Multimodal Data Fusion for Motor Neuron Disease Prognosis Prediction

Florence J Townend

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Centre for Medical Image Computing
University College London

March 14, 2024

I, Florence J Townend, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Abstract!

Acknowledgements

Acknowledgements!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Summary of this thesis first?

Motor neuron disease (MND) is a rare and fatal neurodegenerative disease of the upper (UMN) and lotor motor neurons (LMN). MND is a clinically heterogeneous disease. There are multiple subtypes, all of which have varying survival times and clinical presentations. The most common of these is amyotrophic lateral sclerosis (ALS), which accounts for THIS MANY % of patients. The less common subtypes are Progressive Spinal Muscular Atrophy (PMA), which is characterised by exclusively LMN involvement, Primary Lateral Sclerosis (PLS), which is exclusively UMN involvement, and Progressive Bulbar SOMETHING (PBP). Distinguishing between the various subtypes of MND during diagnosis is important because they have different prognoses and presentations. For example, PLS has a more benign prognosis than ALS because there is reduced respiratory involvement, and patients have a chance of living to a normal lifespan [?]. There is currently no cure for ALS and survival time from diagnosis is usually between 3 to 4 years [?, ?], with 10% of patients living more than 10 years [?]. Rates of ALS within a population are affected by both ancestry and sex. Global ALS incidence is estimated to be 1.68 per 100,000, and European populations experience higher incidence of 1.71 to 1.89 per 100,000 [?]. Additionally, ALS affects 1.3 men for every woman [?].

It is unknown exactly what causes MND, but genetics play a large role. Historically, ALS has been categorised into familial ALS (fALS), meaning there

is a family history of ALS, and sporadic ALS (sALS), meaning no family history. However, 10% of patients with sALS have mutations associated with fALS [?], and genetic contribution of sALS has been estimated as 61% [?]. This hazy border between what differentiates the causes of sALS and fALS has led to the movement of recategorising ALS into "genetically confirmed" and "non-genetically confirmed" [?]. As of 2022, there have been over 40 genes associated with ALS (CITE GOUTMAN 2022B). The most common of these is the C9orf72 hexanucleotide repeat expansion, which occurs in 5-15% of fALS patients [?]. C9orf72 is not unique to ALS; it is also linked to increased risks of frontotemporal dementia, Parkinson's Disease, Huntingdon's Disease, Alzheimer's Disease, Schizophrenia, and bipolar disorder. Apart from genetic risks, there is investigation into environmental risk factors for ALS. A commonly-studied factor is intense physical exercise, including involvement in professional sports and military service [?, ?], and there are multiple theories as to why this might be involvement in increased risk of developing ALS. ADD MORE HERE.

The main symptom of MND is progressive motor loss in any voluntary muscle, which means that any involuntary movement such as pupillary movement is unsually unaffected [?]. This progressive motor loss can manifest as reduced limb movement, difficulty swallowing (dysphagia), difficulty speaking (dysarthria), and impaired respiratory function, which is usually the cause of death for MND patients. At symptom onset, this motor loss is usually focused on one body segment, and the weakness spreads in a predictive pattern to the contralateral side in 85% of patients [?]. The exact mechanism driving this neurodegeneration is not completely understood, and both cellular and molecular processes are being investigated. A specific process "under the microscope" is the aggregation of TDP-43 in the brain, which has been observed in nearly all ALS patients [?].

Recently, there has been increased focus on the non-motor symptoms of MND: cognitive and behavioural changes. These changes are recognised to

occur in 35 to 50% of ALS patients, and some patient characteristics such as genetic factors and site of symptom onset affect the likelihood of cognitive involvement [?, ?]. Frontotemporal dementia (FTD) occurs in 15% of ALS patients, and ALS-FTD is often described as a spectrum [?]. The spectrum also goes two ways, with 12.5% of behavioural-variant FTD patients going on to develop ALS. The most common behavioural symptoms affecting 10% of ALS patients are apathy and loss of sympathy [?]. Other common symptoms are issues with language fluency, social cognitioin, and executive function [?]. Long-term memory and spatial memory are usually spared [?]. Although these cognitive and behavioural symptoms are not the cause of death of MND patients, increased changes may signal faster disease progression and increased caregiver burden, so it is important to recognise and measure these changes clinically.

There is no definitive diagnostic test for MND, and each patient undergoes a tailored investigation of differential diagnosis. The diagnosis process can also be tricky because upper and lower motor neuron involvement may not happen simultaneously. Additionally, there are a number of "ALS mimic diseases", which have similar symptoms to ALS but more benign prognoses, such as SPINAL ATROPHY?. A diagnostic criteria was developed in GIVE THE DATE, called the El Escorial criteria (find a citation for this), which is used for patients with a history of progressive weakness that has spread. Patients are categorised as possible, probable, probable (laboratory supported), or definite ALS. The diagnostic assessment currently does not include cognitive or behavioural changes, but tools are being developed and more commonly used in order to monitor these changes, such as ECAS [?]. Due to the length diagnostic process, general unawareness of MND symptoms, and varied speeds of symptom progression, the delay between symptom onset and diagnosis is on average 12 months and roughly halfway through the disease pathway [?].

Disease progression is clinically monitored using the ALSFRS-R (the revised amytrophic lateral sclerosis functional rating scale) [?]. This is a question-

naire through which patients are scored on their ability to complete disease-related functions, such as swallowing or rolling over in bed. The scores available for each question range from 4, meaning perfect function, to 0, meaning loss of function. These question scores are summed to make the ALSFRS-R score, which has a maximum of 48 and a minimum of 0. ALSFRS-R is often the primary outcome measure in clinical trials, but not without some controversy over whether this is appropriate [?]. Since the questions in the ALSFRS-R span a wide range of domains, there are arguments to limit use of the combined score and instead focus on domain-specific scores, such as the "limb movement score" or the "bulbar score" [?]. MND patients can also be categorised into disease stages, which is supposed to reveal how far along a patient is into their disease course [?]. The most popular staging systems are the King's Clinical Staging System [?] and the ALS Milano-Torino Staging System [?], although they are not yet in widespread clinical use [?]. Common therapies to try to slow the course of the disease include gastrostomy, preventing weight loss, and non-invasive ventilation to prevent respiratory failure [?].

The only treatment available for MND patients in the United Kingdom is Riluzole, an anti-glutamate agent which has been shown to increase median survival by 3 months [?, ?], although there is debate on whether this increase survival is only possible in the advanced stages of ALS which not all patients will reach [?]. Outside of the United Kingdom, Edaravarone is a licensed drug that has shown promise in slowing disease progression. However, the trial has been critised for its restrictive inclusion criteria and there have been concerns raised over its safety [?].

Imaging in MND:

- Only used in the clinical pathway for diagnosis: Ruling out ALS mimics, El Escorial scale, imaging to rule out other conditions

- Neuroimaging-derived measures have been shown to be associated with patient prognosis - insert citations here

- Machine learning for images, and specifically medical imaging, has been shown to be useful for prognosis in other diseases, and to some extent for MND

- Imaging can be difficult in MND later in the disease due to muscle atrophy and difficulty in positioning as well as swallowing and saliva problems

- A growing area of machine learning is multimodal data fusion, where different types of data are combined to improve the performance of the model.

- Imaging has not been included yet into the clinical pathway for prognosis, but perhaps it has a role to play when combined with clinical data through multimodal data fusion AI

This work aims to create a prognostic tool for MND using multimodal data fusion AI, combining clinical, imaging and other data types together at the diagnostic appointment

1.1 Motivation

Prognosis is the prediction of the course of a disease, and is important for both patients and clinicians. For patients, prognosis can help them to plan their lives and make decisions about their care, and for clinicians, prognosis can help them to make decisions about treatment and care, and to plan clinical trials.

- Why MND prognosis? Patients want it, good for clinical trial stratification, good for planning treatment

- Why imaging? Already being taken at diagnosis for differential diagnosis, might as well use it for prognosis - healthcare economics, no extra data collection. Also has shown promise in the literature. A lot fo machine learnign literature on imaging too - good for deep learning because it's so big and complex.

- Why data fusion? MND is multifactorial, so it makes sense to use all the data available to us. Also, it's a growing area of machine learning, and has shown promise in other diseases.

1.2 Project Aims

1. Aim 1: investigate value of imaging in non-ML data fusion prognosis
2. Aim 2: explore methods for multimodal data fusion in the literature and apply to larger multimodal neurodegenerative disease datasets
3. Aim 3: apply to MND prognosis prediction with clinical and imaging data: find the value of imaging and the best way to use it (whole brain, regions of interest, texture analysis, etc
4. Aim 4: Add more modalities such as fluid biomarkers and NLP from radiological reports
5. Aim 5: Create the optimal multimodal data fusion model for MND prognosis and assess added values of different modalities

1.3 Upgrade Thesis Outline

Chapter 2

Literature Review

This chapter contains a review of literature on prognosis of MND, looking at both prognostic factors from various data types, and also attempts to predict prognosis using machine learning.

2.1 Prognostic Factors

In MND research, prognosis is often defined as survival time, but it can also be defined as the rate of progression of the disease, future functional ability, the future need for therapies, or a combination of these [?, ?]. In this section, we will review the literature on prognostic factors in MND, and we will group the factors into four categories: clinical, genetic, fluids, and imaging.

2.1.1 Clinical

A large meta-analysis in 2021 collated research studies on non-genetic factors associated with survival risk in ALS [?]. Hazard ratios (HRs), derived from Cox Proportional Hazards (CPH) models, were calculated for each factor which had at least 3 studies reporting on it. The authors conducted sensitivity analyses and heterogeneity analyses to assess the validity of their findings, and found them to be robust.

Some of the factors associated with shorter survival in ALS are well established in the literature, and clinical information about how the disease first presents is a strong indicator of prognosis.

Firstly, an older age of symptom onset is associated with a higher risk

of death [?]. However, it is common in clinical records for the onset date to be the first date of the month or even the first day of the year if the patient cannot remember the exact date, which leads to age of onset being an imperfect measure. Furthermore, a shorter delay between symptom onset and diagnosis is associated with shorter survival [?] because a shorter delay suggests that the disease is fast-progressing and more obvious to clinicians to diagnose. A delay of more than one year indicates longer survival (HR=0.39) [?] and another study showed a delay of less than one year indicates shorter survival (HR=3.43) [?].

Both the site of symptom onset and the speed at which motor symptoms spreads to other sites are associated with survival. Compared to the most common onset site, spinal, Su and colleagues found that bulbar onset (HR=1.35) and respiratory onset (HR=2.2) are associated with shorter survival [?]. A short interval between the first motor onset and the next site involvement is also associated with shorter survival [?], and the speed of motor symptom progression is a prognostic factor independent of the sites themselves.

Extra-motor symptoms are also associated with shorter survival. Executive dysfunction, the appearance of frontotemporal dementia (FTD), and non-specific dementia are all associated with shorter survival and faster disease progression [?, ?].

Both a smaller ALSFRS-R at diagnosis and a faster rate of ALSFRS-R decline are associated with short survival [?]. he rate of decline in ALSFRS-R from onset to diagnosis is also called the progression rate to baseline, or PRB, and is calculated as

$$PRB = \frac{48 - \text{ALSFRS-R}(t_{diag})}{t_{diag} - t_{onset}}, \tag{2.1}$$

where t_{diag} and t_{onset} are the dates of MND diagnosis and symptom onset respectively, and 48 is the maximum score of ALSFRS-R.

Finally, taking Riluzole is associated with longer survival (HR=0.80), and lower forced vital capacity (FVC) is associated with shorter survival [?].

Due to the heterogeneity of MND, there are factors that have mixed associations in the literature. The El Escorial criteria, used to assist ALS diagnosis, assigns patients into categories associated with the confidence of the diagnosis, from “definite“ to “possible“ [?]. “Definite“ ALS patients have been found to progress faster than “probable“ or “possible“ patients in the meta-analysis and a large multi-centre study [?, ?]. However, in a large cohort of 1,809 Chinese patients, there was no significant relationship between El Escorial and survival [?].

MND is frequently accompanied by rapid weight loss due to feeding and swallowing difficulties, appetite loss, and muscle mass atrophy. A higher body-mass index (BMI) at diagnosis is associated with longer survival in the meta-analysis (HR=0.97) [?], also supported by smaller meta-analysis [?] and a large population study [?]. On the other hand, some studies found that baseline BMI was not important, but rather the rate of BMI decline is a better prognostic factor, both years before disease onset [?] and after diagnosis [?]. More precise measures of body composition, such as MRI of the knees and diaphragm, have found that higher subcutaneous fat is associated with higher ALSFRS-R and lower rate of ALSFRS-R decrease [?].

Statins, a drug that inhibits cholesterol synthesis, has been studied as a prognostic factor in MND. Su and colleagues found no significant effect on survival in their meta-analysis, from three papers reporting non-significant HRs [?]. However, Weisskopf and colleagues found that taking low-potency statins for short durations before diagnosis is protective for survival, but this effect is lost when the duration of statin use is over 3 or the potency of the statin is higher [?]. They suggest that statins might protect ALS survival if used for shorter durations and at lower doses, indicating less severe cardiovascular conditions that could harm survival.

2.1.2 Genetic

Genetic testing after diagnosis is becoming more common because the genetic contribution to MND is better understood and therapeutics are being devel-

oped to target specific genetic mutations [?]. A network meta-analysis on genetic factors associated with survival in ALS found that the C9orf72 repeat expansion is associated with shorter survival(HR=1.6) [?], backed up by other large cohort study [?]. Also associated with shorter survival is ATXN2, a CAG repeat expansion usually associated with spinal onset ALS (HR=3.6), and a mutated FUS (fused in sarcoma) (HR=1.8) [?].

2.1.3 Fluids

Fluid biomarkers are measurements of proteins, metabolites, or other molecules in the blood/serum, cerebrospinal fluid (CSF), or urine that can be used to diagnose and monitor disease.

From the meta-analysis of non-genetic progostic factors, higher levels of creatine kinase and creatinine in serum indicate longer survival [?]. Whereas, higher levels of neurofilament light chain (NfL) in CSF (HR=6.8), NfL in serum (HR=3.7) and albumin in serum (HR=1.52) are harmful prognostic factors.

NfL is the most studied fluid biomarker in ALS. It is a protein that is released into the CSF and blood during the process of neurodegeneration. Although NfL is best measured in CSF through an invasive and difficult procedure, it can also be measured through a simple blood test, albeit in lower concentrations.

NfL levels rise presymptomatically in ALS [?], and the concentration of NfL plateaus around a year after symptom onset [?, ?, ?].

Higher baseline NfL concentration is associated with shorter survival, concluded from consensus of over 20 studies [?]. Dreger and colleagues also found that higher baseline NfL was significantly associated with higher disease aggressiveness, independent of disease accumulation, as estimated by the D50 model [?].

2.1.4 Neuroimaging

Structural MRI is conducted during diagnosis to rule out mimic diseases, but it is not used for monitoring progression due to difficulties in scanning as the

disease progresses. Consequently, the majority of imaging studies in MND are cross-sectional at baseline. This section will discuss the associations between brain imaging measures and prognosis in MND, with the caveat that the vast majority of these studies are only on ALS patients.

ALS imaging studies often suffer from small sample sizes and inadequate patient characterisation, meaning that information on clinical phenotypes and genetic status is often missing, potentially affecting result significance [?]. Furthermore, many studies correlate brain imaging measures with ALSFRS-R, which is a measure dominated by the effects of lower motor neuron degeneration [?]. These limitations result in a large array of inconsistent findings in the literature. In this section, we are explore brain regions implicated in MND prognosis, grouped by their location in the brain.

Whole-brain Measures

Two small studies($N < 35$) reported that lower total grey matter (GM) volume is associated with faster progression, while lower white matter (WM) volume showed no such association [?, ?]. It was concluded that the GM changes occur after diagnosis, making them a potential biomarker for prognosis, and that the WM changes occur before diagnosis, making them a potential biomarker for early diagnosis [?]. However, a study nearly double the size by Trojsi and colleagues found no differences in overall GM or WM damage between fast and slow progressors, both measured by structural MRI and by diffusion tensor imaging (DTI) metrics, such as fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) [?]. Conversely, two separate studies found that lower overall FA is associated with faster progression [?, ?].

These findings highlight the inconsistency of results regarding whole-brain structural changes in ALS progression. More studies have focused on specific brain regions and white matter tracts, which we will discuss in the following sections.

Motor Cortex and Corticospinal Tract

MND, characterised by upper and lower motor neuron degeneration, affects the motor cortex and corticospinal tract (CST). The motor band sign is a hypointensity in the shape of a ribbon at the precentral gyrus. Both a higher baseline intensity and a greater change over 18 months in the motor band sign has found to be associated poor prognosis, measured by shorter survival and faster disease progression respectively [?, ?].

Studies consistently show that higher FA in the CST and a slower rate of FA decline is associated with longer survival, slower progression, and greater baseline function [?, ?, ?, ?, ?]. High FA in tracts like the posterior limb of the internal capsule and right superior longitudinal fascicle is linked to better prognosis, similar to findings in the CST [?, ?]. Furthermore, disease aggressiveness, assessed by the D50 model, correlates with white matter density decreases in tracts connecting frontal, parietal, and occipital lobes, as well as with elevated mean diffusivity (MD) and axial diffusivity (AD) in the fronto-parietal tract [?].

Cortical Thickness

Cortical thickness (CT), measured as the distance between the pial surface and the grey-white matter boundary, has shown mixed associations with MND prognosis. CT loss in the temporal and frontal lobes has been correlated with faster progression in small cohorts ($N < 50$) [?, ?]. However, in a larger cohort of 292 patients, the opposite was found: longer survivors had more widespread CT thinning at diagnosis compared to short survivors [?]. It was reported that shorter survivors then went on to have more extensive changes to CT over time, whereas the CT in longer survivors stayed constant. Finally, Dieckmann and colleagues found no association between CT volumes at baseline and D50 disease aggressiveness [?].

The variation in findings on CT and prognosis may relate to MRI timing. Cross-sectional data suggest baseline low CT associates with short survival, yet longitudinal studies are essential to discern whether the rate of CT thinning,

rather than initial CT, influences outcomes.

Subcortical Structures

Thalamic atrophy, especially in the right thalamus, correlates with disease aggressiveness and progression rate [?, ?]. Additionally, basal ganglia gray matter atrophy in the left caudate and right putamen is associated with faster progression [?, ?]. Initial smaller basal ganglia and amygdala volumes predict shorter survival, although significance diminishes with age of onset adjustment [?]. Furthermore, texture changes in the basal ganglia and hippocampus, detected via DTI analysis, are notable in short-term survivors [?].

The GM volumes of subcortical structures have been associated with cognitive impairment in other diseases [?]. The association between subcortical structures and prognosis in MND could be confounded by the presence of cognitive and behavioural impairment, which is not always reported in the studies and is a prognostic factor in MND [?].

Hippocampus

Studies have reported no associations between hippocampal volume [?, ?] or FA [?] and functional decline, measured by ALSFRS-R, progression rate, and D50 disease aggressiveness. Measuring the hippocampus in other ways, Tae and colleagues found that the shape deviations of the right hippocampus is associated with progression rate [?], and Stoppel and colleagues found that increased hippocampal activation in resting-state fMRI is associated with lower ALSFRS-R [?], albeit in small cohorts of 32 and 12 patients respectively.

Frontal Lobe

Some studies in ALS prognosis have focused on the frontal and fronto-temporal lobes due to their known association with FTD, which is a prognostic factor in MND [?]. Fast disease progression has been associated with GM atrophy [?], decreased FA [?, ?], and decrease functional connectivity [?] in the frontotemporal lobe. ALSFRS-R itself has no association with frontal areas FA [?] but has been correlated with reduced functional connectivity in the left sensorimotor

cortex [?].

Ventricles

Ventricles are enlarged when the brain atrophies, and larger ventricular volume has been associated with lower baseline ALSFRS-R in a study of 112 patients [?].

Brain Stem

It is expected that the brain stem would be a candidate prognostic marker in MND because it is involved in breathing regulation, and the most common cause of death in MND is respiratory failure. However, Steinbach and colleagues found that brain stem GM density has no affect on D50 disease aggressiveness from D50 [?], and no correlation between brain stem FA and ALSFRS-R was found in a study of 253 patients [?]. Although, in a smaller cohort of 60 ALS patients, baseline medulla oblongata volume significantly predicted short versus long survival [?].

Other measures

Other imaging measures have been linked to MND prognosis. Increased brain age, indicating accelerated brain aging, correlates with faster progression in ALS patients, particularly those with cognitive and behavioral impairment [?]. Notably, significant brain age changes were observed only in ALS patients with cognitive and behavioral impairment, suggesting that brain changes may be more pronounced in these individuals.

Magnetic resonance spectroscopy measures brain metabolite concentrations. Lower N-acetylaspartateto choline ratio in the primary motor cortex is associated with shorter survival, even after accounting for ALSFRS-R and FVC [?].

While findings from neuroimaging highlight the consistent involvement of various brain regions and white matter tracts, including the corticospinal tract and subcortical structures, challenges such as inconsistent results and confounding factors show the importance of analysing imaging data with bet-

ter patient characterisation through clinical data, especially cognitive and behavioural impairment.

2.1.5 Spinal Cord Imaging

Imaging of the spinal cord can also be conducted during the differential diagnosis to rule out alternative pathologies [?]. The spinal cord is difficult to image and is affected by many movement artefacts due to its small axial size and the proximity of the lungs and heart. Although it is usually qualitatively interpreted clinically, the cross section area (CSA) of the spinal cord has been quantitatively measured in a number of studies, and has been associated with prognosis in MND.

In a study of 43 ALS patients, Branco and colleagues found significant correlations between baseline ALSFRS-R and cervical spine CSA, and also between disease duration and CSA [?]. Moreover, Grolez and colleagues found that a smaller reduction in cervical spine volume over 3 months is associated with longer survival in a study of 41 patients [?]. However, in a study of 218 MND patients, including ALS, PLS, and PMA, the CSA of the cervical spine only correlated with the baseline ALSFRS-R of PLS and PMA patients, although there was no longitudinal spinal atrophy for the PLS patients [?].

2.1.6 Prognostic Models

There are limited clinical tools for prognosis prediction in MND . The most common way for progression to be assessed is through the rate of decline of ALSFRS-R . Extrapolating this progression rate to predict future ALSFRS-R scores is called the “pre-slope model“, but it is limited by the assumption that ALSFRS-R decline is linear.

The D50 model assumes ALSFRS-R decline is sigmoidal, and works by fitting a sigmoidal curve to a patient’s ALSFRS-R timepoints, yielding an individualised prediction of future ALSFRS-R scores [?, ?]. “Disease aggressiveness“ and “disease accumulation“ are two measures derived from the D50 model, which are the estimated rate of functional loss and the patient’s position

on the D50 curve independent of time, respectively.

A non-linear extension of the D50 model was proposed by Ramamoorthy and colleagues, where Gaussian processes are used to non-parametrically cluster patients into non-linear ALSFRS-R trajectories [?]. They found that many of the patients in their cohort had non-linear ALSFRS-R trajectories (convex, concave, sigmoidal).

Westeneng and colleagues developed the ENCALS model, a multivariable Royston-Parmar model, to predict a survival probability function for an individual ALS patient using data from 14 European ALS centres and over 11,000 patient records [?]. The harmful predictors included in the final model are bulbar onset (HR=1.71), age of onset (HR=1.03), El Escorial definite ALS (HR=1.47), higher PRB (HR=6.33), presence of FTD (HR=1.34), and C9orf72 repeat expansion (HR=1.45). The protective predictors are a longer diagnostic delay (HR=0.52) and a higher FVC (HR=0.99).

2.2 Machine Learning for Prognosis

We have seen that prognostic factors have been identified in MND, and that there are a number of prognostic models that have been developed. However, the prognostic models are not widely used in clinical practice, and there is a need for more accurate and generalisable models to be developed. Machine learning (ML) is A potential solution to this problem is machine learning (ML), which is the use of algorithms to learn from data and make predictions or decisions. In this section, we discuss the literature on ML for MND prognosis using clinical data, imaging data, or a combination of both.

2.2.1 Clinical

As mentioned earlier, prognosis can be defined as the prediction of the future course of a disease. In context of MND, prognosis could be future ALSFRS-R scores, progression rate, survival time, or predicting time until a treatment is needed.

A popular prediction task in ML prognosis is predicting future progres-

sion rate, as calculated by the slope between ALSFRS-R scores. In 2011, the DREAM Phil Bowen Prize4Life ALS Prediction Challenge tasked entrants to use 3 months of clinical trials data to predict the progression rate over the following 9 months [?]. This challenge used the PRO-ACT database, which is the largest publicly available dataset of clinical trials data in MND with over 8,500 patients from multiple trials [?]. Although widely used in ML studies, the trials' inclusion and exclusion criteria led to younger patients with fewer functional impairments, and so results using PRO-ACT have limited generalisability to the MND population.

The challenge results reported that random forest and tree-based decision models performed the best [?], and a post-challenge study found highest performance with ensembles of classical ML models [?]. Some unexpected findings emerged, such as a high variability in individual ALSFRSr scores being a strong predictor of progression rate [?] and previously unidentified progression biomarkers such as blood pressures and uric acid. A 2022 revisit using deep learning models found similar performance to classical ML, suggesting further data or tasks may be needed to demonstrate deep learning's benefits [?].

While the Prize4Life challenge was a great catalyst for ML research in MND, predicting linear decline over 9 months is a flawed task. Condensing 9 months of progression into a single slope oversimplifies the disease course, as evidenced by Ramamoorthy and colleagues finding how non-linear ALSFRS-R trajectories are very common [?]. Furthermore, random forests outperform the pre-slope model in predicting future ALSFRS-R scores, showing that models capable of non-linear calculations are needed for predicting functional decline.

Predicting fast versus slow progression in MND is a common task where patients are labelled based on progression rates [?, ?]. This task often outperforms predicting actual progression rates due to its less granular nature. Training separate models on patient subgroups has also shown to improve performance, either by stratifying on progression rate [?] or on deterioration pattern [?]. The success of these stratification approaches suggests that the het-

erogeneity of MND may be too great for a single model to predict progression rate accurately.

A more data-driven way of grouping patients is to use clustering, which is a type of unsupervised learning where the model groups patients into clusters based on their features. Grollemund and colleagues used UMAP (uniform manifold approximation and projection) to reduce the dimensionality of patient data, and then coarsely divided the lower-dimensional space into tiers of 1-year survival risk [?]. They found that this approach outperformed random forest and logistic regression in predicting 1-year survival, even though the latent space had no knowledge of the survival times of the patients in the training data.

Another goal in MND prognosis is to predict time to treatment, such as time to NIV or time to PEG (percutaneous endoscopic gastrostromy). The IDPP Clef challenge is a recent challenge in 2022 that focused on predicting risks of clinical events and timings in MND [?]. The 4 resulting papers all found that it was comparatively simple to predict the risk of clinical events, but not the timings [?, ?, ?, ?]. Other attempts at predicting time to treatment have focused on only predicting time to NIV [?, ?]. However, predicting the timing of medical interventions is a difficult task due to varying clinic strategies and interpretations of clinical guidelines. Predicting the assessment outcomes that lead to the decision to start treatment may be a more useful task.

Integrating logical rules and explainability methods enhances the clinical relevance of models in MND prognosis. Tavazzi and colleagues used a dynamic bayesian network to simulate disease course according to the MiToS staging system, incorporating clinical and biological logic into their model [?]. This allows the model to not only learn from the data, but also to be guided by clinical and biological sense, which should mitigate spurious conclusions after training.

Müller and colleagues used a deep learning longitudinal neural network to predict respiratory impairment in MND [?]. To overcome the “black-box“

nature of deep learning methods, they employed an explainability method to find the most important features in the model. They found that their model had learned clinically unintuitive relationships, which brought the model's predictions into question. This shows the importance of explainability methods in MND prognosis, and the need for the model's predictions to make sense in the context of the disease.

In predicting MND prognosis with ML and clinical data, there are various approaches to consider. It is crucial to ensure that the model's predictions align with the disease context and to assess their clinical relevance. A survey of 242 Dutch ALS patients revealed a preference for knowing their exact survival time over a survival category (slow, medium, fast) [?]. However, none of the ML studies have attempted to predict exact survival time. This presents a potential future direction for ML in MND prognosis, which could be more meaningful to patients than predicting progression rate.

2.2.2 Imaging

Machine learning has had success within imaging studies of neurodegenerative diseases, such as Alzheimer's disease [?] and Parkinson's disease [?], but less so in MND due to comparatively small sample sizes. Sample sizes are small in MND, and Computer vision ML models require larger cohorts to train the models, due to the complexity of the models and the need for a large number of parameters to be estimated. Few studies have investigated MND progression using imaging independently of clinical data, often addressing the small sample size issue by using image-derived features instead.

Imaging has been used to predict baseline progression rate using white matter connectivity from DTI [?], and to classify patients into neuropathological disease stages using DTI features of ALS-associated tracts [?]. Both of these studies have limited clinical usefulness, since it is not necessary to predict a baseline progression rate, and the neuropathological disease stage was not associated with the popular clinical staging systems, King's and MiToS .

Querin and colleagues used FA of the spinal cord and spinal cord atrophy

to predict survival in a CPH model, and found that MRI parameters were more predictive than clinical features, albeit in a small cohort of 49 patients [?].

In summary, imaging studies in MND have focused on diagnosis and disease understanding rather than prognosis prediction. Despite the many prognostic factors identified in neuroimaging studies, the limited sample sizes in MND imaging research limits the utility of imaging data alone.

2.2.3 Multimodal

Combining imaging and clinical data could increase the amount of information available to a model and address the suboptimal use of imaging data in MND prognosis. Multimodal data fusion is a technique that combines data from various modalities into a single model in order to accomplish this integration. Due to its multifactorial nature and complexity, MND is a good candidate for multimodal data fusion. Integrating different data sources using ML could help to fully capture the complexity of the disease, and to understand the underlying multifactorial mechanisms of MND progression through the interactions between the different modalities within the model. Furthermore, with limited consensus on imaging prognostic markers in MND, a data-driven strategy guided by clinical data could aid in identifying the most essential imaging findings for prognosis. So far, all of the studies with multimodal data fusion in MND have relied on extracted features from the imaging, rather than the images themselves. However, it is possible to use the images themselves in a multimodal model, and this is a promising future direction for MND prognosis.

Combining clinical and imaging data involves concatenating features from different modalities into a single model, feasible when features share the same dimensionality. However, this introduces the curse of dimensionality, potentially leading to overfitting. Kuan and colleagues found a concatenation multimodal survival model performed similarly to a clinical-only model in predicting survival, while Schuster and colleagues showed identical test performance between clinical-only and concatenation multimodal models, possibly due to overfitting and the curse of dimensionality. This suggests the concatenation

method may have limitations in utilizing imaging data effectively.

To mitigate the limitations of concatenation, other studies have used more advanced methods of multimodal data fusion, such as multimodal dimensionality reduction of the joint data. Behler and colleagues used PCA on concatenated features of cognitive, oculomotor, and DTI to cluster patients into neuropathological disease stages [?]. Kmetzsch and colleagues used a deep learning unsupervised method called a variational autoencoder to lower the dimensions of joint miRNA and structural MRI extracted volumes to predict disease progression in FTD, ALS, and ALS-FTD patients [?]. They found that different modalities were important at different stages of the disease. This insight is an example of how multimodal data fusion can unveil more about the disease than unimodal data alone.

Supervised deep learning is a multimodal data fusion method used in MND prognosis. Van der Burgh and colleagues classified sporadic ALS patients into survival categories using clinical characteristics, structural MRI, and diffusion-weighted imaging features [?], and a later extension added simulated TDP-43 accumulation levels [?]. Their model comprised three unimodal neural networks trained separately and a fourth neural network integrating their outputs, showing significantly improved performance compared to unimodal models. However, an innapropriate statistical test was used to compare the performance distributions. An improved approach could involve training all neural networks together, allowing the intermediate weights and feature maps of the different modalities to interact.

In conclusion, prognosis prediction in MND is a complex task with many different approaches and features that can be used. The literature is afflicted with small sample sizes, inconsistent findings, and unclear clinical relevance, but multimodal data fusion is a promising approach for MND prognosis, and has been shown to improve performance in some studies, with the method of data fusion being important. Furthermore, there are many more multimodal data fusion methods that have not been explored in MND, such as graph

neural networks and attention mechanisms, which could be key in accurately predicting MND prognosis.

Chapter 3

Cox Proportional Hazards Model

3.1 Introduction

- Cox proportional hazards model - what is it?
- What is my hypothesis: using clinical and imaging features together in a simple cox regression will need to hugher concordance than imaging and clinical data alone
- What is concordance? How is it calculated?
- Examples of cox models in ALS before:

3.2 Methods

- Clinical Data:
  - Milan data: sample size, some partients died and some were censored and we're given the censor date
  - Features: age at visit, alsfrsr, diagnostic delay, age at diagnosis. Imaging
  - What preprocessing was done:
- Imaging Data:
  - MRI within 12 months of diagnosis

  - Synthsegged into regions - give the regions
  - Normalised (z-score I think, check this)
  - Not all the regions were input into the model - give the regions and why (too many features)

- Demographics of the sample: table
- Experiment details: how the cox model was trained and tested - python package, parameter choices, etc
  - Kaplan meier plots: tertile split
  - Stratified by sex

3.3 Results

Clinical-only:

- Table of results with all results in it
- What are the significant features and their p-values: none
- Hazard ratio plot
- Kaplan Meier plot

Imaging-only:

- Table of results with all results in it
- What are the significant features and their p-values: lateral ventricle and brain stem
- Hazard ratio plot
- Kaplan Meier plot

Clinical and imaging:

- Table of results with all results in it

- What are the significant features and their p-values: ALSFRSr, brain stem, diagnostic delay
- Hazard ratio plot
- Kaplan Meier plot

3.4 Discussion

Do these results make sense?:

- Clinical-only: interesting no significant features, maybe not enough features
- Imaging-only: both significant features and their hazard ratios make sense given the literature (citations)
- Clinical and imaging: significant features and their hazard ratios make sense given the literature (citations)

Differences between inputs:

- Higher concordance with clinical and imaging together as opposed to each by themselves. Better AIC as well.
- Different significant features for clinical and imaging alone and together. Why?

Limitations:

- Small-ish sample size
- However, colinearity of features doesn't diminish the results

Take home message:

- Using imaging-derived feature with clinical features increases concordance, more information into the model is better, makes the clinical measure significant as well when it wasn't before

- Would this result also be sustained when using more sophisticated machine learning models? Instinct says yes because they can handle more features and more complex relationships between features

3.5 Conclusion

- Link to next chapter: we've shown that using imaging and clinical together is better than using them alone
- Let's see if we can improve on this by using more sophisticated machine learning models

Chapter 4

Fusilli: Developing a Data Fusion Python Library

Linking sentence from Cox chapter: We've seen how multimodal data affects a survival Cox model in MND. What about machine learning for multimodal data? This chapter discusses multimodal data fusion in more detail and describe the development of Fusilli, a Python package for multimodal data fusion experimentation and analysis.

4.1 Introduction

Describing multimodal data fusion:

- This is an area of research called multimodal data fusion but it can be called more names than that: multi-view, cross-heterogeneous, etc
- The application areas of multimodal data fusion are very wide reaching: agriculture, disaster management, robotics, healthcare etc
- Additionally, the types of models used in multimodal data fusion can vary a lot, from geometric deep learning to relatively simple neural network architectures.

What did I want to do with data fusion:

- Multimodal data fusion is such a large field with many applications

- We saw in the lit review sections on machine learning that deep learning has not been done much with MND, and deep-learning data fusion has been done once by van der Burgh
- and deep learning data fusion has not been done with tabular and image data in MND
- For my PhD's overall goal of applying multimodal data fusion to motor neuron disease, it is important to have an understanding of the different models and their performances
- This would require a large amount of experimentation and analysis, which would be difficult to do manually
- So I wanted to compare different fusion methods to see what worked best

Issues I came across when starting to look at models:

- Difficulty finding models in the first place: called different names
- As I kept looking, I kept finding lots of and lots of models with different architectures
- Not all papers include code
- Code is not standardised - written in different languages, varying availability, varying guidance
- There are some collections/repos but not what I was looking for

Aims of this work:

- To create a bank of models that can be used for multimodal data fusion
- To create a standardised way of comparing models
- Make it open source for others to use who are in similar positions to me

Chapter aims:

- To describe the development of Fusilli
- To describe the design choices made in developing Fusilli
- To describe the implemented models in Fusilli
- To describe the testing of Fusilli
- To describe the reception of Fusilli

4.2 Fusion Methods

4.2.1 Different types of models

Talk about the review paper and the categories

- Gathered models and then categorised them according to Cui et al. paper - a paper about diagnosis and prognosis deep learning multi modal fusion models.
- The paper included many models with different architectures and categorised based off underlying architecture.
- Usually categorisations are broad with early, late, and intermediate fusion but this paper had more categories which was useful for me.
- Categories figure
- Some models can fall into multiple categories, so chose the category that best described the model.

4.2.2 Finding the models

Literature search - what words I used, not just healthcare but it came up a lot

4.2.3 Models included

Table of models, and benchmarks (uni-modal)

Link to documentation with diagrams of models

4.3 Software Design

4.3.1 Design Goals

List in bullet points the design goals of fusilli and how they were met:

1. Modularity: Easily add models in the future and easy adjustments
2. Beginner-friendly: default parameters for people who don't know how to do deep learning coding, also extensive documentation
3. Expert-friendly: for people who are more familiar with deep learning, they should be able to change the training parameters, modify the models, access trained models for further experiments
4. Wide-ranging: include a wide variety of models
5. Widely-applicable: all models to be used for different tasks, all image dimensions and tabular data
6. Open-source: want people to contribute with their own models. To make sure it doesn't break when people do this, it needs testing.

4.4 Results

4.4.1 Diagram of workflow

4.4.2 Example usage and outputs

Quick-start script with comments going through each line

Figures with output figures from the example notebooks

4.4.3 Documentation

Examples, templates for contributing, etc

4.4.4 Reception

- JOSS paper - under review
- GitHub stars and forks and articles

4.5 Discussion

- Implemented a wide variety of models
- It will help people to know if data fusion is useful for their task and, if so, which model is best for their task
- Limitation and future work: Data inputs - images have to be .pt files and tabular data has to be .csv files. Would be nice to extend data inputs to be jpegs or niis for people less comfortable with Python.
- Limitation and future work: Only 2 modalities - could be extended to more

4.6 Conclusion

TBC...

Chapter 5

Fusilli on MND

5.1 Introduction

- We've seen the results of fusilli on open data
- Now let's do it on MND data
- How different is MND to PD and AD? *Look this up* - what might we expect to see?
- Hypothesis: multimodal data fusion is useful for predicting survival in MND when deployed at the diagnostic appointment
- Motivation - why is this useful? Implications if multimodal data fusion is better, imaging doesn't have to be just used for differential diagnosis and can play a role in prognosis, which is important for clinical trials and patient care.

The aim of this work is to assess the added value of brain ROI volumes in MND prognosis prediction. We hypothesise that the combination of clinical and imaging data will improve the prediction of survival in MND patients. We will do this by using the fusilli package to compare the performance of different tabular-tabular multimodal data fusion methods.

5.2 Data

ALS Biomarker's study and Opsedale San Raffaele. We are using clinical data and brain volumes.

Inclusion criteria: - Diagnosis of MND - Non-missing age, sex, date of diagnosis, date of death, date of onset, features of FTD, ALSFRS-R - MRI within 12 months of diagnosis (either side) - Why within 12 months? Trade-off between having the MRI close to the diagnosis and having enough data to work with

Clinical data features: - Following the same features as the ENCALS model, and working with what we have available - sex, bulbar onset, signs of FTD, ALSFRSr, ALS subtype, diagnostic delay, age at diagnosis

Milan date of death is death or tracheostomy

5.2.1 Clinical Data

- Inclusion criteria
- Features

5.2.2 Imaging Data

- Inclusion criteria: how far away from diagnosis was the MRI? Within 12 months of the diagnosis
- Segmentation using SynthSeg - what is synthseg?
- Features

5.3 Methods

- What are we predicting? Long vs short survival split on the median
- What methods are we using? All the tabular-tabular fusion methods available in fusilli, plus a baseline of just using the clinical data or just using the imaging data
- K-fold cross validation

- To improve the stability of the results, we retrained and reevaluated the models until the mean of the performance of the repetitions converged to *include percentage here*
- Metrics: AUROC and accuracy

5.4 Results

Model comparison figure.

5.5 Discussion

5.5.1 What does it mean??

Interpreting the results.

5.5.2 Limitations

- Limitations on sample size
- Predictive task of classification rather than regression: what if we used a regression task instead? Would that be more useful? It's a harder task so may require more data
- Limitation on using extracted brain volumes rather than raw MRI: what if the regions we've chosen aren't the most important ones? Subcortical regions have shown to have a role in MND, but we haven't included them here. *Look this up - the thalamus stuff*

5.6 Conclusion

First look at multimodal data fusion in MND. What does it mean? What are the implications? What are the next steps?

- If imaging + clinical is useful
  - Let's add modalities
  - Let's mix up the imaging preprocessing: DTI? Sub-cortical segmentation?

- If imaging + clinical isn't useful
  - Let's swap out the imaging for other modalities
  - Let's try different machine learning models
  - Let's mix up the imaging preprocessing: DTI? Sub-cortical segmentation?

Chapter 6

Conclusions and Future Work

6.1 Summary and Conclusions

- What have I done?
- Why is it useful and novel?
- What did I find out?
- What are the implications?
- What are the limitations?

6.2 Future Work

6.2.1 Future project 1: Adding fluid biomarker data

- Motivation: fluid biomarkers are more accessible than MRI etc.
- Feasibility: ALS Biomarkers Study etc.

6.2.2 Future project 2: Altering image preprocessing

- Motivation: might be better to drill down rather than using whole brain, example papers: ..
- Feasibility: Some methods - toolkits, etc.

6.2.3 Future project 3: Spinal cord MRI data fusion

- Motivation: spinal cord is important in MND, example papers to show this
- Feasibility: Access to spinal cord MRI data

6.3 Timeline

- What have I done so far? Papers and conference submissions
- Outcomes for the rest of my PhD
  - Papers
- Gantt chart