

School of Sciences | Department of Computer Science and Engineering

Replace this with the title of the thesis

Replace this with the subtitle (if applicable, if not delete the line)

Master of Science

in enter your discipline

Enter you name here

© Enter the month and year you are submitting

# Acknowledgements

# Abstract

*Write this last and remember what it is about = To provide a brief statement (no more than 1 page long) regarding the work performed. The statement should not go into too many specifics but should provide the reader with enough information to have a good idea of what the project is about.)*

# Table of Contents

[Acknowledgements 2](#_Toc496781209)

[Abstract 3](#_Toc496781210)

[Table of Contents 4](#_Toc496781211)

[1 Introduction 4](#_Toc496781212)

[1.1 Overview 4](#_Toc496781213)

[1.2 Aims and objectives 4](#_Toc496781214)

[1.3 Structure of thesis 4](#_Toc496781215)

[1.4 Summary 4](#_Toc496781216)

[2 Project Scope 4](#_Toc496781217)

[2.1 Introduction 4](#_Toc496781218)

[2.2 Title 1 4](#_Toc496781219)

[2.3 Title 2 4](#_Toc496781220)

[2.4 Title 3 4](#_Toc496781221)

[2.5 Summary 4](#_Toc496781222)

[3 Analysis and Design 4](#_Toc496781223)

[3.1 Introduction 4](#_Toc496781224)

[3.2 Title 1 4](#_Toc496781225)

[3.3 Title 2 4](#_Toc496781226)

[3.4 Title 3 4](#_Toc496781227)

[3.5 Summary 4](#_Toc496781228)

[4 Implementation and Testing 4](#_Toc496781229)

[4.1 Introduction 4](#_Toc496781230)

[4.2 Title 1 4](#_Toc496781231)

[4.3 Title 2 4](#_Toc496781232)

[4.4 Title 3 4](#_Toc496781233)

[4.5 Summary 4](#_Toc496781234)

[5 Conclusions and Future Work 4](#_Toc496781235)

[5.1 Introduction 4](#_Toc496781236)

[5.2 Title 1 4](#_Toc496781237)

[5.3 Title 2 4](#_Toc496781238)

[5.4 Title 3 4](#_Toc496781239)

[5.5 Summary 4](#_Toc496781240)

[Bibliography 4](#_Toc496781241)

[Appendices 4](#_Toc496781242)

[Appendix A 4](#_Toc496781243)

[Appendix B 4](#_Toc496781244)

[Appendix C 4](#_Toc496781245)

[Installation Manual 4](#_Toc496781246)

[Requirements 4](#_Toc496781247)

[Installation procedure 4](#_Toc496781248)

[Configuration 4](#_Toc496781249)

[User Manual 4](#_Toc496781250)

# Introduction

## Overview

Dementia is a progressive and universal condition that primarily affects memory, language and cognitive function, often leading to significant impairment in daily activities. Among the various causes of dementia, Alzheimer’s disease (AD) is the most common, accounting for over 60% of documented cases globally [[1](https://www.alz.org/alzheimers-dementia/what-is-dementia)]. Several major risk factors for AD have been identified, including age, family history, genetics, and certain chronic diseases [[2](https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors?)]. At present, over 55 million people worldwide are living with some form of dementia. This number is expected to rise to 78 million by 2030 and to 139 million by 2050. Much of this increase will occur in developing countries, with the fastest growth in the elderly population projected in Asia, particularly in China, India, and their South Asian and Western Pacific neighbours [[3](https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/#:~:text=Someone%20in%20the%20world%20develops,will%20be%20in%20developing%20countries.)].

Early identification of individuals at risk for AD is crucial for timely intervention, improved management, and the development of potential treatments. Currently, the diagnosis of AD often involves a combination of clinical assessments, neuropsychological evaluations, and advanced neuroimaging techniques, such as structural magnetic resonance imaging (sMRI), functional magnetic resonance imaging (fMRI), computed tomography (CT), and positron emission tomography (PET) scans. The high-throughput neuroimaging data generated by these modalities hold significant potential for the early detection of AD-related structural and functional changes in the brain. However, the high dimensionality and complexity of such data present considerable analytical challenges, requiring sophisticated computational methods to extract meaningful patterns and make accurate predictions [[4](https://braininformatics.springeropen.com/articles/10.1186/s40708-025-00252-3)].

To address these limitations, various machine learning algorithms have been increasingly applied in AD research. These algorithms are capable of handling large, high-dimensional datasets and can identify complex, non-linear relationships that may not be evident through conventional statistical approaches. In particular, **Natural Gradient Boosting (NGBoost)** has emerged as a promising probabilistic machine learning algorithm based on the gradient boosting framework. Unlike conventional classifiers that output fixed class labels, NGBoost predicts full probability distributions over possible outcomes, capturing both the predicted value and the associated uncertainty [[5](https://arxiv.org/pdf/1910.03225)]. This allows NGBoost to model uncertainty in classification tasks, providing more informative and reliable predictions, which is especially valuable in medical decision-making contexts such as AD disease diagnosis. Therefore, the capacity to quantify predictive uncertainty is imperative where early detection of AD is critical yet frequently challenged by overlapping clinical symptoms and heterogeneous patient data.

This master’s thesis aims to explore the capabilities of NGBoost and compare its performance with traditional machine learning models in predicting different classes within the ADNI dataset. Additionally, the study focuses on the explainability of these models to better understand and quantify their behavior, with NGBoost serving as the primary benchmark against which other algorithms are evaluated.

## Aims and objectives

The primary aim of this research is to develop a predictive framework for the early detection of Alzheimer’s disease using advanced machine learning techniques. This study is designed to evaluate the effectiveness of probabilistic and ensemble-based machine learning models in classifying Alzheimer’s disease cases, while also emphasizing model interpretability through explainability tools.

To achieve this aim, the following specific objectives have been established:

1. **To conduct an extensive exploratory data analysis** (EDA) prior to implementing any machine learning models. This is an essential step in understanding the underlying structure and distribution of the dataset, identifying potential anomalies or quality issues, and providing meaningful insights that will inform model selection and evaluation strategies.
2. **To develop and compare** the predictive performance of multiple machine learning algorithms for Alzheimer’s disease classification. These include:
   1. Natural Gradient Boosting (NGBoost), a probabilistic machine learning algorithm capable of producing predictive distributions and quantifying uncertainty.
   2. XGBoost and Random Forest, two widely used ensemble learning methods known for their high predictive performance in various classification tasks.
   3. Logistic Regression, a simple, interpretable baseline model used to assess whether the adoption of more complex algorithms is justified based on performance improvements.
3. **To evaluate the predictive performance** of the developed models and apply machine learning explainability techniques to interpret and explain the models’ decision-making processes. This objective aims to enhance the transparency and clinical applicability of the predictive models by identifying the key features influencing classification outcomes.

## Structure of thesis

## Summary

# Project Scope

## Introduction

## Title 1

## Title 2

## Title 3

## Summary

# Analysis and Design

## Introduction

## Title 1

## Title 2

## Title 3

## Summary

# Implementation and Testing

## Introduction

## Title 1

## Title 2

## Title 3

## Summary

# Conclusions and Future Work

## Introduction

## Title 1

## Title 2

## Title 3

## Summary

# Bibliography

https://link.springer.com/content/pdf/10.1186/s13195-021-00900-w.pdf

# Appendices

## Appendix A

## Appendix B

## Appendix C

# Installation Manual

## Requirements

## Installation procedure

## Configuration

# User Manual