\chapter{Introduction}

\section{Brief Background on Machine Learning and Uncertainty Management}

The world is currently experiencing an era of rapid development in the field of machine learning. This exciting progression has triggered massive transformations across numerous sectors, including healthcare, automotive industries, and crowdsourcing platforms. In healthcare, machine learning algorithms help in diagnosing diseases, predicting patient outcomes, and personalizing treatment plans. In the automotive industry, the emergence of self-driving vehicles is revolutionizing transportation and mobility. In crowdsourcing, machine learning is used to aggregate information sourced from a large number of individuals to solve complex problems or provide insights. However, as we leverage these technologies to drive innovation and efficiency, we are also confronted with a critical challenge: \textit{management of uncertainty and domain fluctuations in machine learning models}. Real-world applications of machine learning are riddled with situations where the provided data can contain missing, incorrect, or uncertain values. This often results in the unpredictable behavior of the machine learning models, leading to fluctuating performance across different domains.

The precision of model predictions is crucial in high-stakes decision-making situations, such as medical diagnosis or autonomous driving. Given the inherent variability of input data and the susceptibility of models to error, accuracy alone is no longer sufficient. It is crucial to also account for uncertainty in the model's predictions. With sufficient knowledge of model and data uncertainty, we can make informed decisions regarding whether to trust the model's predictions or seek additional information before making a final decision. This capability is particularly important in critical situations where inaccurate predictions could have severe repercussions.

\subsection{Uncertainty in Machine Learning Models}

Uncertainty is an integral part of any predictive model, and machine learning models are no exception. Recognizing, quantifying, and managing this uncertainty is a critical aspect of developing robust and reliable machine learning systems. Uncertainties in the field of machine learning are typically divided into two categories: aleatoric and epistemic. Aleatoric uncertainty, also referred to as statistical uncertainty, is linked to the data's inherent noise or variability. This type of uncertainty is often irreducible as it stems from factors such as measurement errors or inherent randomness in the system being modeled. Epistemic uncertainty, on the other hand, is related to our lack of knowledge about the system. This form of uncertainty, also known as model uncertainty, arises from our inability to perfectly specify the parameters of our model or from using a model that does not perfectly capture the true underlying process. The sources of uncertainty in machine learning models are complex and can be attributed to various factors. Measurement errors, missing data, and inherent noise are examples of data-related sources of uncertainty. In contrast, model-related sources of uncertainty include the choice of model structure, or model parameters, and the use of approximations in model computations.

Uncertainty can significantly impact the performance and behavior of machine learning models. Unmanaged uncertainty can lead to reduced model performance as the model struggles to make accurate predictions in the face of noisy data or imperfect model specifications. Uncertainty can also lead to biased predictions, as models may overfit to noisy data or become too confident in their predictions, ignoring the inherent uncertainty in the process. Overconfidence is a particularly dangerous effect of unmanaged uncertainty. When a model is overconfident, it may produce very certain predictions even when they are inaccurate. This can lead to poor decision-making, as users of the model may place too much trust in its predictions. In high-stakes domains such as healthcare or autonomous driving, overconfidence can lead to severe negative outcomes. Therefore, understanding and managing uncertainty is a critical task in machine learning. By providing uncertainty estimates along with predictions, machine learning models can become more reliable and trustworthy, facilitating their use in real-world decision-making tasks.

Recent methods tackle the problem of missing and uncertain data, but they assume the data is drawn from well-known distributions defined by a few parameters. Some studies approach the issue of lack of accurate labels by pre-training a network on a large dataset with noisy labels and fine-tuning it for a smaller target dataset~\cite{oquab\_Learning\_2014}. Semi-supervised techniques have also been proposed where the data with noisy labels is discarded~\cite{zhu\_Learning\_2002}. However, they suffer from model complexity and cannot be applied to large-scale datasets. Even loss functions that are perceived as noise-robust are not completely robust to label noise~\cite{bartlett\_Convexity\_2006}. With the rise and influence of ML in medical applications and the need to translate newly developed techniques into clinical practice, questions about the safety and uncertainty of models have gained more importance. Prior research mostly focused on assessing the correctness of individual decisions and modeling the behavior of individual labelers (human experts or non-experts who assign labels to data)~\cite{raykar\_Supervised\_2009}. Label bias becomes more crucial when we move from using manual delineation as our gold standard to using existing software (e.g., Free Surfer~\cite{fischl\_freesurfer\_2012} for subcortical segmentation tasks).

\section{ Objectives and Scope}

The primary objective of this dissertation is to advance the understanding of uncertainty management and, to a lesser extent, domain fluctuation in machine learning models, with the ultimate aim of enhancing models' performance, reliability, and applicability in a variety of real-world settings. We have developed, and evaluated novel methods for estimating and mitigating the impact of uncertainty on machine learning outcomes. By doing so, we hope to facilitate the more widespread adoption of these technologies across various sectors and real-world applications and, in the process, harness their full potential.

A significant part of this objective is devoted to investigating ways to incorporate uncertainty information into existing techniques with the goal of improving the accuracy of the prediction models as well as providing an additional confidence score that can enhance the models' interpretability and applicability in real-life settings. The challenge facing the machine learning community is to establish a comprehensive understanding of these issues, laying the foundation for the design of new techniques and strategies that manage uncertainty effectively in both supervised and unsupervised settings.

Addressing this challenge, we present several approaches to estimate and mitigate the effect of uncertainty in model outcomes. We propose strategies to manage uncertainty, aiming to make machine learning models more robust, reliable, and trustworthy. We focus primarily on healthcare, autonomous driving, and crowdsourcing, as these are areas where either the stakes are high or their applicability is widespread, and thus accurate uncertainty estimates can significantly enhance decision-making and outcomes.

In healthcare, we aim to improve the performance of machine learning models used for tasks such as disease diagnosis and organ segmentation. In the realm of autonomous driving, we introduce a novel transfer learning approach for driver distraction detection. In the field of crowdsourcing and ensemble learning, we have proposed new label aggregation techniques that take into account the consistency and accuracy of annotators (models in the case of ensemble learning). Further, we investigate the impact of domain fluctuation on model accuracy when segmenting thalamic nuclei and provide a novel technique that facilitates the utilization of low-contrast imaging sequences by developing a model that, albeit trained on advanced imaging technologies, can readily be used on low-contrast yet more widely available imaging sequences without sacrificing too much accuracy.

By keeping a broad scope, we have developed various methods that are not only effective in handling uncertainty and domain fluctuations but are also versatile and adaptable across different domains and applications, and thus have enhanced the utility of these techniques across different domains.

\section{Dissertation's Organization}

This dissertation contributes to ongoing research in uncertainty management and domain fluctuation in machine learning models. The methods proposed in this dissertation take into account the inherent uncertainty and variability in the data, leading to improved model performance. These methods provide a framework that can be extended to other machine learning applications. Chapter 2 presents a novel method, ``crowd-certain'', that provides a more accurate and reliable label aggregation technique, leading to improved overall performance in both crowdsourcing and ensemble learning scenarios. This method takes into account the consistency and accuracy of the annotators as a measure of their reliability, which allows us to obtain a weight that closely follows the annotator's degree of reliability. Chapter 3 proposes a novel hierarchical multilabel classification technique that utilizes the taxonomic relationship between different classes to improve classification accuracy. Further, to reduce the effect of domain fluctuation and improve the generalizability of the model to images obtained from other sources, the proposed technique provides one model trained on multiple large publicly available chest X-ray datasets (CheXpert~\cite{irvin\_CheXpert\_2019}, NIH~\cite{wang\_ChestXRay8\_2017}, and PADCHEST~\cite{bustos\_Padchest\_2020}). Chapter 4 presents a fast and accurate convolutional neural network (CNN) for segmentation of thalamic nuclei that is optimized for various diseases, magnetic field strengths, and image modalities. It demonstrates the potential of the proposed method for improving our understanding of the thalamic nuclei's involvement in neurological diseases. Finally, Chapters 5 and 6 provide a transfer learning approach for the detection of driver distraction and primary cilia cells, respectively. The proposed technique uses a combination of a CNN and a random decision forest to improve classification accuracy.

----- Below needs to be merged into the above -----

## Introduction

### 1. Overview

Machine Learning (ML) is rapidly becoming an integral part of numerous fields, from medical imaging and natural language processing to autonomous driving and crowdsourcing. Its potential to address and solve complex problems is revolutionizing the way we understand and interact with the world around us. This thesis explores innovative ML approaches to address five distinct challenges spanning various domains. Each chapter of this research introduces a novel method, unravels its workings, and evaluates its performance relative to current state-of-the-art methods. Through this, the study seeks to push the boundaries of ML application and contribute to its continuing evolution.

### 2. Background

The advent of ML has marked a significant leap in computational capabilities, enabling us to address problems that were previously considered infeasible or too complex. Crowdsourcing, medical imaging, driver distraction detection, and biological image analysis are fields that have been greatly impacted by this development. However, while advances have been made, several challenges remain. These challenges encompass the reliability of crowdsourced labeling, the differentiation of thoracic diseases, the segmentation of thalamic nuclei, the detection of driver distraction, and the classification of primary cilia in microscopy images. This research is an attempt to address these challenges, extending the capabilities of ML and expanding our knowledge of its potential.

### 3. Problem Statement

The broad problem this thesis addresses is the inherent complexity and variability in real-world data sets and the challenges these present for ML applications. Specifically, it focuses on five separate problems, each pertaining to a distinct domain. Each of these issues, although different in nature, shares a common theme: the need for advanced, robust, and accurate computational methods that can effectively manage and analyze large, complex datasets.

### 4. Research Objectives

The objective of this thesis is to introduce and evaluate novel ML methods for each of the five problem areas. Each method aims to enhance the accuracy, efficiency, and robustness of data analysis in its respective field. By presenting these methods and their respective evaluations, the thesis seeks to contribute to the understanding and development of ML applications.

### 5. Research Questions

This research seeks to answer the following questions:

* Can a novel approach improve the reliability of crowdsourced labeling?
* How can ML techniques enhance the diagnosis of thoracic diseases from chest radiographs?
* Can a CNN-based approach increase the accuracy and efficiency of thalamic nuclei segmentation?
* How can a supervised learning method improve driver distraction detection?
* Can a CNN and random decision forest classifier improve the classification of primary cilia in microscopy images?

### 6. Methodology Overview

Each chapter of the thesis introduces a unique methodology developed to address its respective problem. The methods leverage different ML techniques, including label aggregation, hierarchical multi-label classification, cascaded multi-planar schemes, and convolutional neural networks combined with random decision forests. The performance of each method is then evaluated and compared to existing techniques, enabling a thorough understanding of their effectiveness.

### 7. Scope of the Thesis

This research explores ML applications in various domains. While the techniques introduced in this thesis are generalizable, their evaluations are conducted on specific datasets. Thus, the results and conclusions are subject to the limitations and characteristics of these datasets. The presented methods, however, are scalable and adaptable, potentially opening the door to their broader implementation.

### 8. Thesis Structure

Following this introduction, the thesis is organized into five core chapters, each focusing on a specific ML challenge. The first chapter introduces "Crowd-Certain," a novel approach to label aggregation in crowdsourced and ensemble learning. The second chapter presents a hierarchical multi-label classification method for diagnosing thoracic diseases from chest radiographs. The third chapter unveils a CNN-based method for the segmentation of thalamic nuclei. The fourth chapter focuses on

----- Below needs to be merged into the above -----

\*\*Chapter 1: Introduction\*\*

1. \*\*Introduction\*\*

Machine learning has rapidly transformed the technological landscape, providing novel solutions to complex problems across diverse fields. The breadth of machine learning applications spans from data analysis and natural language processing to more recent advancements in computer vision, radiology, neuroimaging, and biological image analysis. This thesis focuses on several distinct applications, demonstrating the versatility and potential of machine learning techniques in addressing varied challenges.

2. \*\*Background and Motivation\*\*

The exponential growth of digital data across all fields has necessitated the development of innovative strategies for data analysis and interpretation. This burgeoning demand has been significantly met by the advent of machine learning, which harnesses computational power to uncover hidden patterns, make predictions, and facilitate decision-making. Leveraging these capabilities, this thesis explores diverse applications of machine learning, including crowdsourced data analysis, medical imaging diagnostics, driver distraction detection, and biological image analysis.

3. \*\*Statement of the Problem\*\*

While machine learning offers promising solutions, it also poses unique challenges in each application. For instance, the variability in reliability among crowd annotators complicates the creation of robust systems using crowdsourced data. Similarly, the high degree of similarity among radiographic manifestations of different thoracic diseases complicates the development of accurate diagnostic systems. In driver distraction detection, the challenge lies in accurately classifying human behaviors from dashboard camera images. In biological image analysis, the task of detecting and classifying primary cilia within microscopy images is hindered by the similar characteristics of non-cilia elements and imaging noise.

4. \*\*Research Objectives\*\*

This thesis aims to develop novel machine learning methods to address the aforementioned challenges. The objectives include the development of an improved label aggregation technique for crowdsourced data, the creation of accurate classification methods for thoracic diseases, the detection of driver distraction using dashboard camera images, and the automated detection and classification of primary cilia within microscopy images.

5. \*\*Research Questions\*\*

- How can we improve label aggregation techniques for crowdsourced data to create more reliable and robust systems?

- Can deep learning techniques enhance the diagnosis of thoracic diseases from chest radiographs?

- How can machine learning methods detect driver distraction using dashboard camera images?

- Can an automated process efficiently detect and classify primary cilia within microscopy images?

6. \*\*Methodology Overview\*\*

Each chapter of this thesis presents a different methodology, developed in response to the specific problem at hand. These include a novel approach for label aggregation, hierarchical multi-label classification methods for thoracic disease diagnosis, a supervised learning method for driver distraction detection, and a convolutional neural random forest classifier for primary cilia detection.

7. \*\*Scope of the Thesis\*\*

This thesis encompasses several applications of machine learning, each dealing with distinct problems and methodologies. While the primary focus is on the development of machine learning techniques, the implications of these methods extend to diverse areas, including data science, radiology, road safety, and biology.

8. \*\*Thesis Structure\*\*

The subsequent chapters provide an in-depth discussion of each of the mentioned topics. Chapters 2 through 6 respectively discuss label aggregation in crowdsourced data, thoracic disease diagnosis from chest radiographs, driver distraction detection using dashboard camera images, segmentation of thalamic nuclei, and primary cilia detection in microscopy images.

In conclusion, this thesis offers novel insights into the practical applications of machine learning across various domains, addressing some of the most pressing challenges in each field. By developing novel methodologies and exploring their potential implications, this work seeks to contribute significantly to the ongoing discourse on machine learning and its myriad applications.

------------- Chapters -----------

# Crowd

Crowdsourcing has revolutionized data collection and labeling, creating an influx of data in many fields, including computer vision and natural language processing. Although this large-scale participatory model promises a cost-effective solution for gathering massive amounts of data, it does not come without challenges. One significant hurdle is the variability in the reliability of the annotators. In most cases, crowd annotators are not domain experts and can deliver inconsistent and sometimes inaccurate labels. These discrepancies pose considerable challenges in creating a reliable and robust system that can use this data effectively and accurately.

Repeated labeling and label aggregation are common strategies employed to mitigate these discrepancies. They involve the labeling of the same data by multiple annotators and subsequent combination of these labels to estimate a single, aggregated label. Majority voting and other such methods are typically used to derive the aggregated label, aiming to mitigate biases and inconsistencies. However, real-world datasets often fail to align with the assumptions and constraints of many label aggregation techniques due to factors such as label inaccuracies, class imbalances, or sheer size.

Recognizing this limitation, this chapter presents the development and evaluation of a novel approach for label aggregation: Crowd-Certain. This technique stands out through its ability to consistently outperform existing methods by leveraging the consistency of annotators versus a trained classifier to determine a reliability score for each annotator. In doing so, it offers superior performance, robustness, and computational efficiency.

The central problem addressed in this chapter is the challenge associated with crowdsourced labeling, which can be dynamic, uncertain, and inherently varied in quality. Given the lack of domain expertise and potential discrepancies in cognitive abilities among crowd annotators, the acquired labels often suffer from a lack of quality and reliability. A potential solution to this issue is label aggregation, the process of inferring an aggregated label from a multi-label set. Despite its potential, label aggregation has proven to be a challenging endeavor due to problem-specific characteristics, the quality of labels, and the amount of data available. This is particularly evident in real-world datasets, which often present challenges such as labeling inaccuracies, class imbalances, and overwhelming sizes.

This chapter will delve into the detailed workings of Crowd-Certain, providing an understanding of its core features and capabilities. We will explore how Crowd-Certain calculates annotator reliability, its utilization of predicted probabilities, and how it circumvents the need for recurrent simulation processes. Our empirical evaluation and comparative study against ten other label aggregation techniques will be presented, underlining Crowd-Certain's impressive performance. Further discussion will include the examination of different confidence score measurement techniques and their performance in relation to our method.

# Taxonomy

The rapid advancement of medical imaging technology has amplified the significance and dependency on radiology in the diagnosis of various diseases, particularly those related to the thorax, such as heart and lung disorders. However, the high degree of similarity in the radiographic manifestations of different thoracic diseases has often led to diagnostic inaccuracies, thus escalating the need for efficient and accurate diagnostic systems. This chapter focuses on employing deep learning techniques to accurately diagnose thoracic diseases from chest radiographs, which, despite being a challenging task, holds the potential to improve patient outcomes considerably.

Chest Radiography (CXR), being the most common radiological examination, presents a pivotal tool in the swift detection and subsequent treatment of thoracic diseases. The fundamental challenge in interpreting CXRs lies in the differentiation of various thoracic diseases, the similarities among which often result in misinterpretation. Overcoming this barrier through the development of a precise system to identify and localize common thoracic diseases would provide indispensable assistance to radiologists, minimizing diagnostic errors.

While significant strides have been made in the realm of natural language processing, enabling the collection of extensive annotated datasets, the training of Convolutional Neural Networks (CNNs), and thereby the learning of intricate relationships between image objects, the need for vast amounts of labeled data for their training presents its own set of challenges. Nonetheless, deep learning techniques have gained substantial popularity in medical imaging, primarily due to their capacity to execute complex tasks with minimal human intervention.

The primary thrust of this chapter is to address the hurdles of multi-label classification in the diagnosis of thoracic diseases. Herein, we introduce two innovative hierarchical multi-label classification methods that leverage the taxonomy of pathologies to improve both the accuracy and interpretability of disease classifications. These techniques cater to scenarios where ground truth is available (termed "loss") and when it is not (termed "logit"), thereby enhancing their adaptability to new tasks. The overarching aim is to provide an additional layer of decision support to radiologists, thereby fostering better patient outcomes. This chapter discusses the "loss" and "logit" methods, their evaluations on various chest radiograph datasets, and their potential benefits and limitations.

# ChatGPT

The chapter in discussion primarily focuses on the development and evaluation of a fast and accurate method for the segmentation of thalamic nuclei, employing a convolutional neural network (CNN) based approach. Thalamic nuclei play an integral part in numerous brain functions, and their accurate segmentation is key to the study of several neurological diseases.

The novel method proposed in this chapter, a cascaded multi-planar scheme with a modified residual U-Net architecture, aims to address the challenges faced by existing segmentation techniques. It seeks to offer a more efficient and accurate approach that can work effectively on various types of imaging data, including conventional and white-matter-nulled (WMn) magnetization prepared rapid gradient echo (MPRAGE) data. These improvements facilitate the study of healthy subjects as well as patients with Multiple Sclerosis (MS) and Essential Tremor (ET), thereby expanding the scope of this tool to a broad range of applications.

The significance of this study lies in the potential implications of its findings. The proposed method has demonstrated remarkable performance on several fronts, such as speed, accuracy, and versatility, as evidenced by the performance metrics Dice similarity coefficient and volume similarity index (VSI). Furthermore, it was shown to be resistant to noise, able to deliver reliable results with signal-to-noise ratios (SNRs) as low as half the baseline.

The practical utility of the method is underscored by its application to the study of thalamic nuclei atrophy in MS patients. Not only does this novel approach outperform current state-of-the-art methods, but it does so in a fraction of the time, offering significant improvements in efficiency. This tool presents an opportunity to expand our understanding of the role of thalamic nuclei in various neurological diseases, a prospect of great interest to researchers and clinicians alike.

The comprehensive approach employed in this study takes into consideration the segmentation of thalamic nuclei from structural MRI data across different field strengths and disease conditions. The modified residual U-Net in a cascaded multi-planar scheme was validated and compared to existing segmentation methods, demonstrating significant improvements. Its application to the more commonly acquired CSFn-MPRAGE, by fine-tuning the network trained on WMn-MPRAGE data, showcases the robustness and versatility of this novel tool.

The findings from this chapter, including the validation of the proposed method on data from healthy subjects and patients with MS and ET, pave the way for significant advancements in neuroimaging. This method provides researchers with a powerful tool that combines speed, accuracy, and versatility, demonstrating the considerable potential of machine learning techniques in the study of brain structures and diseases. Future work should continue to explore the potential applications of this method, particularly in the context of other neurological conditions and imaging techniques.

# Drive-Net

This chapter aims to highlight the significance and findings of a study conducted to detect driver distraction using automated methods. The study's importance is underscored by the high rate of motor vehicle accidents caused by driver distraction. The top ten causes of driver distraction are identified, illustrating the pervasive nature of this issue. Amidst this backdrop, the advent of automated detection methods for driver distraction, particularly utilizing dashboard camera images, represents a promising strategy in mitigating the damage caused by distracted driving.

The research presented in this chapter revolves around the development of a novel supervised learning method dubbed "Drive-Net". This method leverages the potential of machine learning by combining a Convolutional Neural Network (CNN) and a Random Decision Forest to classify images of drivers. In the context of burgeoning research on object detection and human behavior detection within computer vision literature, Drive-Net marks an innovative contribution. This study employs the recent advancements in machine learning, specifically deep learning, to address the complex challenge of human behavior detection, particularly driver distraction.

In order to validate the efficacy of Drive-Net, the researchers undertook a comprehensive comparative analysis against two popular machine learning approaches, namely a Recurrent Neural Network (RNN) and a Multilayer Perceptron (MLP). The comparison was based on the methods' performance on a publicly available database comprising approximately 22,426 images, manually annotated by an expert.

The results generated through this study are remarkable. Drive-Net outperforms both RNN and MLP by achieving a detection accuracy rate of 95%, which is 2% higher than the best results previously recorded using the same database. The superiority of Drive-Net's performance substantiates the efficacy of the proposed method, presenting a credible solution to the pertinent issue of driver distraction.

This chapter offers valuable insights into the potential of automated methods in combating the escalating problem of distracted driving, further contributing to the broader field of computer science. The success of Drive-Net in detecting driver distraction from dashboard camera images signifies a noteworthy stride in the utilization of machine learning in real-world applications, potentially instigating a positive transformation in road safety measures. The results bear implications not only for researchers in computer science but also for policymakers and stakeholders in road safety.

# Cilia

This chapter of the thesis introduces and explores the challenge of accurate detection and classification of primary cilia within microscopy images, a task that holds significant importance for numerous biological studies, including diagnosing primary ciliary dyskinesia. Manual detection and classification of these organelles prove to be time-consuming and highly susceptible to subjective bias due to the similar characteristics of non-cilia elements, imaging noise, bleed-through, and clutter within the image. Furthermore, large-scale data processing exacerbates these issues, necessitating the development of an efficient, automated process.

To address this, the chapter proposes the implementation of a convolutional neural random forest classifier. This classifier combines the advantages of a convolutional neural network (CNN), renowned for their performance in many computer vision applications, with the divide-and-conquer strategy employed by decision trees. The aim of this novel method is to classify primary cilia based on whether they are located within an aquaporin 2 (AQP2) expressed region or elsewhere within the microscopy images.

The chapter proceeds with a detailed examination of the proposed classifier, setting it apart from traditional CNNs and random forests. It outlines how the classifier's random decision forest provides the final predictions, using features derived from the CNN, effectively reducing the uncertainty typically associated with routing decisions in decision trees. This approach unifies the impressive feature learning abilities of CNNs with the practicality of decision trees, proposing a unique solution to the task of classifying primary cilia.

This classifier's performance is compared with that of an unsupervised k-means classifier and a supervised multi-layer perceptron (MLP) classifier, benchmarking its effectiveness. The results presented in the chapter demonstrate superior classification accuracy for the proposed method over these traditional classifiers, offering promising insights for future applications in biological studies.

In conclusion, this chapter presents a pioneering exploration into the automated detection and classification of primary cilia in microscopy images. Through the implementation of a convolutional neural random forest classifier, it proposes an innovative solution to a complex problem, contributing to the broader field of machine learning in biological image analysis. The insights derived from this research have the potential to significantly enhance the diagnosis of primary ciliary dyskinesia and related research, showcasing the intersection of computer science and biology.