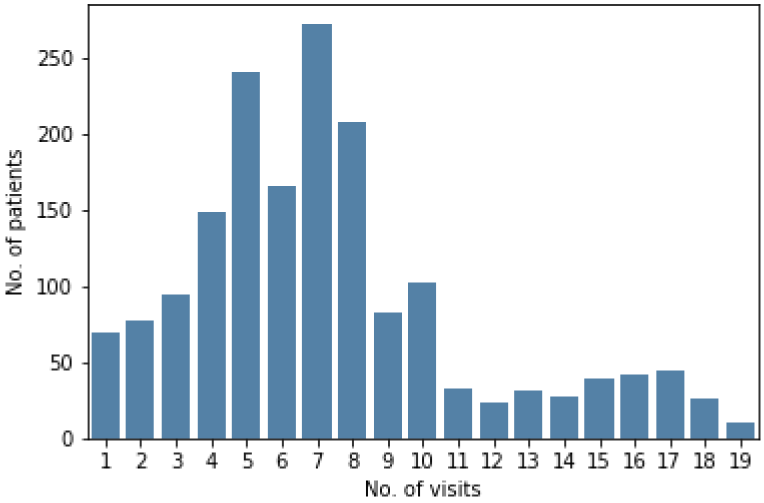


Figure 6.1: Number of patients with x number of visits

The missing values for the MRI-based features, including ventricles, are imputed with a linear averaging over the prior and the following visits if they exist. If these missing spots are not surrounded by the known values from the prior and later visits of the same patient, with an exception for ventricles, we simply fill all others with zero.

For the patient-level features of modality B, such as education and APOϵ4, we repeated the values from baseline, if they exist, for the other records of the patient. The progression features which are visit-based, i.e. $Abeta$, $Ptau$ and $Hippocampus-volume$, are imputed using the same linear assumption in Equation 6.2.

As it is described in Equation 6.2, for patient p, if a record in time t with missing MRI-based values X_{pt} is surrounded by two known records at time t_1 and t_2, we calculate X_{pt} by a linear weighted average of the surrounded known visits of the same patient.

$$x_{pt} = \frac{x_{t_2} - x_{t_1}}{t_2 - t_1} \tag{6.2}$$

The previous computational model [167] suggests an exponential volume change progression. However, for the short periods of visits, we found the linear assumption a reasonably simple method to prevent the injection of bias by the imputed data. After imputation, the remaining records with no MRI-based information, i.e., the edge records not surrounded by two known records, were removed. Two records are also removed for having alphas that were 100 times greater than the maximum alpha of the rest of the records. The assumption is that either the records are capturing some miscalculation, noise, or extreme outliers. The number of patients and records during each of these steps is in Table 6.4.

Table 6.4: The number of patients and records in each pre-processing step.

	Original	**Imputed**	**Dropped**	**Final**
Patients	1737	1047	23	**1714**
Records	12741	2185	3111	**9630**

The diagnosis information, which is used in our last experiment, is imputed by repeating the last known diagnosis of the patient for the following missed sessions. We trained the network with the imputed and actual records. However, all the train, validation, and test evaluations are reported solely based on the prediction for the actual records.

6.6 Evaluation Method

6.6.1 Data Splitting

The data is split into training and validation using a five-fold cross-validation approach. The division is conducted on patients rather than visits so that no records

Table 6.5: The descriptive analysis of candidate features for modality B. The Gini-index is calculated using Equation 6.10, the SDR-Alpha is the standard deviation reduction in alpha-ventricles if the records will be divided into two groups by the feature on the specified breakpoint. The breakpoints are selected using the regression tree process on D1-D2 datasets. The data normalization and feature imputation for all these features are done using Equation 6.1 and 6.2, respectively.

Feature	**Gini-index**	**SDR-Alpha**	**Missing (%)**	**Break Point**
ADAS13	0.485	2.477	26.640	-0.426
TAU	0.238	4.588	78.548	1.216
Diagnosis	0.450	1.539	0.094	-0.485
PTAU	0.495	0.398	56.650	-0.145
Education	0.234	0.648	0	1.258
WholeBrain	0.391	1.247	21.909	-1.607
MMSE	0.491	-0.003	25.840	0.449
APOE4	0.165	0.038	0.090	1.468
Age	0.159	0.663	0	-1.509
Hippocampus	0.279	0.080	32.203	0.970

from the patient in the training data are used for validation. Each visit is represented by the MRI-based measures, discussed in Section 6.3.2, and is passed to the model. The model is expected to predict the ventricle volume change incline, i.e., α, which approximates the brain deterioration rate. With a linear assumption, as mentioned in Equation 6.7, the ventricle volumes in future dates are predicted by multiplying α with the period length (in months).

In order to compare the results of this model with the ones from other TADPOLE participants, the patients' feature is extracted from their last record. The predicted α is then used for calculation of patients' ventricle volumes in the following potential 60 month-by-month sessions. At the time of competition, the test set was not available, so they asked for monthly predictions for the next five years. The evaluation process is then conducted by comparing the actual measurements during two years after the competition, called D4, with the corresponding record in our submission [8].

[8] The source code for evaluation functions is available on GitHub: `https://github.com/noxtoby/TADPOLE/tree/master/evaluation`

6.6.2 Evaluation Metrics

We calculated the same metrics for our prediction as to the ones on the leaderboard. This way, we can compare the results from our model to the previous competitors. The main metric for ranking the submissions in the competition is the mean absolute error (MAE), but they provide complimentary comparison using Weighted Error Score (WES) and Coverage Probability Accuracy (CPA) metrics, as well. Therefore, to have a more concise report, we report all the three metrics for the test-set evaluation, but only MAE for the train and validation evaluation.

6.6.2.1 Mean Absolute Error (MAE)

As mentioned, this is the main metric that submitted results are ranked on. This metric punishes the mistakes in a linear behavior. In other words, making big mistakes in few records, for example, 50% error in 1% of the records, is equivalent to making small mistakes in a large number of the records, 1% error in 50% of the records in the example. The formulation for this calculation is shown in Equation 6.3, where N is the number of records for prediction, and $\tilde{V_i}$ and V_i are the predicted and actual values for Ventricles/ICV, respectively.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\tilde{V_i} - V_i| \tag{6.3}$$

6.6.2.2 Weighted Error Score (WES)

This error metric accounts for the confidence of the predictions. Unlike the linear nature of MAE, this metric punishes errors with high-confidence more than the lower ones. Of course, this means that if all the confidence-intervals are similar, this is going to be the same as MAE. Equation 6.5 shows how this metric is being calculated, where N is the number of records, and $\tilde{C}_-$ and $\tilde{C}_+$ are the low and high bounds of the prediction intervals, and $\tilde{V_i}$ and V_i are the predicted and actual values for

Ventricles/ICV, respectively.

$$\tilde{C_i} = (C_+ - C_-)^{-1} \tag{6.4}$$

$$WES = \frac{\sum_{i=1}^{N} \tilde{C_i}|\tilde{V_i} - V_i|}{\sum_{i=1}^{N} \tilde{C_i}} \tag{6.5}$$

6.6.2.3 Coverage Probability Accuracy (CPA)

The CPA error metric accounts for the accurate coverage expectation from a confidence interval. As mentioned in Subsection 6.8.2, we calculate a 50% confidence interval for each prediction. This means that we expect the actual value to fall in this interval 50% of the time. Clearly, the big intervals have more coverage than the smaller ones. Therefore, the CPA error metric punishes higher and lower coverage than the target, i.e. 50%, equally. This means that the interval should neither be too vast nor too narrow to keep the coverage as close to 0.5 as possible. This error metric is measured by the Equation 6.6, where $\tilde{C}_-$ and $\tilde{C}_+$ are the low and high bounds of the prediction intervals, and V_i is the actual value for Ventricles/ICV.

$$\begin{aligned} CPA &= |\text{actual coverage probability} - \text{nominal coverage probability}| \\ &= |\frac{count(\tilde{C}_- < V_i < \tilde{C}_+)}{N} - 0.5| \end{aligned} \tag{6.6}$$

6.7 Learning Models: Convolutional Neural Network

6.7.1 Single Modal Architecture

In order to capture the abnormal patterns of cortical volumes, we design a six-level convolutional neural network. Each level contains two consecutive 1D-convolutions followed by a max-pooling layer. We select a kernel size of three for the convolution blocks and a pooling size of two for each pooling operation. The patterns are then passed to a block with two fully-connected layers for the regression analysis and prediction of α.

As it is demonstrated in Figure 6.2, the input to the CNN architecture is a batch of 1D four-channels tensor, and the output is a single value. The activation function for all the convolution blocks is a Leaky ReLU function with a negative slope of 0.1. After the six convolution levels, the values are flattened into a 1D tensor and go through two fully connected layers. The activation functions in these layers are linear due to the regression nature of the learning task. This linear function lets the network to be able to learn any real value (positive and negative). The expected output of the learning model is a single number, which represents the expected brain deterioration or ventricle growth rate.

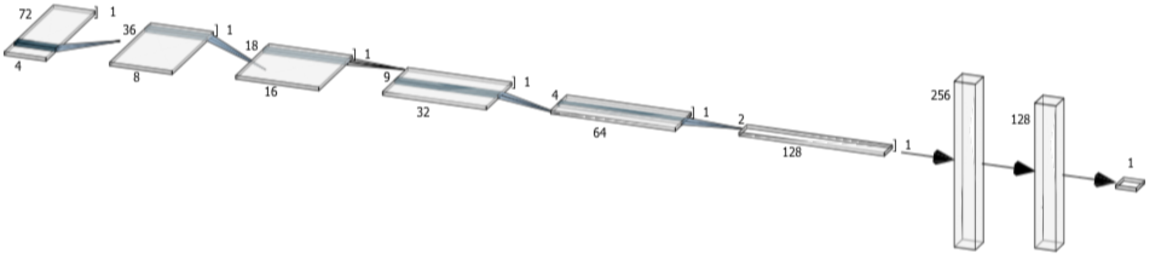
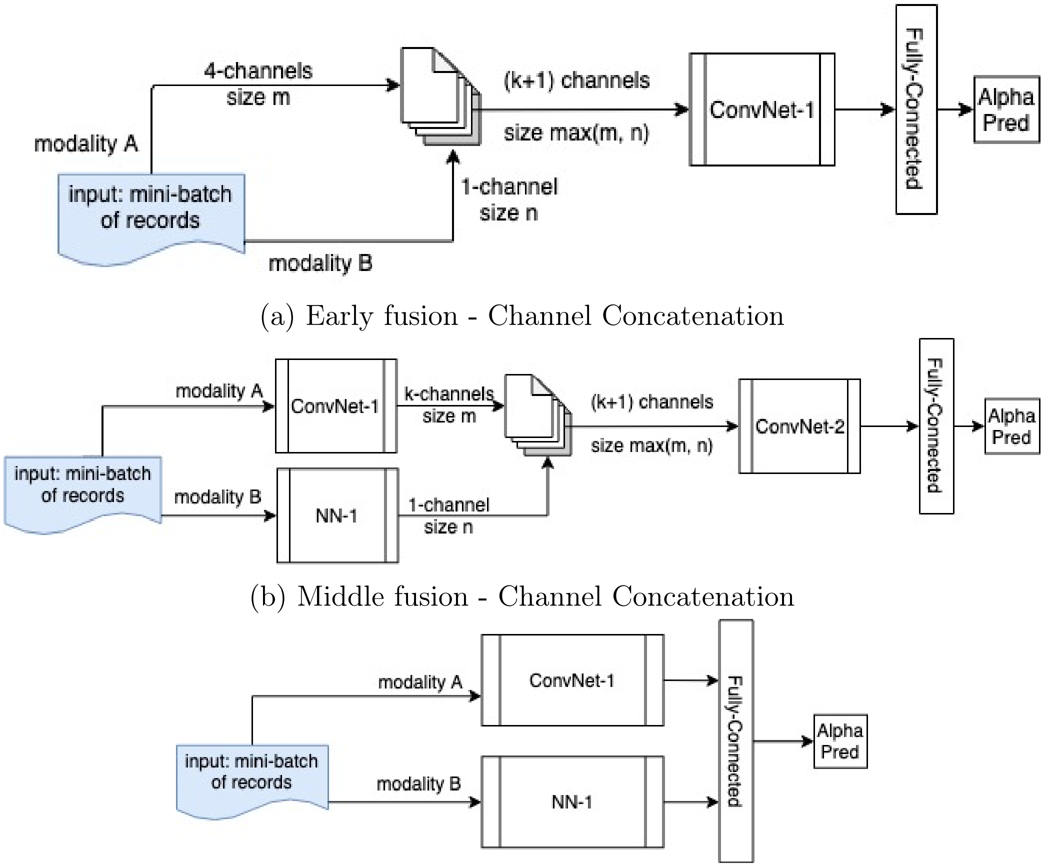


Figure 6.2: Convolutional neural network architecture for ventricle volume growth prediction. The architecture consists of 6 convolution level. In each level two convolution (kernel=1x3) is applied consequently. The a pooling with kernel=1x2 downsamples for the next level.

The network weights are initiated by Xavier normal approach [179], and biases by zeros. We use the L1 for the network loss function as it is described in Equation 6.3 and the Adagrad algorithm [180] was as our optimization approach. The learning rate is evolving through the experiment, starting from 0.01. When the validation is not improved for three consecutive epochs -i.e., plateau, the learning rate reduces by a factor of 0.5. The threshold for learning rate decrements is 0.0001.

6.7.2 Multimodal CNN: A Fusion Model

Learning feature and extracting from multiple modalities using deep learning was initially proposed by Ngiam et al. [34] for two high dimensional sources of data such as audio and video. In that paper, they demonstrated the higher performance of a fused representation for a learning task over learning on two independent networks. Since then, various architectures are proposed for the fusion of information from multiple modalities of data. Among those, early, middle and late fusions are the most common ones in a variety of applications [30, 132, 181]. We implement early, middle, and late fusion architectures, Figure 6.3, to explore the potential of these fusion approaches for integration of data from modality A and B in neural deterioration progression prediction.



(a) Early fusion - Channel Concatenation

(b) Middle fusion - Channel Concatenation

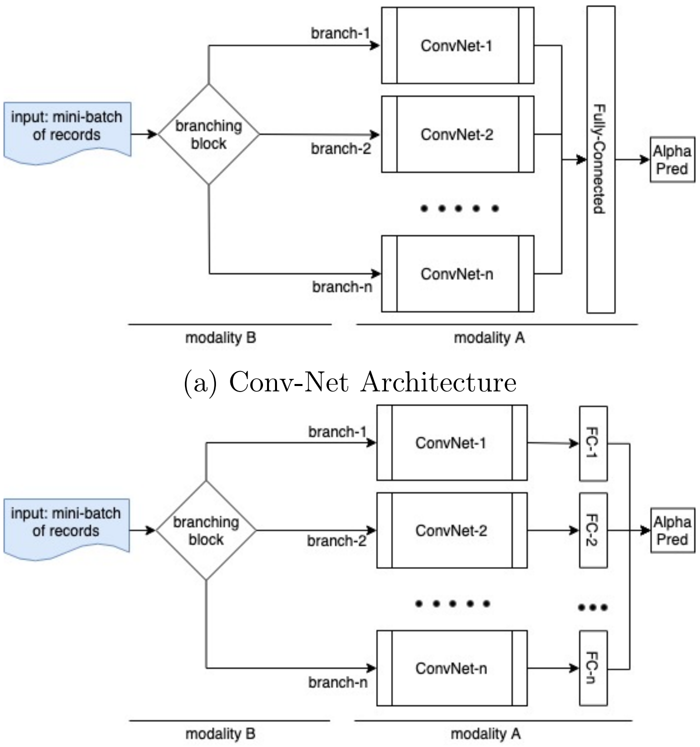
(c) Late fusion architecture

Figure 6.3: Fusion-Structure CNN architectures

6.7.3 Multimodal CNN: A Tree-Structure Model

In the previous part, we assume one model is sufficient to predict the α for all the patients. From the prior studies, we know that the biomarker abnormality pattern is different in each stage of the disease [151, 170, 174]. Prior to this, we also knew that some individual-level features of the patients -e.g., age, education, and APOE4 genetic factor, have a correlation with the AD progression and brain deterioration rate [167]. We assume that in an ideal scenario with sufficient data, a CNN/RNN model can learn the patterns either from the modal A alone or through a fusion can learn. However, in the absence of sufficient data, we use some prior analysis to lead the model. This is the idea behind the tree-structure CNN in this section.

We propose two tree-structure architectures; convolution and full branching, which are demonstrated in Figure 6.4.



(a) Conv-Net Architecture

(b) Full-Net-Architecture

Figure 6.4: Tree-Structure CNN architectures.

6.8 Experiment Setting

The model is trained and validated in a five-fold cross-validation setting. The division is conducted on patients rather than visits so that no records from the patient in the training data are used for validation. Each visit is represented by the MRI-based measures, discussed in Section 6.4.1, and is passed to the model. The model is expected to predict the ventricle volume change rate -i.e., α, which approximates the brain deterioration rate. With a linear assumption, as mentioned in Equation 6.7, the ventricle volumes in future dates are predicted by multiplying α with the period length (in months).

The last visit of each patient is used to test the model and predict their ventricle volumes in the following potential 60 month-by-month sessions. This is because, at the time of competition, the test set was not available, so they asked for monthly predictions for the next five years. The evaluation process is then conducted by comparing the closest record in the actual measurements, called D4, which was measured during two years after the competition[9]. A summary of the experiment pipeline is shown in Figure 6.5.

6.8.1 Initiation and Hyperparameter Setting

The hyperparameters of the network are the ones that had the best validation performance on the single modal architecture. These settings are constant for most experiments unless we mention otherwise. Table 6.6, demonstrates these hyperparameters and the corresponding values.

Table 6.6: These parameters are chosen during the validation phase. There are other setting in the network design (such as dropout=0), which was not explored and can be analyzed in the future.

Parameter	**Value**
Depth of the network	6
Filter count for the first layer	8
Learning rate	0.0001
Convolution kernel size	3
Pooling size	2
Optimization method	Relu with negative slope=0.1

[9] The source code for evaluation functions is available on GitHub: `https://github.com/noxtoby/TADPOLE/tree/master/evaluation`

6.8.2 Confidence Interval Calculation

We get five different trained models from a five-fold cross-validation approach. This means that for the last visit of each patient, we have five different predictions for α. The average for these predictions are reported as predicted alpha or $\tilde{\alpha}$, and the confidence interval is being measured by Equation 6.8, assuming that alpha follows t-test conditions. In this equation, $t_{CL=0.25}$ is the t-factor for confidence interval of 0.25, and $\sigma_{samples}$ is the standard deviation of the five predictions for α. In this equation, df is the degree of freedom in the samples. As we have five samples from five-fold experiments, we consider having four degrees of freedom.

As it is shown in Equation 6.7, the actual ventricle/ICV growth or α is calculated using the ventricles/ICV of two visits with volume change of $\Delta(V)$ across a time interval $\Delta(time)$. Inherently, the confidence interval for the ventricle/ICV is the predicted interval for alpha multiplied by a factor of time. In other words, the distance from the current point increases uncertainty. As Equation 6.9 shows, we further increase this uncertainty by another square-root factor of time-based on the experimental values during the validation phase.

$$\alpha = \Delta(V)/\Delta(time) \tag{6.7}$$

$$\tilde{C}_{\alpha} = \tilde{\alpha} \pm t_{CL=0.25} \times \frac{\sigma_{samples}}{\sqrt{df}} \tag{6.8}$$

$$\tilde{C}_{V} = \tilde{C}_{\alpha} \times \Delta(time) \times \sqrt{\Delta(time)} \tag{6.9}$$

6.9 Single Modal Experiments

The best result on the TADPOLE competition live leaderboard for the ventricle prediction task belongs to a disease progression model [151]. The team's predictions were evaluated after a year upon the acquisition of new measurements for the patients [176]. On average, the (five) teams with neural network approaches perform reasonably well in ventricle prediction and are in the top half of the leaderboard. Yet, none of the five submitted approaches beats the winner. All of the five methods have a recurrent neural network architecture. Table 6.7 demonstrates the rank, features, methods, and performance of these teams as well as the winner. As Table 6.7 shows, our convolutional neural net model outperforms the leaderboard winner and other neural-network-based submissions for ventricle volume prediction task [10].

Table 6.7: Summary of methods, features, and evaluation of the winner as well as all the teams who used deep learning approach for ventricle prediction task in TADPOLE competition. The results are based on our last visit on June 20th, 2020. More information about the features and methods can be found in the competition paper [28].

Vent Rank	**Team Name**	**Method**	**Vents MAE**	**Vents WES**	**Vents CPA**
1.5	EMC1 Std/ Custom[a] [151]	DPM SVM & 2D-spline	0.41	0.29	0.43
7.0	CN2L-Neural Network	Three-layer RNN	0.44	0.44	0.27
13.0	CBIL [20]	Multimodal LSTM	0.46	0.46	0.09
16.0	CN2L-Average	Average of Three-layer RNN & Random Forrest	0.49	0.49	0.33
24.0	BravoLab	LSTM	0.58	0.58	0.41
25.0	BGU-LSTM	LSTM	0.60	0.60	0.23
	Benchmark[b]	Mixed effects modeling	0.56	0.56	0.5
	Crowdsourced[c] [28]	Consensus Median	0.38	0.33	0.09
	Our Result	Convolutional Neural Net	**0.39**	0.3	**0.1**

[a] Winner of the competition in ventricles-icv prediction part.
[b] This is the best benchmark out of five benchmark approaches.
[c] This is the best crowd-sourced results out of three methods of aggregation -i.e., mean, median, best.

[10] The complete live leaderboard can be found at `https://tadpole.grand-challenge.org/D4_Leaderboard/`.

We further assess the robustness and sensitivity of the results to the unknown/random factors through extensive sets of experiments discussed in the following subsections and summarized in Table 6.8.

6.9.1 Robustness Analysis

We evaluated the robustness of the model by observing the behavior of the model for other random values for the seeds, the number of epochs, and the order of the sub-regions in the feature vector. Table 6.8 displays the result of the model with different data-seed and model-seeds. Finally, we can see with a small number of training epochs, we have relatively good results, and more epochs only slightly improve the results.

Table 6.8: Results' robustness: The MAE is not highly sensitive to the random seeds and the number of epochs. The CPA and WES have more volatility, but it is still in a comparably range with the original run and the results from the winner. It is important to mention that in train and validation sets, the predictions are done for the next visit, which is on average six month ahead but on the test (D4) this time-span is up to five years ahead. Therefore, the loss of D4 is a factor of five or so more than train-validation.

Experiment	**vents Train MAE**	**vents Val MAE**	**vents MAE**	**vents WES**	**vents CPA**
Winner			0.41	0.29	0.43
CNN-Orig-Exp	0.08	0.08	0.39	0.30	0.10
Dataseeds (10 different values)	0.08 ± 0.00	0.08 ± 0.00	0.39 ± 0.0047	0.30 ± 0.032	0.13 ± 0.0626
DataSeedsPreds-combined			**0.39**	0.30	0.08
ModelSeeds (10 different values)	0.08 ± 0.00	0.08 ± 0.00	0.39 ± 0.0032	0.31 ± 0.0135	0.04 ± 0.0320
ModelSeedsPreds-combined			**0.39**	0.31	0.01
Epochs (20, 50, 100, 200, 300, 400, 500, 700, 1000)	0.08 ± 0.00	0.08 ± 0.00	0.39 ± 0.0000	0.31 ± 0.0163	0.03 ± 0.0255
EpochsPreds-combined			**0.39**	0.31	0.07

6.9.2 Feature Ordering

The volume-based representation maps the three-dimensional spatial information into a one-dimensional space. Therefore, the ordering of the features cannot be preserved. Having the highest volatility, Table 6.9, in comparison to the other factors, Table 6.8, it is important to note the impact of feature ordering in keeping some spatial patterns. Nevertheless, the worst result is not drastically changing the performance and is still comparable to the winner performance. Despite being obvious, it is worth mentioning that we use the same feature ordering for the test set as the one that the network is being trained on.

Table 6.9: The MAE has a higher sensitivity to changes in the ordering of the regions. However, the range of changes are still comparable with the original results and the ones from the winner. This means that despite having some optimum regional layouts, not knowing the best layout will not worsen the performance, drastically.

Experiment	**vents Train MAE**	**vents Val MAE**	**vents MAE**	**vents WES**	**vents CPA**
CNN-Orig-Exp	0.08	0.08	0.39	0.30	0.10
RegionOrders (10 different random orders)	0.08 ± 0.00	0.08 ± 0.00	0.39 ± 0.0070	0.28 ± 0.0166	0.04 ± 0.0308
RegionOrder-combined			**0.39**	0.27	0.10

6.9.3 Interpretation Limitation

To have a better interpretation of the performance, we should take into account multiple aspects of the problem. Firstly, the disease progression curve might be different from the neural damage part. This means that for this particular task, the ratio of information to noise is much less when we only take the brain's volumetric information rather than the rest of the measures. Secondly, the ventricle growth is a consequence of brain atrophy, but this atrophy can be either Alzheimer's or other types of dementia.

Lastly, the disease pattern has both cross-sectional and longitudinal patterns. By selecting a convolution net over the recurrent neural net, we concentrated on a cross-sectional pattern and replaced the temporal information we could capture from the data with a simple linear model assumption. Thus, this model is performing better than the available temporal models for the near future, e.g., less than five years, which is our data. Still, the error on average grows over time as it is observable in Figure 6.6.

6.10 Multimodal Experiments

The single modal CNN, as mentioned in 6.9, already outperforms the leaderboard. The single modal prediction is only aware of the current state of the brain regions' volume. A physician, however, would typically take other patient's information into account for making a prediction. This includes the history and individual characteristics of the patient. To reflect this practice into our computation model, we hypothesize that the extra information provided by other modalities can improve the prediction performance. Subsection 6.10.1 explains the hypotheses and some preliminary evaluation. The following subsections evaluate multiple approaches for including the other information from modalities into the current model.

6.10.1 Hypotheses Analysis

According to the previous studies in the literature, the disease progression rate gets worse in the later stages [167]. Furthermore, the individual characteristics such as the $APOE\epsilon4$ genetic factor have an association with the progression rate [167]. Following this domain knowledge, we make two hypotheses:

- H1: The neural deterioration progression rate looks differently for the patients in different stages.

- H2: The progression slope or overall curve shape is different for individuals with different characteristics, such as genetic factors.

The Ventricle-ICV and Alpha distribution for the records of three different AD stages is demonstrated in Figure 6.7. Particularly, the difference between Figures 6.7a and 6.7c supports the first hypothesis. It is also in line with the discriminative approach suggested by the TADPOLE winner [151].

Figures 6.8c and 6.8d demonstrate how median of Ventricles-ICV volume and progression rate are different in each genetic group. The Figures 6.8b and 6.8a further demonstrates the progression differences with respect to the patients' initial (baseline) diagnoses.

These initial analyses clearly suggest that at least the patient stage and genetic factor can provide some beneficial information regarding the curve shape.

As we discussed, the CNN model we designed does not have an understanding of the temporal aspect, i.e., the patient's history. We expect that providing a complete picture of the patient's current status compensates for part of that weakness.

The differences between MCI subgroups in Figure 6.8 further suggests that the initial three clinical diagnosis labels, i.e., cognitive normal (CN), mild cognitive impairment (MCI), Alzheimer's Disease (AD), may not be sufficient for conveying the patient's stage. Similarly, the $APOE\epsilon4$ shows the impact of individual genetic differences. One can also take other individual differences into account.

To better capture the stage information, we include the features mentioned in a prior work [151] to the diagnosis, and we use Age, $APOE\epsilon4$ and $Education$, as mentioned in another work [167], for capturing the individual differences.

6.10.2 Multi-modal: Fusion Analysis

We implemented the fusion framework discussed in Section 6.7.2 using all the features in modality B. Table 6.10 shows the results of different levels of fusion - look at Figure 6.3, for a convolutional neural network with original depth of six. As the results demonstrate, there is no monotonic relationship between fusion depth and performance. This is one example of the cases that we cannot predict what the most effective fusion level for integrating information of the modalities is. It is important to note that fusion at depth five in this example performs significantly worse than the original single modality. Even the late fusion method, which is the most popular fusion approach in heterogeneous modalities, is not adding more value to the single modal method. This is one reason that we call the common fusion approaches **blind method**.

Table 6.10: The results for different levels of fusion combining all of the modality B features. These results are based on training for **30 epochs**, and by adding batch normalization and dropout with probability 0.2 to the convolution architecture. All the numbers in the table are rounded to two decimal points and sorted ascending based on D4-MAE. As it is shown on the table, the best result on D4 (test data) belongs to ADAS13 as the branching feature, which significantly outperforms the winner and our original result.

Fusion Type	**Fusion Depth**	**Train-MAE**	**Val-MAE**	**D4-MAE**	**D4-WES**	**D4-CPA**
Middle Fusion	2	0.08	0.08	0.37	0.30	0.17
Early Fusion	0	0.08	0.08	0.37	0.30	0.03
Middle Fusion	3	0.08	0.08	0.38	0.31	0.03
Late Fusion		0.08	0.08	0.38	0.38	0.10
Middle Fusion	1	0.08	0.08	0.40	0.34	0.15
Middle Fusion	4	0.08	0.08	0.41	0.29	0.03
Middle Fusion	5	0.22	0.23	0.53	0.32	0.04
OrigCNN		0.09	0.09	0.38	0.31	0.03
Winner				0.41	0.29	0.43

6.10.3 Tree-Structures: Auxiliary Feature Analysis and Evaluation

Decision tree and neural networks have some fundamental differences in the mechanisms they utilize for learning the statistical patterns. There are some advantages and disadvantages to each of these approaches. Using the architectures explained in Section 6.7.3, we benefit from some of the decision tree mechanisms to improve the previous convolutional neural network architecture in the absence of a large dataset. Table 6.11 shows the results for a binary-tree with a deterministic branching method, and the Table 6.12, shows the results using a non-deterministic method such as a weighted-Sigmoid block for branching. The results are more interpretable and have some hint for a future model integration method.

$$Gini - index = 1 - \sum_{i=1}^{n}(p_i)^2 \tag{6.10}$$

Table 6.11: The results for a binary branching using modality B features. These results are based on training for **30 epochs**, and by adding batch normalization and dropout with probability 0.2 to the convolution architecture. Each row of the table demonstrates an experiment using one of the features from modality B to do the branching. The breakpoint is selected through the standard regression model tree process by selecting the value that provides maximum standard deviation reduction. All the numbers in the table are rounded to two decimal points and sorted ascending based on D4-MAE. As it is shown on the table, the best result on D4 (test data) belongs to APOE4, WholeBrain, and ADAS13 as the branching feature, which outperforms the winner and our original result.

Feature	**Gini-index**	**SDR-Alpha**	**Miss (%)**	**Break Point**	**Arch**	**Train-MAE**	**Val-MAE**	**D4-MAE**	**D4-WES**	**D4-CPA**
APOE4	0.17	0.04	0.090	1.47	conv	0.09	0.08	**0.37**	0.33	0.09
WholeBrain	0.39	1.25	21.91	-1.61	conv	0.08	0.08	**0.37**	0.32	0.03
ADAS13	0.49	2.48	26.64	-0.43	conv	0.08	0.08	**0.37**	0.35	0.09
TAU	0.24	4.59	78.55	1.22	conv	0.08	0.08	0.38	0.36	0.13
Education	0.23	0.65	0	1.26	full	0.12	0.12	0.38	0.32	0.15
Age	0.16	0.67	0	-1.51	conv	0.08	0.08	0.38	0.34	0.33
Age	0.16	0.67	0	-1.51	full	0.10	0.09	0.38	0.34	0.34
Hippocampus	0.28	0.08	32.20	0.97	conv	0.09	0.10	0.38	0.33	0.11
WholeBrain	0.39	1.25	21.91	-1.61	full	0.10	0.10	0.40	0.34	0.19
PTAU	0.49	0.40	56.65	-0.14	conv	0.09	0.09	0.40	0.29	0.08
PTAU	0.49	0.40	56.65	-0.14	full	0.10	0.10	0.40	0.37	0.24
ADAS13	0.49	2.48	26.64	-0.43	full	0.10	0.10	0.40	0.39	0.21
TAU	0.24	4.59	78.55	1.22	full	0.09	0.10	0.41	0.33	0.01
MMSE	0.49	-0.00	25.84	0.45	conv	0.09	0.09	0.41	0.35	0.32
Education	0.23	0.65	0	1.26	conv	0.10	0.11	0.41	0.32	0.04
Diagnosis	0.45	1.54	0.09	-0.48	conv	0.08	0.08	0.42	0.37	0.04
Hippocampus	0.28	0.08	32.20	0.97	full	0.13	0.12	0.44	0.36	0.11
APOE4	0.17	0.04	0.090	1.47	full	0.12	0.12	0.45	0.33	0.05
Diagnosis	0.45	1.54	0.09	-0.48	full	0.13	0.13	0.49	0.49	0.20
MMSE	0.49	-0.00	25.84	0.45	full	0.13	0.14	0.56	0.33	0.21
OrigCNN						0.09	0.09	0.38	0.31	0.03
Winner								0.41	0.29	0.43

6.11 Conclusion and Future Works

In this chapter, we suggested a volume-based convolutional neural network, which uses engineered features instead of raw intensity values of structural MRI for predicting ventricle volume change in AD patients. We demonstrated that this model outperforms the prior winner of the TADPOLE competition. Additionally, we did further robustness analysis and observed comparable performance with the winner despite changing the number of epochs, data, model random seeds, and even feature order. Also, we observed the lowest sensitivity of the model in the number of epochs, and the highest sensitivity belonged to the feature ordering.

After the result analysis, we hypothesized that deterioration for patients at a different stage and different individual characteristics follow a different distribution. We used fusion as well as a more careful tree-structure convolution neural network (tree-CNN) methods for integrating the modality information. Getting the best result from tree-CNN with branch factor ADAS13 validated the hypothesis and importance of the patient stage. It further confirmed that a careful knowledge-based feature fusion could provide better results than a blind fusion method. We further observed that a branching feature that has the highest standard deviation reduction is not necessarily the best branching feature. However, a trade-off between data size and SDR can demonstrate a suitable branch feature.

In our future work, we will further analyze the models with multiple features in modality B of tree-CNN. We will also examine more than two branches, and its effect on performance increase or decrease. Another factor is the data size in each group and the balance between the size of branches. Due to the limitation of the available data, we need some form of simulated data for further controlled analyses[11].

[11] An extension of this work covering some of these analyses will be submitted as a journal article.

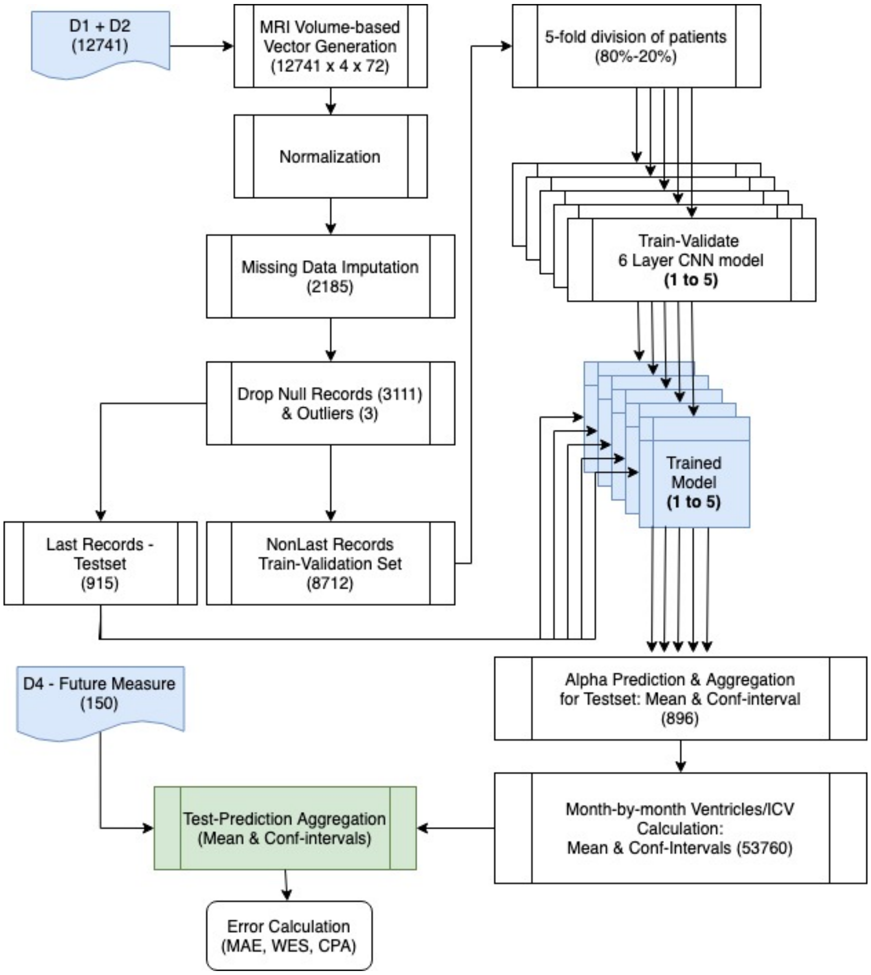


Figure 6.5: A summary of experiment pipeline for evaluation of volume-based convolutional neural network for TADPOLE AD data. Training five models using five-fold data, we calculate the average expectation and confidence interval for alpha (ventricles/ICV growth). The measure is then used to make a monthly prediction file for the next five years. D4 is then compared with the closest records to the actual measurement-date of the corresponding patients in the prediction file.

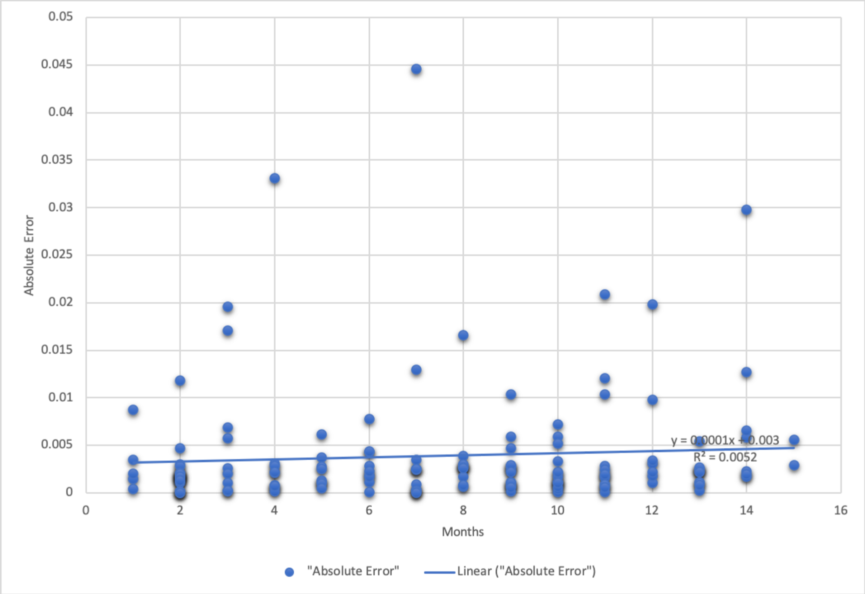
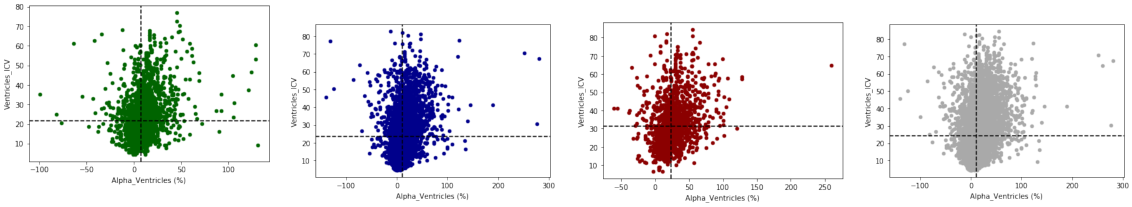


Figure 6.6: The absolute error is on average lower for the near future than later.



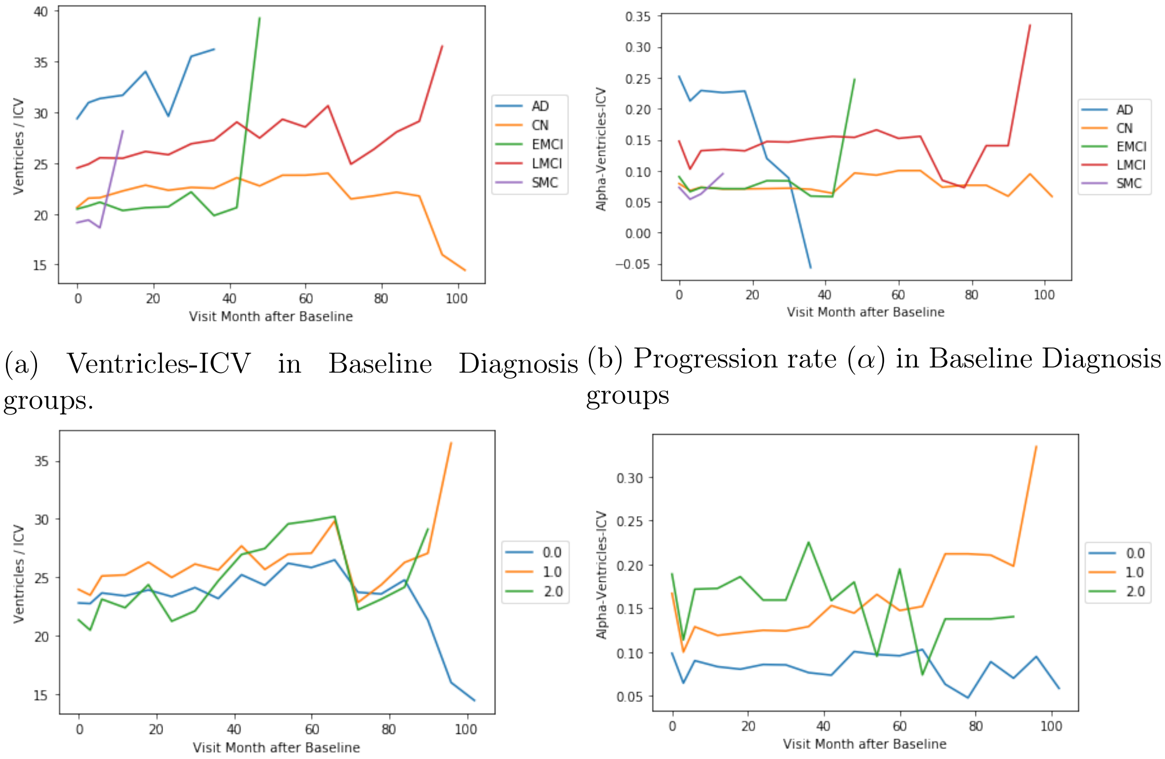
(a) Cognitive-normal records

(b) MCI records

(c) AD records

(d) All records

Figure 6.7: Distribution of Ventricles/ICV vs Alpha-Ventricles-ICV, i.e % change in Ventricles/ICV from one session to the next. The distribution is shown for each categories of visit according to the physician diagnosis for that visit. Additionally, the median line for Ventricles/ICV and Alpha-Ventricles-ICV is shown in black dashed-line. The distribution shift is more observable when comparing the normal and AD patients.



(a) Ventricles-ICV in Baseline Diagnosis groups.

(b) Progression rate (α) in Baseline Diagnosis groups

(c) Ventricles-ICV in $APOE\epsilon4$ genetic factor groups.

(d) Progression rate (α) in $APOE\epsilon4$ genetic factor groups.

Figure 6.8: Comparison of Median *Ventricle/ICV* (left) and *Alpha-Ventricles-ICV* (right) changes over the visits in each Diagnosis-at-baseline groups (top) and APOE groups (bottom). A general lower incline for median ventricles-icv and alpha is observable for APOE4=0 and CN groups. In AD patients, despite positive incline on Ventricles-ICV, the Alpha median is reducing, which is different from MCI in this sense. Moreover, EMCI and SMC is more similar to CN in the early months, while LMCI is closer to the AD pattern in that time.

Table 6.12: Instead of a deterministic threshold, we used similar architectures with a Sigmoid block. This block provides a weight for the corresponding branch making a nondeterministic threshold. The input for the Sigmoid block in each row is one of the *Diagnosis, ADAS13, APOE4, MMSE* features, and the number of branches tried for these features is between 2 to 6. These results are based on training for **30 epochs**, and by adding batch normalization and dropout with probability 0.2 to the convolution architecture. All the numbers in the table are rounded to two decimal points and sorted ascending based on D4-MAE. As it is shown in the table, the best result on D4 (test data) belongs to diagnosis with two branches, which outperforms the previous results. These results have informative hints for the future works.

Feature	**Miss (%)**	**Branch-Count**	**Arch**	**Train-MAE**	**Val-MAE**	**D4-MAE**	**D4-WES**	**D4-CPA**
Diagnosis	0.09	2	full	0.10	0.09	**0.36**	0.33	0.01
ADAS13	26.64	3	conv	0.12	0.13	0.37	0.29	0.27
APOE4	0.090	2	full	0.09	0.09	0.37	0.40	0.00
Diagnosis	0.09	3	conv	0.09	0.09	0.38	0.34	0.28
ADAS13	26.64	6	conv	0.24	0.24	0.39	0.30	0.37
Diagnosis	0.09	3	full	0.14	0.15	0.39	0.33	0.33
MMSE	25.84	5	full	0.20	0.20	0.41	0.29	0.09
MMSE	25.84	3	full	0.16	0.16	0.42	0.30	0.27
ADAS13	26.64	3	full	0.11	0.11	0.42	0.30	0.11
MMSE	25.84	2	full	0.10	0.11	0.42	0.36	0.11
Diagnosis	0.09	5	full	0.16	0.17	0.43	0.30	0.07
ADAS13	26.64	2	conv	0.10	0.10	0.46	0.26	0.02
ADAS13	26.64	6	full	0.29	0.29	0.47	0.32	0.43
Diagnosis	0.09	2	conv	0.22	0.24	0.47	0.31	0.03
MMSE	25.84	3	conv	0.12	0.12	0.50	0.33	0.14
APOE4	0.090	4	conv	0.32	0.32	0.50	0.36	0.39
ADAS13	26.64	2	full	0.14	0.14	0.50	0.36	0.33
APOE4	0.090	2	conv	0.30	0.30	0.52	0.30	0.04
MMSE	25.84	5	conv	0.25	0.25	0.52	0.29	0.40
MMSE	25.84	6	conv	0.32	0.33	0.54	0.29	0.35
MMSE	25.84	6	full	0.33	0.32	0.54	0.31	0.35
Diagnosis	0.09	5	conv	0.10	0.10	0.54	0.40	0.29
ADAS13	26.64	4	conv	0.32	0.30	0.60	0.36	0.42
ADAS13	26.64	5	full	0.14	0.14	0.62	0.52	0.39
MMSE	25.84	2	conv	0.18	0.19	0.62	0.29	0.09
Diagnosis	0.09	6	full	0.40	0.39	0.69	0.40	0.37
ADAS13	26.64	5	conv	0.15	0.15	0.70	0.32	0.41
MMSE	25.84	4	full	0.25	0.23	0.79	0.33	0.29
Diagnosis	0.09	4	full	0.18	0.17	0.84	0.44	0.24
APOE4	0.090	4	full	0.37	0.36	0.95	0.35	0.26
Diagnosis	0.09	4	conv	0.27	0.27	1.00	0.41	0.37
ADAS13	26.64	4	full	0.36	0.37	1.06	0.31	0.35
Diagnosis	0.09	6	conv	0.39	0.39	1.14	0.53	0.41
MMSE	25.84	4	conv	0.54	0.55	1.45	0.36	0.34
OrigCNN				0.09	0.09	0.38	0.31	0.03
Winner						0.41	0.29	0.43

CHAPTER 7: Conclusion

This dissertation defined a framework for analyzing the multimodal regression applications from the information abstraction perspective. We defined a dimensional-based framework to study conventional methods for different combinations of low-level and high-level modalities. We designed a pipeline to use domain knowledge to integrate heterogeneous features from multiple modalities effectively.

Our framework used the representation dimensionality as a quick description of a high-level representation vs. a low-level one. However, we demonstrate that this notion is not precisely equivalent to abstraction in Chapter 2. For example, in the case of treatment planning in Chapter 5, we demonstrated a domain-driven abstraction-leverage in CT-scan representation without changing the dimensionality.

We further discussed the modality integration method and the popular fusion approach for homogeneous and heterogeneous modality combinations. However, we discussed through the case of Alzheimer's disease progression prediction in Chapter 6 that a blind fusion might not always improve the performance. We suggested a tree-structure convolution neural network (tree-CNN) that can enhance information integration performance for some specific distributions along the proposed pipeline. We demonstrated this method's higher performance compared to a blind fusion framework for the disease progression prediction in Alzheimer's Disease.

7.1 Future Perspective

The tree-CNN architecture we suggested is a pipeline of classification followed by a regression model. While the model tree is a simple classification method, demonstrated in our architecture, this can be further extended with more complex classification/clustering methods. The core idea we are proposing and will be expanding in our future work is a model integration perspective instead of blind data and feature fusion. This topic is specifically crucial for the regression task, which requires a more precise model, and the target linearity, monotony, and topology directly impact the task complexity.

In our work, we used the simplest forms of classification and convolution neural networks. The model could be explored through more complex models of classic classification and deep learning architectures, e.g., adversarial networks or attention-based methods. These tools can further refine the models we suggested both in a bottom-up and in a top-down way.

In our future works, we plan to evaluate this with some *simulated data* to quantitatively study the situation in which a model-integration is more appropriate than a data-integration method. Specifically, we want to explore the role of low-dim feature importance in distribution desegregation, the number of branches, and its trade-off with the data-size. This helps us to move toward a more quantified pipeline for using domain-knowledge in designing an efficient model.

Finally, the contribution of our method is mainly studied from the optimization perspective [182] in this dissertation. The domain-driven pipeline can be studied from other aspects, such as interpretability, usability from the physicians' perspective, and the other practical applications. This can be further researched in future works.

REFERENCES

[1] J. He, "Learning from data heterogeneity: Algorithms and applications," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 0, pp. 5126–5130, 2017.

[2] G. Zhong, L.-N. Wang, X. Ling, and J. Dong, "An overview on data representation learning: From traditional feature learning to recent deep learning," *The Journal of Finance and Data Science*, vol. 2, no. 4, pp. 265–278, 2016.

[3] J. Landgrebe and B. Smith, "Making AI meaningful again," *Synthese*, pp. 1–23, 2019.

[4] M. Naumov, "On the Dimensionality of Embeddings for Sparse Features and Data," *arXiv preprint arXiv:1901.02103*, pp. 1–8, 2019.

[5] P. J. Blanco, M. Discacciati, and A. Quarteroni, "Modeling dimensionally-heterogeneous problems: Analysis, approximation and applications," *Numerische Mathematik*, vol. 119, no. 2, pp. 299–335, 2011.

[6] G. Marcus, "Deep Learning: A Critical Appraisal," *arXiv preprint*, pp. 1–24, 2018.

[7] Y. Bengio, G. Mesnil, Y. Dauphin, and S. Rifai, "Better mixing via deep representations," in *30th International Conference on Machine Learning, ICML 2013*, vol. 28, pp. 552–560, 2013.

[8] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, J. Pfrommer, A. Pick, R. Ramamurthy, M. Walczak, J. Garcke, C. Bauckhage, and J. Schuecker, "Informed Machine Learning – A Taxonomy and Survey of Integrating Knowledge into Learning Systems," *arXiv preprint arXiv:1903.12394*, pp. 1–20, 2019.

[9] S. Behnke, *Hierarchical neural networks for image interpretation*, vol. 2766. Springer, 2003.

[10] L. Yuan, Y. Ge, W. R. Lee, F. F. Yin, J. P. Kirkpatrick, and Q. J. Wu, "Quantitative analysis of the factors which affect the interpatient organ-At-risk dose sparing variation in IMRT plans," *Medical Physics*, vol. 39, no. 11, pp. 6868–6878, 2012.

[11] O. Nwankwo, H. Mekdash, D. S. K. Sihono, F. Wenz, and G. Glatting, "Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: Validation of a KBRT algorithm for prostate cancer treatment planning," *Radiation Oncology*, vol. 10, no. 1, pp. 1–5, 2015.

[12] B. Emami, "Tolerance of Normal Tissue to Therapeutic Radiation," *Reports of radiotherapy and Oncology*, vol. 1, no. 1, pp. 35–48, 2013.

[13] H. Kiiveri and T. P. Speed, "Structural Analysis of Multivariate Data: A Review," *Sociological Methodology*, vol. 13, p. 209, 1982.

[14] W. Guo, J. Wang, and S. Wanga, "Deep multimodal representation learning: a survey," *IEEE Access*, vol. 7, pp. 1–1, 2019.

[15] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[16] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, "Survey on deep learning for radiotherapy," *Computers in Biology and Medicine*, vol. 98, no. March, pp. 126–146, 2018.

[17] L. Sun, S. Zhang, H. Chen, and L. Luo, "Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning," *Frontiers in Neuroscience*, vol. 13, no. JUL, pp. 1–9, 2019.

[18] H. Choi and D. S. Lee, "Generation of structural MR images from amyloid PET: Application to MR-less quantification," *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1111–1117, 2018.

[19] D. Sun, M. Wang, and A. Li, "A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 3, pp. 841–850, 2019.

[20] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, B. T. T. Yeo, and A. D. N. Initiative, "Predicting Alzheimer's disease progression using deep recurrent neural networks," *BioRxiv*, p. 755058, 2019.

[21] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 538–546, 2019.

[22] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G. R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019.

[23] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173–183, 2018.

[24] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis," *Neuroinformatics*, vol. 16, no. 3-4, pp. 295–308, 2018.

[25] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis," *Medical image analysis*, vol. 42, no. 1995, pp. 60–88, 2017.

[26] R. Zemouri, N. Zerhouni, and D. Racoceanu, "Deep learning in the biomedical applications: Recent and future status," *Applied Sciences (Switzerland)*, vol. 9, no. 8, 2019.

[27] Z. Aslan, "On The Use of Deep Learning Methods on Medical Images," *The International Journal of Energy & Engineering Sciences IJEES-V3*, vol. 2, pp. 1–15, 2018.

[28] R. V. Marinescu, N. P. Oxtoby, A. L. Young, E. E. Bron, A. W. Toga, M. W. Weiner, F. Barkhof, N. C. Fox, P. Golland, S. Klein, and D. C. Alexander, "TADPOLE challenge: Accurate alzheimer's disease prediction through crowdsourced forecasting of future data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11843 LNCS, no. Icv, pp. 1–10, 2019.

[29] Y. Ge and Q. J. Wu, "Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches," *Medical Physics*, vol. 46, no. 6, pp. 2760–2775, 2019.

[30] T. Baltrusaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.

[31] N. Srivastava and R. Salakhutdinov, "Multimodal learning with Deep Boltzmann Machines," *Journal of Machine Learning Research*, vol. 15, pp. 2949–2980, 2014.

[32] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Information Fusion*, vol. 38, pp. 43–54, 2017.

[33] Y. Li, M. Yang, and Z. M. Zhang, "A Survey of Multi-View Representation Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. PP, no. 8, p. 1, 2018.

[34] J. Ngiam, A. Khosla, and M. Kim, "Multimodal deep learning," *Machine Learning Research*, vol. 85, pp. 1–9, 2011.

[35] K. Liu, Y. Li, N. Xu, and P. Natarajan, "Learn to Combine Modalities in Multimodal Deep Learning," *arXiv preprint arXiv:1805.11730*, 2018.

[36] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[37] Z. Che, *Deep Learning Models For Temporal Data in Health Care*. PhD thesis, University of Southern California, 2018.

[38] B. Yang, A. J. Ma, and P. C. Yuen, "Learning domain-shared group-sparse representation for unsupervised domain adaptation," *Pattern Recognition*, vol. 81, pp. 615–632, 2018.

[39] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[40] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. Wichmann, and W. Brendel, "Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," *Iclr*, no. c, pp. 1–20, 2019.

[41] M. Ponsen, M. E. Taylor, and K. Tuyls, "Abstraction and generalization in reinforcement learning: A summary and framework," in *International Workshop on Adaptive and Learning Agents*, pp. 1–32, 2009.

[42] R. Ilin, T. Watson, and R. Kozma, "Abstraction hierarchy in deep learning neural networks," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, no. May 2017, pp. 768–774, 2017.

[43] R. Kozma, R. Ilin, and H. T. Siegelmann, "Evolution of Abstraction Across Layers in Deep Learning Neural Networks," *Procedia Computer Science*, vol. 144, no. January, pp. 203–213, 2018.

[44] Y. Bengio, "Learning Deep Architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[45] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

[46] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[47] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," *arXiv preprint arXiv:1703.00810*, pp. 1–19, 2017.

[48] M. I. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, A. Courville, and R. D. Hjelm, "MINE: Mutual Information Neural Estimation," *arXiv preprint*, 2018.

[49] J. Jo and Y. Bengio, "Measuring the tendency of CNNs to Learn Surface Statistical Regularities," *arXiv preprint*, no. 1, 2017.

[50] L. Baldassarre, M. Pontil, and J. Mourão-Miranda, "Sparsity is better with stability: Combining accuracy and stability for model selection in brain decoding," *Frontiers in Neuroscience*, vol. 11, no. FEB, 2017.

[51] A. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik, and S. Savarese, "Taskonomy: Disentangling Task Transfer Learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3712–2722, 2018.

[52] A. Gonzalez-garcia, "Image-to-image translation for cross-domain disentanglement," in *Advances in Neural Information Processing Systems*, pp. 1287–1298, 2018.

[53] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," *arXiv preprint*, pp. 1–24, 2018.

[54] C. Wong, F. Deligianni, M. Berthelot, J. Andreu-perez, B. Lo, and G.-z. Yang, "Deep Learning for Health Informatics," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 21, no. 1, pp. 4–21, 2017.

[55] L. Wei, S. Osman, M. Hatt, and I. El Naqa, "Machine learning for radiomics-based multimodality and multiparametric modeling," *Quarterly Journal of Nuclear Medicine and Molecular Imaging*, vol. 63, no. 4, pp. 323–338, 2019.

[56] X. Luo, K. Mori, and T. M. Peters, "Advanced Endoscopic Navigation: Surgical Big Data, Methodology, and Applications," *Annual Review of Biomedical Engineering*, vol. 20, pp. 221–251, 2018.

[57] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, "Biosignal-Based Spoken Communication: A Survey," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 12, pp. 2257–2271, 2017.

[58] S. B. Somai, W. B. Abdessalem Karaa, and H. H. Ben Ghezala, "Deep learning and semantic medical image processing and retrieval: Datasets, software, applications and perspectives," in *Proceedings of the 31st International Business Information Management Association Conference, IBIMA 2018: Innovation Management and Education Excellence through Vision 2020*, pp. 2596–2609, 4 2018.

[59] L. Yue, D. Tian, W. Chen, X. Han, and M. Yin, "Deep learning for heterogeneous medical data analysis," *World Wide Web*, no. July 2019, pp. 2715–2737, 2020.

[60] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of Deep Learning and Reinforcement Learning to Biological Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

[61] H. Muller and D. Unay, "Retrieval from and Understanding of Large-Scale Multi-modal Medical Datasets: A Review," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2093–2104, 2017.

[62] D. Durstewitz, G. Koppe, and A. Meyer-Lindenberg, “Deep neural networks in psychiatry,” *Molecular Psychiatry*, vol. 24, no. 11, pp. 1583–1598, 2019.

[63] T. Jo, K. Nho, and A. J. Saykin, “Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data,” *Frontiers in Aging Neuroscience*, vol. 11, no. August, 2019.

[64] S. Prange, E. Metereau, and S. Thobois, “Structural Imaging in Parkinson’s Disease: New Developments,” *Current Neurology and Neuroscience Reports*, vol. 19, no. 8, 2019.

[65] T. Bhattacharya, T. Brettin, J. H. Doroshow, Y. A. Evrard, E. J. Greenspan, A. L. Gryshuk, T. T. Hoang, C. B. Lauzon, D. Nissley, L. Penberthy, E. Stahlberg, R. Stevens, F. Streitz, G. Tourassi, F. Xia, and G. Zaki, “AI Meets Exascale Computing: Advancing Cancer Research With Large-Scale High Performance Computing,” *Frontiers in Oncology*, vol. 9, no. October, pp. 1–8, 2019.

[66] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, “Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls,” *NeuroImage*, vol. 145, pp. 137–165, 2017.

[67] R. Feng, M. Badgeley, J. Mocco, and E. K. Oermann, “Deep learning guided stroke management: a review of clinical applications,” *Journal of neurointerventional surgery*, vol. 10, no. 4, pp. 358–362, 2018.

[68] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, “Data Augmentation for Brain-Tumor Segmentation: A Review,” *Frontiers in Computational Neuroscience*, vol. 13, no. December, pp. 1–18, 2019.

[69] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, and C. P. Langlotz, “Deep learning in neuroradiology,” *American Journal of Neuroradiology*, vol. 39, no. 10, pp. 1776–1784, 2018.

[70] E. Victor, Z. M. Aghajan, A. R. Sewart, and R. Christian, “Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation,” *Psychological Assessment*, vol. 31, no. 8, pp. 1019–1027, 2019.

[71] A. Cheerla and O. Gevaert, “Deep learning with multimodal representation for pancancer prognosis prediction,” *Bioinformatics*, vol. 35, no. 14, pp. i446–i454, 2019.

[72] J. Yap, W. Yolland, and P. Tschandl, “Multimodal skin lesion classification using deep learning,” *Experimental Dermatology*, vol. 27, no. 11, pp. 1261–1267, 2018.

[73] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal Assessment of Parkinson's Disease: A Deep Learning Approach," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1618–1630, 2019.

[74] Z. Hu, Z. Zhang, H. Yang, Q. Chen, and D. Zuo, "A deep learning approach for predicting the quality of online health expert question-answering services," *Journal of Biomedical Informatics*, vol. 71, pp. 241–253, 2017.

[75] P. Luo, Y. Li, L.-P. Tian, and F.-X. Wu, "Enhancing the prediction of disease–gene associations with multimodal deep learning," *Bioinformatics*, vol. 35, no. 19, pp. 3735–3742, 2019.

[76] X. Niu, F. Zhang, J. Kounios, and H. Liang, "Improved prediction of brain age using multimodal neuroimaging data," *Human Brain Mapping*, vol. 41, no. 6, pp. 1626–1643, 2020.

[77] Y. Li, F. Meng, and J. Shi, "Learning using privileged information improves neuroimaging-based CAD of Alzheimer's disease: a comparative study," *Medical and Biological Engineering and Computing*, vol. 57, no. 7, pp. 1605–1616, 2019.

[78] C. F. Liu, S. Padhy, S. Ramachandran, V. X. Wang, A. Efimov, A. Bernal, L. Shi, M. Vaillant, J. T. Ratnanather, A. V. Faria, B. Caffo, M. Albert, and M. I. Miller, "Using deep Siamese neural networks for detection of brain asymmetries associated with Alzheimer's Disease and Mild Cognitive Impairment," *Magnetic Resonance Imaging*, vol. 64, no. July, pp. 190–199, 2019.

[79] K.-H. Thung, P.-T. Yap, and D. Shen, "Multi-stage Diagnosis of Alzheimer's Disease with Incomplete Multimodal Data via Multi-task Deep Learning," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 160–168, Springer, 2017.

[80] S. M. Plis, M. F. Amin, A. Chekroud, D. Hjelm, E. Damaraju, H. J. Lee, J. R. Bustillo, K. H. Cho, G. D. Pearlson, and V. D. Calhoun, "Reading the (functional) writing on the (structural) wall: Multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia," *NeuroImage*, vol. 181, pp. 734–747, 2018.

[81] M. P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Deep Learning with Edge Computing for Localization of Epileptogenicity Using Multimodal rs-fMRI and EEG Big Data," *Proceedings - 2017 IEEE International Conference on Autonomic Computing, ICAC 2017*, pp. 83–92, 2017.

[82] A. M. Chiarelli, P. Croce, A. Merla, and F. Zappasodi, "Deep learning for hybrid EEG-fNIRS brain-computer interface: Application to motor imagery classification," *Journal of Neural Engineering*, vol. 15, no. 3, 2018.

[83] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, “A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2018.

[84] L. Zhengyi, Z. Hui, Y. Dandan, and X. Shuiqing, “Multimodal deep learning network based hand ADLs tasks classification for prosthetics control,” *Proceedings of 2017 International Conference on Progress in Informatics and Computing, PIC 2017*, no. 2015, pp. 91–95, 2017.

[85] E. A. Bernal, X. Yang, Q. Li, J. Kumar, S. Madhvanath, P. Ramesh, and R. Bala, “Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors,” *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 107–118, 2018.

[86] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, “Deep Learning-Based Image Segmentation on Multimodal Medical Imaging,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 162–169, 2019.

[87] J. W. Jeong, M. H. Lee, F. John, N. L. Robinette, A. J. Amit-Yousif, G. R. Barger, S. Mittal, and C. Juhász, “Feasibility of Multimodal MRI-Based Deep Learning Prediction of High Amino Acid Uptake Regions and Survival in Patients With Glioblastoma,” *Frontiers in Neurology*, vol. 10, no. December, pp. 1–8, 2019.

[88] H. Li, A. Li, and M. Wang, “A novel end-to-end brain tumor segmentation method using improved fully convolutional networks,” *Computers in Biology and Medicine*, vol. 108, no. March, pp. 150–160, 2019.

[89] Z. Tang, P. T. Yap, and D. Shen, “A New Multi-Atlas Registration Framework for Multimodal Pathological Images Using Conventional Monomodal Normal Atlases,” *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2293–2304, 2019.

[90] T. K. Yoo, J. Y. Choi, J. G. Seo, B. Ramasubramanian, S. Selvaperumal, and D. W. Kim, “The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: a preliminary experiment,” *Medical and Biological Engineering and Computing*, vol. 57, no. 3, pp. 677–687, 2019.

[91] R. Karthik, U. Gupta, A. Jha, R. Rajalakshmi, and R. Menaka, “A deep supervised approach for ischemic lesion segmentation from multimodal MRI using Fully Convolutional Network,” *Applied Soft Computing Journal*, vol. 84, p. 105685, 2019.

[92] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, “Fully Convolutional Networks for Monocular Retinal Depth Estimation and Optic

Disc-Cup Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1417–1426, 2019.

[93] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, "Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[94] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, no. Isbi, pp. 1449–1453, 2018.

[95] X. Cheng, L. Zhang, and Y. Zheng, "Deep similarity learning for multimodal medical images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 6, no. 3, pp. 248–252, 2018.

[96] C. Zhang, Y. Song, S. Liu, S. Lill, C. Wang, Z. Tang, Y. You, Y. Gao, A. Klistorner, M. Barnett, and W. Cai, "MS-GAN: GAN-Based Semantic Segmentation of Multiple Sclerosis Lesions in Brain Magnetic Resonance Imaging," *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 2019.

[97] D. Liu, H. Zhang, M. Zhao, X. Yu, S. Yao, and W. Zhou, "Brain Tumor Segmention Based on Dilated Convolution Refine Networks," *Proceedings - 2018 IEEE/ACIS 16th International Conference on Software Engineering Research, Management and Application, SERA 2018*, pp. 113–120, 2018.

[98] Y. Zhuge, A. V. Krauze, H. Ning, J. Y. Cheng, B. C. Arora, K. Camphausen, and R. W. Miller, "Brain tumor segmentation using holistically nested neural networks in MRI images," *Medical Physics*, vol. 44, no. 10, pp. 5234–5243, 2017.

[99] M. H. Le, J. Chen, L. Wang, Z. Wang, W. Liu, K.-T. T. Cheng, and X. Yang, "Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks," *Physics in Medicine & Biology*, vol. 62, no. 16, p. 6497, 2017.

[100] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical Image Retrieval using Deep Convolutional Neural Network," *Neurocomputing*, vol. 266, pp. 8–20, 2017.

[101] T. Che, Y. Zheng, J. Cong, Y. Jiang, Y. Niu, W. Jiao, B. Zhao, and Y. Ding, "Deep Group-Wise Registration for Multi-Spectral Images from Fundus Images," *IEEE Access*, vol. 7, pp. 27650–27661, 2019.

[102] G. Gonella, E. Binaghi, P. Nocera, and C. Mordacchini, "Investigating the behaviour of machine learning techniques to segment brain metastases in radiation therapy planning," *Applied Sciences (Switzerland)*, vol. 9, no. 16, 2019.

[103] H. Li, P. Boimel, J. Janopaul-Naylor, H. Zhong, Y. Xiao, E. Ben-Josef, and Y. Fan, “Deep convolutional neural networks for imaging data based survival analysis of rectal cancer,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, no. Isbi, pp. 846–849, 2019.

[104] G. Haskins, J. Kruecker, U. Kruger, S. Xu, P. A. Pinto, B. J. Wood, and P. Yan, “Learning deep similarity metric for 3D MR–TRUS image registration,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 3, pp. 417–425, 2019.

[105] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, “A deep learning model integrating FCNNs and CRFs for brain tumor segmentation,” *Medical image analysis*, vol. 43, pp. 98–111, 2018.

[106] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, “Deep Learning Framework for Alzheimer’s Disease Diagnosis via 3D-CNN and FSBi-LSTM,” *IEEE Access*, vol. 7, pp. 63605–63618, 2019.

[107] L. Zhang, L. Lu, R. M. Summers, E. Kebebew, and J. Yao, “Convolutional Invasion and Expansion Networks for Tumor Growth Prediction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 638–648, 2018.

[108] C. H. Pham, C. Tor-Díez, H. Meunier, N. Bednarek, R. Fablet, N. Passat, and F. Rousseau, “Multiscale brain MRI super-resolution using deep 3D convolutional networks,” *Computerized Medical Imaging and Graphics*, vol. 77, 2019.

[109] I. Banerjee, A. Crawley, M. Bhethanabotla, H. E. Daldrup-Link, and D. L. Rubin, “Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma,” *Computerized Medical Imaging and Graphics*, vol. 65, pp. 167–175, 2018.

[110] L. Xu, G. Tetteh, J. Lipkova, Y. Zhao, H. Li, P. Christ, M. Piraud, A. Buck, K. Shi, and B. H. Menze, “Automated Whole-Body Bone Lesion Detection for Multiple Myeloma on 68 Ga-Pentixafor PET/CT Imaging Using Deep Learning Methods,” *Contrast Media and Molecular Imaging*, vol. 2018, 2018.

[111] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J. B. Clavier, and P. Meyer, “Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network,” *Computers in Biology and Medicine*, vol. 95, no. November 2017, pp. 43–54, 2018.

[112] Y. Yoo, L. Y. Tang, T. Brosch, D. K. Li, S. Kolind, I. Vavasour, A. Rauscher, A. L. MacKay, A. Traboulsee, and R. C. Tam, “Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls,” *NeuroImage: Clinical*, vol. 17, no. October 2017, pp. 169–178, 2018.

[113] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, M. Weiner, P. Aisen, R. Petersen, C. Jack, W. Jagust, J. Trojanowki, A. Toga, L. Beckett, R. Green, A. Saykin, J. Morris, L. Shaw, J. Kaye, J. Quinn, L. Silbert, B. Lind, R. Carter, S. Dolen, L. Schneider, S. Pawluczyk, M. Beccera, L. Teodoro, B. Spann, J. Brewer, H. Vanderswag, A. Fleisher, J. Heidebrink, J. Lord, S. Mason, C. Albers, D. Knopman, K. Johnson, R. Doody, J. Villanueva-Meyer, M. Chowdhury, S. Rountree, M. Dang, Y. Stern, L. Honig, K. Bell, B. Ances, M. Carroll, M. Creech, E. Franklin, M. Mintun, S. Schneider, A. Oliver, D. Marson, R. Grifth, D. Clark, D. Geldmacher, J. Brockington, E. Roberson, M. N. Love, H. Grossman, E. Mitsis, R. Shah, L. deToledo Morrell, R. Duara, D. Varon, M. Greig, P. Roberts, M. Albert, C. Onyike, D. D'Agostino, S. Kielb, J. Galvin, B. Cerbone, C. Michel, D. Pogorelec, H. Rusinek, M. de Leon, L. Glodzik, S. D. Santi, P. Doraiswamy, J. Petrella, S. Borges-Neto, T. Wong, E. Coleman, C. Smith, G. Jicha, P. Hardy, P. Sinha, E. Oates, G. Conrad, A. Porsteinsson, B. Goldstein, K. Martin, K. Makino, M. Ismail, C. Brand, R. Mulnard, G. Thai, C. Mc-Adams-Ortiz, K. Womack, D. Mathews, M. Quiceno, A. Levey, J. Lah, J. Cellar, J. Burns, R. Swerdlow, W. Brooks, L. Apostolova, K. Tingus, E. Woo, D. Silverman, P. Lu, G. Bartzokis, N. Graf-Radford, F. Parftt, T. Kendall, H. Johnson, M. Farlow, A. M. Hake, B. Matthews, J. Brosch, S. Herring, C. Hunt, C. Dyck, R. Carson, M. MacAvoy, P. Varma, H. Chertkow, H. Bergman, C. Hosein, S. Black, B. Stefanovic, C. Caldwell, G. Y. R. Hsiung, H. Feldman, B. Mudge, M. Assaly, E. Finger, S. Pasternack, I. Rachisky, D. Trost, A. Kertesz, C. Bernick, D. Munic, M. M. Mesulam, K. Lipowski, S. Weintraub, B. Bonakdarpour, D. Kerwin, C. K. Wu, N. Johnson, C. Sadowsky, T. Villena, R. S. Turner, K. Johnson, B. Reynolds, R. Sperling, K. Johnson, G. Marshall, J. Yesavage, J. Taylor, B. Lane, A. Rosen, J. Tinklenberg, M. Sabbagh, C. Belden, S. Jacobson, S. Sirrel, N. Kowall, R. Killiany, A. Budson, A. Norbash, P. L. Johnson, T. Obisesan, S. Wolday, J. Allard, A. Lerner, P. Ogrocki, C. Tatsuoka, P. Fatica, E. Fletcher, P. Maillard, J. Olichney, C. DeCarli, O. Carmichael, S. Kittur, M. Borrie, T. Y. Lee, R. Bartha, S. Johnson, S. Asthana, C. Carlsson, S. Potkin, A. Preda, D. Nguyen, P. Tariot, A. Burke, N. Trncic, S. Reeder, V. Bates, H. Capote, M. Rainka, D. Scharre, M. Kataki, A. Adeli, E. Zimmerman, D. Celmins, A. Brown, G. Pearlson, K. Blank, K. Anderson, L. Flashman, M. Seltzer, M. Hynes, R. Santulli, K. Sink, L. Gordineer, J. Williamson, P. Garg, F. Watkins, B. Ott, H. Querfurth, G. Tremont, S. Salloway, P. Malloy, S. Correia, H. Rosen, B. Miller, D. Perry, J. Mintzer, K. Spicer, D. Bachman, N. Pomara, R. Hernando, A. Sarrael, N. Relkin, G. Chaing, M. Lin, L. Ravdin, A. Smith, B. A. Raj, and K. Fargher, "Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images," *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018.

[114] Y. Liu, S. Stojadinovic, B. Hrycushko, Z. Wardak, S. Lau, W. Lu, Y. Yan, S. B. Jiang, X. Zhen, R. Timmerman, L. Nedzi, and X. Gu, "A deep convolutional neural network-based automatic delineation strategy for multiple brain

metastases stereotactic radiosurgery," *PLoS ONE*, vol. 12, no. 10, pp. 1–17, 2017.

[115] Y. Hao, M. Usama, J. Yang, M. S. Hossain, and A. Ghoneim, "Recurrent convolutional neural network based multimodal disease risk prediction," *Future Generation Computer Systems*, vol. 92, pp. 76–83, 2019.

[116] S. Perek, N. Kiryati, G. Zimmerman-Moreno, M. Sklair-Levy, E. Konen, and A. Mayer, "Classification of contrast-enhanced spectral mammography (CESM) images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 2, pp. 249–257, 2019.

[117] L. Wang, S. Wang, R. Chen, X. Qu, Y. Chen, S. Huang, and C. Liu, "Nested dilation networks for brain tumor segmentation based on magnetic resonance imaging," *Frontiers in Neuroscience*, vol. 13, no. APR, pp. 1–14, 2019.

[118] S. Zhang, Q. Dong, W. Zhang, H. Huang, D. Zhu, and T. Liu, "Discovering hierarchical common brain networks via multimodal deep belief network," *Medical Image Analysis*, vol. 54, pp. 238–252, 2019.

[119] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. Ben Ayed, "HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1116–1126, 2019.

[120] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2018.

[121] S. Kumar, S. Conjeti, A. G. Roy, C. Wachinger, and N. Navab, "Infinet : Fully Convolutional Networks for Infant Brain MRI Segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 145–148, IEEE, 2018.

[122] W. Gu, X. Gu, J. Gu, B. Li, Z. Xiong, and W. Wang, "Adversary guided asymmetric hashing for cross-modal retrieval," *ICMR 2019 - Proceedings of the 2019 ACM International Conference on Multimedia Retrieval*, pp. 159–167, 2019.

[123] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," *Medical physics*, vol. 46, no. 1, pp. e1–e36, 2019.

[124] J. R. Quinlan, "Learning with continuous classes," *Australian Joint Conference on Artificial Intelligence*, vol. 92, pp. 343–348, 1992.

[125] S. Victor, "Telling Tales : A Review of C . K . Riessman ' s Narrative Methods for the Human Sciences," *The Weekly Qualitative Report*, vol. 2, no. 29, pp. 172–176, 2009.

[126] Z. Yu, "High Accuracy Postal Address Extraction From Web Pages," *Master Thesis: Dalhousie University.*, no. March, 2007.

[127] D. Toth, S. Miao, T. Kurzendorfer, C. A. Rinaldi, R. Liao, T. Mansi, K. Rhode, and P. Mountney, "3D/2D model-to-image registration by imitation learning for cardiac procedures," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 8, pp. 1141–1149, 2018.

[128] G. Carneiro, T. Peng, C. Bayer, and N. Navab, "Automatic Quantification of Tumour Hypoxia from Multi-Modal Microscopy Images Using Weakly-Supervised Learning Methods," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1405–1417, 2017.

[129] D. Nie, L. Wang, Y. Gao, and D. Sken, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2016-June, pp. 1342–1345, 2016.

[130] D. Cunefare, C. S. Langlo, E. J. Patterson, S. Blau, A. Dubra, J. Carroll, and S. Farsiu, "Deep learning based detection of cone photoreceptors with multimodal adaptive optics scanning light ophthalmoscope images of achromatopsia," *Biomedical Optics Express*, vol. 9, no. 8, p. 3740, 2018.

[131] N. Zhang, Y. Cao, B. Liu, and Y. Luo, "Improved multimodal representation learning with skip connections," *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*, pp. 654–662, 2017.

[132] J.-M. Pérez-Rúa, V. Vielzeuf, S. Pateux, M. Baccouche, and F. Jurie, "MFAS: Multimodal Fusion Architecture Search," in *IEEE Conference on computer vision and pattern recognition*, pp. 6966–6975, 2019.

[133] H. R. V. Joze, A. Shaban, M. L. Iuzzolino, and K. Koishida, "MMTM: Multimodal Transfer Module for CNN Fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13289–13299, 2020.

[134] M. Tavakoli H., B. Ru, T. Xie, M. Hadzikadic, Q. J. Wu, and Y. Ge, "Dose Prediction for Prostate Radiation Treatment: Feasibility of a Distance-Based Deep Learning Model," *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pp. 2379–2386, 2019.

[135] D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal, and S. Jiang, "Dose Prediction with U-net: A Feasibility Study for Predicting Dose Distributions from Contours using Deep Learning on Prostate IMRT Patients," *Arxiv*, vol. m, pp. 8–15, 2017.

[136] Z. Zhang, M. W. Beck, D. A. Winkler, B. Huang, W. Sibanda, and H. Goyal, "Opening the black box of neural networks: methods for interpreting neural network models in clinical applications," *Annals of Translational Medicine*, vol. 6, no. 11, pp. 216–216, 2018.

[137] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Interpretable machine learning: definitions, methods, and applications," *preprint arXiv*, pp. 1–11, 2019.

[138] S. Bang, P. Xie, H. Lee, W. Wu, and E. Xing, "Explaining a black-box using Deep Variational Information Bottleneck Approach," *arXiv preprint arXiv:1902.06918*, 2019.

[139] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Augu, pp. 1135–1144, 2016.

[140] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, pp. 4765–4774, 2017.

[141] N. Prasad and K. Palla, "The Role of Context in the Prediction of Acute Hypotension in Critical Care," in *SAIL: Symposium on Artificial Intelligence for Learning Health Systems, 2020*, 10 2020.

[142] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*, vol. 11141 LNCS, (Cham), pp. 270–279, Springer, 2018.

[143] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[144] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. International Conference on Learning Representations*, 2014.

[145] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI," *arXiv preprint*, 2016.

[146] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, 2019.

[147] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *arXiv preprint arXiv:1712.04621*, 2017.

[148] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*, pp. 117–122, 2018.

[149] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Medical Physics*, vol. 44, no. 10, pp. 5221–5233, 2017.

[150] B. D. de Vos, J. M. Wolterink, P. A. de Jong, M. A. Viergever, and I. Išgum, "2D image classification for 3D anatomy localization: employing deep convolutional neural networks," *Medical Imaging 2016: Image Processing*, vol. 9784, no. March 2016, p. 97841Y, 2016.

[151] V. Venkatraghavan, E. E. Bron, W. J. Niessen, and S. Klein, "Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling," *NeuroImage*, vol. 186, no. August 2018, pp. 518–532, 2019.

[152] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. O'Reilly, and Y. Tong, "Probabilistic Attribute Tree in Convolutional Neural Networks for Facial Expression Recognition," *arXiv preprint arXiv:1812.07067*, vol. 1, no. c, 2018.

[153] N. Srivastava and R. Salakhutdinov, "Discriminative transfer learning with tree-based priors," *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

[154] B. Emami, J. Lyman, A. Brown, L. Cola, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, "Tolerance of normal tissue to therapeutic irradiation," *International Journal of Radiation Oncology* Biology* Physics*, vol. 21, no. 1, pp. 109–122, 1991.

[155] C. Burman, G. J. Kutcher, B. Emami, and M. Goitein, "Fitting of normal tissue tolerance data to an analytic function," *International Journal of Radiation Oncology* Biology* Physics*, vol. 21, no. 1, pp. 123–135, 1991.

[156] S. M. Bentzen, L. S. Constine, J. O. Deasy, A. Eisbruch, A. Jackson, L. B. Marks, R. K. Ten Haken, and E. D. Yorke, "Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues," *International Journal of Radiation Oncology* Biology* Physics*, vol. 76, no. 3, pp. S3–S9, 2010.

[157] D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal, and S. Jiang, "A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning," *Scientific Reports*, vol. 9, no. 1, p. 1076, 2019.

[158] R. Mahmood, A. Babier, A. McNiven, A. Diamant, and T. C. Y. Chan, "Automated Treatment Planning in Radiation Therapy using Generative Adversarial Networks," in *Machine Learning Research 85*, pp. 1–15, 2018.

[159] X. Chen, K. Men, Y. Li, J. Yi, and J. Dai, "A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning," *Medical Physics*, pp. 56–64, 2018.

[160] Y. Bengio, "The Consciousness Prior," *arXiv preprint*, 2017.

[161] B. Wu, F. Ricchetti, G. Sanguineti, M. Kazhdan, P. Simari, M. Chuang, R. Taylor, R. Jacques, and T. McNutt, "Patient geometry-driven information retrieval for IMRT treatment plan quality control," *Medical Physics*, vol. 36, no. 12, pp. 5497–5505, 2009.

[162] X. Zhu, Y. Ge, T. Li, D. Thongphiew, F. Yin, and Q. J. Wu, "A planning quality evaluation tool for prostate adaptive IMRT based on machine learning," *Medical physics*, vol. 38, no. 2, pp. 719–726, 2011.

[163] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.

[164] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[165] K. Yan, X. Wang, L. Lu, and R. M. Summers, "DeepLesion: Automated Deep Mining, Categorization and Detection of Significant Radiology Image Findings using Large-Scale Clinical Lesion Annotations," *arXiv preprint*, pp. 1–9, 2017.

[166] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. C. Morris, R. C. Petersen, A. J. Saykin, L. M. Shaw, A. W. Toga, and J. Q. Trojanowski, "Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials," *Alzheimer's and Dementia*, vol. 13, no. 4, pp. e1–e85, 2017.

[167] K. Ito, B. Corrigan, Q. Zhao, J. French, R. Miller, H. Soares, E. Katz, T. Nicholas, B. Billing, R. Anziano, and T. Fullerton, "Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database," *Alzheimer's and Dementia*, vol. 7, no. 2, pp. 151–160, 2011.

[168] R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox, "Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 7, pp. 4703–4707, 2002.

[169] L. Harper, G. G. Fumagalli, F. Barkhof, P. Scheltens, J. T. O'Brien, F. Bouwman, E. J. Burton, J. D. Rohrer, N. C. Fox, G. R. Ridgway, and J. M. Schott, "MRI visual rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases," *Brain*, vol. 139, no. 4, pp. 1211–1225, 2016.

[170] G. Martí-Juan, G. Sanroma-Guell, and G. Piella, "A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease," *Computer Methods and Programs in Biomedicine*, vol. 189, p. 105348, 2020.

[171] T. Wang, R. G. Qiu, and M. Yu, "Predictive Modeling of the Progression of Alzheimer's Disease with Recurrent Neural Networks," *Scientific Reports*, vol. 8, no. 1, pp. 1–12, 2018.

[172] M. Mehdipour Ghazi, M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, and L. Sørensen, "Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling," *Medical Image Analysis*, vol. 53, pp. 39–46, 2019.

[173] L. E. Givon, L. J. Mariano, D. O'Dowd, J. M. Irvine, and A. R. Schneider, "Cognitive Subscore Trajectory Prediction in Alzheimer's Disease," *arXiv preprint arXiv:1706.08491*, pp. 1–7, 2017.

[174] N. Bhagwat, J. D. Viviano, A. N. Voineskos, and M. M. Chakravarty, "Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data," *PLoS Computational Biology*, vol. 14, no. 9, 2018.

[175] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.

[176] R. V. Marinescu, N. P. Oxtoby, A. L. Young, E. E. Bron, A. W. Toga, M. W. Weiner, F. Barkhof, N. C. Fox, A. Eshaghi, T. Toni, M. Salaterski, V. Lunina, M. Ansart, S. Durrleman, P. Lu, S. Iddi, D. Li, W. K. Thompson, M. C. Donohue, A. Nahon, Y. Levy, D. Halbersberg, M. Cohen, H. Liao, T. Li, K. Yu, H. Zhu, J. G. Tamez-Pena, A. Ismail, T. Wood, H. C. Bravo, M. Nguyen, N. Sun, J. Feng, B. T. T. Yeo, G. Chen, K. Qi, S. Chen, D. Qiu, I. Buciuman, A. Kelner, R. Pop, D. Rimocea, M. M. Ghazi, M. Nielsen, S. Ourselin, L. Sorensen, V. Venkatraghavan, K. Liu, C. Rabe, P. Manser, S. M. Hill, J. Howlett, Z. Huang, S. Kiddle, S. Mukherjee, A. Rouanet, B. Taschler, B. D. M. Tom, S. R. White, N. Faux, S. Sedai, J. d. V. Oriol, E. E. V. Clemente, K. Estrada, L. Aksman, A. Altmann, C. M. Stonnington, Y. Wang, J. Wu, V. Devadas, C. Fourrier, L. L. Raket, A. Sotiras, G. Erus, J. Doshi, C. Davatzikos, J. Vogel, A. Doyle, A. Tam, A. Diaz-Papkovich, E. Jammeh, I. Koval, P. Moore, T. J. Lyons, J. Gallacher, J. Tohka, R. Ciszek, B. Jedynak, K. Pandya, M. Bilgel, W. Engels, J. Cole, P. Golland, S. Klein, and D. C. Alexander, "The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up," *arXiv preprint*, 2020.

[177] C. R. Jack and D. M. Holtzman, "Biomarker modeling of alzheimer's disease," *Neuron*, vol. 80, no. 6, pp. 1347–1358, 2013.

[178] M. Baumgart, H. M. Snyder, M. C. Carrillo, S. Fazio, H. Kim, and H. Johns, "Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective," *Alzheimer's and Dementia*, vol. 11, no. 6, pp. 718–726, 2015.

[179] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

[180] J. C. Duchi, P. L. Bartlett, and M. J. Wainwright, "Randomized smoothing for (parallel) stochastic optimization," *Proceedings of the IEEE Conference on Decision and Control*, vol. 12, pp. 5442–5444, 2012.

[181] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Information Fusion*, vol. 46, no. June 2018, pp. 147–170, 2019.

[182] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–79, 2004.